ΕΘΝΙΚΌ ΜΕΤΣΌΒΙΟ ΠΟΛΥΤΕΧΝΕΊΟ
ΣΧΟΛΉ ΗΛΕΚΤΡΟΛΌΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Τεχνική και οικονομική ανάλυση της μεταφοράς και φιλοξενίας των υπηρεσιών πληροφορικής μιας τηλεπικοινωνιακής επιχείρησης στο Cloud

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Γ. Τουμπανιάρης

Δημήτριος Ε. Κουσούλης

Επιβλέπων : Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021

ΕΘΝΙΚΌ ΜΕΤΣΌΒΙΟ ΠΟΛΥΤΕΧΝΕΊΟ
ΣΧΟΛΉ ΗΛΕΚΤΡΟΛΌΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Τεχνική και οικονομική ανάλυση της μεταφοράς και φιλοξενίας των υπηρεσιών πληροφορικής μιας τηλεπικοινωνιακής επιχείρησης στο Cloud

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Παναγιώτης Γ. Τουμπανιάρης

Δημήτριος Ε. Κουσούλης

Επιβλέπων : Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16η Ιουλίου 2021.

.....	.....	.....
Ιωάννης Ψαρράς	Δημήτριος Ασκούνης	Χρυσόστομος Δούκας
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τους συγγραφείς και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

.....

Παναγιώτης Γ. Τουμπανιάρης

Δημήτριος Ε. Κουσούλης

Διπλωματούχοι Ηλεκτρολόγοι Μηχανικοί και Μηχανικοί Υπολογιστών Ε.Μ.Π.

Περίληψη

Ο στόχος αυτής της διπλωματικής είναι να αναλύσει και να σχεδιάσει τη "νεφοποίηση" (cloudification), δηλαδή την μεταφορά των συστημάτων και υπηρεσιών στο cloud, σε έναν βασικό τομέα της αγοράς που επηρεάστηκε τόσο στο φορτίο υπηρεσιών όσο και στις συνθήκες εργασίας των υπαλλήλων της - μιας Τηλεπικοινωνιακής εταιρείας (TelCo).

Στη παρούσα διπλωματική, θα:

A. Πραγματοποιήσουμε μια Τεχνική Ανάλυση στην τρέχουσα κατάσταση της εταιρείας αναφορικά με:
    a. τη δομή της εταιρείας.
    b. τις υπηρεσίες πληροφορικής.

B. Σχεδιάσουμε μια λύση που βασίζεται στην Πληροφορική Cloud για κάθε τύπο υπηρεσιών πληροφορικής, πιο συγκεκριμένα:
    a. Ορίσουμε τις Cloud Native υπηρεσίες και τις βέλτιστες πρακτικές του Cloud.
    b. Μεταφράσουμε τις τρέχουσες υπηρεσίες σε υπηρεσίες Cloud Ready / Native.
    c. Εκτελέσουμε ανάλυση κόστους και ανάλογη σύγκριση με το κόστος ενός φυσικού/ιδιόκτητου κέντρου δεδομένων.

C. Δημιουργήσουμε ένα πλάνο Μεταφοράς στο Cloud, δηλαδή θα αναλύσουμε:
    a. τις στρατηγικές.
    b. τους παράγοντες κόστους.

Λέξεις Κλειδιά: Τεχνική Ανάλυση, Οικονομική Ανάλυση, Μεταφορά στο Νέφος, Τηλεπικοινωνιακή Εταιρία, Υπολογιστικό Νέφος.

Abstract

The target of this thesis it to analyze and plan the cloudification of an Enterprise in a core sector that was affected in both the service load as well as the working conditions of its employees, a Telecommunication Company (TelCo).

We will

A. Perform a Technical Analysis of the current state of the company.
    a. Analyze the company structure.
    b. Analyze the IT services.

B. Design a Cloud Computing based solution for every type of IT services.
    a. Define Cloud native services and best practices.
    b. Translate the current services to Cloud Ready / Native services
    c. Perform a Financial analysis and comparison to on-premise Data Center costs.

C. Provide a Cloud Migration plan.
    a. Analyze Strategies.
    b. Analyze Cost Vectors.

Keywords: Financial Analysis, Technical Analysis, Cloud Migration, Telecommunication Company, Cloud Computing

Ευχαριστίες

Θα θέλαμε να ευχαριστήσουμε θερμά τον επιβλέπων καθηγητή κ. Ψαρρά για την ευκαιρία και την εμπιστοσύνη που μας έδειξε στην επιλογή και συγγραφή της παρούσας διπλωματικής εργασίας.

Θα θέλαμε τέλος να ευχαριστήσουμε τις οικογένειές μας για την υπομονή και υποστήριξή τους καθ' όλη τη διάρκεια των σπουδών μας.

Πίνακας περιεχομένων

Πρόλογος

Κατά τη διάρκεια της πρόσφατης πανδημίας COVID-19, κατέστη σαφές ότι οι επιχειρήσεις οποιουδήποτε μεγέθους πρέπει να επανεκτιμήσουν το λειτουργικό τους μοντέλο προκειμένου να μπορέσουν να προσαρμοστούν στις νέες προκλήσεις. Η πανδημία εισήγαγε απότομα την ανάγκη ευελιξίας, τόσο στον φόρτο εργασίας των συστημάτων πληροφορικής, όσο και στις συνθήκες εργασίας των εργαζομένων και της συμπεριφοράς των καταναλωτών. Και ενώ το παλιό, "παραδοσιακό" μοντέλο λειτουργίας ακυρώθηκε μέσα σε λίγες εβδομάδες, το νέο μοντέλο, μας παρουσιάζει πολλές προκλήσεις καθώς και ευκαιρίες, αλλά πάνω απ' όλα, μας υπενθυμίζει ότι πολλές από αυτές τις αλλαγές καθυστέρησαν να πραγματοποιηθούν.

Ο μετασχηματισμός cloud διαμορφώνεται τη τελευταία δεκαετία, αλλά πραγματικά απογειώθηκε το 2020, όταν η ανάγκη για ψηφιακές υπηρεσίες και απομακρυσμένη εργασία αυξήθηκε εκθετικά και σε παγκόσμιο επίπεδο. Οι πελάτες απαιτούν περισσότερες από τις, παραδοσιακά φυσικές υπηρεσίες να προσφέρονται μέσω του Διαδικτύου ενώ ταυτόχρονα οι εργαζόμενοι υποχρεώθηκαν να εκτελούν τα καθήκοντά τους χωρίς να έχουν πρόσβαση στις εγκαταστάσεις του γραφείου.

Ο στόχος αυτής της διπλωματικής είναι να αναλύσει και να σχεδιάσει τη "νεφοποίηση" (cloudification), δηλαδή την μεταφορά των συστημάτων και υπηρεσιών στο cloud, σε έναν βασικό τομέα της αγοράς που επηρεάστηκε τόσο στο φορτίο υπηρεσιών όσο και στις συνθήκες εργασίας των υπαλλήλων της - μιας εταιρείας τηλεπικοινωνιών (TelCo).

Στη παρούσα διπλωματική, θα:

D. Αναλύσουμε την τρέχουσα κατάσταση της εταιρείας αναφορικά με:
    a. τη δομή της εταιρείας.
    b. τις υπηρεσίες πληροφορικής.

E. Σχεδιάσουμε μια λύση που βασίζεται στο Cloud για κάθε τύπο υπηρεσιών πληροφορικής, πιο συγκεκριμένα:
    a. Ορίσουμε τις Cloud Native υπηρεσίες και τις βέλτιστες πρακτικές του Cloud.
    b. Μεταφράσουμε τις τρέχουσες υπηρεσίες σε υπηρεσίες Cloud Ready / Native.
    c. Εκτελέσουμε ανάλυση κόστους και ανάλογη σύγκριση με το κόστος ενός φυσικού/ιδιόκτητου κέντρου δεδομένων.

F. Δημιουργήσουμε ένα πλάνο μετεγκατάστασης, δηλαδή θα αναλύσουμε:
    a. τις στρατηγικές.
    b. τους παράγοντες κόστους.

Δεν θα εμβαθύνουμε στην οργανωτική δομή της εταιρείας όμως θα αναλύσουμε σημαντικές ομάδες υπηρεσιών πληροφορικής όπως το Κέντρο επικοινωνίας (Contact Center). Παράλληλα, μικρότερες υπηρεσίες όπως οι νομικές υπηρεσίες, θα ενσωματωθούν σε μεγαλύτερες ομάδες. Επίσης, οι

υπηρεσίες υποστήριξης δικτύου τηλεπικοινωνιών, τόσο για σταθερές όσο και για κινητές συσκευές, είναι "engineering” υπηρεσίες και απαιτείται να βρίσκονται όσο το δυνατόν πιο κοντά στο φυσικό δίκτυο. Για αυτό το λόγο, οι συγκεκριμένες υπηρεσίες είναι εκτός του εύρους αυτής της διπλωματικής.

Τέλος, επιλέγουμε την AWS (Amazon Web Services) ως πάροχο υπηρεσιών Cloud, λόγω της υψηλής διαθεσιμότητας και του τεράστιου χαρτοφυλακίου cloud υπηρεσιών της.

Ανάλυση Παρούσας Κατάστασης «As Is»

Δομή μίας επιχείρησης

Καθώς ολόκληρη η δομή της επιχείρησης είναι εκτός του πεδίου αυτής της διπλωματικής, έχουμε συγκεντρώσει την ανάλυση, στις οργανωτικές μονάδες που σχετίζονται με την πληροφορική και τα αποτελέσματά μας βασίστηκαν σε δεδομένα από πραγματικές επιχειρήσεις. Λαμβάνοντας υπόψη ότι αυτά τα δεδομένα είναι στρατηγικής φύσης και δεν μπορούν να δημοσιευτούν, έχει δοθεί ιδιαίτερη προσοχή στην απολύμανση (sanitization) αυτών και ενώ οι αριθμοί που ακολουθούν προσεγγίζουν σε μεγάλο βαθμό την πραγματικότητα, δεν μπορούν να αποδοθούν άμεσα σε καμία μεμονωμένη Εταιρεία Τηλεπικοινωνιών (TelCo).

Σε οποιαδήποτε TelCo επιχείρηση, μπορούμε να προσδιορίσουμε δύο κύριες κατηγορίες υπηρεσιών: Υπηρεσίες τεχνολογίας πληροφοριών (IT) και υπηρεσίες δικτύου (NW). Οι υπηρεσίες δικτύου περιλαμβάνουν όλες τις υπηρεσίες που υποστηρίζουν τη λειτουργία του δικτύου τόσο για σταθερές όσο και για κινητές υπηρεσίες.

Οι σταθερές υπηρεσίες περιλαμβάνουν υπηρεσίες διαχείρισης DSLAM, υπηρεσίες διαχείρισης εξοπλισμού πελατών (CPE), υπηρεσίες παρακολούθησης συνδεσιμότητας, υπηρεσίες χαρτογράφησης φυσικού δικτύου (Copper / Fiber), υπηρεσίες σχεδιασμού και λειτουργίας και υπηρεσίες σύνδεσης στο διαδίκτυο (συνδεσιμότητα backbone).

Οι υπηρεσίες κινητής τηλεφωνίας περιλαμβάνουν υπηρεσίες λειτουργίας πύργου κινητής (παρακολούθηση συνδεσιμότητας, ασφάλεια, ισχύ, πρόσβαση), παρακολούθηση υγείας δικτύου, υπηρεσίες παρακολούθησης χωρητικότητας δικτύου και υπηρεσίες ενοποίησης, 5G και δρομολόγησης κειμένου (υπηρεσίες SMS).

Αυτές οι υπηρεσίες πρέπει να παραμείνουν κοντά στο φυσικό δίκτυο και δεν είναι αποτελεσματικό να μεταφερθούν στο Cloud, εκτός εάν υπάρχει κέντρο δεδομένων κάποιου παρόχου cloud στην ίδια περιοχή με το φυσικό δίκτυο. Τη στιγμή της γραφής αυτής της διπλωματικής, κανένας πάροχος δημοσίου cloud δεν έχει φυσική παρουσία στην Ελλάδα, επομένως αυτές οι υπηρεσίες θα παραμείνουν εκτός του πεδίου αυτής της διπλωματικής.

Οι υπηρεσίες πληροφορικής αποσυνδέονται από το δίκτυο και αλληλεπιδρούν με αυτό μόνο μέσω άλλων υπηρεσιών δικτύου και ως εκ τούτου, δεν απαιτούν εγγύτητα με το φυσικό δίκτυο. Θα εστιάσουμε την ανάλυσή μας σε αυτές τις υπηρεσίες στην επόμενη ενότητα.

Υπηρεσίες πληροφορικής

Οι ακόλουθες μονάδες ταυτοποιήθηκαν, με αναλυτικότητα έως και 1% του TCO, με τυχόν μικρότερες μονάδες να ενοποιούνται σε μεγαλύτερες.

Ενοποιημένη λίστα	
Τμήμα	**% του TCO**
Διαχείριση πελατών	23,00%
Υπηρεσίες χρέωσης	14,98%
Επιχειρηματική ευφυΐα	14,29%
Υπηρεσία διαχείρισης παραγγελιών	10,45%
Κέντρο επικοινωνίας	8,01%
Υποστήριξη Backend	6,97%
Εφαρμογές Frontend	6,97%
Υπηρεσίες Middleware	6,97%
Σύστημα Διαχείρισης Πόρων (ERP)	5,92%
Υπηρεσίες διαχείρισης συμβάντων	1,39%
Υπηρεσίες διαχείρισης περιεχομένου	1,05%
Σύνολο	100,00%

Πίνακας 1 : Ενοποιημένη λίστα τμημάτων

Όπως μπορούμε να δούμε στον Πίνακα 1, οι οργανωτικές μονάδες που συντομεύθηκαν για το κόστος, είναι:

1. **Διαχείριση πελατών**, συμπεριλαμβανομένης της διαχείρισης σχέσεων πελατών (CRM), όπως το Siebel και η διαχείριση επιχειρησιακών σχέσεων (BRM), οι οποίες αποτελούν τη ραχοκοκαλιά οποιουδήποτε τύπου επιχείρησης παρόχου υπηρεσιών και ως εκ τούτου, καταναλώνουν μακράν το μεγαλύτερο μέρος του TCO.

   Ένα λογισμικό CRM βοηθά τις επιχειρήσεις να διατηρούν επαφή με τους πελάτες, να βελτιστοποιούν τις διαδικασίες και να αυξάνουν τα κέρδη. Το BRM, από την άλλη πλευρά, είναι ένα στρατηγικό επιχειρηματικό λογισμικό που εστιάζει στη διαχείριση λογαριασμών και τη συνεργασία στην εμπορική διαδικασία. Ενσωματώνεται με συστήματα CRM για την παροχή αμφίδρομης ροής σχετικών δεδομένων (push and pull).

   Αυτές είναι συνήθως τεράστιες μονολιθικές εφαρμογές με μεγάλες ποσότητες ιστορικών δεδομένων και επίσης μεγάλο βαθμό εξατομίκευσης.

2. **Υπηρεσίες χρέωσης (Billing)**, συμπεριλαμβανομένης της χρέωσης, της λογιστικής και άλλων χρηματοοικονομικών υπηρεσιών, όπως αξιολόγηση, τιμολόγηση και δημιουργία PDF για πελάτες.
   Σε αυτές συνήθως συμπεριλαμβάνονται τοπικές εφαρμογές που ακολουθούν την τοπική νομοθεσία και ενδέχεται να μην διαθέτουν εκδόσεις που να είναι συμβατές με cloud.

3. **Επιχειρηματική ευφυΐα (Business Intelligence),** συμπεριλαμβανομένων όλων των υπηρεσιών που σχετίζονται με Big Data για τη συλλογή, επεξεργασία και αποθήκευσή τους, καθώς και υπηρεσίες Business Intelligence (BI) για την ανάλυση μεγάλου όγκου δεδομένων με σκοπό την εξαγωγή αποτελεσμάτων που μπορούν να χρησιμοποιηθούν από την επιχείρηση για την βελτιστοποίηση των προϊόντων και υπηρεσιών της. Υπηρεσίες όπως η προγνωστική ανάλυση, το always on marketing και άλλες, ανήκουν επίσης σε αυτήν την κατηγορία.

   Αυτές είναι συνήθως εφαρμογές που μπορούν εύκολα να μεταφερθούν στο cloud, καθώς έχουν αναπτυχθεί πρόσφατα οπότε έχουν εκδόσεις συμβατές με cloud, καθώς και εκδόσεις βελτιστοποιημένες για το cloud.

4. **Υπηρεσία διαχείρισης παραγγελιών,** συμπεριλαμβανομένων όλων των υπηρεσιών front-end και αυτοματισμού για την επεξεργασία παραγγελιών και υπηρεσιών από πελάτες.
   Αυτές είναι τυπικές εφαρμογές που μπορούν να μεταφερθούν απευθείας στο cloud, καθώς δεν έχουν συγκεκριμένες απαιτήσεις.

5. **Κέντρο επικοινωνίας (Contact Center),** συμπεριλαμβανομένων των:
   Computer Telephony Integration (CTI) - λογισμικό που επιτρέπει να έχουμε όλες τις δυνατότητες ενός τηλεφωνικού κέντρου (εγγραφή κλήσης / απόκρισης, δρομολόγηση κλπ.) χωρίς τη χρήση μιας συμβατικής τηλεφωνικής συσκευής. Ενσωματώνεται επίσης με λογισμικό που σχετίζεται με τους πελάτες, παρέχοντας συγκεκριμένα δεδομένα και ιστορικό πελατών απευθείας στο σταθμό εργασίας του αντιπροσώπου στην αρχή της εισερχόμενης κλήσης, προκειμένου να μεγιστοποιηθεί η αποδοτικότητα και να βελτιωθεί η ποιότητα υπηρεσίας.

   Intelligent Workload Distribution(IWD) - μια λύση που δημιουργεί μια λίστα εργασιών με κεντρική διαχείριση και προτεραιότητα. Επιτρέπει την παρουσίαση της εργασίας στον κατάλληλο πόρο την κατάλληλη ώρα και τοποθεσία. Συλλέγει εργασίες σε πραγματικό χρόνο από συστήματα πολλαπλών πηγών, δίνει προτεραιότητα ή επαναπροσδιορίζει τις εργασίες βάσει επιχειρηματικών κανόνων και, στη συνέχεια, κατανέμει τις εργασίες στον πιο κατάλληλο πόρο / πράκτορα, μειώνοντας τον χρόνο και το κόστος.

Interactive Voice Response (IVR) - μια τεχνολογία τηλεφωνικού συστήματος που επιτρέπει στους εισερχόμενους καλούντες να έχουν πρόσβαση σε πληροφορίες μέσω ενός συστήματος φωνητικής απόκρισης προ-εγγεγραμμένων μηνυμάτων χωρίς να μιλούν με ένα φυσικό πρόσωπο.

Αυτό περιλαμβάνει επίσης άλλα συστήματα, όπως συστήματα έξυπνης συνομιλίας κειμένου και ενσωματώσεις κοινωνικών μέσων.

Ιδιαίτερη προσοχή πρέπει να δοθεί σε αυτές τις υπηρεσίες, λόγω της ευαισθησίας τους στη καθυστέρηση. Στις περισσότερες περιπτώσεις, τα 250ms είναι η μέγιστη αποδεκτή καθυστέρηση, ενώ η ITU-T G.114 συνιστά μέγιστη καθυστέρηση 150ms. Για την περίπτωσή μας, τα κέντρα δεδομένων των δημόσιων παρόχων cloud στη Γερμανία, την Ιταλία και τη Γαλλία παρέχουν αρκετά χαμηλό χρόνο καθυστέρησης, όπως μπορούμε να δούμε στον παρακάτω πίνακα.

AWS	
Περιγραφή περιοχής	**Καθυστέρηση (ms)**
Μιλάνο	57
Φρανκφούρτη	62
Παρίσι	67
Λονδίνο	78
Στοκχόλμη	83
Ιρλανδία	83
AZURE	
Περιγραφή περιοχής	**Καθυστέρηση (ms)**
Παρίσι	81
Φρανκφούρτη	85
Ζυρίχη	87
Λονδίνο	93
Κάρντιφ	94
Ιρλανδία	97
Ολλανδία	101
Νορβηγία	112
GCP	
Περιγραφή περιοχής	**Καθυστέρηση (ms)**
Φρανκφούρτη	62
Eemshaven (NL)	71
St. Ghislain (BE)	73
Λονδίνο	77

6. **Υποστήριξη Backend,** συμπεριλαμβανομένων των υπηρεσιών ασφαλείας, διαχείρισης ταυτότητας και πρόσβασης, εσωτερικές πύλες πληροφοριών και συνεργασίας, υπηρεσίες ανθρώπινου δυναμικού και νομικές υπηρεσίες.

   Μερικές από αυτές (για παράδειγμα, το IAM και υπηρεσίες Ασφαλείας) πρέπει να παραμείνουν εν μέρει εντός του χώρου, καθώς χρησιμοποιούνται επίσης από τις υπηρεσίες δικτύου. Για αυτές τις υπηρεσίες, θα ακολουθηθεί μια υβριδική προσέγγιση cloud. Οι υπόλοιπες υπηρεσίες μπορούν να μεταφερθούν με ασφάλεια, καθώς δεν υπάρχει εξάρτηση από την καθυστέρηση (latency) ή την εγγύτητα.

7. **Εφαρμογές Frontend,** συμπεριλαμβανομένων των πυλών που εισέρχονται οι πελάτες (ιστότοπος της εταιρείας, πύλη για νέους) και πύλες καμπάνιας.

   Αυτοί είναι ιδανικοί υποψήφιοι για τη μεταφορά στο cloud, καθώς τέτοιες εφαρμογές ιστού είναι ανεκτικές στη καθυστέρηση και θα επωφεληθούν σε μεγάλο βαθμό από υπηρεσίες cloud όπως το Content Delivery Network (CDN) ή την αυτόματη κλιμάκωση (auto scaling) κλπ.

8. **Υπηρεσίες Middleware,** συμπεριλαμβανομένου του Enterprise service bus (ESB) και οποιουδήποτε API / File abstraction layer.

   Οι εφαρμογές που εξαρτώνται από αυτές τις υπηρεσίες Middleware, θα μεταφερθούν στο cloud, αλλά ορισμένες μπορεί επίσης να παραμείνουν υπό προυποθέσεις, όπως οι υπηρεσίες δικτύου. Θα ακολουθήσουμε μια υβριδική προσέγγιση για αυτές.

9. **Σύστημα Διαχείρισης Πόρων (ERP),** συμπεριλαμβανομένης οποιασδήποτε εφαρμογής Enterprise Resource Planning όπως το SAP.

   Συνήθως αυτές είναι επίσης μεγάλες μονολιθικές εφαρμογές, αλλά με λιγότερη εξατομίκευση από τις CRM και BRM εφαρμογές , οπότε θα είναι πολύ πιο εύκολη η μετεγκατάσταση τους. Επιπλέον, όλοι οι μεγάλοι πάροχοι ERP έχουν ήδη λύσεις φιλοξενούμενες στο cloud που μπορούν να καταναλωθούν «ως υπηρεσία» (SaaS).

10. **Υπηρεσίες διαχείρισης συμβάντων (Incident Management),** συμπεριλαμβανομένων όλων των εσωτερικών συστημάτων έκδοσης εισιτηρίων.

    Αυτές είναι τυπικές εφαρμογές που μπορούν να μεταφερθούν απευθείας στο cloud, καθώς δεν έχουν συγκεκριμένες απαιτήσεις.

11. **Διαχείριση περιεχομένου (Content Management),** συμπεριλαμβανομένων όλων των συστημάτων διαχείρισης περιεχομένου που καταναλώνουν οι πελάτες, όπως Internet TV.

    Αυτές οι υπηρεσίες, εξαρτώνται σε μεγάλο βαθμό από το εύρος ζώνης, οπότε πρέπει να παραμείνουν στο χώρο της εταιρίας, τουλάχιστον εν μέρει. Το ίδιο το περιεχόμενο θα πρέπει να παραμείνει εντός του χώρου, κοντά στο τοπικό δίκτυο, για να αξιοποιήσει τη χωρητικότητα και να παρέχει περιεχόμενο τοπικά και δωρεάν, ενώ το σύστημα ελέγχου μπορεί να μεταφερθεί στο cloud ώστε να δημιουργηθεί μια υβριδική εγκατάσταση.

Σχεδιασμός Μελλοντικής Κατάστασης «Το Be»

Υπηρεσίες Cloud

Ο όρος "υπηρεσίες cloud" αναφέρεται σε ένα ευρύ φάσμα υπηρεσιών κατ' απαίτηση (on-demand) που παρέχονται μέσω του διαδικτύου σε επιχειρήσεις και καταναλωτές. Αυτές οι υπηρεσίες προορίζονται να παρέχουν απλή, οικονομική και αποδοτική πρόσβαση σε εφαρμογές και πόρους, χωρίς να απαιτείται εσωτερική τεχνολογική υποδομή.

Οι προμηθευτές και οι πάροχοι υπηρεσιών cloud διαχειρίζονται όλες τις πτυχές των υπηρεσιών cloud. Διατίθενται στους πελάτες μέσω των διακομιστών (server) των παρόχων, γεγονός που εξαλείφει την ανάγκη μιας επιχείρησης να φιλοξενεί τις εφαρμογές εντός του χώρου της.

Το cloud computing έχει αλλάξει ριζικά τον τρόπο λειτουργίας των επιχειρήσεων, παρέχοντας σημαντικά οφέλη σε σχέση με τις "παραδοσιακές" επιλογές των ιδιόκτητων κέντρων δεδομένων, όπως:

1. **Ευελιξία** : Η ικανότητα πρόσβασης σε ένα διαφορετικό σύνολο τεχνολογιών που μας επιτρέπει να καινοτομούμε πιο γρήγορα και να χτίζουμε σχεδόν οτιδήποτε μπορούμε να φανταστούμε. Μπορούμε να παρέχουμε γρήγορα πόρους όπως απαιτείται - από υπηρεσίες υποδομής όπως υπολογιστική ισχύ, αποθήκευση πληροφοριών και βάσεις δεδομένων έως το Internet of Things, το Machine Learning, τα data lakes και τα analytics, μεταξύ άλλων υπηρεσιών. Η ανάπτυξη οποιουδήποτε τύπου υπηρεσίας ή πόρου πληροφορικής γίνεται μέσα σε λίγα λεπτά.

2. **Ελαστικότητα** : Με το cloud computing, αποφεύγουμε το αρχικό κόστος του υπερβολικού πλήθους πόρων για τη διαχείριση μελλοντικών αιχμών στην επιχειρηματική δραστηριότητα. Αντί για αυτό, παρέχουμε μόνο τους πόρους που απαιτούνται ενώ μπορούμε να αυξήσουμε άμεσα ή να μειώσουμε αυτούς τους πόρους με βάση τις τρέχουσες επιχειρηματικές απαιτήσεις μας, στις περισσότερες περιπτώσεις - μέσα σε λίγα λεπτά.

3. **Εξοικονόμηση κόστους** : Το cloud καθιστά εφικτή την ανταλλαγή κεφαλαιουχικών δαπανών (CAPEX) για λειτουργικά έξοδα (OPEX), εξαλείφοντας το κόστος κατοχής και λειτουργίας κέντρων δεδομένων και φυσικών διακομιστών και πληρώνοντας μόνο για τους πόρους που καταναλώνουμε. Επιπλέον, το μεταβλητό κόστος είναι σημαντικά χαμηλότερο από αυτό που θα πληρώναμε για να το κάνουμε μόνοι μας, λόγω των οικονομιών κλίμακας.

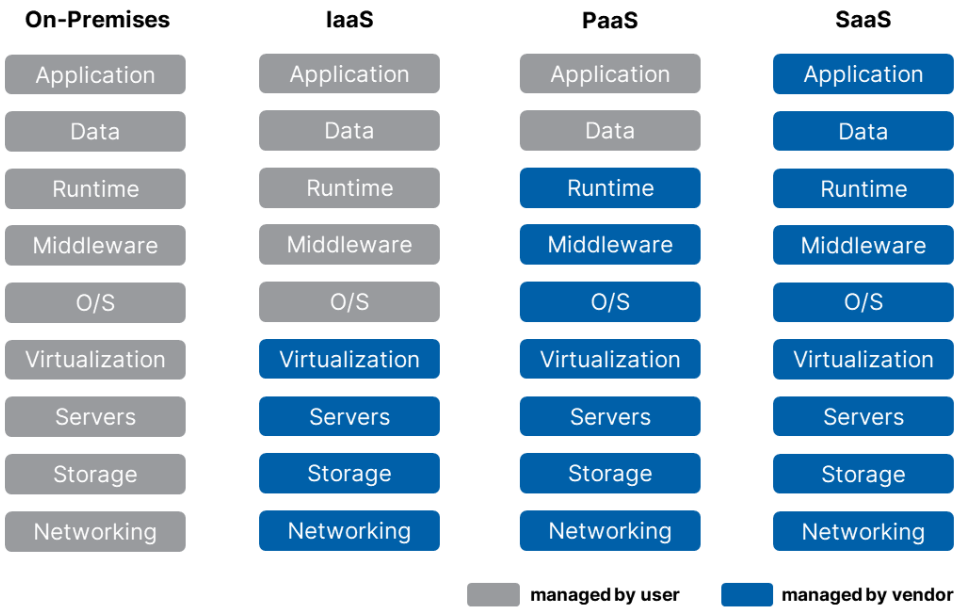
4. **Ανάπτυξη εφαρμογών σε παγκόσμια κλίμακα**: Σε λίγα λεπτά, οι εφαρμογές μας μπορούν να αναπτυχθούν παγκοσμίως σε μια νέα γεωγραφική περιοχή. Με την ανάπτυξη των υπηρεσιών μας σε πολλές φυσικές τοποθεσίες, είμαστε σε θέση να διατηρούμε μια εγγύτητα με τους τελικούς χρήστες (πελάτες), μειώνοντας έτσι τον χρόνο καθυστέρησης και βελτιώνοντας την συνολική εμπειρία του πελάτη.

Υπάρχουν τρεις βασικοί τύποι υπολογιστικού νέφους και ο καθένας προσφέρει ένα διαφορετικό επίπεδο ελέγχου, ευελιξίας και διαχείρισης για την κάλυψη των μοναδικών αναγκών κάθε ενδιαφερομένου.

- **Υποδομή ως υπηρεσία (IaaS)** : Το IaaS παρέχει τα θεμελιώδη δομικά στοιχεία για το cloud computing. Παρέχει συνήθως πρόσβαση σε δυνατότητες δικτύωσης, υπολογιστές (εικονικούς ή αποκλειστικούς) και χώρο αποθήκευσης δεδομένων. Το IaaS μας επιτρέπει να έχουμε τον μεγαλύτερο βαθμό ελέγχου και ευελιξίας έναντι των πόρων πληροφορικής μας.

- **Πλατφόρμα ως υπηρεσία (PaaS): Το** PaaS εξαλείφει την ανάγκη διαχείρισης της υποκείμενης υποδομής (συνήθως υλικού και λειτουργικών συστημάτων) και παρέχει περισσότερο χρόνο για να επικεντρωθούμε στην ανάπτυξη και διαχείριση εφαρμογών. Αυτό αυξάνει την αποδοτικότητά μας επειδή δεν είμαστε πλέον υπεύθυνοι για την απόκτηση πόρων, τον προγραμματισμό χωρητικότητας, τη συντήρηση λογισμικού, την ενημέρωση κώδικα ή οποιαδήποτε άλλη χρονοβόρα δραστηριότητα που σχετίζεται με τη λειτουργία των εφαρμογών. Αυτός ο τύπος υπηρεσίας είναι ιδανικός για προγραμματιστές που θέλουν να δημιουργήσουν, να δοκιμάσουν και να αναπτύξουν τον κώδικά τους χωρίς την ανάγκη παροχής και διαχείρισης της υποκείμενης υποδομής.

- **Λογισμικό ως υπηρεσία (SaaS): Το** SaaS είναι ένα πλήρως λειτουργικό προϊόν που διαχειρίζεται και υποστηρίζεται από τον πάροχο cloud υπηρεσιών. Συνήθως, όταν οι χρήστες αναφέρονται στο SaaS, αναφέρονται σε εφαρμογές τελικού χρήστη (όπως ηλεκτρονικό ταχυδρομείο μέσω διαδικτύου). Με την προσφορά SaaS, οι πελάτες δεν είναι υπεύθυνοι για τη συντήρηση ή τη διαχείριση της υποκείμενης υποδομής της υπηρεσίας, απλώς πρέπει να εξετάσουν πώς θα χρησιμοποιήσουν το λογισμικό. Αυτός ο τύπος είναι ο πιο αναγνωρισμένος τύπος υπηρεσίας cloud με μια ποικιλία υπηρεσιών, όπως αποθήκευση αρχείων, δημιουργία αντιγράφων ασφαλείας, ηλεκτρονικά μηνύματα ηλεκτρονικού ταχυδρομείου και εργαλεία διαχείρισης έργου.



Εικόνα 1 : On-Prem, IaaS, PaaS, SaaS (PentaSecurity)

Cloud Native

Ζούμε σε μια εποχή ραγδαίας τεχνολογικής ανάπτυξης, στην οποία οι εφαρμογές έχουν γίνει όλο και πιο περίπλοκες ενώ παράλληλα, οι χρήστες έχουν αυξήσει τις προσδοκίες και τις απαιτήσεις τους. Οι χρήστες αναμένουν άμεση απόκριση, προηγμένες δυνατότητες και μηδενικό χρόνο διακοπής. Τα ζητήματα απόδοσης, τα επαναλαμβανόμενα σφάλματα και η αδυναμία γρήγορης μετακίνησης δεν είναι πλέον αποδεκτά. Αυτές οι απαιτήσεις οδήγησαν στην ανάπτυξη και χρήση τεχνολογιών και μεθοδολογιών που προσφέρουν, ευελιξία, αξιοπιστία, αυξημένη ικανοποίηση των πελατών και αποδοτικότητα κόστους. (RedHat, 2020) (RackSpace, 2020) (SDX Central, 2016)

Το cloud-native αναφέρεται στην προσέγγιση που χρησιμοποιείται για τη δημιουργία εφαρμογών που βασίζονται στο cloud. Οι εφαρμογές και οι υπηρεσίες που είναι Cloud Native διαφέρουν από τους παλαιούς ομολόγους τους, επειδή είναι ειδικά σχεδιασμένες για το cloud από την αρχή. Μπορούν να αναπτυχθούν και να επισκευαστούν γρηγορότερα, να έχουν πιο ρευστή αρχιτεκτονική όπως επίσης να τοποθετηθούν και να μετακινηθούν εύκολα σε διάφορα περιβάλλοντα.

Οι τεχνολογίες Cloud Native επιτρέπουν στις επιχειρήσεις να αναπτύξουν και να εκτελέσουν επεκτάσιμες εφαρμογές σε μοντέρνα, δυναμικά περιβάλλοντα όπως δημόσια, ιδιωτικά και υβριδικά clouds. Αυτές οι τεχνολογίες δίνουν τη δυνατότητα στους μηχανικούς να κάνουν αυτοματοποιημένες αλλαγές υψηλής επίπτωσης συχνά και προβλέψιμα, με ελάχιστη προσπάθεια.

Οι εφαρμογές και οι υπηρεσίες που πληρούν τις προαναφερθείσες απαιτήσεις ονομάζονται Cloud-Native.

Οι cloud native εφαρμογές έχει μερικά κοινά στοιχεία όπως:

- **Containerized:** Τα containers είναι μία λογική συσκευασία λογισμικού που περιέχει όλα τα απαραίτητα στοιχεία για την εκτέλεση μιας εφαρμογής, όπως κώδικα εφαρμογής, χρόνο εκτέλεσης, αρχεία διαμόρφωσης και εξαρτήσεις μεταξύ τους.

- **Microservices:** Πρόκειται για μια αρχιτεκτονική προσέγγιση για την ανάπτυξη και παροχή εφαρμογών. Πριν από αυτήν την τεχνολογία, οι εφαρμογές ή τα συστήματα αποτελούσαν ένα ενιαίο κομμάτι χρησιμοποιώντας μια βάση κώδικα για όλες τις λειτουργίες και τις υπηρεσίες τους. Αυτή η προσέγγιση ονομάζεται μονολιθική αρχιτεκτονική και αυτές οι εφαρμογές έχουν ένα μεγάλο μειονέκτημα όσον αφορά την ανάπτυξη κώδικα. Με αποτέλεσμα σε περίπτωση αλλαγών, ολόκληρα τα συστήματα / υπηρεσίες να πρέπει να επαναληφθούν από το μηδέν, οδηγώντας σε μη διαθεσιμότητα υπηρεσιών, επιπλέον προσπάθεια και μεγάλες καθυστερήσεις. Τα microservices αναπτύσσονται σε containers και εφαρμόζουν τη μέθοδο "διαίρει και βασίλευε" σε ολόκληρη την εφαρμογή, δηλαδή απομονώνουν τη κάθε λειτουργία μίας εφαρμογής ως μία υπηρεσία, μέσα στο δικό της container ξεχωριστά. Στη συνέχεια τα containers συνδέονται μεταξύ τους μέσω APIs. Αυτή η μέθοδος προσφέρει χαλαρά συνδεδεμένες εφαρμογές που επιτρέπουν στους μηχανικούς να δημιουργήσουν και να αναπτύξουν τον κώδικά τους χωρίς τον κίνδυνο να προκαλέσουν ζημιά ή να κλείσουν ολόκληρη την εφαρμογή.

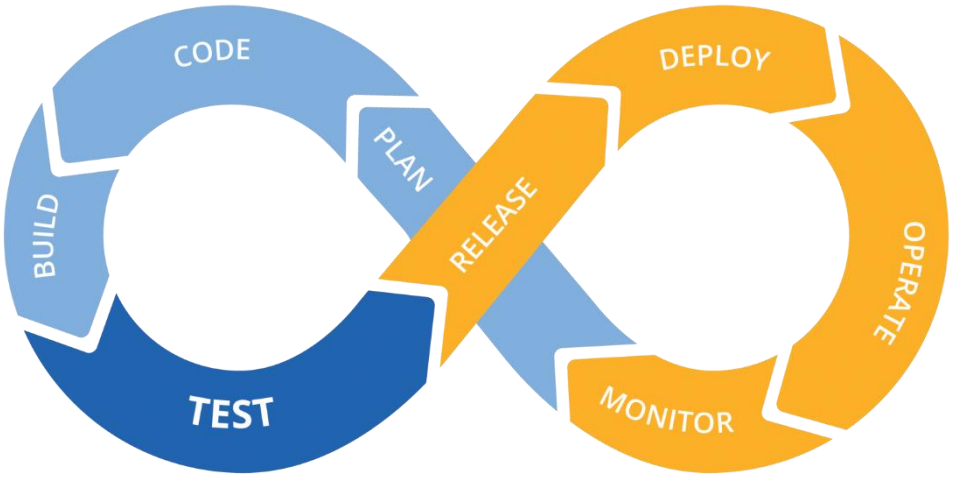
- **Υπολογισμός χωρίς διακομιστές (Serverless Computing):** Αυτή είναι η τεχνολογία που επιτρέπει στους μηχανικούς και τους προγραμματιστές να εκτελούν τις διαδικασίες και τις λειτουργίες τους χωρίς την ανάγκη παροχής των υποκείμενων πλατφορμών. "Χωρίς διακομιστή", δεν σημαίνει ότι κανένας διακομιστής δεν εμπλέκεται στις διαδικασίες υπολογισμών - σημαίνει ότι δεν υπάρχει πλέον ανάγκη συντήρησης των διακομιστών που απαιτούνται για τη εκτέλεση των λειτουργιών. Όλη η βαριά εργασία της διαχείρισης λειτουργιών πραγματοποιείται εκτός της επιχείρησης. Αυτό οδηγεί σε μεγαλύτερη απόδοση, ασφάλεια και μεγαλύτερη εξοικονόμηση κόστους.

- **Συνεχής ολοκλήρωση / Συνεχής παράδοση:** Ονομάζεται επίσης CI / CD Pipeline είναι μια μεθοδολογία, η οποία απεικονίζεται σε μια μορφή ροής εργασίας που επιτρέπει τη συχνή παράδοση εφαρμογών στους πελάτες εισάγοντας συνεχή αυτοματισμό και παρακολούθηση σε όλα τα στάδια του κύκλου ζωής και ανάπτυξης εφαρμογών - από

τις φάσεις ολοκλήρωσης και δοκιμών έως την ανάπτυξη και παρακολούθηση. Αυτή η προσέγγιση προσφέρει έναν αυτοματοποιημένο, αξιόπιστο και γρηγορότερο τρόπο για τη συνεχή διόρθωση σφαλμάτων, τη δοκιμή και την ανάπτυξη κώδικα, χωρίς τον κίνδυνο να καταστραφεί ή να επηρεαστεί η εφαρμογή.

Για την εφαρμογή και τη διευκόλυνση αυτής της διαδικασίας, απαιτείται συνεργασία και συνδυασμένη προσπάθεια μεταξύ διαφορετικών ομάδων μηχανικών. Τόσο η ανάπτυξη λογισμικού (σχέδιο, κώδικας, κατασκευή, δοκιμή) όσο και οι λειτουργίες πληροφορικής (απελευθέρωση, ανάπτυξη, λειτουργία, παρακολούθηση) συνδυάζονται σε μια ενοποιημένη διαλειτουργική ομάδα που ονομάζεται DevOps

Το DevOps αναφέρεται επίσης ως ένα εργασιακό περιβάλλον, ένα σύνολο πρακτικών και κουλτούρας που πρέπει να έχει αυτή η συνδυασμένη ομάδα, για να επικοινωνεί συνεχώς και να συνεργάζεται με στόχο την γρήγορη και συχνή παροχή υπηρεσιών λογισμικού και υποδομών (infrastructure), μέσω αυτοματοποιημένων και τυποποιημένων διαδικασιών.



Εικόνα 2 : CI / CD Pipeline

Ασφάλεια από τον σχεδιασμό (Security by Design)

Σήμερα, τα δεδομένα μπορούν να επηρεάσουν τα πάντα, από τις τάσεις της αγοράς, τις πολιτικές εκλογές έως την έναρξη ή τον τερματισμό των πολέμων. Για αυτούς τους λόγους, τα δεδομένα έχουν

γίνει πιο πολύτιμα από τον χρυσό και η σωστή χρήση και η ασφάλεια αυτών, δεν ήταν ποτέ πιο σημαντική. Η ανάγκη φυσικής και λογικής ασφάλειας της υποδομής φιλοξενίας, των υπηρεσιών και των δεδομένων των πελατών, αυξάνεται συνεχώς. Οι επιχειρήσεις προσπαθούν να βρουν πιο ασφαλείς τρόπους πρόσβασης στις εφαρμογές τους, να μεταφέρουν και να αποθηκεύσουν τα δεδομένα τους και να προστατεύσουν τα κέντρα δεδομένων (data centers) τους, μειώνοντας παράλληλα το κόστος και τις προσπάθειές τους για παροχή και εποπτεία.

Παρακάτω, αναφέρουμε μερικές από τις βασικές υπηρεσίες και πρακτικές για καλύτερη ασφάλεια της υποδομής, των λογαριασμών και των υπηρεσιών, όχι μόνο για μια εταιρεία TelCo, αλλά και για οποιαδήποτε μεσαία και μεγάλη επιχείρηση.

Μία από τις κύριες υπηρεσίες που αφορούν τον έλεγχο πρόσβασης χρήστη και την παροχή λογαριασμών στις υπηρεσίες και τους πόρους AWS, είναι η υπηρεσία AWS Identity and Access Management (IAM). Το AWS Identity and Access Management (IAM) μας επιτρέπει να ορίζουμε μεμονωμένους λογαριασμούς χρηστών με δικαιώματα σε πόρους AWS.

Για προνομιακούς λογαριασμούς, το AWS Multi-Factor Authentication (MFA) είναι διαθέσιμο, με επιλογές για έλεγχο ταυτότητας βάσει λογισμικού και/ή υλικού. Αυτή η δυνατότητα προσφέρει ένα επιπλέον επίπεδο ελέγχου ταυτότητας του χρήστη, παρέχοντας διακριτικά ασφαλείας ή κωδικούς πρόσβασης μέσω μιας εφαρμογής για κινητά ή μίας σύνδεσης υλικής συσκευής. Το IAM μπορεί επίσης να χρησιμοποιηθεί για να παραχωρήσει ομόσπονδη πρόσβαση(federated access) στην AWS Management Console και AWS API service στους υπαλλήλους και τις εφαρμογές μέσω των υπαρχόντων συστημάτων διαχείρισης ταυτότητας, όπως το Microsoft Active Directory.

Το Identity federation είναι μια σχέση εμπιστοσύνης μεταξύ δύο μερών που επιτρέπει τον έλεγχο ταυτότητας των χρηστών και τη διαβίβαση των πληροφοριών που απαιτούνται για την εξουσιοδότηση της πρόσβασής τους σε πόρους. Το Federation είναι μια ευρέως χρησιμοποιούμενη προσέγγιση για την ανάπτυξη συστημάτων ελέγχου πρόσβασης, που διαχειρίζονται κεντρικά τους χρήστες σε ένα μοναδικό IdP (Identity Provider) και διαχειρίζονται την πρόσβασή τους σε πολλές εφαρμογές και υπηρεσίες που ενεργούν ως Service Provider (SP).

Ένα ισχυρό εργαλείο IAM είναι η υπηρεσία καταλόγου AWS, επίσης γνωστή ως AWS Managed Microsoft Active Directory (AD), η οποία επιτρέπει τη χρήση διαχειριζόμενης υπηρεσίας καταλόγου Active Directory (AD) σε AWS για directory-aware φορτία και πόρους. Το AWS Managed Microsoft AD είναι ενσωματωμένο σε έναν υπάρχοντα Microsoft Active Directory και δεν απαιτεί συγχρονισμό ή αναπαραγωγή δεδομένων από την εσωτερική υπηρεσία Active Directory προς το cloud. Μπορούμε να χρησιμοποιήσουμε τα τυπικά εργαλεία διαχείρισης της υπηρεσίας καταλόγου Active Directory και τις ενσωματωμένες δυνατότητες της, όπως Group Policies και μοναδική σύνδεση (Single Sign On), ενώ ταυτόχρονα μειώνονται τα γενικά έξοδα διατήρησης δεδομένων καταχωρημένων σε πολλούς καταλόγους.

Το Single Sign-On (SSO) είναι μια τεχνολογία που συνδυάζει πολλές οθόνες σύνδεσης εφαρμογών σε μία. Το SSO απαιτεί από έναν χρήστη να εισάγει τα διαπιστευτήρια σύνδεσής του (όνομα χρήστη, κωδικό πρόσβασης κλπ) μία φορά, σε μία σελίδα, προκειμένου να έχει πρόσβαση σε όλες τις SaaS εφαρμογές. Το AWS SSO προσφέρει τη δυνατότητα διαχείρισης της πρόσβασης πολλαπλών λογαριασμών και εφαρμογών AWS, κεντρικά, διαμορφώνοντας και διατηρώντας όλα τα απαραίτητα δικαιώματα για τους λογαριασμούς της εταιρείας αυτόματα, χωρίς πρόσθετη ρύθμιση. Επιπλέον, προσφέρει δικαιώματα χρήστη βάσει ρόλων εργασίας (Role Based Access Control - RBAC) και προσαρμογή αυτών των αδειών όσον αφορά την ευθυγράμμιση με συγκεκριμένες απαιτήσεις ασφαλείας. Ο πιο συνηθισμένος κανόνας και η βέλτιστη πρακτική της δημιουργίας και διαχείρισης δικαιωμάτων σε λογαριασμούς χρηστών ή ομάδων ονομάζεται "Least Privilege" και σημαίνει ότι τα άτομα μπορούν να έχουν δικαιώματα (επίπεδα πρόσβασης και άδειες), τόσα όσα θεωρούνται επαρκή για τον σκοπό ή τον εργασιακό ρόλο τους. (AWS, 2021) (AWS, 2021) (Cloudflare, 2021)

Διαχωρισμένοι λογαριασμοί (Segregated Accounts)

Αντί να αναπαράγετε τη δομή αναφοράς του οργανισμού, συνιστάται η οργάνωση του φόρτου εργασίας σε ξεχωριστούς λογαριασμούς και οι λογαριασμοί αυτοί να ομαδοποιούνται με βάση τη λειτουργία, τις απαιτήσεις συμμόρφωσης ή ένα κοινό σύνολο ελέγχων. Οι λογαριασμοί είναι ένα αυστηρό σύνορο στο AWS και συνιστάται ο διαχωρισμός του φόρτου εργασίας παραγωγής από το φόρτο εργασίας της ανάπτυξης και της δοκιμής. (AWS, 2020)

- **Χρησιμοποίηση λογαριασμών για τμηματοποίηση του φόρτου εργασίας**: Ξεκινάμε έχοντας υπόψη την ασφάλεια και την υποδομή για να επιτρέψουμε στον οργανισμό να καθιερώσει τυποποιημένες διασφαλίσεις καθώς αυξάνεται ο φόρτος εργασίας. Αυτή η προσέγγιση καθορίζει όρια και διατηρεί τον έλεγχο. Συνιστάται ο διαχωρισμός του φόρτου εργασίας σε επίπεδο λογαριασμού για την απομόνωση των περιβαλλόντων παραγωγής από τα περιβάλλοντα ανάπτυξης και δοκιμών, καθώς και για τον καθορισμό ενός ισχυρού λογικού ορίου μεταξύ των φόρτων εργασίας που επεξεργάζονται δεδομένα με διαφορετικούς βαθμούς ευαισθησίας, όπως ορίζονται από τις απαιτήσεις εξωτερικής συμμόρφωσης (όπως PCI- DSS ή GDPR) από τα φορτία εργασίας που δεν το κάνουν.

- **Ασφαλίζουμε όλους τους λογαριασμούς AWS** : Η προστασία ενός λογαριασμού AWS συνεπάγεται ορισμένα βήματα, συμπεριλαμβανομένης της ασφάλειας και της μη χρήσης του root χρήστη, καθώς και τη διατήρηση των τρεχόντων στοιχείων επικοινωνίας. Το AWS Organizations μας επιτρέπει να διαχειριζόμαστε κεντρικά τους λογαριασμούς καθώς ο φόρτος εργασίας αυξάνεται και κλιμακώνεται στο AWS. Το AWS Organizations επιτρέπει

στους χρήστες να διαχειρίζονται λογαριασμούς, να διαμορφώνουν ρυθμίσεις ασφαλείας και να διαμορφώνουν υπηρεσίες σε πολλούς λογαριασμούς.

- **Κεντρική διαχείριση λογαριασμών** : Το AWS Organizations αυτοματοποιεί τη δημιουργία και τη διαχείριση λογαριασμών AWS, καθώς και τον έλεγχο αυτών των λογαριασμών μόλις δημιουργηθούν. Μας δίνει τη δυνατότητα να οργανώσουμε λογαριασμούς σε οργανωτικές μονάδες (Organizational Units), όπως θα συζητηθεί σε μια μεταγενέστερη ενότητα, οι οποίες μπορεί να αντιπροσωπεύουν διαφορετικά περιβάλλοντα ανάλογα με τις απαιτήσεις και τον σκοπό του φόρτου εργασίας.

- **Κεντρικός έλεγχος** : Ελέγχουμε τι μπορούν να κάνουν οι λογαριασμοί AWS περιορίζοντας την πρόσβαση σε συγκεκριμένες υπηρεσίες, περιοχές και ενέργειες υπηρεσιών. Το AWS Organizations μας δίνει τη δυνατότητα να εφαρμόζουμε προστατευτικά φράγματα σε επίπεδο οργανισμού, μονάδας οργάνωσης ή λογαριασμού που ισχύουν για όλους τους χρήστες και τους ρόλους διαχείρισης ταυτότητας και πρόσβασης AWS (IAM) μέσω πολιτικών ελέγχου υπηρεσιών (Service Control Policies - SCP). Παραδείγματος χάρη, μπορούμε να εφαρμόσουμε ένα SCP που εμποδίζει τους χρήστες να ξεκινήσουν πόρους σε περιοχές που δεν τους έχει επιτραπεί ρητά. Το AWS Control Tower απλοποιεί τη διαδικασία δημιουργίας και διαχείρισης πολλαπλών λογαριασμών. Συγχρόνως αυτοματοποιεί τη δημιουργία λογαριασμών στο AWS Organization ενώ ταυτόχρονα αυτοματοποιεί την παροχή υπηρεσιών, εφαρμόζει προστατευτικά κιγκλιδώματα (συμπεριλαμβανομένης της πρόληψης και της ανίχνευσης) και παρέχει έναν πίνακα ελέγχου για προβολή.

- **Κεντρική διαχείριση υπηρεσιών και πόρων** : Το AWS Organizations μας επιτρέπει να διαχειριζόμαστε κεντρικά υπηρεσίες AWS που ισχύουν για όλους τους λογαριασμούς μας. Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε το AWS CloudTrail για να καταγράψουμε κεντρικά όλες τις ενέργειες που εκτελούνται σε ολόκληρο τον οργανισμό μας και να αποτρέψουμε την απενεργοποίηση της καταγραφής από τους λογαριασμούς μελών. Επιπλέον, μπορούμε να συγκεντρώσουμε κεντρικά δεδομένα για κανόνες που ορίζονται με το AWS Config, το οποίο μας επιτρέπει να ελέγξουμε το φόρτο εργασίας για συμμόρφωση (Compliance) και να ανταποκριθούμε γρηγορότερα σε αλλαγές. Η αυτοματοποίηση υποδομής ως κώδικας (Infrastructure as Code - IaC) μας δίνει τη δυνατότητα να παρέχουμε έναν νέο λογαριασμό αυτόματα για να ικανοποιήσουμε τις απαιτήσεις ασφαλείας μας.

Ασφαλές δίκτυο

Οι χρήστες, εσωτερικοί και εξωτερικοί του οργανισμού μας, μπορούν να βρίσκονται οπουδήποτε. Πρέπει να εγκαταλείψουμε τα παραδοσιακά μοντέλα εμπιστοσύνης σε οποιονδήποτε και σε οτιδήποτε με πρόσβαση στο δίκτυο. Όταν τηρούμε την αρχή της ασφάλειας σε όλα τα επίπεδα, εφαρμόζουμε μια στρατηγική Zero Trust. Η ασφάλεια Zero Trust είναι ένα μοντέλο στο οποίο τα στοιχεία της εφαρμογής ή τα microservices θεωρούνται διακριτά το ένα από το άλλο και κανένα στοιχείο δεν εμπιστεύεται κανένα άλλο στοιχείο. (AWS, 2020)

Οι βέλτιστες πρακτικές του Cloud Security περιλαμβάνουν:

- **Δημιουργία πολλαπλών επιπέδων δικτύου** : Για τμηματοποίηση στοιχείων όπως υπολογιστικά συστήματα, συμπλέγματα βάσεων δεδομένων και λειτουργίες που μοιράζονται απαιτήσεις προσβασιμότητας, μπορούν να χρησιμοποιηθούν υποδίκτυα (subnets) για τη δημιουργία λογικών επιπέδων. Για παράδειγμα, ένα σύμπλεγμα βάσεων δεδομένων που δεν απαιτεί πρόσβαση στο Διαδίκτυο θα πρέπει να τοποθετηθεί σε υποδίκτυα που δεν έχουν καμία διαδρομή προς ή από το Διαδίκτυο. Αυτά τα υποδίκτυα θα πρέπει επίσης να διαμορφωθούν με γνώμονα την αρχή σχεδιασμού «ελάχιστης πρόσβασης», που σημαίνει ότι η επικοινωνία μεταξύ υποδικτύων και περιουσιακών στοιχείων θα πρέπει να επιτρέπεται κατά περίπτωση και κατευθυντικά, ώστε να μην επιτρέπεται η πρόσβαση σε επίπεδο υποδικτύου - και επομένως ένας εισβολέας που έχει αποκτήσει πρόσβαση σε ένα από αυτά, να θέτει σε κίνδυνο κάποιο άλλο. Αυτή η πολυεπίπεδη προσέγγιση στη διαμόρφωση ελέγχου πρόσβασης δικτύου μετριάζει τον αντίκτυπο μιας εσφαλμένης διαμόρφωσης ενός επιπέδου που θα επέτρεπε την ακούσια ή κακόβουλη πρόσβαση.

- **Έλεγχος κυκλοφορίας σε όλα τα επίπεδα** : Κατά το σχεδιασμό της τοπολογίας του δικτύου μας, πρέπει να λάβουμε υπόψη τις απαιτήσεις συνδεσιμότητας κάθε στοιχείου. Για παράδειγμα, εάν ένα στοιχείο απαιτεί σύνδεση στο Διαδίκτυο (τόσο εισερχόμενα όσο και εξερχόμενα), ή συνδεσιμότητα σε εικονικά ιδιωτικά clouds (Virtual Private Network - VPC), υπηρεσίες αιχμής (edge services) ή εξωτερικά κέντρα δεδομένων, πολλαπλοί έλεγχοι θα πρέπει να εφαρμόζονται με εις βάθος αμυντικό τρόπο τόσο στην εισερχόμενη όσο και στην εξερχόμενη κίνηση, συμπεριλαμβανομένης της χρήσης ομάδων ασφαλείας (stateful inspection firewalls) σε ένα σύστημα, λίστες ελέγχου πρόσβασης δικτύου (Access Control List - ACLs) σε επίπεδο υποδικτύου, υποδίκτυα, και πίνακες διαδρομής. Κάθε υποδίκτυο μπορεί να έχει τον δικό του πίνακα διαδρομών, ο οποίος καθορίζει τους κανόνες δρομολόγησης για τη διαχείριση των διαδρομών που λαμβάνονται από την κίνηση εντός του υποδικτύου.

  Όταν δημιουργείται μια δομή , μια βάση δεδομένων ή άλλη υπηρεσία, κάθε διεπαφή δικτύου έχει τη δική της ομάδα ασφαλείας. Αυτό το τείχος προστασίας λειτουργεί σε ξεχωριστό επίπεδο από το λειτουργικό σύστημα και μπορεί να χρησιμοποιηθεί για τον καθορισμό

κανόνων για την επιτρεπόμενη κυκλοφορία εισερχόμενων και εξερχόμενων. Επιπλέον, μπορούν να καθοριστούν σχέσεις μεταξύ ομάδων ασφαλείας. Για παράδειγμα, συστήματα σε μια ομάδα ασφαλείας επιπέδου βάσης δεδομένων δέχονται επισκεψιμότητα μόνο από συστήματα εντός του επιπέδου εφαρμογής, με βάση τις ομάδες ασφαλείας που έχουν εκχωρηθεί στις αντίστοιχες δομές. Επίσης, ένα υποδίκτυο μπορεί να συσχετιστεί με ένα δικτυακό (Access Control List - ACL), το οποίο λειτουργεί ως stateless τείχος προστασίας. Το δικτυακό ACL πρέπει να διαμορφωθεί έτσι ώστε να περιορίζει τον τύπο κυκλοφορίας που επιτρέπεται μεταξύ των επιπέδων.

Η συνδεσιμότητα με το Διαδίκτυο θα πρέπει επίσης να ελέγχεται, σε καμία περίπτωση δεν επιτρέπεται η ελεύθερη πρόσβαση. Η απαίτηση θα πρέπει να τεκμηριωθεί, να είναι στη λίστα επιτρεπόμενων χρησιμοποιώντας έναν εξερχόμενο διακομιστή μεσολάβησης (proxy server), ο οποίος θα πρέπει να είναι το μόνο στοιχείο που μπορεί να έχει άμεση πρόσβαση στο Διαδίκτυο και η συνδεσιμότητα πρέπει να καταγράφεται. Οποιοδήποτε μη εξουσιοδοτημένο αίτημα πρέπει να εγείρει ειδοποίηση ασφαλείας και να διερευνάται.

- **Διεξαγωγή επιθεωρήσεων και παροχή προστασίας** : Σε κάθε επίπεδο, πρέπει να επιθεωρούμε και να φιλτράρουμε την κίνηση. Ένα τείχος προστασίας εφαρμογών ιστού (Web Application Firewall - WAF) μπορεί να βοηθήσει στην προστασία συστημάτων που επικοινωνούν μέσω πρωτοκόλλων που βασίζονται σε HTTP από κοινές επιθέσεις. Με το φιλτράρισμα και την παρακολούθηση της κίνησης HTTP μεταξύ μιας εφαρμογής ιστού και του Διαδικτύου, ένα WAF βοηθά στην προστασία των εφαρμογών ιστού. Προστατεύει συνήθως εφαρμογές ιστού από πλαστογράφηση μεταξύ ιστοτόπων, δέσμες ενεργειών μεταξύ ιστοτόπων (XSS), συμπερίληψη αρχείων και έγχυση SQL (SQL Injection), μεταξύ άλλων επιθέσεων. Το WAF είναι ένα πρωτόκολλο layer 7 (στο μοντέλο Open Systems Interconnection), το οποίο δεν προορίζεται να προστατεύσει από όλους τους τύπους επιθέσεων. Αυτή η μέθοδος μετριασμού επίθεσης χρησιμοποιείται συνήθως σε συνδυασμό με μια σειρά εργαλείων (Layer 4 Firewalls, Intrusion Prevention Systems) που μαζί παρέχουν μια ολοκληρωμένη άμυνα ενάντια σε μια ποικιλία φορέων επίθεσης.

  Όταν ένα WAF αναπτύσσεται μπροστά από μια εφαρμογή ιστού, δημιουργεί ένα εμπόδιο μεταξύ της εφαρμογής ιστού και του Διαδικτύου. Ενώ ένας διακομιστής μεσολάβησης προστατεύει την ταυτότητα ενός υπολογιστή-πελάτη (End Client) μέσω της χρήσης ενός ενδιάμεσου, ένα WAF ενεργεί ως αντίστροφος διακομιστής μεσολάβησης, προστατεύοντας τον διακομιστή από την έκθεση απαιτώντας από τους πελάτες να περάσουν από το WAF πριν φτάσουν στον διακομιστή.

  Το WAF διέπεται από ένα σύνολο κανόνων, που συχνά αναφέρονται ως πολιτικές (policies). Αυτές οι πολιτικές έχουν σχεδιαστεί για την προστασία αδυναμιών των εφαρμογών

φιλτράροντας την κακόβουλη επισκεψιμότητα. Η αξία ενός WAF οφείλεται εν μέρει στην ταχύτητα και την ευκολία με την οποία μπορούν να τροποποιηθούν οι πολιτικές, επιτρέποντας ταχύτερη απόκριση σε διάφορους φορείς επίθεσης. Για παράδειγμα, ο περιορισμός της κίνησης μπορεί να εφαρμοστεί γρήγορα κατά τη διάρκεια μίας επίθεσης άρνησης υπηρεσιών (Distributed Denial of Service - DDoS) τροποποιώντας τις πολιτικές WAF.

Ένα σύστημα πρόληψης εισβολών (Intrusion Prevention System - IPS) είναι ένας τύπος ασφάλειας δικτύου που παρακολουθεί και αποτρέπει την εμφάνιση εντοπισμένων απειλών. Τα συστήματα πρόληψης εισβολής παρακολουθούν συνεχώς το δίκτυο, αναζητώντας και καταγράφοντας πιθανά κακόβουλα συμβάντα. Το IPS ειδοποιεί τους διαχειριστές συστήματος για αυτά τα συμβάντα και λαμβάνει προληπτικά μέτρα, όπως το κλείσιμο σημείων πρόσβασης και τη διαμόρφωση τείχους προστασίας, για την αποτροπή μελλοντικών επιθέσεων. Επιπλέον, οι λύσεις IPS μπορούν να χρησιμοποιηθούν για τον εντοπισμό παραβιάσεων των εταιρικών πολιτικών ασφαλείας, αποτρέποντας τους υπαλλήλους και τους επισκέπτες του δικτύου να παραβιάζουν τους κανόνες που περιέχονται σε αυτές τις πολιτικές.

- **Αυτοματοποίηση προστασίας δικτύου** : Στο cloud οι προηγούμενες υπηρεσίες μπορούν να συνεργαστούν για να προειδοποιήσουν και να μετριάσουν τους κινδύνους ασφαλείας χωρίς καθυστερήσεις και ανθρώπινη παρέμβαση και συνήθως προσφέρονται ως υπηρεσία. Χρησιμοποιώντας πληροφορίες για την απειλή και ανίχνευση ανωμαλιών, τα εργαλεία ανίχνευσης και πρόληψης εισβολών μπορούν να προσαρμοστούν και να μετριάσουν τον αντίκτυπο των τρεχουσών απειλών. Το τείχος προστασίας εφαρμογών ιστού είναι ένα παράδειγμα του πώς μπορούμε να αυτοματοποιήσουμε την προστασία του δικτύου, χρησιμοποιώντας τη λύση AWS WAF Security Automations για να αποκλείσουμε αυτόματα αιτήματα που προέρχονται από γνωστές διευθύνσεις IP ενός κακόβουλου παράγοντα.

Για επικοινωνία μεταξύ λογαριασμών, συνιστάται η χρήση επιπέδου απομόνωσης, όπως το AWS Private Link. Πρόκειται για μια υπηρεσία που δημιουργεί μια σύνδεση μεταξύ ενός εξισορροπητή φόρτωσης δικτύου (Network Load Balancer - NLB) και ενός τελικού σημείου VPC (VPC endpoint), χρησιμοποιώντας το δίκτυο κορμού του παρόχου cloud, το οποίο είναι κρυφό και απρόσιτο σε όλους τους καταναλωτές. Αυτό μας επιτρέπει να διατηρούμε όλους τους λογαριασμούς ανεξάρτητους από το δίκτυο, χωρίς να χρειάζεται να έχουμε αλληλεπικαλυπτόμενες περιοχές IP σε όλους τους λογαριασμούς ή να διατηρούμε ένα σύστημα διαχείρισης IP ενώ ταυτόχρονα δεν χρειάζεται προσεκτική σχεδίαση της δρομολόγησης. Το τελικό σημείο VPC βρίσκεται και απευθύνεται στο δίκτυο του συστήματος προέλευσης, ενώ το NLB κατοικεί και απευθύνεται στο δίκτυο του συστήματος προορισμού. Αυτό παρέχει, εκτός από την ανεξαρτησία του δικτύου, ένα επιπλέον

επίπεδο ασφάλειας, καθώς η επικοινωνία πρέπει να καθιερωθεί ρητά χρησιμοποιώντας προκαθορισμένα πρότυπα IaC.

Προστασία δεδομένων / GDPR

Τον Απρίλιο του 2016, ένας νέος ευρωπαϊκός νόμος περί απορρήτου γνωστός με τη συντομογραφία «GDPR» (General Data Protection Regulation) θεσπίστηκε, προκειμένου να επιβάλει και να εναρμονίσει ορισμένους νόμους περί προστασίας δεδομένων σε ολόκληρη την Ευρωπαϊκή Ένωση και τα κράτη μέλη της. Ο GDPR ισχύει για όλους τους οργανισμούς που επεξεργάζονται δεδομένα προσωπικού χαρακτήρα στην Ευρωπαικη Ένωση (ΕΕ), είτε έχουν εγκατάσταση στην ΕΕ είτε επεξεργάζονται προσωπικά δεδομένα κατοίκων της ΕΕ ή όταν προσφέρουν αγαθά ή υπηρεσίες σε άτομα στην ΕΕ ή όταν παρακολουθούν τη συμπεριφορά τους στην ΕΕ. Οποιαδήποτε πληροφορία σχετίζεται με ένα ταυτοποιημένο ή αναγνωρίσιμο φυσικό πρόσωπο θεωρείται προσωπικό δεδομένο. Η παραβίαση ή η μη συμμόρφωση αυτού του νόμου περί απορρήτου μπορεί να οδηγήσει σε πρόστιμο έως και 20 εκατομμύρια ευρώ ή στο 4% των παγκόσμιων ετήσιων εσόδων της εταιρείας από το προηγούμενο οικονομικό έτος - όποιο ποσό είναι υψηλότερο. (European Union, 2021)

Όπως έχουμε ήδη υπογραμμίσει, η ασφάλεια των δεδομένων είναι υψίστης σημασίας, επομένως πρέπει να εφαρμοστούν τεχνικές κρυπτογράφησης, τόσο σε δεδομένα **σε κατάσταση ηρεμίας** όσο και σε δεδομένα **κατά τη μεταφορά** .

Για συμμόρφωση με τους κανονισμούς και προστασία δεδομένων, τα κρυπτογραφημένα δεδομένα σε κατάσταση ηρεμίας είναι κρίσιμα. Συμβάλλει στην ασφάλεια ευαίσθητων δεδομένων που είναι αποθηκευμένα σε δίσκους διασφαλίζοντας ότι κανένας χρήστης ή εφαρμογή δεν μπορεί να το διαβάσει χωρίς έγκυρο κλειδί. Το AWS προσφέρει μια ποικιλία επιλογών για κρυπτογράφηση σε κατάσταση ηρεμίας και διαχείρισης κλειδιών. Για παράδειγμα, μπορούμε να κρυπτογραφήσουμε δεδομένα προτού εγγραφούν σε μη πτητικό χώρο αποθήκευσης (non-volatile storage), όπως κρυπτογράφηση των τόμων AWS EBS volumes ή να ρυθμίσουμε τον κάδο AWS S3 για κρυπτογράφηση πλευράς διακομιστή (Server-Side Encryption - SSE) χρησιμοποιώντας κρυπτογράφηση AES-256 ή ακόμη και κρυπτογράφηση πλευράς πελάτη (Client-Side Encryption) που προσφέρει τη δυνατότητα κρυπτογράφησης δεδομένων πριν από την αποστολή τους σε κάδο S3. Η κρυπτογράφηση δεδομένων σε επίπεδο EC2 Instance είναι επίσης εφικτή, έχοντας κρυπτογράφηση σε επίπεδο συστήματος (disk-level) ή σε επίπεδο συστήματος αρχείων (file system-level).

Από την άλλη πλευρά, η AWS κατέβαλε επίσης πολλή προσπάθεια για την ασφάλεια και την κρυπτογράφηση δεδομένων κατά τη μεταφορά από το ένα σύστημα στο άλλο, συμπεριλαμβανομένων πόρων εντός ή εκτός του AWS. Με τη δημιουργία ενός λογαριασμού AWS, παρέχεται μια λογικά απομονωμένη ενότητα του AWS Cloud που ονομάζεται Amazon Virtual Private Cloud (VPC). Εκεί, μπορούμε να ξεκινήσουμε πόρους AWS μέσα σε ένα καθορισμένο εικονικό δίκτυο. Έχουμε πλήρη έλεγχο του εικονικού περιβάλλοντος δικτύωσης, συμπεριλαμβανομένης της δυνατότητας διαμόρφωσης του δικού μας εύρους διευθύνσεων IP, υποδικτύων, πινάκων διαδρομών και πύλης δικτύου.

Επιπλέον, μπορούμε να δημιουργήσουμε μια σύνδεση εικονικού ιδιωτικού δικτύου (Virtual Private Network - VPN) μεταξύ του εταιρικού κέντρου δεδομένων και του Amazon VPC μας, επιτρέποντάς μας να αξιοποιήσουμε το AWS Cloud ως επέκταση του εταιρικού κέντρου δεδομένων μας. Όσον αφορά την προστατευμένη επικοινωνία μεταξύ Amazon VPC και εταιρικών ιδιόκτητων κέντρων δεδομένων, το AWS διαθέτει μια ποικιλία επιλογών σύνδεσης VPN ανάλογα με τις επιχειρηματικές και τεχνικές ανάγκες της εταιρείας. Το AWS Client VPN επιτρέπει την ασφαλή πρόσβαση σε πόρους AWS μέσω υπηρεσιών VPN που βασίζονται σε πελάτες (client-based). Στο AWS Marketplace, μπορούμε να αγοράσουμε μια συσκευή VPN λογισμικού τρίτου μέρους (third-party) που μπορούμε να εγκαταστήσουμε σε μια δομή Amazon EC2 στο Amazon VPC μας.

Εναλλακτικά, μπορούμε να δημιουργήσουμε μια σύνδεση VPN IPsec μεταξύ του VPC και του απομακρυσμένου δικτύου μας για την ασφαλή επικοινωνία. Μπορούμε να χρησιμοποιήσουμε το AWS Direct Connect για να δημιουργήσουμε μια αποκλειστική ιδιωτική σύνδεση από ένα απομακρυσμένο δίκτυο στο Amazon VPC μας. Αυτή η σύνδεση μπορεί να συνδυαστεί με ένα AWS Site-to-Site VPN για τη δημιουργία μιας κρυπτογραφημένης ιδιωτικής σύνδεσης χρησιμοποιώντας σήραγγες IPsec (IPsec tunnels). (AWS, 2021)

Σχέδια αποκατάστασης καταστροφών (Disaster Recovery) και επιχειρησιακής συνέχειας (Business Continuity)

Ένα σχέδιο επιχειρησιακής συνέχειας είναι ένα ολοκληρωμένο σχέδιο που περιγράφει πώς μια επιχείρηση θα συνεχίσει να λειτουργεί σε περίπτωση καταστροφής. Αυτό το σχέδιο έχει περιεκτικό πεδίο εφαρμογής, αλλά εστιάζει σε συγκεκριμένα σενάρια που θα μπορούσαν να οδηγήσουν σε λειτουργικούς κινδύνους. Ο στόχος του σχεδιασμού συνέχειας των επιχειρήσεων είναι η διατήρηση κρίσιμων λειτουργιών, έτσι ώστε η επιχείρησή μας να μπορεί να συνεχίσει να λειτουργεί κανονικά ακόμη και σε περίπτωση ασυνήθιστων περιστάσεων.

Όταν εφαρμόζεται σωστά, ένα σχέδιο επιχειρησιακής συνέχειας θα πρέπει να επιτρέπει στους πελάτες να συνεχίζουν να λαμβάνουν υπηρεσίες με ελάχιστη διακοπή κατά τη διάρκεια ή αμέσως μετά από μια καταστροφή. Ένα διεξοδικό σχέδιο πρέπει επίσης να λαμβάνει υπόψη τις απαιτήσεις των επιχειρηματικών εταίρων και των πωλητών.

Το σχέδιο συνέχειας πρέπει να υπάρχει ως γραπτό έγγραφο που περιγράφει τις κρίσιμες λειτουργίες της επιχείρησης. Αυτό ενδέχεται να περιλαμβάνει μια λίστα κρίσιμων προμηθειών και επιχειρηματικών λειτουργιών, καθώς και αντίγραφα κρίσιμων αρχείων και στοιχεία επικοινωνίας για βασικούς υπαλλήλους. Οι πληροφορίες που περιέχονται στο πλάνο πρέπει να επιτρέπουν στην επιχείρηση να συνεχίσει τις κανονικές της εργασίες το συντομότερο δυνατό μετά από μία καταστροφή.

Ένα πρόγραμμα αποκατάστασης καταστροφών ή δεδομένων είναι ένα πιο εξειδικευμένο στοιχείο ενός μεγαλύτερου σχεδίου επιχειρηματικής συνέχειας. Περιστασιακά, το πεδίο εφαρμογής ενός σχεδίου αποκατάστασης καταστροφών συμπυκνώνεται για να επικεντρώνεται αποκλειστικά στα συστήματα δεδομένων και πληροφοριών μιας επιχείρησης. Με πιο απλά λόγια, ένα σχέδιο αποκατάστασης καταστροφών χρησιμοποιείται για την αποθήκευση δεδομένων σε περίπτωση καταστροφής με μοναδικό σκοπό την γρήγορη ανάκτησή τους. Έχοντας υπόψη αυτόν τον στόχο, τα σχέδια αποκατάστασης καταστροφών αναπτύσσονται συνήθως για να καλύψουν τις συγκεκριμένες απαιτήσεις του τμήματος πληροφορικής (IT) για την επανέναρξη των δραστηριοτήτων - που τελικά επηρεάζουν ολόκληρη την επιχείρηση.

Ανάλογα με τη φύση της καταστροφής, το σχέδιο μπορεί να περιλαμβάνει τα πάντα, από την ανάκτηση ενός μόνο αρχείου έως την ανάκτηση ενός ολόκληρου κέντρου δεδομένων. Επειδή η πλειονότητα των επιχειρήσεων βασίζεται σε μεγάλο βαθμό στη συλλογή ψηφιακών πληροφοριών και των αντίστοιχων τεχνολογιών, ένα σχέδιο αποκατάστασης καταστροφών αποτελεί κρίσιμο στοιχείο του επιτυχημένου σχεδιασμού επιχειρησιακής συνέχειας.

Σε ορισμένες περιπτώσεις, ο προγραμματισμός αποκατάστασης καταστροφών μπορεί επίσης να αναφέρεται σε πρωτόκολλα που υπάρχουν εκτός του τμήματος τεχνολογίας πληροφοριών. Για παράδειγμα, τα σχέδια αποκατάστασης καταστροφών θα μπορούσαν να περιλαμβάνουν διαδικασίες για το προσωπικό ανάκτησης για τον εντοπισμό μιας εφεδρικής επιχειρηματικής τοποθεσίας, προκειμένου να συνεχίσει τις κρίσιμες εργασίες. Αυτό θα μπορούσε να είναι επωφελές σε περίπτωση φυσικής καταστροφής, όπως πλημμύρες, καθιστώντας τις υπάρχουσες επιχειρηματικές εγκαταστάσεις άχρηστες. Επιπλέον, το σχέδιο μπορεί να περιλαμβάνει οδηγίες για τον τρόπο αποκατάστασης της επικοινωνίας μεταξύ προσωπικού έκτακτης ανάγκης εάν οι συνήθεις γραμμές επικοινωνίας δεν είναι διαθέσιμες. Εάν το τμήμα πληροφορικής μας αναπτύσσει ένα σχέδιο που εστιάζει στην πληροφορική, είναι ζωτικής σημασίας να ενσωματώσουμε όλα τα πρωτόκολλα ανάκτησης εκτός πληροφορικής στο ευρύτερο σχέδιο επιχειρηματικής συνέχειας.

Συνοψίζοντας, η αποκατάσταση καταστροφών είναι η διαδικασία αποκατάστασης δεδομένων, διακομιστών, αρχείων, εφαρμογών λογισμικού και λειτουργικών συστημάτων μετά από ένα καταστροφικό συμβάν και περιέχεται στο σχέδιο αποκατάστασης καταστροφών. Αντίθετα, η επιχειρησιακή συνέχεια αναφέρεται στον τρόπο με τον οποίο μια επιχείρηση συνεχίζει να λειτουργεί σε περίπτωση τεχνολογικής αποτυχίας ή διακοπής λειτουργίας και περιέχεται στο σχέδιο επιχειρηματικής συνέχειας.

Οι κύριες υπηρεσίες, οι οποίες είναι απολύτως κρίσιμες για τη συνεχή λειτουργία της επιχείρησης, θα πρέπει να αποτελούν μέρος του Σχεδίου Επιχειρηματικής Συνέχειας (BC) και να έχουν αυστηρά σχέδια αποκατάστασης καταστροφών (DR), ενώ οι μη κρίσιμες υπηρεσίες μπορούν να έχουν λιγότερο αυστηρά σχέδια DR, εάν χρειάζονται.

RTO / RPO και MTPoD

Το RTO σημαίνει Recovery Time Objective και είναι ένας όρος που αναφέρεται στο χρονικό διάστημα που μια εφαρμογή μπορεί να μην είναι διαθέσιμη χωρίς να προκαλέσει σημαντική διακοπή της επιχείρησης. Ορισμένες εφαρμογές μπορεί να μην είναι διαθέσιμες για μέρες χωρίς να προκαλέσουν σημαντική ζημιά. Αντιθέτως, ορισμένες κρίσιμες εφαρμογές μπορεί να μην είναι διαθέσιμες για λίγα δευτερόλεπτα χωρίς να προκαλούν ενόχληση στους υπαλλήλους, θυμό στο πελάτη ή ζημιά στην επιχείρηση.

Το RTO δεν είναι συνώνυμο με το χρονικό διάστημα μεταξύ απώλειας και ανάκτησης, ο στόχος λαμβάνει επιπλέον υπόψη τα βήματα που απαιτούνται από την ομάδα πληροφορικής για την επαναφορά της εφαρμογής και των δεδομένων της. Εάν η ομάδα τεχνολογίας πληροφοριών έχει επενδύσει σε υπηρεσίες ανακατεύθυνσης για κρίσιμες εφαρμογές, το RTO μπορεί να εκφραστεί με ασφάλεια σε δευτερόλεπτα.

Το RPO σημαίνει Στόχος Σημείου Ανάκτησης και αναφέρεται στην ανοχή απώλειας του οργανισμού, δηλαδή το ποσό των δεδομένων που μπορεί να χαθούν χωρίς να προκαλέσει σημαντική ζημιά στον οργανισμό. Ο στόχος καθορίζεται ως προς το χρονικό διάστημα μεταξύ του συμβάντος απώλειας και του πιο πρόσφατου προηγούμενου αντιγράφου ασφαλείας.

Αν δημιουργήσουμε αντίγραφα ασφαλείας όλων ή των περισσότερων δεδομένων μας με τακτικά προγραμματισμένες διαδικασίες 24 ωρών, θα χάσουμε τα δεδομένα αξίας 24 ωρών στο χειρότερο σενάριο. Αυτό είναι αποδεκτό για ορισμένες εφαρμογές, για άλλες όμως, δεν ισχύει.

Η μέγιστη ανεκτή περίοδος διακοπής (Maximum Tolerable Period of Disruption - MTPoD) είναι η χρονική περίοδος μετά από μια καταστροφή κατά τη διάρκεια της οποίας η βιωσιμότητα ενός οργανισμού θα επιδεινωθεί ανεπανόρθωτα εάν η παραγωγή δεν συνεχιστεί. Κατά τη διεξαγωγή μιας ανάλυσης επιχειρηματικών επιπτώσεων (Business Impact Analysis - BIA) και την ανάπτυξη ενός σχεδίου αποκατάστασης καταστροφών / επιχειρησιακής συνέχειας, το MTPoD είναι κρίσιμο.

Οι παραπάνω τρεις μετρήσεις είναι πολύ σημαντικές για να προσδιορίσουν την κρίσιμη σημασία κάθε υπηρεσίας και να σχεδιάσουν ενέργειες για τις στρατηγικές δημιουργίας αντιγράφων ασφαλείας, την αποκατάσταση καταστροφών και την επιχειρηματική συνέχεια.

Οι υπηρεσίες που έχουν MTPoD και RTO σε ημέρες (μία ή περισσότερες), συνήθως έχουν μόνο ορισμένες προεπιλεγμένες στρατηγικές δημιουργίας αντιγράφων ασφαλείας. Η διαδικασία επαναφοράς πρέπει να τεκμηριωθεί και η διάρκεια αποκατάστασης πρέπει να εμπίπτει στο RTO.

Οι υπηρεσίες που έχουν MTPoD και RTO μετρήσιμα σε ώρες, συνήθως έχουν απλά πλάνα αποκατάστασης καταστροφών που αντιστοιχούν στην αναδημιουργία της υποδομής σε δευτερεύουσα τοποθεσία όταν συμβεί μια καταστροφή. Έτσι, η στρατηγική δημιουργίας αντιγράφων ασφαλείας θα πρέπει να λαμβάνει υπόψη την άμεση διαθεσιμότητα των αντιγράφων στη δευτερεύουσα τοποθεσία.

Οι υπηρεσίες που έχουν MTPoD και RTO μετρήσιμα σε λίγα λεπτά, συνήθως έχουν πολύ πιο αυστηρά σχέδια αποκατάστασης καταστροφών και αποτελούν επίσης μέρος του Σχεδίου Επιχειρηματικής Συνέχειας. Συνήθως έχουν ενεργό ή παθητικό εφεδρικό διακομιστή (Standby server) σε άλλη περιοχή (σε απόσταση άνω των 300 KMs), κάνουν χρήση της διαδικτυακής αναπαραγωγής δεδομένων και ο χρόνος ανάκτησης μπορεί να κυμαίνεται από σχεδόν πραγματικό χρόνο έως λίγα λεπτά.

Όλα τα σχέδια πρέπει να αποτελούν μέρος περιοδικών δοκιμών καταστροφών, προκειμένου να δοκιμάζονται και να αποδεικνύεται η ετοιμότητα και η αποτελεσματικότητα τους.

Στρατηγικές αντιγράφων ασφαλείας (Back Up Strategies)

Οι στρατηγικές δημιουργίας αντιγράφων ασφαλείας θα πρέπει επίσης να τεκμηριώνονται καλά, να εφαρμόζονται και να δοκιμάζονται περιοδικά, κάνοντας επαναφορές και δοκιμές ακεραιότητας δεδομένων. Έχουν σχεδιαστεί σύμφωνα με το απαιτούμενο RPO και την κρισιμότητα της υπηρεσίας.

Έτσι, ένα RPO 24 ωρών, σημαίνει ότι η Επιχείρηση μπορεί να δεχτεί την απώλεια δεδομένων αξίας το πολύ 24 ωρών και μια κατάλληλη στρατηγική δημιουργίας αντιγράφων ασφαλείας θα λαμβάνει πλήρη αντίγραφα ασφαλείας κάθε μέρα στις 00:00 (ή οποιονδήποτε άλλο χρόνο χαμηλού φορτίου).

Ένα RPO 1 ώρας θα μας ανάγκαζε να δημιουργήσουμε μια πιο περίπλοκη στρατηγική δημιουργίας αντιγράφων ασφαλείας, για παράδειγμα πλήρη αντίγραφα ασφαλείας σε χρόνους χαμηλού φορτίου και σταδιακά αντίγραφα ασφαλείας κάθε 1 ώρα διαφορετικά, αλλά αυτό εξαρτάται σε μεγάλο βαθμό από τον τύπο των δεδομένων για τα οποία δημιουργούνται αντίγραφα ασφαλείας. Για παράδειγμα, τα containers είναι αμετάβλητα, οπότε η διατήρηση ενός αντιγράφου ασφαλείας όποτε αλλάζουν είναι αρκετή, καθώς και με αρχεία διαμόρφωσης που πρέπει να ενημερώνονται μέσω ενός αποθετηρίου κώδικα (code repository). Αντιθέτως, οι βάσεις δεδομένων αλλάζουν συνέχεια, επομένως εκεί πρέπει να χρησιμοποιούνται τα στενότερα διαστήματα αντιγράφων ασφαλείας.

Ένα RPO μικρότερο από 1 ώρα, ή ακόμη και μηδέν σημαίνει ότι θα πρέπει να χρησιμοποιείται πρακτικά σύγχρονη διαδικτυακή αναπαραγωγή δεδομένων, έτσι ώστε να μην υπάρχει σχεδόν καμία απώλεια.

Οι στρατηγικές δημιουργίας αντιγράφων ασφαλείας πρέπει επίσης να λαμβάνουν υπόψη τη διαθεσιμότητα τους, επομένως τα αντίγραφα ασφαλείας δεν πρέπει να βρίσκονται κοντά στα αρχικά δεδομένα (αντίγραφα ασφαλείας εκτός ιστότοπου), θα πρέπει να αντιγράφονται απευθείας σε δευτερεύοντες ιστότοπους σε περίπτωση κρίσιμων υπηρεσιών αποστολής και το μέσο δημιουργίας αντιγράφων ασφαλείας θα πρέπει να διασφαλίζει ότι τα δεδομένα θα είναι διαθέσιμα εντός της περιόδου RTO. Για παράδειγμα, ένα εφεδρικό Tape που αποστέλλεται σε απόσταση 500 km εκτός τοποθεσίας χωρίς συσκευή ανάγνωσης ταινιών ή με πολύ αργή συνδεσιμότητα δεν είναι έγκυρη λύση για περίπτωση της μίας ώρας RTO.

Δευτερεύοντες ιστότοποι (Cross-Region Secondary Sites)

Σε περίπτωση κρίσιμων εφαρμογών αποστολής, όπου απαιτείται αυστηρό σχέδιο DR, η ύπαρξη δευτερευόντων ιστότοπων είναι υψίστης σημασίας. Αυτό είναι εξαιρετικά εύκολο να επιτευχθεί στο cloud, χρησιμοποιώντας την ήδη υπάρχουσα Υποδομή ως Κώδικα (IaC) που χρησιμοποιήθηκε για την ανάπτυξη του πρωτεύοντος, για τη δημιουργία ενός δευτερεύοντος και τη δημιουργία αντιγραφής δεδομένων μεταξύ περιοχών.

Μπορούμε να εντοπίσουμε τα ακόλουθα σενάρια:

1. Cold Standbys: Πρόκειται για αντίγραφα υποδομής των κύριων ιστότοπων που παραμένουν απενεργοποιημένα και μόνο τα δεδομένα αντιγράφονται. Αυτά έχουν το χαμηλότερο κόστος, ενώ είναι σε θέση να καταστούν πλήρως λειτουργικά, είτε χειροκίνητα είτε αυτόματα, μέσα σε μερικά λεπτά.

2. Hot Standbys: Πρόκειται για αντίγραφα υποδομής των πρωταρχικών ιστότοπων, τα οποία είναι απενεργοποιημένα κατά το ήμισυ, πράγμα που σημαίνει ότι δεν λειτουργούν στην πλήρη επεξεργαστική ισχύ τους. Είναι πιο ακριβά από τα cold standbys, διατηρώντας παράλληλα πολλές από τις διαθέσιμες λειτουργίες σε λίγα δευτερόλεπτα και φτάνοντας σε πλήρη χωρητικότητα μέσα σε λίγα λεπτά.

3. Ενεργές δευτερεύουσες τοποθεσίες (Active secondary sites): Πρόκειται για πλήρως λειτουργικά αντίγραφα των κύριων ιστότοπων. Είναι πάντα ενεργοποιημένα με την ίδια ισχύ με το πρωτεύον και διατίθενται εντός δευτερολέπτων, συνήθως με διαφάνεια στον τελικό χρήστη. Αυτή είναι η πιο ακριβή λύση, διπλασιάζοντας το κόστος ενός πρωτεύοντος ιστότοπου (primary site) με την προσθήκη του κόστους αναπαραγωγής δεδομένων, παρέχει όμως ανακατεύθυνση σε σχεδόν πραγματικό χρόνο.
   Αυτοί οι ιστότοποι επιτρέπουν τη ρύθμιση παραμέτρων εξισορρόπησης φορτίου Active-Active σε περίπτωση που τις υποστηρίζουν οι εφαρμογές.

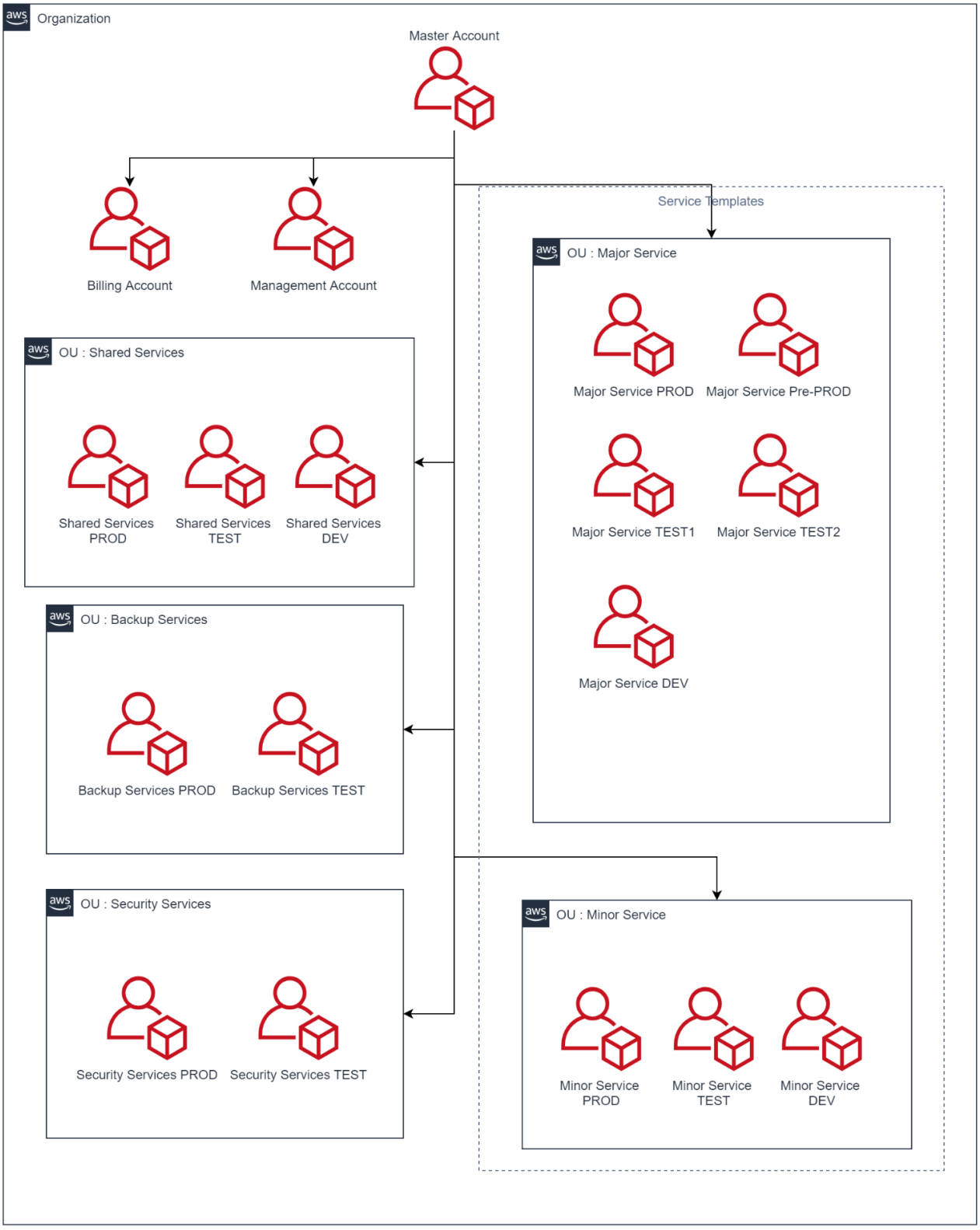
Σχεδιασμός Οργανωτικής Δομής

Η υπηρεσία AWS Organizations επιτρέπει σε μια εταιρεία να εφαρμόσει μια δομή δέντρου οργάνωσης στο cloud, η οποία βοηθά στην κεντρική διαχείριση, τη διακυβέρνηση και τη διαχείριση κόστους. Παρέχει έναν τρόπο γρήγορης κλιμάκωσης του φόρτου εργασίας, διατηρώντας τις υπηρεσίες ταυτόχρονα διαχωρισμένες και ασφαλείς. Επίσης καθιστά δυνατό τον κεντρικό έλεγχο όλων των θυγατρικών λογαριασμών, των εξουσιοδοτημένων δικαιωμάτων και των ελέγχων πρόσβασης, επωφελούμενο από τις εκπτώσεις μεγάλου όγκου και διατηρώντας παράλληλα μια κεντρική αλλά λεπτομερή διαχείριση κόστους.

Η δομή είναι παρόμοια με το Active Directory Forest, με κύριο λογαριασμό (root account), θυγατρικούς λογαριασμούς ή οργανωτικές μονάδες ως κλαδους. Μια οργανωτική μονάδα (OU) μπορεί να διατηρεί έναν ή περισσότερους θυγατρικούς λογαριασμούς. Τα δικαιώματα πρόσβασης, οι άδειες και οι πολιτικές μπορούν να εφαρμοστούν σε επίπεδο λογαριασμού ή οργανωτικών μονάδων (Organizational Units - OU).

Έχοντας αυτό κατά νου, θα στοχεύσουμε να δημιουργήσουμε μια δομή όπου οι επιχειρησιακές μονάδες έχουν αντίστοιχες OU στο cloud, παρακολουθώντας τον δικό τους προϋπολογισμό και την πραγματική κατανάλωση, και επιτρέποντας στο προσωπικό που χειριζόταν τις υπηρεσίες του να συνεχίσει να τις λειτουργεί στο cloud χωρίς να έχει πρόσβαση σε άλλες υποδομές, ακολουθώντας επίσης τον κανόνα «Least Privilege».

Μπορείτε να δείτε ένα διάγραμμα που απεικονίζει τη βασική δομή του Οργανισμού, στο ακόλουθο σχήμα, το οποίο εμφανίζει επίσης κοινόχρηστες υπηρεσίες και περιβάλλοντα που θα αναλυθούν στην επόμενη ενότητα.



Εικόνα 3 : Διάγραμμα οργάνωσης

Σχεδιασμός υπηρεσιών πληροφορικής

Οι υπηρεσίες πληροφορικής μπορούν να κατηγοριοποιηθούν σε δύο άξονες:

1. Ανάλογα με το μέγεθος της υπηρεσίας σε κύριες (Major) και δευτερεύουσες (Minor)
2. Ανάλογα με τον επιχειρηματικό αντίκτυπο σε Mission Critical, Standard και Basic

Οι κύριες υπηρεσίες έχουν συνήθως μεγάλες απαιτήσεις υποδομής και περιβαλλόντων δοκιμής και ανάπτυξης, ενώ οι δευτερεύουσες υπηρεσίες έχουν συνήθως απλούστερες απαιτήσεις υποδομής και απαιτούν ένα από κάθε περιβάλλον ανάπτυξης, δοκιμής και παραγωγής.

Οι υπηρεσίες Mission Critical είναι υπηρεσίες των οποίων η μη διαθεσιμότητα θα έχει σημαντικό αντίκτυπο στην εταιρεία και θα καταστρέψει την επιχειρηματική συνέχεια. Αυτό θα είχε ως αποτέλεσμα μεγάλη απώλεια εσόδων και τυχόν κίνδυνοι πρέπει να εντοπιστούν και να αποκατασταθούν αμέσως. Αυτές οι υπηρεσίες απαιτούν μεγάλη γεωγραφική διαθεσιμότητα και χρόνους RPO / RTO μετρούμενους σε λεπτά. Παραδείγματα τέτοιων υπηρεσιών είναι τα CRM, ERP και CTI.

Για αυτές τις υπηρεσίες θα εξετάσουμε το ενδεχόμενο χρησιμοποίησης γεωγραφικού πλεονασμού (Regional Redundancy) και θα σχεδιάσουμε τη λύση με τέτοιο τρόπο, ώστε οι υπηρεσίες να παραμείνουν λειτουργικές, με ελάχιστη, εάν υπάρξει, μη διαθεσιμότητα σε περίπτωση γεωγραφικής καταστροφής. Τα hot standbys θα τοποθετηθούν σε μια δεύτερη περιοχή, οι βάσεις δεδομένων και οι αποθηκευτικοί χώροι θα αναπαραχθούν στη δεύτερη περιοχή και οι παγκόσμιες υπηρεσίες (global services) θα χρησιμοποιηθούν για την αυτόματη εκτέλεση του failover σε περίπτωση καταστροφής.

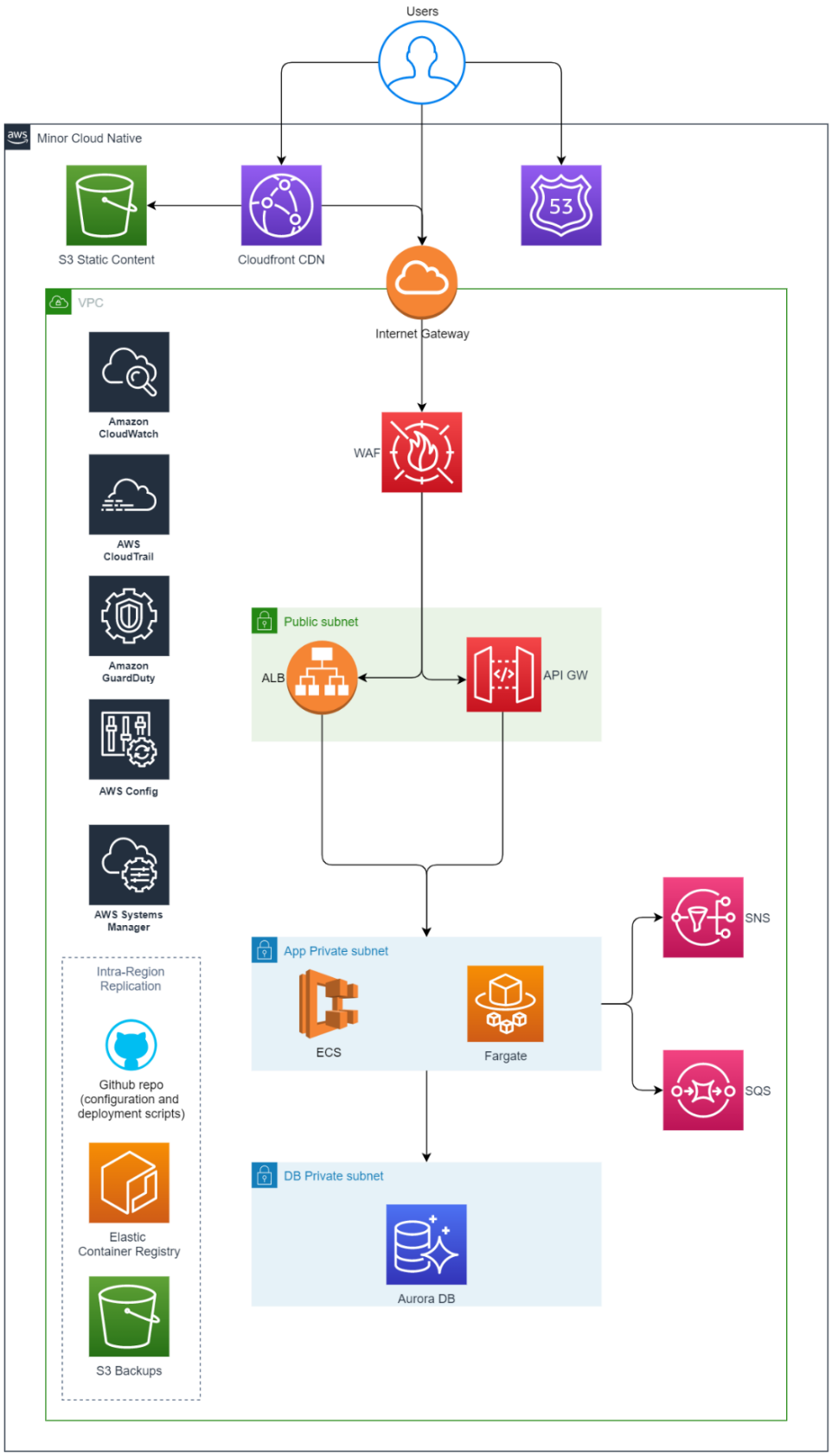
Οι τυπικές υπηρεσίες είναι υπηρεσίες των οποίων η μη διαθεσιμότητα θα έχει μεσαίο αντίκτυπο στην εταιρεία και δεν θα σπάσει τη συνέχεια της επιχείρησης. Τυχόν απώλεια εσόδων που προκύπτει από αυτή τη μη διαθεσιμότητα δεν είναι αρκετά σημαντική ώστε να ληφθεί υπόψη το πρόσθετο κόστος της γεωγραφικής διαθεσιμότητας και οι κίνδυνοι εντοπίζονται, αξιολογούνται και γίνονται αποδεκτοί ή αποκαθίστανται βάσει του μεσαίου επιχειρηματικού αντίκτυπου. Το RPO / RTO μετράται σε ώρες.

Για αυτές τις υπηρεσίες θα εξετάσουμε το ενδεχόμενο χρήσης μόνο πολλαπλών ζωνών διαθεσιμότητας, προκειμένου να μην επηρεαστούμε από τις διακοπές λειτουργίας του κέντρου δεδομένων του παρόχου. Τα αντίγραφα ασφαλείας θα αντιγραφούν σε μια δεύτερη περιοχή και θα εφαρμοστεί μια διαδικασία αποκατάστασης καταστροφών όπου ολόκληρο το περιβάλλον παραγωγής θα αναπτυχθεί εκ νέου στη δεύτερη περιοχή και θα δημιουργηθούν αντίγραφα ασφαλείας, σε περίπτωση εκτεταμένης γεωγραφικής μη διαθεσιμότητας, σύμφωνα με το απαιτούμενο RTO. Αυτό είναι εφικτό με χρήση της υποδομής ως κώδικα και των cloud native εφαρμογών, και μπορεί ακόμη και να αυτοματοποιηθεί εάν χρειαστεί.

Οι βασικές (Basic) υπηρεσίες είναι υπηρεσίες των οποίων η μη διαθεσιμότητα θα έχει ελάχιστο αντίκτυπο στην εταιρεία και καμία επίπτωση στην Επιχειρηματική συνέχεια.

Για αυτές τις υπηρεσίες, τα σχέδια είναι τα ίδια με τις τυπικές υπηρεσίες, αλλά δεν υπάρχει ανάγκη για υψηλή διαθεσιμότητα, επομένως θα βρίσκονται σε ένα μόνο κέντρο δεδομένων (ζώνη διαθεσιμότητας) μιας μεμονωμένης περιοχής και τα αντίγραφα ασφαλείας τους θα βρίσκονται σε μία μόνο περιοχή. Καθώς τα σχέδια είναι τα ίδια, δεν θα παρέχουμε ξεχωριστά σχέδια για τις βασικές υπηρεσίες.

1. Cloud Native – Minor – Standard Υπηρεσία



Εικόνα 4 : Cloud Native δευτερεύουσα τυπική υπηρεσία

Στο παραπάνω διάγραμμα μπορούμε να δούμε μια τυπική αρχιτεκτονική για μια standard υπηρεσία Cloud Native, η οποία είναι μια υπηρεσία που δεν έχει υψηλές απαιτήσεις υποδομής και δεν χρειάζεται πλεονασμό σε επίπεδο περιοχής. Πρέπει να επισημάνουμε ότι όλες οι υπηρεσίες στο διάγραμμα είναι διαθέσιμες εντός της περιοχής, αυτόματα, σε 3 ζώνες διαθεσιμότητας (Availability Zones).

Οι υπηρεσίες που χρησιμοποιούνται είναι οι εξής:

- **Route 53** : Παγκόσμια υπηρεσία συστήματος ονοματοδοσίας Διαδικτύου (Global Domain Name System - DNS), με σχεδόν 100% διαθεσιμότητα, καθώς εκτείνεται σε όλες τις περιοχές AWS και παρέχει επίσης διαχείριση της κυκλοφορίας και φιλτράρισμα με βάση τη γεωγραφική τοποθεσία καθώς και την εξισορρόπηση φορτίου και ανακατεύθυνση κίνησης σε γεωγραφικές περιοχές, χρησιμοποιώντας ελέγχους σωστής λειτουργίας (health Checks).

- **Cloudfront** : Ένα παγκόσμιο δίκτυο παράδοσης περιεχομένου (Content Delivery Network - CDN) με στενή ενσωμάτωση στις υπηρεσίες AWS και πολλές άκρες δικτύου (Edge Locations), με την πρόσθετη δυνατότητα παροχής τερματισμού SSL, εκτέλεσης λειτουργιών lambda για τη μετατροπή αιτημάτων και απαντήσεων και τη διαχείριση επισκεψιμότητας βάσει κανόνων .

- **S3** : Αποθήκευση αντικειμένων με υψηλή διαθεσιμότητα και αξιοπιστία. Τα αντικείμενα αποθηκεύονται και ανακτώνται με αιτήματα HTTP / HTTPS ή μέσω AWS API. Στην αρχιτεκτονική μας, το S3 bucket χρησιμοποιείται για την αποθήκευση στατικών αντικειμένων (S3 Static Content) για παράδοση μέσω Cloudfront, καθώς και στο backend για αποθήκευση αντιγράφων ασφαλείας, αρχείων καταγραφής κλπ.

- **Internet Gateway** : Μια υπηρεσία πύλης χωρίς διακομιστή για σύνδεση από και προς το Διαδίκτυο.

- **WAF** : Ένα τείχος προστασίας εφαρμογών Ιστού που προστατεύει από τις πιο συνηθισμένες επιθέσεις, όπως scripting μεταξύ ιστότοπων (XSS) και εισαγωγή SQL(SQL Injection). Μπορεί να χρησιμοποιήσει κανόνες που προβλέπονται από την AWS ή από τρίτο πάροχο που ενημερώνονται αυτόματα.

- **Elastic Load Balancer** : Μια διαχειριζόμενη υπηρεσία εξισορρόπησης φορτίου που παρέχει είτε Application Load Balancers (ALB) που λειτουργούν στο επίπεδο 7 είτε Network Load Balancers (NLB) που λειτουργούν στο επίπεδο 4. Και οι δύο παρέχουν δυνατότητες

εξισορρόπησης φορτίου με ελέγχους σωστής λειτουργίας, τερματισμό SSL, με τον ALB να παρέχει επιπλέον δρομολόγηση βάσει URL και εκφόρτωση TLS.

- **API Gateway** : Μια διαχειριζόμενη υπηρεσία για δημοσίευση, συντήρηση, παρακολούθηση και προστασία API. Εκφορτώνει εργασίες όπως διαχείριση της κυκλοφορίας, υποστήριξη κοινής χρήσης πόρων (Cross-Origin Resource Sharing - CORS), εξουσιοδότηση, έλεγχος πρόσβασης, περιορισμός, παρακολούθηση και διαχείριση εκδόσεων API από τους διακομιστές, τα containers ή τις λειτουργίες πίσω από αυτό.

- **ECS** : Το Elastic Container Service είναι μια διαχειριζόμενη πλατφόρμα ενορχήστρωσης των containers που παρέχεται από την AWS με στενή ενοποίηση με τις υπόλοιπες υπηρεσίες AWS. Δεν είναι συμβατό με το Kubernetes.

- **Fargate** : Μία υπολογιστική πλατφόρμα για την δημιουργία containers κατ 'απαίτηση με τρόπο χωρίς διακομιστή (serverless). Διαθέσιμο και επεκτάσιμο, μπορεί να χρησιμοποιηθεί από το ECS απευθείας για ενορχήστρωση μικρού έως μεσαίου φόρτου εργασίας.

- **Aurora DB** : Μια επεκτάσιμη σχεσιακή βάση δεδομένων (Relational Database) που βασίζεται σε μηχανές βάσεων δεδομένων AWS που είναι συμβατές με MySQL και PostgreSQL, αλλά παρέχουν υψηλότερες επιδόσεις και καλύτερη ενσωμάτωση με τις υπηρεσίες AWS.

- **Simple Notification Service** (SNS): Μια διαχειριζόμενη υπηρεσία ειδοποιήσεων για επικοινωνία Application to Application (A2A) με χρήση pub/sub τοπολογίας και επικοινωνίας Application to Person (A2P) χρησιμοποιώντας κοινά κανάλια όπως email, mobile push και SMS. Η pub/sub τοπολογία επιτρέπει "θέματα" με πολλούς "συνδρομητές" που θα λάβουν όλοι το μήνυμα / ειδοποίηση "δημοσιευμένο" από έναν ή περισσότερους "εκδότες".

- **Simple Queue Service** (SQS): Μια υπηρεσία διαχειριζόμενης ουράς για την ταξινόμηση και την παράδοση μηνυμάτων με εγγυημένη παράδοση. Τα μηνύματα που προστίθενται στην ουρά μπορούν να ανακτηθούν (pull) από άλλες υπηρεσίες, εφαρμογές σε VM ή container και λειτουργίες Lambda και θα παραμείνουν κρυμμένα για μια προκαθορισμένη χρονική περίοδο (ενώ βρίσκονται υπό επεξεργασία) και είτε διαγράφονται από τον μηχανισμό ανάκτησης (retriever) μετά την επεξεργασία ή επαναφορά στην ουρά εάν έχει ξεπεραστεί το χρονικό όριο. Αυτό σημαίνει ότι εάν κάτι δεν πήγε καλά με την επεξεργασία, τα μηνύματα θα είναι διαθέσιμα για λήψη από άλλη μονάδα. Τέτοια αποτυχημένα μηνύματα μπορούν στη συνέχεια να προστεθούν στην συλλογή αποτυχημένων μηνυμάτων για περαιτέρω έλεγχο. Αυτός ο μηχανισμός παρέχει ένα στρώμα αποσύνδεσης μεταξύ υπηρεσιών, συστατικών εφαρμογών και microservices.

- **CloudWatch** : Μια διαχειριζόμενη υπηρεσία παρακολούθησης και παρατηρησιμότητας, η οποία συλλέγει δεδομένα παρακολούθησης και λειτουργίας με τη μορφή αρχείων καταγραφής, μετρήσεων και συμβάντων, και παρέχει οπτικοποιήσεις, συναγερμούς και πληροφορίες για την υποδομή και τις εφαρμογές.

- **CloudTrail** : Μια διαχειριζόμενη υπηρεσία που παρέχει το ιστορικό συμβάντων της δραστηριότητας λογαριασμών AWS, συμπεριλαμβανομένων των ενεργειών που πραγματοποιούνται μέσω της κονσόλας διαχείρισης AWS, των SDK AWS, των εργαλείων γραμμής εντολών και άλλων υπηρεσιών. Μπορεί να προωθήσει αρχεία καταγραφής στο CloudWatch.

- **GuardDuty** : Μια διαχειριζόμενη υπηρεσία εντοπισμού απειλών. Παρακολουθεί συνεχώς για κακόβουλη δραστηριότητα και μη εξουσιοδοτημένη συμπεριφορά, χρησιμοποιώντας Τεχνητή Νοημοσύνη (Artificial Inteligence) και Μηχανική Μάθηση (Machine Learning) σε όλα τα αρχεία καταγραφής για την ανίχνευση ανωμαλιών. Μπορεί επίσης να χρησιμοποιήσει το Lambda για αυτόματη απάντηση σε απειλές.

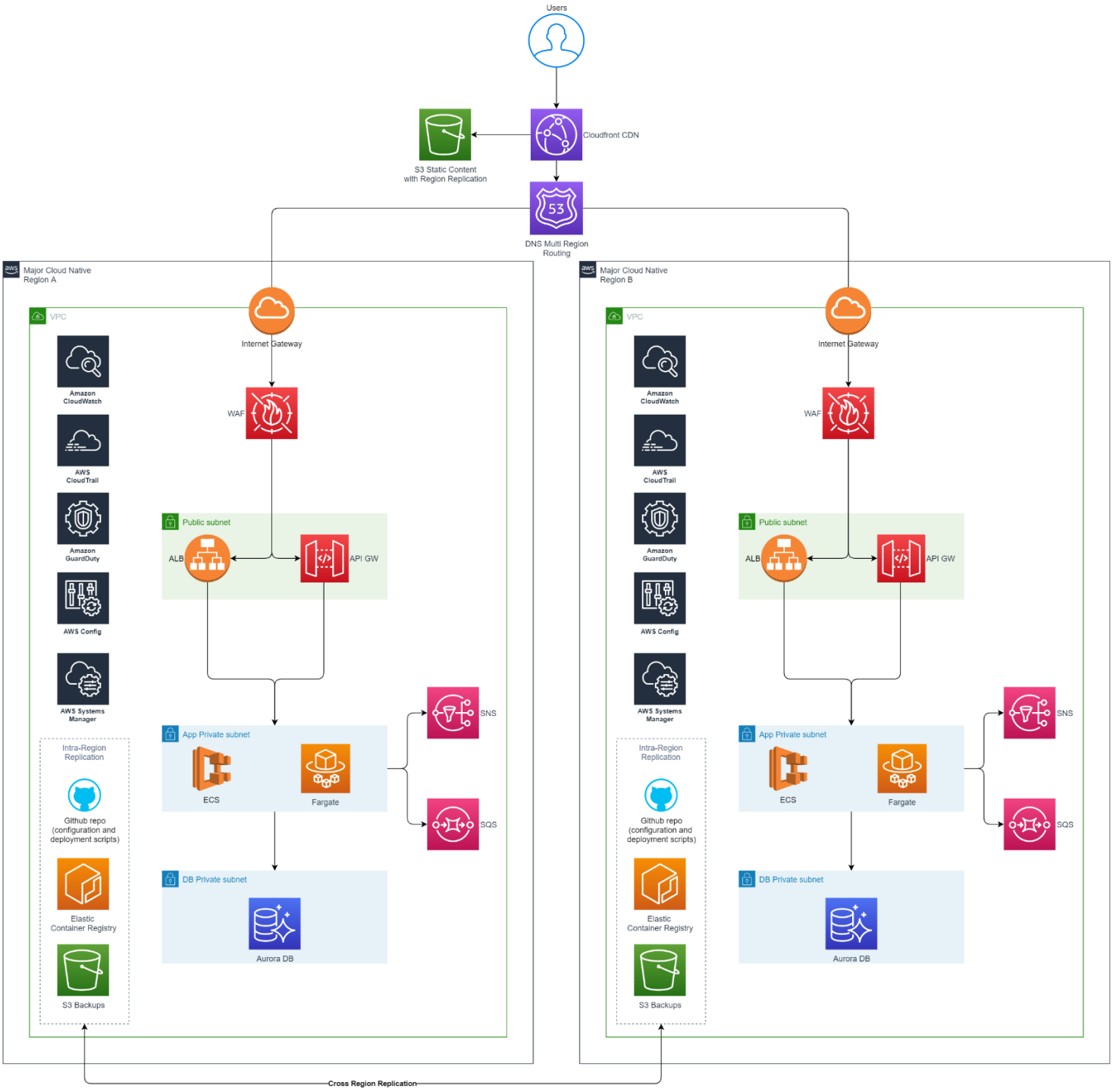
- **Config** : Μια διαχειριζόμενη υπηρεσία που επιτρέπει την αξιολόγηση και τον έλεγχο των διαμορφώσεων των πόρων στον λογαριασμό. Βασίζεται σε κανόνες και χρησιμοποιεί αυτούς τους κανόνες για να αξιολογεί συνεχώς τις διαμορφώσεις και τις σχέσεις μεταξύ πόρων καθώς και να εντοπίζει τυχόν μη συμμόρφωση ή αλλαγές. Αυτή είναι η ευκολότερη μέθοδος για τον εντοπισμό μη προστατευόμενων κάδων S3 ή βάσεων δεδομένων με δημόσια endpoints που μέχρι στιγμής έχουν οδηγήσει σε πολλές διαρροές ασφαλείας.

- **Systems manager** : Ένας διαχειριζόμενος κόμβος λειτουργιών που παρέχει συγκεντρωτικές λειτουργίες και αυτοματοποίηση λειτουργικών εργασιών χρησιμοποιώντας runbooks. Περιλαμβάνει έναν διαχειριστή συμβάντων, έναν διαχειριστή εφαρμογών με δυνατότητα AppConfig που αυτοματοποιεί την ανάπτυξη και διαχείριση διαμόρφωσης εφαρμογών, ένα χώρο αποθήκευσης παραμέτρων που μπορεί να χρησιμοποιηθεί για την αποθήκευση παραμέτρων εφαρμογής, έναν διαχειριστή ενημερώσεων κώδικα για την αυτοματοποίηση επιδιορθώσεων στα παράθυρα συντήρησης και το Session Manager που επιτρέπει την ασφαλή πρόσβαση στην υποδομή χρησιμοποιώντας την κονσόλα AWS.

- **Elastic Container Registry** : Μία διαχειριζόμενη υπηρεσία για την αποθήκευση, διαχείριση, κοινή χρήση και ανάπτυξη container-based εικόνων και αντικειμένων. Ως μία πλήρως διαχειριζόμενη SaaS υπηρεσία, μπορεί να καταναλωθεί χωρίς επιβάρυνση διαχείρισης και είναι πλήρως κλιμακούμενη.

- **Github / AWS CodeCommit:** Το Github είναι μια third-party υπηρεσία ελέγχου και καταγραφής έκδοσης κώδικα, που επίσης παρέχεται ως SaaS. Συνδυάζεται με το οικοσύστημα AWS και προτείνεται λόγω της εξοικείωσης του προγραμματιστή με την υπηρεσία. Μια εξαιρετική εναλλακτική λύση είναι το CodeCommit της AWS που είναι μια πλήρως διαχειριζόμενη υπηρεσία, βασίζεται στο GIT, αλλά φυσικά παρέχει το υψηλότερο επίπεδο διασύνδεσης με το AWS.

2. Cloud Native – Minor – Mission Critical Υπηρεσία



Εικόνα 5 : Cloud Native minor Mission Critical Service

Στο παραπάνω διάγραμμα μπορούμε να δούμε μια τυπική αρχιτεκτονική για μια υπηρεσία Cloud Native minor Mission Critical, η οποία είναι μια υπηρεσία που δεν έχει υψηλές απαιτήσεις υποδομής αλλά χρειάζεται γεωγραφικό πλεονασμό. Το υψηλό επίπεδο πλεονασμού επιτυγχάνεται σε δύο διαφορετικές περιοχές που διαχωρίζονται γεωγραφικά περισσότερο από 100 χιλιόμετρα και έτσι παρέχουν υψηλό επίπεδο απομόνωσης καταστροφών. Η κυκλοφορία κατανέμεται σε όλες τις περιοχές μέσω του Route 53 και μπορεί να κατανέμεται δυναμικά μεταξύ των γεωγραφικών περιοχών με βάση τις ακόλουθες πολιτικές :

- **Πολιτική δρομολόγησης Failover** - Αυτή είναι η απλούστερη πολιτική δρομολόγησης, η Route 53 πραγματοποιεί ελέγχους σωστής λειτουργίας και δρομολογεί την κυκλοφορία από την κύρια στην εφεδρική περιοχή σε περίπτωση αποτυχίας.
- **Πολιτική δρομολόγησης γεωγραφικής θέσης** - Αυτή η πολιτική δρομολόγησης κατανέμει την κυκλοφορία με βάση την τοποθεσία του χρήστη σύμφωνα με προκαθορισμένους κανόνες που ορίζει ο διαχειριστής. Χρησιμοποιώντας αυτήν την πολιτική, μπορούμε να επιλέξουμε ποιες γεωγραφικές τοποθεσίες θα εξυπηρετούνται από ποια περιοχή AWS.
- **Πολιτική δρομολόγησης Geoproximity** - Αυτή η πολιτική δρομολόγησης διανέμει την κυκλοφορία με βάση την τοποθεσία των πόρων και μπορεί, προαιρετικά, να μεταφέρει την κίνηση από πόρους σε μια περιοχή, σε πόρους σε άλλη.
- **Πολιτική δρομολόγησης καθυστέρησης** - Αυτή η πολιτική δρομολόγησης κατανέμει την κυκλοφορία βάσει του χρόνου μεταφοράς δεδομένων από τον χρήστη στην τοποθεσία των πόρων. Χρησιμοποιώντας αυτήν την πολιτική, μπορούμε να δρομολογήσουμε την κίνηση προς την περιοχή που παρέχει τον καλύτερο χρόνο μεταφοράς με λιγότερο χρόνο μετ' επιστροφής (round trip time).
- **Πολιτική δρομολόγησης απαντήσεων πολλαπλών τιμών** - Αυτή η πολιτική δρομολόγησης κατανέμει την κυκλοφορία τυχαία σε έως και οκτώ υγιείς εγγραφές DNS. Αυτή είναι η πιο βασική μορφή εξισορρόπησης φορτίου round-robin, αλλά με το πρόσθετο χαρακτηριστικό του ελέγχου σωστής λειτουργίας των προορισμών.
- **Σταθμισμένη πολιτική δρομολόγησης** - Αυτή η πολιτική δρομολόγησης κατανέμει την επισκεψιμότητα μεταξύ πόρων με βάση το βάρος που εκχωρούμε σε κάθε πόρο. Χρησιμοποιώντας αυτήν την πολιτική δρομολόγησης μπορούμε να δρομολογήσουμε επισκεψιμότητα σε διαφορετικές εκδόσεις μιας εφαρμογής για δοκιμή ή για σε περιπτώσεις μετεγκατάστασης.

Πρέπει να επισημάνουμε ότι όλες οι υπηρεσίες του διαγράμματος είναι διαθέσιμες σε κάθε περιοχή εκτείνοντας αυτόματα σε τρεις ζώνες διαθεσιμότητας.

Οι υπηρεσίες που χρησιμοποιούνται είναι οι εξής:

- **Route 53** : Παγκόσμιες υπηρεσίες DNS, με σχεδόν 100% διαθεσιμότητα, καθώς εκτείνονται σε όλες τις περιοχές AWS, και παρέχουν επίσης διαχείριση της κυκλοφορίας με βάση τη γεωγραφική ή το φιλτράρισμα με βάση τη γεωγραφική τοποθεσία και την εξισορρόπηση φορτίου / ανακατεύθυνση σε διάφορες περιοχές χρησιμοποιώντας ελέγχους σωστής λειτουργίας. Σε αυτή τη ρύθμιση θα χρησιμοποιηθεί για τη δρομολόγηση της κίνησης μεταξύ των περιοχών.

- **Cloudfront** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **S3** : Αποθήκευση αντικειμένων με υψηλή διαθεσιμότητα και αξιοπιστία. Τα αντικείμενα αποθηκεύονται και ανακτώνται με αιτήματα HTTP / HTTPs ή μέσω AWS API. Στην αρχιτεκτονική μας, το S3 bucket χρησιμοποιείται για την αποθήκευση στατικών αντικειμένων (S3 Static Content) για παράδοση μέσω Cloudfront, καθώς και στο backend για αποθήκευση αντιγράφων ασφαλείας, αρχείων καταγραφής κλπ. Το AWS S3 επιτρέπει την αναπαραγωγή μεταξύ περιοχών, η οποία θα χρησιμοποιηθεί για να διατηρείται αυτόματα η αποθήκευση αντικειμένων ενημερωμένη και στις δύο περιοχές.

- **Internet Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **WAF** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Elastic Load Balancer** : Ίδιο με το «1. Cloud Native - Minor - Standard Service ».

- **API Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **ECS** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Fargate** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Aurora DB** : Μια επεκτάσιμη σχεσιακή βάση δεδομένων που βασίζεται σε AWS Database Engine και είναι συμβατή με MySQL και PostgreSQL, αλλά παρέχει υψηλότερη απόδοση και καλύτερη ενσωμάτωση με υπηρεσίες AWS. Σε αυτήν την περίπτωση θα χρησιμοποιήσουμε το Aurora Global Database, το οποίο θα εκτείνεται στις 2 περιοχές, με ένα πρωτεύον και ένα δευτερεύον instance. Μόνο μία περιοχή μπορεί να χειριστεί και τις δύο εγγραφές ενώ και οι δύο περιοχές μπορούν να χειριστούν τις αναγνώσεις.

- **Simple Notification Service**(SNS): Όπως στο «1. Cloud Native - Minor - Standard Service».

- **Simple Queue Service**(SQS): Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **CloudWatch** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **CloudTrail** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **GuardDuty** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

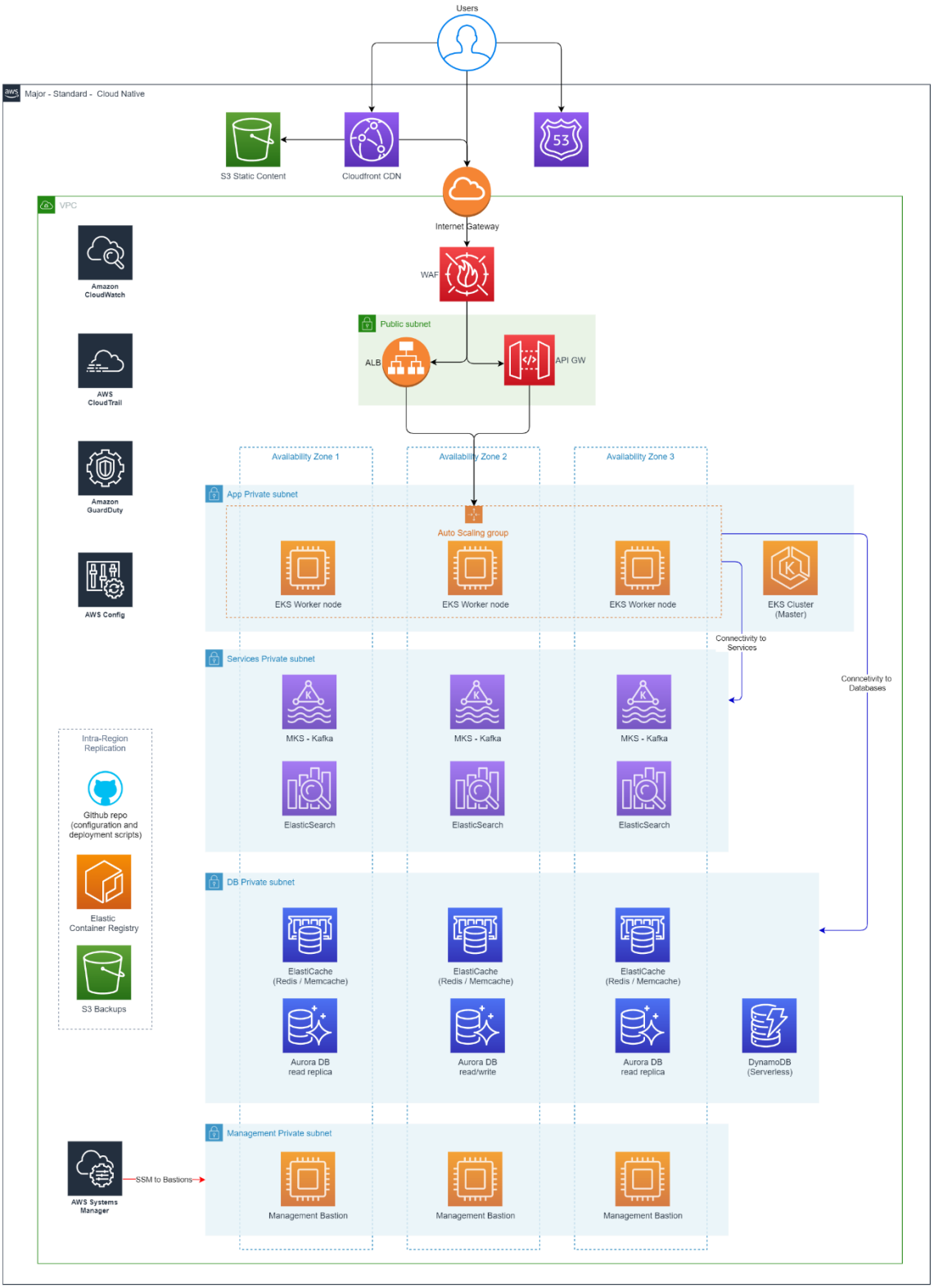
- **Config** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Systems Manager** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Elastic Container Registry** : Μία διαχειριζόμενη υπηρεσία για την αποθήκευση, διαχείριση, κοινή χρήση και ανάπτυξη container-based εικόνων και αντικειμένων. Ως μία πλήρως διαχειριζόμενη SaaS υπηρεσία, μπορεί να καταναλωθεί χωρίς επιβάρυνση διαχείρισης και είναι πλήρως κλιμακούμενη. Το ECR μπορεί να αναπαράγει αυτόματα το μητρώο σε πολλές περιοχές, μια δυνατότητα που θα χρησιμοποιηθεί σε αυτήν την περίπτωση.

- **Github / AWS CodeCommit:** Το Github είναι μια third-party υπηρεσία ελέγχου και καταγραφής έκδοσης κώδικα, που επίσης παρέχεται ως SaaS. Συνδυάζεται με το οικοσύστημα AWS και προτείνεται λόγω της εξοικείωσης του προγραμματιστή με την υπηρεσία. Μια εξαιρετική εναλλακτική λύση είναι το CodeCommit της AWS που είναι μια πλήρως διαχειριζόμενη υπηρεσία, βασίζεται στο GIT, αλλά φυσικά παρέχει το υψηλότερο επίπεδο διασύνδεσης με το AWS. Και οι δύο υπηρεσίες είναι διαθέσιμες σε πολλές περιοχές, αλλά το CodeCommit χρειάζεται μη αυτόματη (scripting) αναπαραγωγή (replication) αποθετηρίου.

3. Cloud Native – Major – Standard Υπηρεσία



Εικόνα 6 : Cloud Native - Major - Τυπική υπηρεσία

Στο παραπάνω διάγραμμα μπορούμε να δούμε μια τυπική αρχιτεκτονική για μια βασική τυπική υπηρεσία Cloud Native, η οποία είναι μια υπηρεσία που έχει υψηλές απαιτήσεις υποδομής αλλά δεν χρειάζεται γεωγραφικό πλεονασμό. Διαχωρίσαμε τις υπηρεσίες στις τρεις ζώνες διαθεσιμότητας και τοποθετήσαμε πόρους σε καθεμία από αυτές για να παρέχουμε υψηλή διαθεσιμότητα. Οι ανάγκες υψηλής απόδοσης / χωρητικότητας των μεγάλων εφαρμογών είναι καταλληλότερες για φιλοξενία σε υποδομές με διακομιστές, οι οποίες, ενώ αυξάνουν κάπως τα λειτουργικά γενικά έξοδα, παρέχουν επαρκή χωρητικότητα σε μεγάλες κλίμακες.

Οι εφαρμογές εξακολουθούν να φιλοξενούνται σε containers, αλλά αυτή τη φορά ο ενορχηστρωτής είναι το βιομηχανικό πρότυπο Kubernetes, σε μια διαχειριζόμενη έκδοση AWS, η Elastic Kubernetes Service (EKS). Ο χειρισμός των μηνυμάτων γίνεται από μια διαχειριζόμενη υπηρεσία Kafka και οι ανάγκες αναζήτησης και ευρετηρίασης (indexing) καλύπτονται από μια διαχειριζόμενη υπηρεσία Elasticsearch.

Οι ανάγκες βάσης δεδομένων διατίθενται σε τρεις εκδόσεις, το ElasticCache για προσωρινή αποθήκευση, την αποθήκευση συνεδρίας και οποιαδήποτε άλλη προσωρινή ανάγκη αποθήκευσης DB, το Aurora DB για ανάγκες σχεσιακής DB και το DynamoDB για ανάγκες NoSQL.

Οι υπηρεσίες που χρησιμοποιούνται είναι οι εξής:

- **Route 53** : Ίδιο με το «στο 1. Cloud Native - Minor - Standard Service ».
- **Cloudfront** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **S3** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Internet Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **WAF** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Elastic Load Balancer** : Μια διαχειριζόμενη υπηρεσία εξισορρόπησης φορτίου που παρέχει είτε Application Load Balancer (ALB) που λειτουργεί στο επίπεδο 7 είτε Network Load Balancer (NLB) που λειτουργεί στο επίπεδο 4. Και οι δύο παρέχουν δυνατότητες εξισορρόπησης φορτίου με ελέγχους σωστής λειτουργίας, τερματισμό SSL, με τον ALB να παρέχει επιπλέον δρομολόγηση βάσει URL και εκφόρτωση TLS. Σε αυτήν την περίπτωση, τα ALB και NLB μπορούν να δημιουργηθούν αυτόματα από την EKS κατά τη δημιουργία μιας υπηρεσίας "Ingress" ή "Load Balance".
- **API Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service "

- **Elastic Kubernetes Service (EKS)** : Μια διαχειριζόμενη πλατφόρμα ενορχήστρωσης containers που παρέχεται από την AWS, βασισμένη στο Kubernetes ανοιχτού κώδικα με στενή ενοποίηση με τις υπόλοιπες υπηρεσίες AWS. Τα περισσότερα από τα εργαλεία που έχουν σχεδιαστεί για το Kubernetes θα λειτουργούν στο EKS, με διαχειριζόμενη διαθεσιμότητα και δυνατότητα κλιμάκωσης για τους κόμβους επιπέδου ελέγχου και οι κόμβοι μπορούν να λειτουργούν είτε σε EC2 (VMs - όπως στην περίπτωσή μας) είτε στο Fargate.

- **Elastic Cloud Compute (EC2)** : Μηχανή υπολογισμού (Compute Engine) για την εκτέλεση VM στο cloud. Υπάρχουν πολλοί τύποι και μεγέθη διαθέσιμα, υπολογισμός, μνήμη ή δίκτυο βελτιστοποιημένο, σε αρχιτεκτονικές x86 και ARM. Στην περίπτωσή μας, χρησιμοποιούνται για να φιλοξενήσουν τους κόμβους του συμπλέγματος (cluster) EKS, αλλά και εσωτερικά για τους κόμβους MKS, ElasticSearch και ElastiCache.

- **Aurora DB** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **DynamoDB** : Μια διαχειριζόμενη βάση δεδομένων κλειδιού-τιμής και εγγράφων, πολλαπλών περιοχών, πολλαπλών ενεργών κόμβων με υψηλή ανθεκτικότητα και ενσωματωμένη ασφάλεια, διαχειριζόμενη δημιουργία αντιγράφων ασφαλείας και επαναφορά, καθώς και προσωρινή αποθήκευση στη μνήμη. Είναι μια υπηρεσία χωρίς διακομιστή και μπορεί να κλιμακωθεί αυτόματα σε "σχεδόν απεριόριστη απόδοση και αποθήκευση".

- **ElastiCache** : Μια διαχειριζόμενη αποθήκευση δεδομένων στη μνήμη, συμβατή με το Redis ή το MemCached. Αυτές οι αποθήκες δεδομένων μπορούν να χρησιμοποιηθούν για την αποθήκευση δεδομένων που χρειάζονται συχνά πρόσβαση, όπως δεδομένα συνεδρίας (session data) και δεδομένα παρακολούθησης (tracking data). Καθώς ολόκληρη η βάση δεδομένων είναι αποθηκευμένη στη μνήμη, μπορεί να παρέχει χρόνους απόκρισης μικροδευτερολέπτων, αλλά με πολύ χαμηλότερη ανθεκτικότητα, καθώς τα δεδομένα δεν αποθηκεύονται σε μόνιμη αποθήκευση, στην προκαθορισμένη διαμόρφωση.

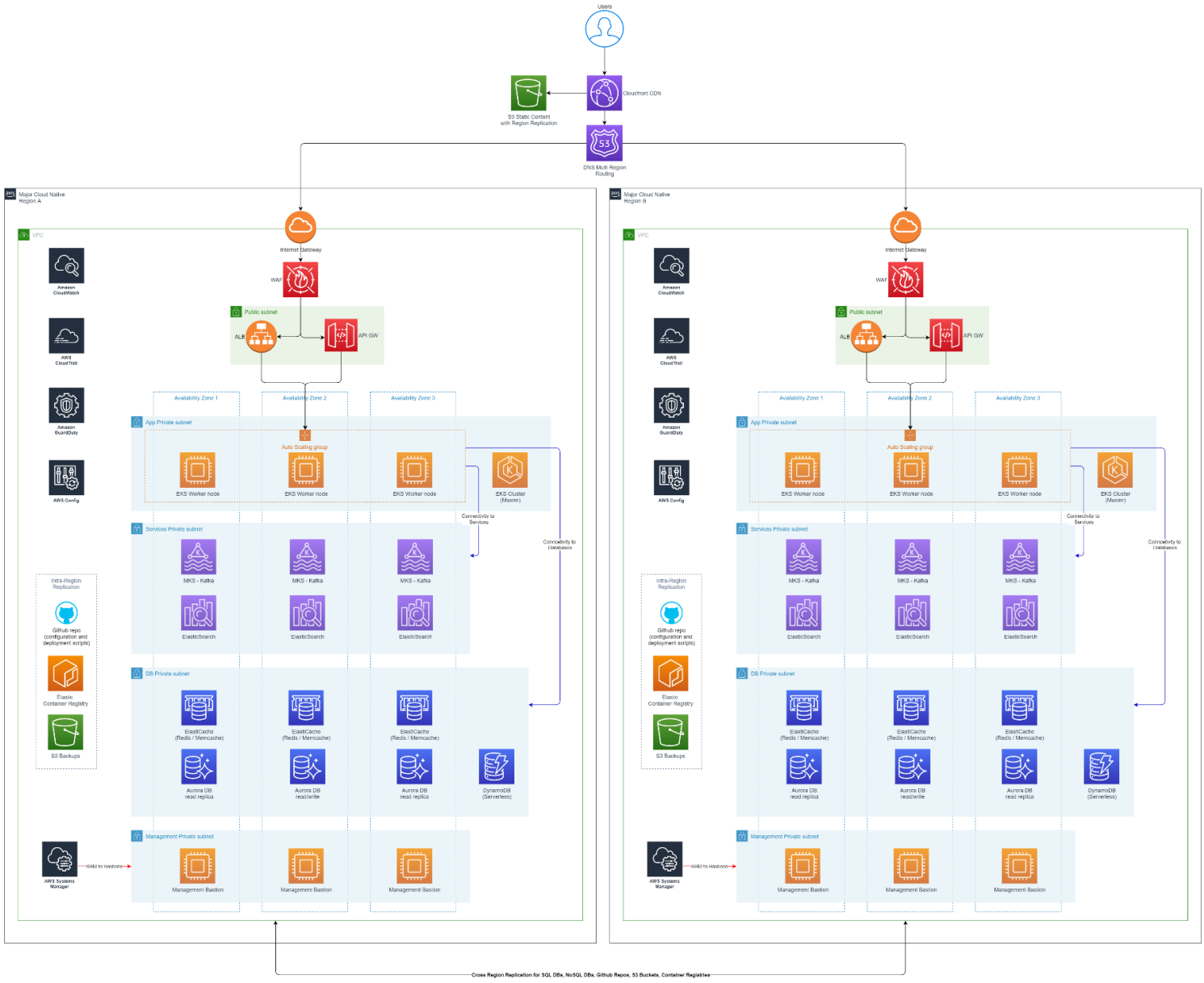
- **ElasticSearch** : Μια διαχειριζόμενη υπηρεσία ElasticSearch. Το Elasticsearch είναι μια μηχανή διανομής, αναζήτησης και ανάλυσης για όλους τους τύπους δεδομένων, συμπεριλαμβανομένων κειμένων, αριθμητικών, γεωχωρικών, δομημένων και μη δομημένων. Μπορεί να χρησιμοποιηθεί για ευρετηρίαση (Indexing) και παροχή δυνατοτήτων αναζήτησης σε ένα ευρύ φάσμα εφαρμογών, από εφαρμογές ιστότοπου έως Business Analytics. Το AWS ElasticSearch, βασισμένο στον ανοιχτό κώδικα ElasticSearch είναι συμβατό με άλλα συχνά χρησιμοποιούμενα προϊόντα Elastic όπως το Kibana και το Logstash.

- **Διαχειριζόμενη ροή (Streaming) για Apache Kafka (MKS)** : Μια διαχειριζόμενη υπηρεσία Kafka. Το Apache Kafka είναι μια πλατφόρμα ανοιχτού κώδικα για τη δημιουργία αγωγών

(pipelines) και εφαρμογών ροής δεδομένων σε πραγματικό χρόνο και χρησιμοποιείται για τη δημοσίευση (write) και την συνδρομή (read) σε ροές συμβάντων, συμπεριλαμβανομένης της συνεχούς εισαγωγής / εξαγωγής δεδομένων από άλλα συστήματα, για την αποθήκευση ροών των συμβάντων με αξιοπιστία για όσο χρονικό διάστημα απαιτείται και για την επεξεργασία ροών γεγονότων τη στιγμή που συμβαίνουν ή αναδρομικά. Το MSK παρέχει και τρέχει αυτόματα τα συμπλέγματα (clusters), παρακολουθεί την κατάσταση του συμπλέγματος και αντικαθιστά αυτόματα τους ελλαττωματικούς κόμβους χωρίς διακοπή λειτουργίας και ασφαλίζει το σύμπλεγμα κρυπτογραφώντας δεδομένα σε κατάσταση ηρεμίας.

- **CloudWatch** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **CloudTrail** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **GuardDuty** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Διαμόρφωση** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Διαχείριση συστημάτων** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Μητρώο ελαστικών containers** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Github / AWS CodeCommit** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

4. Cloud Native – Major – Mission Critical Υπηρεσία



Εικόνα 7 : Cloud Native - Major - Critical Service αποστολής

Στο παραπάνω διάγραμμα μπορούμε να δούμε μια τυπική αρχιτεκτονική για μια υπηρεσία Cloud Native major Mission Critical, η οποία είναι μια υπηρεσία που έχει υψηλές απαιτήσεις υποδομής και χρειάζεται πλεονασμό σε επίπεδο περιοχής. Όπως στο "3." έχουμε χωρίσει τις υπηρεσίες στις τρεις ζώνες διαθεσιμότητας και έχουμε τοποθετήσει πόρους σε καθεμία από αυτές, για να παρέχουμε υψηλή διαθεσιμότητα. Όπως στο "3." οι εφαρμογές εξακολουθούν να φιλοξενούνται σε container ενορχηστρωμένα από την EKS. Ο χειρισμός των μηνυμάτων γίνεται από μια διαχειριζόμενη υπηρεσία Kafka και οι ανάγκες αναζήτησης και ευρετηρίασης καλύπτονται από μια διαχειριζόμενη αναζήτηση ElasticSearch.

Οι ανάγκες βάσης δεδομένων διατίθενται σε τρεις εκδόσεις, το ElasticCache για προσωρινή αποθήκευση, την αποθήκευση συνεδρίας και οποιαδήποτε άλλη προσωρινή ανάγκη αποθήκευσης DB, το Aurora DB για ανάγκες σχεσιακών DB και το DynamoDB για ανάγκες NoSQL.

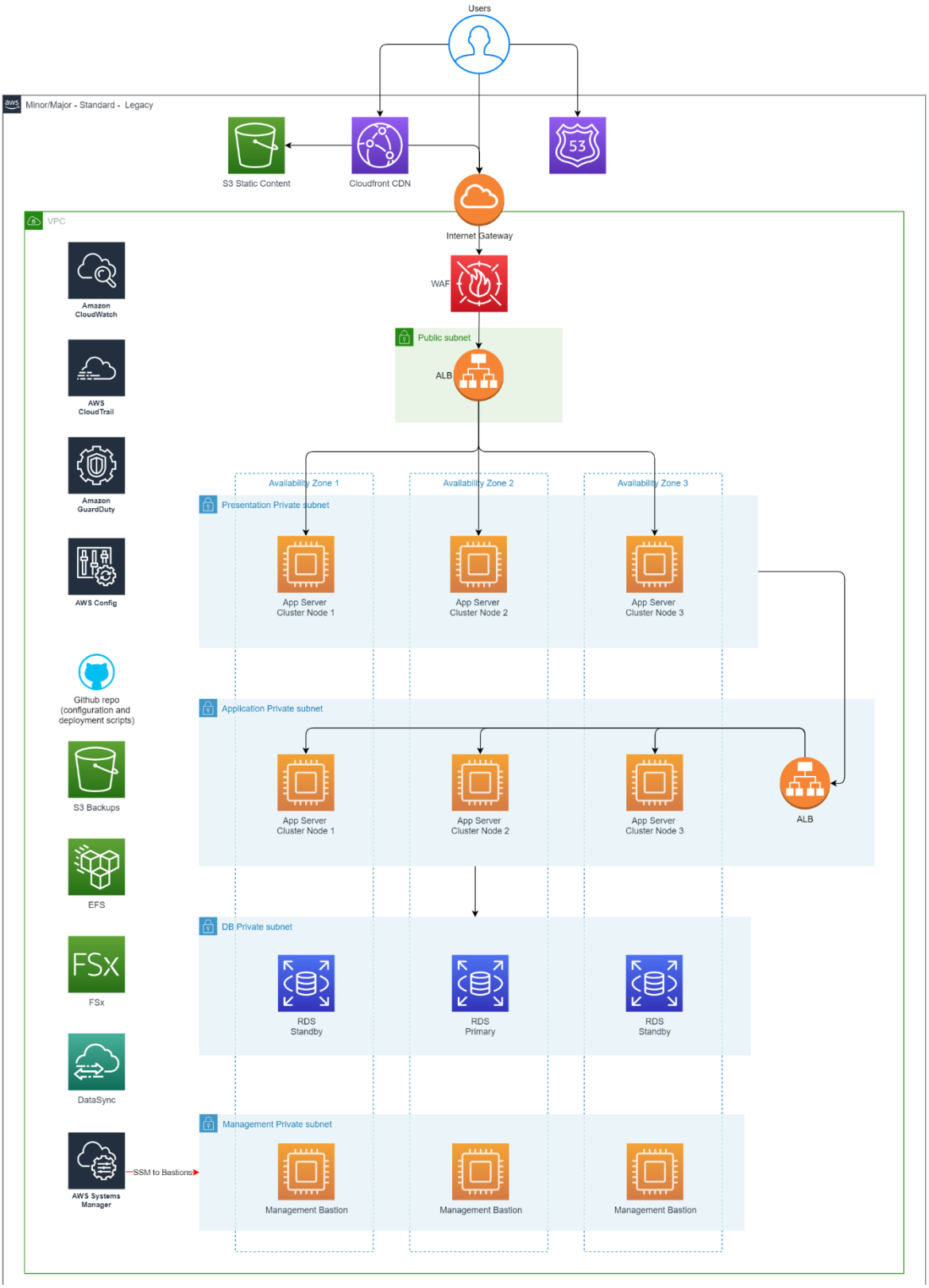
Όπως στο «2. Cloud Native - Minor - Mission Critical Service ", η κίνηση κατανέμεται σε όλες τις περιοχές μέσω της Route53 και μπορεί να κατανέμεται δυναμικά μεταξύ των περιοχών με βάση τις προαναφερθείσες πολιτικές.

Οι υπηρεσίες που χρησιμοποιούνται είναι οι εξής:

- **Route 53**: Όπως στο «2. Cloud Native - Minor - Mission Critical Service ».
- **Cloudfront** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **S3** : Όπως στο «2. Cloud Native - Minor - Mission Critical Service ».
- **Internet Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **WAF** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Elastic Load Balancer** : Όπως στο «3. Cloud Native - Major - Τυπική υπηρεσία ”.
- **API Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Υπηρεσία Elastic Kubernetes (EKS)** : Όπως στο «3. Cloud Native - Major - Τυπική υπηρεσία ”.
- **Elastic Cloud Compute (EC2)** : Όπως στο «3. Cloud Native - Major - Τυπική υπηρεσία ”.
- **Aurora DB** : Όπως στο «2. Cloud Native - Minor - Mission Critical Service ».
- **DynamoDB** : Όπως στο «3. Cloud Native - Major - Τυπική υπηρεσία ”.

- **ElastiCache** : Όπως στο «3. Cloud Native - Major - Τυπική υπηρεσία ”.
- **ElasticSearch** : Όπως στο «3. Cloud Native - Major - Τυπική υπηρεσία ”.
- **Διαχειριζόμενη ροή για Apache Kafka (MKS)** : Όπως στο «3. Cloud Native - Major - Τυπική υπηρεσία ”.
- **CloudWatch** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **CloudTrail** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **GuardDuty** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Διαμόρφωση** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Διαχείριση συστημάτων** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Μητρώο ελαστικών containers** : Όπως στο «2. Cloud Native - Minor - Mission Critical Service ».
- **Github / AWS CodeCommit** : Όπως στο «2. Cloud Native - Minor - Mission Critical Service ».

5. Legacy – Minor/Major – Standard Υπηρεσία



Εικόνα 8 : Legacy - Minor / Major - Τυπική εξυπηρέτηση

Στο παραπάνω διάγραμμα μπορούμε να δούμε μια τυπική αρχιτεκτονική για μια τυπική υπηρεσία Legacy, η οποία είναι μια υπηρεσία που έχει στοιχεία που δεν είναι Cloud Native, πράγμα που σημαίνει ότι δεν μπορούν να χρησιμοποιηθούν ούτε containers, ούτε λειτουργίες χωρίς διακομιστές. Σε αυτήν την περίπτωση χρησιμοποιείται μια πιο κλασική αρχιτεκτονική VM, ενώ εξακολουθεί να επωφελείται από υπηρεσίες cloud όπως διαχειριζόμενος αποθηκευτικός χώρος, διαχειριζόμενες βάσεις δεδομένων και διαχειριζόμενοι εξισορροπητές φόρτωσης (Load Blancers). Καθώς η εφαρμογή υποστηρίζει μόνο VM, δεν υπάρχει διάκριση μεταξύ μεγάλων ή δευτερευουσών υπηρεσιών, η διαφορά είναι μόνο στο μέγεθος των VM, που δεν απεικονίζεται στο διάγραμμα. Έτσι, η ίδια αρχιτεκτονική εφαρμόζεται και στις δύο περιπτώσεις.

Στις περισσότερες περιπτώσεις παλαιών εφαρμογών δεν μπορεί να χρησιμοποιηθεί οποιοδήποτε είδος ελαστικότητας, επομένως λείπουν ομάδες αυτόματης κλιμάκωσης από το διάγραμμα. Εκτός αν η εφαρμογή δεν υποστηρίζει καν πολλές παρουσίες (instances), τα VM μπορούν να εκτείνονται σε ζώνες διαθεσιμότητας για υψηλή διαθεσιμότητα και να διαμορφώνουν το ALB ώστε να παράγει συμβάντα συναγερμού σε περίπτωση που κάποιο instance γίνει ελλαττωματικό. Σε περίπτωση που η εφαρμογή μπορεί να λειτουργήσει μόνο με ενεργό-παθητικό (active-passive) τρόπο, τότε απαιτείται ένας πρόσθετος διακομιστής για να ενεργήσει ως «φύλακας» (watchdog), ο οποίος θα παρακολουθεί την υγεία της εφαρμογής και θα αλλάξει την ενεργή παρουσία χρησιμοποιώντας σενάρια, σε περίπτωση που το ενεργό VM παρουσιάσει πρόβλημα.

Ως έσχατη λύση, ένα περιβάλλον συμπλέγματος μπορεί να αναπαραχθεί στο cloud, χρησιμοποιώντας υπάρχουσες τεχνολογίες συμπλέγματος, όπως το Microsoft Windows Server Failover Cluster, το Veritas Cluster για το Unix / Linux ή το λογισμικό "keepalived", το οποίο δεν θα εκμεταλλευτεί τις τεχνολογίες cloud και θα εισάγει πρόσθετο κόστος.

Οι ανάγκες της βάσης δεδομένων θα πρέπει να καλύπτονται από τις διαχειριζόμενες βάσεις δεδομένων Amazon RDS, οι οποίες υποστηρίζουν μηχανές MS SQL, Oracle, MySQL, PostgreSQL και MariaDB και καλύπτουν τις περισσότερες περιπτώσεις χρήσης. Μια δημοφιλής εξαίρεση είναι το IBM DB2, το οποίο θα πρέπει να εγκατασταθεί σε παρουσίες EC2 και να είναι πλήρως διαχειριζόμενο από εμάς.

Οι υπηρεσίες που χρησιμοποιούνται είναι οι εξής:

- **Route 53**: Όπως στο «1. Cloud Native - Minor - Standard Service. "
- **Cloudfront** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **S3** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **Internet Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service ».
- **WAF** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Elastic Load Balancer** : Ανάλογα με την περίπτωση, το Load Balancer σε αυτό το σενάριο θα εκτελέσει εξισορρόπηση φορτίου μεταξύ των παρουσιών ή θα κατευθύνει την κυκλοφορία σε μία μόνο παρουσία με μη-αυτόματη ανακατεύθυνση, χρησιμοποιώντας έναν τρίτο διακομιστή ελέγχου υγείας, αλλά δεν θα χρησιμοποιήσει ποτέ τις λειτουργίες αυτόματης κλιμάκωσης. Μπορεί να είναι είτε ALB όταν υποστηρίζεται το Layer 7 LB, και ο τερματισμός SSL / HTTPS μπορεί να πραγματοποιηθεί στο ALB, είτε Network Load Balancer (NLB), εάν τα πακέτα πρέπει να φτάσουν αμετάβλητα στην εφαρμογή.

- **Elastic Cloud Compute (EC2)** : Μηχανή υπολογισμού για την εκτέλεση VM στο cloud. Υπάρχουν πολλοί τύποι και μεγέθη διαθέσιμα, βελτιστοποιημένα για υπολογισμό, μνήμη ή δίκτυο, με αρχιτεκτονικές x86 και ARM. Στην περίπτωσή μας, χρησιμοποιούνται για τη φιλοξενία των εφαρμογών και αναλόγως, για έναν ανεξάρτητο διακομιστή ελέγχου υγείας ή ακόμα και για τις προσαρμοσμένες βάσεις δεδομένων. Οι παρουσίες EC2 υποστηρίζονται από τόμους Elastic Block Storage (EBS) που είναι ένα σύστημα αποθήκευσης Block Storage Attached Network (SAN) και Elastic File Storage Shares (EFS), το οποίο είναι ένας τύπος NFS Network Attached Storage (NAS) για διακομιστές unix / linux ή αντίστοιχα Amazon FSx για Windows File Server που είναι τύπου SMB NAS για διακομιστές Windows.

- **EBS** : Μια πλήρως διαχειριζόμενη, εύκολη στη χρήση, υψηλής απόδοσης υπηρεσία αποθήκευσης block που έχει σχεδιαστεί για χρήση με το Amazon Elastic Compute Cloud (EC2) τόσο για ταχύτητα μεταγωγής όσο και για πολύ υψηλό αριθμό συναλλαγών σε οποιαδήποτε κλίμακα. Οι τόμοι του EBS αναπαράγονται σε μία ζώνη διαθεσιμότητας και κλιμακώνονται έως τα Petabytes. Τα στιγμιότυπα EBS με αυτοματοποιημένες πολιτικές κύκλου ζωής είναι διαθέσιμα για χρήση για δημιουργίας αντιγράφων ασφαλείας και μπορούν να αναπαραχθούν σε διάφορες περιοχές ή ακόμη και να χρησιμοποιηθούν για τη δημιουργία νέων παρουσιών, όμοια με το πρωτότυπο.

- **EFS** : Ένα πλήρως διαχειριζόμενο, απλό, χωρίς διακομιστή, set and forget, σύστημα αρχείων συμβατό με NFSv4 που μας επιτρέπει να μοιραζόμαστε δεδομένα αρχείων χωρίς να παρέχουμε ή να διαχειριζόμαστε χώρο αποθήκευσης. Το EFS παρέχει αυτόματα επεκτάσιμη αποθήκευση και απόδοση. Στο βασικό μοντέλο η απόδοση εξαρτάται από τον χώρο αποθήκευσης (η απόδοση αυξάνεται ανάλογα με τον αποθηκευτικό χώρο που χρησιμοποιείται), αλλά η απόδοση μπορεί επίσης να παρέχεται σε μηνιαία βάση με επιπλέον χρέωση, εάν δεν χρησιμοποιείται αρκετός χώρος αποθήκευσης.

- **Amazon FSx για Windows File Server** : Ένας πλήρως διαχειριζόμενος, εξαιρετικά αξιόπιστος και επεκτάσιμος χώρος αποθήκευσης αρχείων που είναι προσβάσιμος μέσω του βιομηχανικά πρωτοτυποποιημένου πρωτοκόλλου block μηνυμάτων διακομιστή (SMB) που είναι το προεπιλεγμένο πρωτόκολλο κοινής χρήσης αρχείων για Windows. Είναι

ενσωματωμένο σε Windows Server, παρέχοντας ένα ευρύ φάσμα διαχειριστικών δυνατοτήτων, όπως Quotas χρηστών, επαναφορά αρχείων τελικού χρήστη και ενοποίηση με το Microsoft Active Directory (AD). Προσφέρει επιλογές ανάπτυξης single-AZ και multi-AZ, πλήρη διαχείριση αντιγράφων ασφαλείας και κρυπτογράφηση δεδομένων σε κατάσταση ηρεμίας και μεταφοράς. Μπορεί επίσης να χρησιμοποιηθεί σε Unix / Linux, αλλά το NFS έχει ένα μικρό πλεονέκτημα απόδοσης και εγγενή ενσωμάτωση.

- **Amazon RDS** : Μια πλήρως διαχειριζόμενη υπηρεσία σχεσιακής βάσης δεδομένων που παρέχει οικονομικά αποδοτική και επεκτάσιμη χωρητικότητα, ενώ αυτοματοποιεί χρονοβόρες εργασίες διαχείρισης, όπως παροχή υλικού, ρύθμιση βάσης δεδομένων, ενημέρωση κώδικα και αντιγράφων ασφαλείας. Οι εφεδρείες με ή χωρίς δυνατότητα ανάγνωσης (standby/read replicas) μπορούν να χρησιμοποιηθούν σε πολλές ζώνες διαθεσιμότητας.

- **CloudWatch** : Μια διαχειριζόμενη υπηρεσία παρακολούθησης και παρατηρησιμότητας, η οποία συλλέγει δεδομένα παρακολούθησης και λειτουργίας με τη μορφή αρχείων καταγραφής, μετρήσεων και συμβάντων και παρέχει οπτικοποιήσεις, συναγερμούς και πληροφορίες για την υποδομή και τις εφαρμογές. Σε αυτήν την περίπτωση, ο πράκτορας (Agent) Cloudwatch θα πρέπει να εγκατασταθεί στα VM που φιλοξενούν τις εφαρμογές, προκειμένου να συγκεντρώσει αρχεία καταγραφής και μετρήσεις.

- **CloudTrail** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **GuardDuty** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

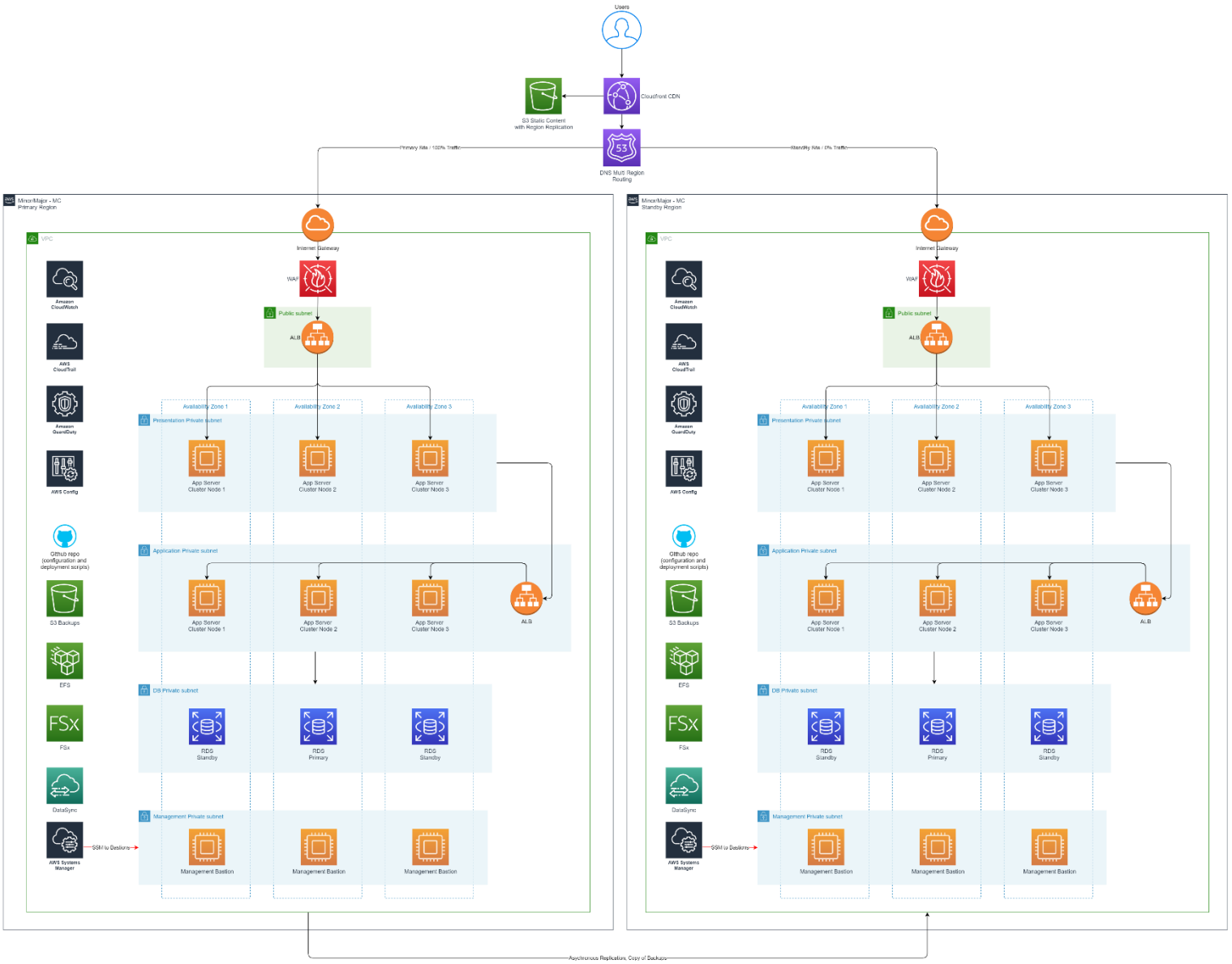
- **Διαμόρφωση** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Διαχείριση συστημάτων** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Github / AWS CodeCommit:** Το Github είναι μια third-party υπηρεσία ελέγχου και καταγραφής έκδοσης κώδικα, που επίσης παρέχεται ως SaaS. Συνδυάζεται με το οικοσύστημα AWS και προτείνεται λόγω της εξοικείωσης του προγραμματιστή με την υπηρεσία. Μια εξαιρετική εναλλακτική λύση είναι το CodeCommit της AWS που είναι μια πλήρως διαχειριζόμενη υπηρεσία, βασίζεται στο GIT, αλλά φυσικά παρέχει το υψηλότερο επίπεδο διασύνδεσης με το AWS. Και οι δύο υπηρεσίες είναι διαθέσιμες σε πολλές περιοχές, αλλά το CodeCommit χρειάζεται μη αυτόματη (scripting) αναπαραγωγή (replication) αποθετηρίου. Παρόλο που οι Εφαρμογές δεν είναι Cloud Native, εξακολουθεί να είναι καλή πρακτική να διατηρούνται τα αρχεία παραμετροποίησης σε ένα αποθετήριο και να

αναπτύσσονται οι αλλαγές χειροκίνητα ή να ενσωματωθεί χειροκίνητα η εφαρμογή σε έναν αγωγό CI / CD για να αυτοματοποιηθούν οι διαδικασίες όσο το δυνατόν περισσότερο.

6. Legacy – Minor/Major – Mission Critical Υπηρεσία



Εικόνα 9 : Legacy - Minor / Major - Mission Critical Service

Στο παραπάνω διάγραμμα μπορούμε να δούμε μια τυπική αρχιτεκτονική για μια υπηρεσία Legacy Mission Critical, που είναι μια υπηρεσία που έχει στοιχεία τα οποία δεν είναι Cloud Native, κάτι που σημαίνει ότι δεν μπορούν να χρησιμοποιηθούν ούτε containers, ούτε λειτουργίες χωρίς διακομιστές, αλλά εξακολουθεί να είναι κρίσιμη για την επιχειρησιακή συνέχεια και πρέπει να υποστεί ελάχιστο χρόνο διακοπής. Σε αυτήν την περίπτωση η αρχιτεκτονική είναι ίδια με το «5. Legacy - Minor / Major - Standard Service ", αλλά εκτείνεται σε δύο περιοχές.

Σε αντίθεση με τις κρίσιμες λύσεις Cloud Native Mission, η Route 53 μπορεί να χρησιμοποιηθεί μόνο με την πολιτική δρομολόγησης Failover (αυτή είναι η απλούστερη πολιτική δρομολόγησης, η Route53 εκτελεί ελέγχους υγείας και δρομολογεί την κυκλοφορία από την κύρια στην δευτερεύουσα “stand-by” σε περίπτωση αποτυχίας της κύριας γεωγραφικής περιοχής) ή την πολιτική σταθμισμένης δρομολόγησης (αυτή η πολιτική δρομολόγησης κατανέμει την επισκεψιμότητα μεταξύ πόρων με βάση το βάρος που αντιστοιχίζουμε σε κάθε πόρο και μπορεί να χρησιμοποιηθεί για την εκτέλεση αναβαθμίσεων εφαρμογών τύπου Blue/Green).

Οι ανάγκες της βάσης δεδομένων πρέπει να καλύπτονται από τις διαχειριζόμενες βάσεις δεδομένων Amazon RDS, οι οποίες υποστηρίζουν μηχανές MS SQL, Oracle, MySQL, PostgreSQL και MariaDB, οι οποίες καλύπτουν τις περισσότερες περιπτώσεις χρήσης και θα δημιουργούνται αντίγραφα ανάγνωσης (read replicas) σε διαφορετικές γεωγραφικές περιοχές. Μια δημοφιλής εξαίρεση είναι το IBM DB2, το οποίο θα πρέπει να εγκατασταθεί σε παρουσίες EC2 και να είναι πλήρως αυτοδιαχειριζόμενο, συμπεριλαμβανομένης της αναπαραγωγής μεταξύ περιοχών. Εάν δεν υποστηρίζεται η αναπαραγωγή, τότε πρέπει να εφαρμοστεί ημι-χειροκίνητη διαδικασία δημιουργίας αντιγράφων ασφαλείας / αντιγραφής / επαναφοράς ή πρέπει να χρησιμοποιηθεί εργαλείο τρίτου παρόχου (3rd party), με επιπλέον κόστος.

Οι υπηρεσίες που χρησιμοποιούνται είναι οι εξής:

- **Route 53**: Παγκόσμιες υπηρεσίες DNS, με σχεδόν 100% διαθεσιμότητα, καθώς εκτείνονται σε όλες τις γεωγραφικές περιοχές που έχει παρουσία η AWS και παρέχουν επίσης διαχείριση της κυκλοφορίας. Σε αυτήν τη ρύθμιση θα χρησιμοποιηθεί για τη δρομολόγηση της κυκλοφορίας βάσει της πολιτικής Failover ή σταθμισμένης δρομολόγησης.

- **Cloudfront** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **S3** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Internet Gateway** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **WAF** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Elastic Load Balancer** : Όπως και στο «5. Legacy - Minor / Major - Τυπική υπηρεσία ».

- **Elastic Cloud Compute (EC2)** : Όπως στο «5. Legacy - Minor / Major - Τυπική υπηρεσία ».

- **EBS** : Μια πλήρως διαχειριζόμενη, εύκολη στη χρήση, υψηλής απόδοσης υπηρεσία αποθήκευσης block που έχει σχεδιαστεί για χρήση με το Amazon Elastic Compute Cloud (EC2) τόσο για ταχύτητα μεταγωγής όσο και για πολύ υψηλό αριθμό συναλλαγών σε οποιαδήποτε κλίμακα. Οι τόμοι του EBS αναπαράγονται σε μία ζώνη διαθεσιμότητας και κλιμακώνονται έως τα Petabytes. Τα στιγμιότυπα EBS με αυτοματοποιημένες πολιτικές κύκλου ζωής είναι διαθέσιμα για δημιουργία αντιγράφων ασφαλείας και μπορούν να αντιγραφούν σε διάφορες περιοχές ή ακόμη και να χρησιμοποιηθούν για τη δημιουργία νέων παρουσιών, όμοια με το πρωτότυπο. Σε αυτήν την περίπτωση, τα στιγμιότυπα EBS θα πρέπει να αντιγραφούν στην δευτερεύουσα γεωγραφική περιοχή χρησιμοποιώντας έναν replicated κάδο S3, καθώς τα στιγμιότυπα αποθηκεύονται αυτόματα στο S3.

- **EFS** : Ένα πλήρως διαχειριζόμενο, απλό, χωρίς διακομιστή, set and forget, σύστημα αρχείων συμβατό με NFSv4 που μας επιτρέπει να μοιραζόμαστε δεδομένα αρχείων χωρίς να παρέχουμε ή να διαχειριζόμαστε χώρο αποθήκευσης. Η αναπαραγωγή μεταξύ γεωγραφικών περιοχών μπορεί να επιτευχθεί χρησιμοποιώντας το AWS Datasync, μια άλλη υπηρεσία AWS που επιτρέπει τη μεταφορά μεταξύ περιοχών ή λογαριασμών.

- **Amazon FSx για Windows File Server** : Ένας πλήρως διαχειριζόμενος, εξαιρετικά αξιόπιστος και επεκτάσιμος χώρος αποθήκευσης αρχείων που είναι προσβάσιμος μέσω του βιομηχανικά πρωτοτυποποιημένου πρωτοκόλλου block μηνυμάτων διακομιστή (SMB) που είναι το προεπιλεγμένο πρωτόκολλο κοινής χρήσης αρχείων για Windows. Όπως και το EFS, η αναπαραγωγή μεταξύ περιοχών μπορεί να επιτευχθεί χρησιμοποιώντας το AWS Datasync.

- **Amazon RDS** : Μια πλήρως διαχειριζόμενη υπηρεσία σχεσιακής βάσης δεδομένων που παρέχει οικονομικά αποδοτική και επεκτάσιμη χωρητικότητα, η οποία αυτοματοποιεί χρονοβόρες εργασίες διαχείρισης, όπως παροχή υλικού, ρύθμιση βάσης δεδομένων, ενημέρωση κώδικα και αντίγραφα ασφαλείας. Οι εφεδρείες με ή χωρίς δυνατότητα ανάγνωσης (standby/read replicas) μπορούν να χρησιμοποιηθούν σε πολλές ζώνες διαθεσιμότητας, και στην περίπτωσή μας σε όλες τις γεωγραφικές περιοχές.

- **CloudWatch** : Όπως στο «5. Legacy - Minor / Major - Τυπική υπηρεσία ».

- **CloudTrail** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **GuardDuty** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Διαμόρφωση** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Διαχείριση συστημάτων** : Όπως στο «1. Cloud Native - Minor - Standard Service ».

- **Github / AWS CodeCommit** : Όπως στο «5. Legacy - Minor / Major - Τυπική υπηρεσία ».

7. Υπηρεσίες τύπου SaaS

Υπάρχουν υπηρεσίες που παρέχονται ήδη ως SaaS, όπου η εταιρεία δεν συμμετέχει καθόλου στη φιλοξενία ή στις λειτουργίες. Αυτές οι υπηρεσίες είναι ως επί το πλείστον υπηρεσίες πληροφορικής τύπου Office, με τις πιο κοινές να είναι το λογισμικό ηλεκτρονικού ταχυδρομείου και γραφείου (Word Processor, Spreadsheet, Presentation κλπ). Οι κύριοι πάροχοι είναι η Microsoft με το Microsoft 365 και η Google με το Google Workspace που καταλαμβάνουν σχεδόν το 100% της αγοράς. (Forbes, 2020)

Αυτές οι σουίτες παρέχουν όλα όσα χρειάζονται για να εκτελούν οι εργαζόμενοι κοινές καθημερινές εργασίες και να συνεργάζονται με συναδέλφους, να μοιράζονται και να διαχειρίζονται τον καθημερινό φόρτο εργασίας. Φιλοξενούνται στην υποδομή του παρόχου και ο πελάτης είναι υπεύθυνος μόνο για τη διαχείριση της πρόσβασης και την εγκατάσταση οποιουδήποτε λογισμικού που μπορεί να χρειαστεί σε συσκευές τελικού χρήστη - εάν υπάρχει. Η λύση της Google βασίζεται εξ ολοκλήρου στον ιστό, ενώ η Microsoft απαιτεί την εγκατάσταση της αυτόνομης σουίτας λογισμικού για να επωφεληθούμε την πλήρη λειτουργικότητα της.

Πολλοί προμηθευτές λογισμικού έχουν ήδη αρχίσει να παρέχουν τις λύσεις τους ως SaaS, οπότε υπάρχει η δυνατότητα να καταναλώσουμε CRM, BRM, ERP, εφαρμογές λογιστικής, διαχείριση περιουσιακών στοιχείων (asset management) και άλλους τύπους λογισμικού ως υπηρεσία.

Αυτές οι υπηρεσίες χρειάζονται ελάχιστη προσπάθεια για την μεταφορά, καθώς μόνο οι ενσωματώσεις ελέγχου ταυτότητας / εξουσιοδότησης πρέπει να μετεγκατασταθούν, ενώ θα συνεχίσουν να παρέχονται όπως πριν.

Διάφορες εφαρμογές SaaS επιτρέπουν στους τελικούς χρήστες να διαχειρίζονται τις ρυθμίσεις των SSO τους ανεξάρτητα από τον προμηθευτή. Για να αποφύγουμε την απώλεια πρόσβασης σε αυτές τις περιπτώσεις, καλό είναι να δημιουργήσουμε μια δοκιμαστική δομή και να συμπεριλάβουμε μια μέθοδο για έλεγχο ταυτότητας ως διαχειριστής της εφαρμογής. Η δοκιμαστική δομή επιτρέπει την επικύρωση της μετεγκατάστασης χωρίς να επηρεάζει τυχόν υπάρχοντες χρήστες. Αυτό θα επιτρέψει

στην υπηρεσία SaaS να λειτουργεί χρησιμοποιώντας τις υπηρεσίες ταυτότητας που έχουν μετεγκατασταθεί στο cloud.

Σε αυτήν την κατηγορία μπορούμε επίσης να προσθέσουμε εταιρικές υπηρεσίες τελικού χρήστη, όπως Desktop-as-a-Service ή εφαρμογές αυτοεξυπηρέτησης γραφείου IT (office IT).

Το Desktop-as-a-Service (DaaS), είναι μια υπηρεσία που βασίζεται σε cloud και παρέχει με ασφάλεια εικονικές εφαρμογές και εικονικούς επιτραπέζιους υπολογιστές σε οποιαδήποτε συσκευή ή τοποθεσία. Ασφαλείς εφαρμογές SaaS αλλά και εφαρμογές παλαιού τύπου (legacy applications), καθώς και πλήρεις εικονικοί επιτραπέζιοι υπολογιστές που βασίζονται σε Windows, παρέχονται και παραδίδονται στο εργατικό δυναμικό, μέσω αυτής της λύσης εικονικοποίησης επιφάνειας εργασίας (Desktop Virtualization Interface). Το DaaS παρέχει ένα απλό και προβλέψιμο μοντέλο συνδρομής pay-as-you-go, το οποίο επιτρέπει την αύξηση ή μείωση σύμφωνα με τις εκάστοτε ανάγκες. Αυτή η υπηρεσία «με το κλειδί στο χέρι» είναι απλή στη διαχείριση, καθώς εξαλείφει πολλές από τις εργασίες διαχείρισης IT που σχετίζονται με λύσεις επιτραπέζιου υπολογιστή. (Citrix, 2021)

Τα πλεονεκτήματα του DaaS είναι τα εξής:

1. **Βελτιωμένη προσβασιμότητα:** Η υπηρεσία είναι διαθέσιμη οπουδήποτε, οποτεδήποτε και με οποιονδήποτε τρόπο και επιτρέπει στους χρήστες να έχουν απομακρυσμένη πρόσβαση στους επιτραπέζιους υπολογιστές τους μέσω υπολογιστή, φορητού υπολογιστή, tablet ή ακόμα και smartphone, παρέχοντάς τους μια, άνευ προηγουμένου, ελευθερία και ευελιξία. Μπορούν να εργαστούν από οπουδήποτε αρκεί να διαθέτουν υπολογιστή και αξιόπιστη σύνδεση στο Διαδίκτυο.

2. **Οι κεφαλαιουχικές δαπάνες (CAPEX) μειώνονται:** Με το παραδοσιακό μοντέλο συνδρομής, το Desktop as a Service (DaaS) διακόπτει τον κύκλο των επενδύσεων σε επιτραπέζιο υλικό, διακομιστές και αδειοδότηση, απελευθερώνοντας κεφάλαια που είχαν δαπανηθεί προηγουμένως για απόσβεση περιουσιακών στοιχείων, τα οποία μπορούν να χρησιμοποιηθούν για επενδύσεις και πρωτοβουλίες υψηλότερης αξίας. Δεν υπάρχει πλέον ανάγκη για ακριβό, εταιρικό υλικό (hardware) τελικού χρήστη, καθώς σχεδόν οποιοσδήποτε φορητός ή επιτραπέζιος υπολογιστής μπορεί να εκπληρώσει το ρόλο του πελάτη (client) DaaS. Οι χρήστες μπορούν επίσης να χρησιμοποιήσουν τις δικές τους συσκευές, εάν αυτές που τους παρέχει η εταιρεία δεν τους καλύπτουν ή εξυπηρετούν.

3. **Χαμηλότερο λειτουργικό κόστος:** Το DaaS μειώνει ένα μεγάλο μέρος του διαχειριστικού όγκου επιτρέποντας την ενοποίηση ή την ανακατανομή λειτουργικών πόρων, μειώνοντας

παράλληλα τις απαιτήσεις χώρου, ισχύος και ψύξης. Οι συμβάσεις υποστήριξης για επιχειρησιακές συσκευές τελικού χρήστη μπορούν επίσης να εξαλειφθούν.

4. **Μεγαλύτερη ευελιξία και απόκριση:** Η δυναμική φύση του DaaS, με την ταχεία παροχή και την εγγενή επεκτασιμότητα, σε συνδυασμό με το μοντέλο Opex χαμηλού κόστους, το καθιστά ιδανικό για οργανισμούς που απαιτείται να επεκταθούν γρήγορα ή να ανταποκριθούν σε ευκαιρίες, είτε πρόκειται για μεταφορά των επιτραπέζιων υπολογιστών (desktops) στο διαδίκτυο ή για την προώθηση νέων εφαρμογών σε μια δομή φιλοξενούμενων επιφανειών εργασίας.

5. **Βελτιωμένη ασφάλεια:** Το DaaS μετατοπίζει το βάρος ασφαλείας από τις μεμονωμένες συσκευές και το μεταφέρει σε μια ασφαλή υποδομή κέντρου δεδομένων. Τα δεδομένα δεν εκτίθενται πλέον σε μια τοπική συσκευή αλλά αποθηκεύονται - και δημιουργούνται αντίγραφα ασφαλείας τακτικά - σε ένα ασφαλές φιλοξενούμενο περιβάλλον. Είναι επίσης κρυπτογραφημένο και προσβάσιμο μόνο μέσω πρωτοκόλλων ελέγχου ταυτότητας πολλαπλών παραγόντων (multi factor authentication), σύμφωνα με τις αυστηρές οδηγίες ασφαλείας της εταιρείας, ανεξάρτητα από τη μέθοδο πρόσβασης σε αυτό.

6. **Καλύτερη εναρμόνιση με τις απαιτήσεις της επιχείρησης**: Οι επιχειρήσεις ξοδεύουν κεφάλαια συχνά για να καλύψουν τις υψηλές απαιτήσεις πόρων, αλλά μετά από το διάστημα αιχμής/ μεγάλου φόρτου, οι πόροι που αποκτήθηκαν δεν έχουν πλέον χρησιμότητα για την εταιρία, κάτι που είναι αποφέρει ζημιά. Το DaaS μπορεί να κλιμακωθεί και να προσαρμοστεί στις μεταβαλλόμενες ανάγκες, διασφαλίζοντας ότι πληρώνουμε μόνο για αυτό που χρησιμοποιούμε. Οι πόροι των εικονικών επιτραπέζιων υπολογιστών μπορούν να αλλάζουν ανάλογα με τον ρόλο του εκάστοτε χρήστη, κάτι που παρέχει ένα επιπλέον επίπεδο εναρμόνισης με τις πολιτικές καθώς και αποδοτικότητας κόστους (cost efficiency).

7. **Αυξημένη ανθεκτικότητα και αξιοπιστία:** Οι χρήστες μπορούν να βασίζονται σε αδιάλειπτη πρόσβαση και εγγυημένη απόδοση καθώς οι εικονικοί επιτραπέζιοι υπολογιστές παραδίδονται από παρόχους λύσεων που συνήθως εγγυώνται 99,99% συνεχούς λειτουργίας, βάσει ενός συνδυασμού υψηλής ποιότητας υποδομής, ασφάλειας και υποστήριξης.

8. **Βελτίωση της επιχειρησιακής συνέχειας**: Το DaaS είναι μια ελκυστική εναλλακτική λύση στην παραδοσιακή επιλογή ανάκτησης καταστροφών (Disaster Recovery) μιας δευτερεύουσας stand-by τοποθεσίας, λόγω της ευρείας διαθεσιμότητας ενός εικονικού επιτραπέζιου υπολογιστή και του κεντρικού backup δεδομένων. Επιπλέον, αντιμετωπίζει το αυξανόμενο ζήτημα των καιρικών / ταξιδιωτικών διαταραχών, επιτρέποντας στους εργαζόμενους να εργάζονται από απόσταση ή από το σπίτι.

9. **Αυξημένη οικονομική προβλεψιμότητα**: Η μηνιαία τιμολόγηση με σταθερή χρέωση ανά χρήστη, παρέχει βεβαιότητα κόστους και απλοποιεί τον προϋπολογισμό και τις προβλέψεις. Αυτό μπορεί να έχει ευεργετική επίδραση στις ταμειακές ροές και να βοηθήσει στον στρατηγικό σχεδιασμό. Διαφορετικές ομάδες χρηστών δύνανται να έχουν διαφορετικές ανάγκες μεγέθους, αλλά το κόστος θα παραμείνει προβλέψιμο για κάθε ομάδα.

10. **Αυξημένη συνέπεια και ετοιμότητα για το μέλλον:** Το προσωπικό σε έναν οργανισμό επωφελείται από μια τυποποιημένη βασική επιφάνεια εργασίας (συγχρόνως εξακολουθεί να επιτρέπεται η εξατομίκευση), τις ίδιες εκδόσεις εφαρμογών, τακτική ανανέωση και μια συνεχιζόμενη επένδυση στην back-end υποδομή. Όλα αυτά συμβάλλουν στη βέλτιστη εμπειρία χρήστη.

11. **Βελτιώσεις παραγωγικότητας:** Ο αυξημένος χρόνος λειτουργίας, η βελτιωμένη απόδοση και οι αυξημένες ευκαιρίες συνεργασίας μπορούν επίσης να αυξήσουν την παραγωγικότητα. Η ικανότητα εργασίας από οπουδήποτε, σε υλικό που μπορεί να προσαρμοστεί για να καλύψει τις ατομικές ανάγκες των χρηστών, καθιστά την εργασία ακόμη πιο παραγωγική.

12. **Ένα προφίλ πιο φιλικό προς το περιβάλλον:** Η μειωμένη ζήτηση για νέο εξοπλισμό, η μειωμένη κατανάλωση ενέργειας, η εκτεταμένη διάρκεια ζωής του υπολογιστή και η πιο ευέλικτη εργασία, συμβάλλουν στη μείωση του αποτυπώματος άνθρακα μιας εταιρείας και στη συμμόρφωση με τους περιβαλλοντικούς στόχους.

Οι υπηρεσίες που μπορούν να προσφερθούν σε ένα μοντέλο SaaS και δύνανται να βοηθήσουν στη μείωση των γενικών εξόδων διαχείρισης τελικών χρηστών, μεταφέροντάς τους κάποιες καθημερινές εργασίες, περιλαμβάνουν:

- **Την επαναφορά κωδικού πρόσβασης:** Η επαναφορά ενός ξεχασμένου ή ληγμένου κωδικού πρόσβασης είναι μια από τις πιο συχνά απαιτούμενες δραστηριότητες του γραφείου εξυπηρέτησης και μπορεί εύκολα να αυτοματοποιηθεί. Χρησιμοποιώντας πολλαπλές μεθόδους ελέγχου ταυτότητας - τηλέφωνο, SMS, email, εφαρμογή ελέγχου ταυτότητας - οι περισσότεροι πάροχοι ταυτότητας έχουν αναπτύξει λύσεις για ασφαλή επαναφορά κωδικού πρόσβασης από τους ίδιους τους τελικούς χρήστες.

- **Παραγγελία και εγκατάσταση λογισμικού:** Χρησιμοποιώντας προκαθορισμένα πακέτα και μια αυτοματοποιημένη διαδικασία έγκρισης και παράδοσης, οι χρήστες μπορούν να «παραγγείλουν» το λογισμικό που χρειάζονται μέσα σε ένα εσωτερικό «κατάστημα» λογισμικού και αυτό θα εγκατασταθεί αυτόματα από τον ίδιο μηχανισμό που παρέχει τα πακέτα ενημέρωσης.

- **Διαχείριση δικαιωμάτων πρόσβασης:** Οι τελικοί χρήστες μπορούν να ζητήσουν πρόσβαση σε υπηρεσίες ή άλλα στοιχεία χρησιμοποιώντας μια αυτοματοποιημένη πλατφόρμα διαχείρισης πρόσβασης, η οποία θα ζητήσει έγκριση από τη διαχείριση (management) και, στη συνέχεια, θα προσθέσει αυτόματα τον χρήστη στις αντίστοιχες ομάδες ασφαλείας του παρόχου ταυτότητας (για παράδειγμα Active Directory). Επιπλέον, η πλατφόρμα μπορεί να ζητήσει την υλοποίηση της διασύνδεσης δικτύου (κανόνες τείχους προστασίας), εάν απαιτείται, είτε δημιουργώντας ένα ticket για μη αυτόματη εφαρμογή είτε με άμεση ενσωμάτωση σε ένα «δίκτυο καθορισμένο από λογισμικό» (Software Defined Network- SDN). Στην περίπτωση του DaaS, αυτή η διαδικασία μπορεί να αυτοματοποιηθεί πλήρως, προσθέτοντας επίσης αυτόματα τον χρήστη στις αντίστοιχες ομάδες ασφαλείας δικτύου.

- **Απομακρυσμένη, ζωντανή ή κατ' απαίτηση αίθουσα διδασκαλίας και εκπαίδευση:** Οι υπάλληλοι μπορούν να επιλέξουν το δικό τους ρυθμό εκπαίδευσης, αλλά και να αποκτήσουν εκπαίδευση υψηλής ποιότητας από την ασφάλεια των σπιτιών τους. Τα ζωντανά σεμινάρια μπορούν να αποθηκευτούν εύκολα για μελλοντική αναπαραγωγή / αναθεώρηση, ερωτήσεις μπορούν να υποβληθούν ηλεκτρονικά χωρίς να διακοπεί ο παρουσιαστής και οι εργαζόμενοι μπορούν να επιστρέψουν στις καθημερινές τους εργασίες μέσα σε λίγα λεπτά.

Υπηρεσίες που θα παραμείνουν σε λειτουργία στο ιδιόκτητο κέντρο δεδομένων

Όπως αναφέρθηκε προηγουμένως, ορισμένες υπηρεσίες θα παραμείνουν εντός του ιδιόκτητου κέντρου. Αυτές περιλαμβάνουν:

- Υπηρεσίες που πρέπει να βρίσκονται πολύ κοντά στο φυσικό σταθερό δίκτυο, όπως η διαχείριση DSLAM, οι υπηρεσίες μέτρησης επιδόσεων και παρακολούθησης, οι υπηρεσίες IPTV που θα επιφέρουν σημαντικό κόστος εάν παρέχονται μέσω του δημόσιου δικτύου κλπ.

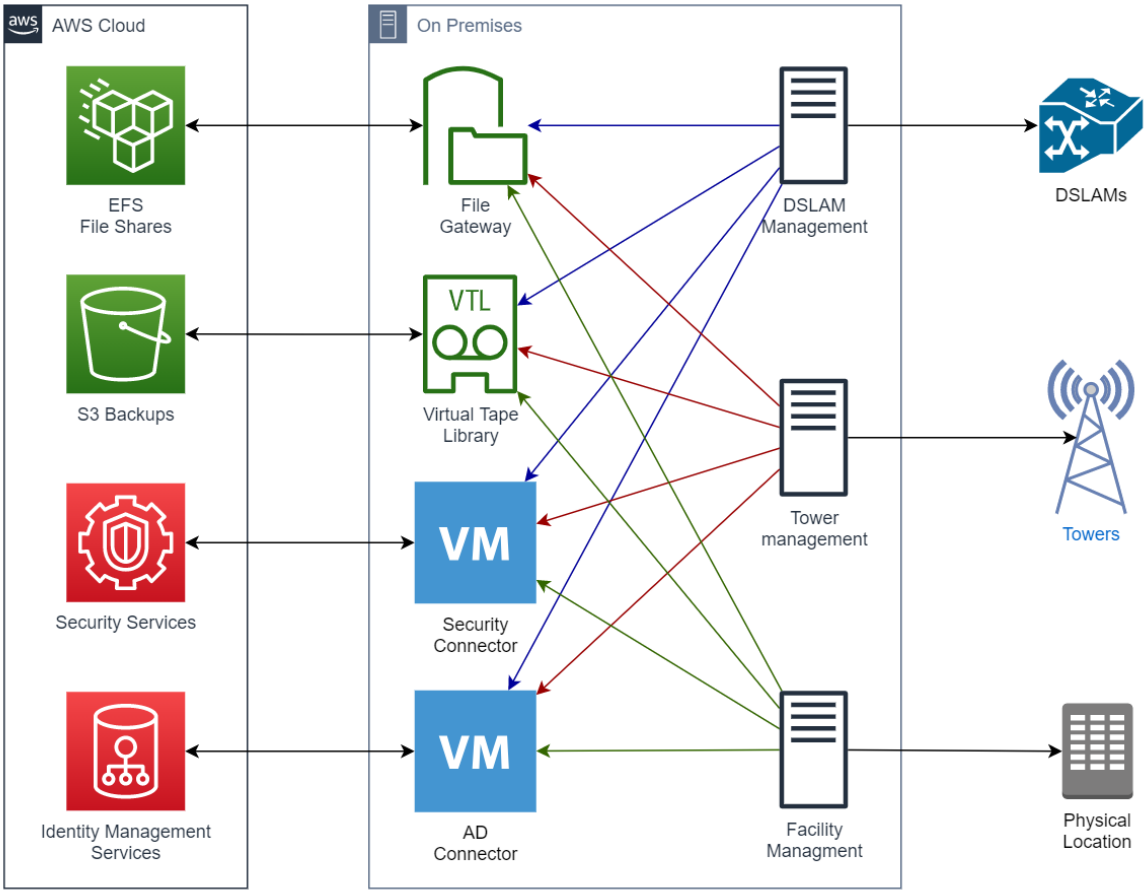
- Υπηρεσίες που πρέπει να βρίσκονται κοντά στο δίκτυο κινητής τηλεφωνίας, όπως λογισμικό διαχείρισης εξοπλισμού πύργων.

- Υπηρεσίες που υποστηρίζουν τις φυσικές εγκαταστάσεις, όπως διαχείριση πρόσβασης, διαχείριση Generators / UPS, συναγερμών και εγγραφής και διαχείρισης CCTV.

- Υπηρεσίες που απαιτούνται ως επέκταση των Cloud Migrated υπηρεσιών, όπως σύνδεσμοι (connectors) αποθήκευσης για δημιουργία αντιγράφων ασφαλείας, AD Connectors για το

Active Directory, σύνδεσμοι για τους σαρωτές ασφαλείας, σαρωτές κακόβουλου λογισμικού κλπ.

Αυτές οι υπηρεσίες θα λειτουργούν σε περιβάλλον υβριδικού cloud και θα επωφεληθούν από τη στενή ενοποίηση με τις Υπηρεσίες Cloud .

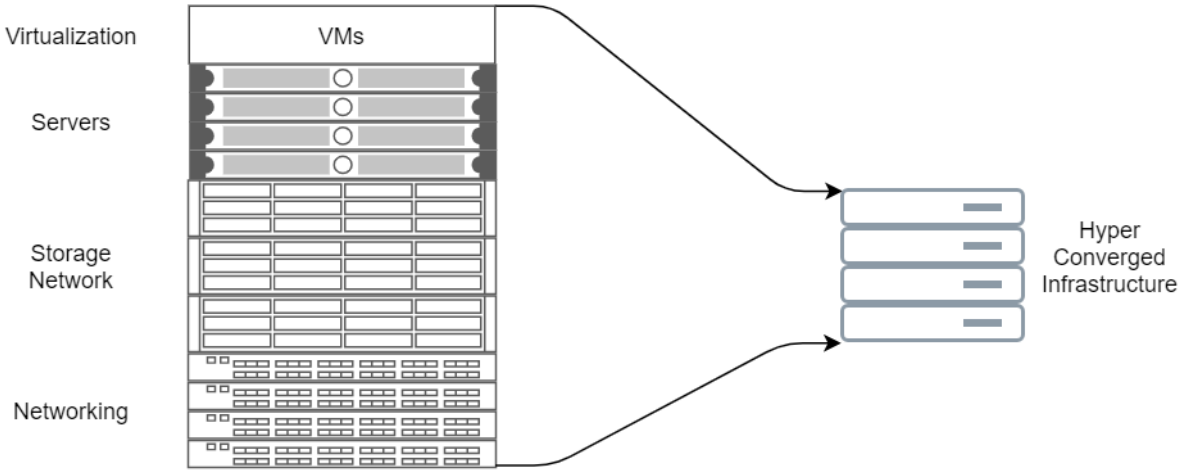


Εικόνα 10 : Σχεδιασμός υβριδικής αρχιτεκτονικής

Ανάλυση και σύγκριση κόστους

Για να μπορέσουμε να συγκρίνουμε το κόστος της υποδομής που στεγάζεται σε ένα ιδιόκτητο κέντρο δεδομένων με το κόστος κατανάλωσης υποδομής σε ένα δημόσιο κέντρο δεδομένων παρόχου cloud, θα αναλύσουμε το κόστος ενός παρόμοιου συνόλου στοιχείων εντός και εκτός του cloud.

Για αυτήν την άσκηση υποθέτουμε ότι η εσωτερική υποδομή βασίζεται στην πλέον μοντέρνα αρχιτεκτονική «Hyper-converged». Η υπερσυγκεντρωμένη υποδομή (hyper-converge-infrastructure HCI) είναι μια υποδομή πληροφορικής που καθορίζεται μέσω λογισμικού και εικονικοποιεί όλα τα στοιχεία των συμβατικών υλικών συστημάτων. Το HCI περιλαμβάνει, τουλάχιστον, εικονικοποιημένους υπολογιστικούς πόρους (μέσω ενός hypervisor) καθώς και αποθήκευση που καθορίζεται από λογισμικό και εικονικοποιημένη δικτύωση (δικτύωση καθορισμένη από λογισμικό - SDN). Το HCI συνήθως εκτελείται σε εμπορικούς διακομιστές off-the-shelf (COTS). (Wikipedia, 2021)



Εικόνα 11 : Hyper Convergence

Το Hyper Convergence είναι μια μετάβαση από τα διακριτά υλικά συστήματα που σύνδεονται μεταξύ τους, σε ένα περιβάλλον λογισμικού, στο οποίο όλες οι λειτουργικές πτυχές λειτουργούν σε εμπορικούς διακομιστές (COTS). Τα συστήματα διακομιστών με Direct-Attached Storage (DAS) χρησιμοποιούνται συνήθως σε περιβάλλοντα HCI, σε αντίθεση με τα παραδοσιακά Storage Attached Networks (SAN) και Network Attached Storages (NAS) που έχουμε στη συμβατική αρχιτεκτονική. Το HCI προσφέρει τη δυνατότητα άμεσης ενσωμάτωσης σε μια ομάδα παρόμοιων συστημάτωνπου βρίσκονται σε ένα κέντρο δεδομένων. Όλοι οι πόροι του φυσικού κέντρου δεδομένων διαχειρίζονται μέσω μιας κεντρικής πλατφόρμας διαχείρισης. Οι ανεπάρκειες του "παραδοσιακού" κέντρου δεδομένων εξαλείφονται και το συνολικό κόστος ιδιοκτησίας (TCO) για κέντρα δεδομένων μειώνεται,

χάρη στην ενοποίηση όλων των λειτουργικών στοιχείων σε επίπεδο εποπτών (hypervisors), αλλά και στην ομοσπονδιακή διοίκηση (federated administration).

Η Hyper-Converged Αρχιτεκτονική εξαλείφει την ανάγκη για διακριτή αποθήκευση, περίπλοκες τοπολογίες δικτύου, σχέδια υψηλής διαθεσιμότητας ή πολλαπλές κονσόλες διαχείρισης.

Τα στελέχη πληροφορικής θεωρούν ολοένα και περισσότερο το HCI ως έναν αποτελεσματικό τρόπο για να αυξήσουν την επιχειρηματική ευελιξία τους, αναφέροντας την αυξημένη επιχειρηματική ευελιξία, την παραγωγικότητα του προσωπικού πληροφορικής, τη λειτουργική αποδοτικότητα και τον ταχύτερο χρόνο εργασιών ως πλεονεκτήματα της λύσης. Το HCI παρέχει επίσης μείωση των λειτουργικών δαπανών (OPEX) και ρίσκου, ενώ ταυτόχρονα αυξάνει την ευελιξία των επιχειρήσεων. (Azeem & Sharma, 2017)

Για τους σκοπούς της παρούσας σύγκρισης, επιλέξαμε μια εικονική υποδομή 1'000 εικονικών servers (VMs) με τις αντίστοιχες προδιαγραφές όπως παρουσιάζονται στον παρακάτω πίνακα:

	Ανά διακομιστή	Σύνολο (για 1000 διακομιστές)
vCores	4	4'000
Μνήμη (GB)	16 GB	16'000 GB
Αποθήκευση (GB)	500 GB	500'000 GB

Πίνακας 3 : Προδιαγραφές εικονικού διακομιστή

a. Υποδομή εντός ιδιόκτητων εγκαταστάσεων

Διακομιστές

Για τους διακομιστές επιλέξαμε καθιερωμένους διακομιστές rack από έναν από τους κορυφαίους παρόχους υποδομής, την DELL EMC. Το μοντέλο που επιλέξαμε είναι ο "PowerEdge R7525 Rack Server" με διαμόρφωση δύο CPU με 24 πυρήνες ο καθένας, συνολικά 48 πυρήνες, 256 GB μνήμης και 6553.6 GB πλεονάζοντος (redundant) αποθηκευτικού χώρου υψηλής απόδοσης. Ο διακομιστής περιλαμβάνει επίσης πλεονάζουσα τροφοδοσία, προσαρμογέα δικτύου Quad 25Gbps, λειτουργική μονάδα για έλεγχο εξ' αποστάσεως, επαγγελματική εξυπηρέτηση (ProSupport - επόμενης ημέρας στο

χώρο του κέντρου δεδομένων) για 5 χρόνια και άδειες για VMWare hyper-converged συστήματα για 5 χρόνια.

Εικόνα 12 : Διακομιστής Dell PowerEdge R7525 Rack

Κατά τον υπολογισμό του αριθμού των διακομιστών που απαιτούνται, πρέπει να λάβουμε υπόψη τους ακόλουθους παράγοντες:

1. Αστοχίες υλικού: Σε περίπτωση βλάβης υλικού, οι υπόλοιποι διακομιστές θα πρέπει να είναι σε θέση να χειριστούν τον πρόσθετο φόρτο εργασίας έως ότου αποκατασταθούν οι αποτυχημένοι διακομιστές.

2. Αύξηση χρήσης: Σε περίπτωση περιοδικής ή μόνιμης αύξησης χρήσης, οι διακομιστές θα πρέπει να είναι σε θέση να χειρίζονται το πρόσθετο φορτίο (αύξηση προδιαγραφών VM ή εισαγωγή πρόσθετων διακομιστών) έως ότου μπορεί να αγοραστεί, να διαμορφωθεί και να εγκατασταθεί η νέα υποδομή. Αυτή η διαδικασία, με βάση τις διαδικασίες της εταιρείας, μπορεί να διαρκέσει από μερικές ημέρες έως μερικούς μήνες (εκτίμηση κόστους, εξασφάλιση του προϋπολογισμού, παραγγελία αγοράς, παράδοση).

3. Νέες υπηρεσίες / αναβαθμίσεις παλαιών υπηρεσιών: Σε περίπτωση εισαγωγής νέων υπηρεσιών ή αναβάθμισης παλαιών υπηρεσιών, ο αριθμός των διακομιστών πρέπει να είναι σε θέση να χειριστεί το πρόσθετο φορτίο έως ότου, όπως παραπάνω, να αγοραστεί, να διαμορφωθεί και να εγκατασταθεί η νέα υποδομή.

Οι διακομιστές θα πρέπει να έχουν ένα μέγιστο όριο χρήσης το οποίο θα υπερκαλύπτει το απαιτούμενο όριο των παραπάνω συνθηκών, και συνεπώς πρέπει να γίνει υπερ-παροχής (overprovision). Στο συγκεκριμένο παράδειγμα μας, υπάρχει over-provisioning 120% και επομένως η

εκμετάλλευση των πόρων κατά την διάρκεια της κανονικής λειτουργίας κατά 83,3%, το οποίο είναι ένα ασφαλές ποσοστό και θα χρησιμοποιηθεί στην περίπτωσή μας.

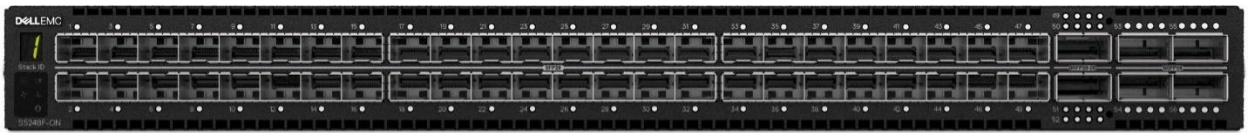
Βάσει των παραπάνω, ο αριθμός των απαιτούμενων διακομιστών θα είναι ο υψηλότερος, υπολογισμένος με βάση όλες τις απαιτήσεις (Πυρήνες, Μνήμη, Αποθήκευση):

	Σύνολο (για 1000 διακομιστές)	Σύνολο (Overprovisioned)	Ανά φυσικό διακομιστή	Απαιτούνται διακομιστές
vCores	4'000	4'800	48	**100**
Μνήμη (GB)	16'000 GB	19'200 GB	256 GB	75
Αποθήκευση (GB)	500'000 GB	600'000 GB	6'553 GB	92

Πίνακας 4 : Ο αριθμός των απαιτούμενων διακομιστών.

Διακόπτες δικτύου (Switches)

Για τους διακόπτες δικτύου επιλέξαμε το Dell EMC Networking S5248F-ON από τον ίδιο προμηθευτή, το οποίο έχει 48 διαχειριζόμενες θύρες 25 Gbit με επιπλέον 4 x 100 Gigabit QSFP28 + 2 x 200 Gigabit QSFP28-DD θύρες.



Εικόνα 13 : Διακόπτης Dell S-Series S5248F-ON

Οι διακομιστές έχουν διαμορφωθεί με κάρτες δικτύου quad-25Gpbs, καθώς η υποδομή Hyperconverged βασίζεται σε μεγάλο βαθμό στο υποκείμενο δίκτυο για τις περισσότερες από τις βασικές λειτουργίες της, όπως η αναπαραγωγή αποθηκευτικού χώρου (storage replication) σε διαφορετικούς κόμβους, μετεγκατάσταση εικονικών μηχανών (VM migration) για εξισορρόπηση φορτίου και ανοχή σφαλμάτων, καθώς και επικοινωνιακές ανάγκες μεταξύ των VMs επομένως ένα δίκτυο υψηλής απόδοσης και πλεονασμού είναι υψίστης σημασίας για την εύρυθμη λειτουργία.

Κάθε μία από τις θύρες quad των διακομιστών θα συνδεθεί σε διαφορετικό διακόπτη για πλεονασμό και οι διακόπτες θα διασυνδέονται ανεξάρτητα χρησιμοποιώντας τις θύρες 100Gbit και 200Gbit, και ως εκ τούτου, το δίκτυο μπορεί να επιβιώσει την αποτυχία έως και 3 διακοπτών ταυτόχρονα, αν και με υποβαθμισμένη απόδοση.

Για τον υπολογισμό του αριθμού των απαιτούμενων διακοπτών, υπολογίζουμε τον συνολικό αριθμό των θυρών διακομιστή και τους διαιρούμε με τον αριθμό των θυρών ανά διακόπτη, ο οποίος θα παρέχει τον απόλυτο ελάχιστο απαιτούμενο αριθμό διακοπτών:

- 100 διακομιστές * 4 θύρες ανά διακομιστή = 400 θύρες συνολικά
- 400 θύρες / 48 θύρες ανά διακόπτη = 8,33 → στρογγυλοποίηση στον επόμενο ακέραιο = 9

Σε αυτό το σημείο θα πρέπει να εξετάσουμε την καλύτερη πρακτική του να έχουμε ελεύθερες θύρες σε κάθε διακόπτη για μελλοντική επέκταση ή σε περίπτωση δυσλειτουργίας μίας θύρας και να προσθέσουμε έναν πρόσθετο διακόπτη (10 διακόπτες συνολικά) που θα μειώσει τη χρήση των θυρών σε περίπου 400 / (48 * 10) = 83,3%, το οποίο συμβαδίζει με την πολιτική υπέρ-παροχής διακομιστών 120%.

Δρομολογητές (Routers)

Για τους δρομολογητές επιλέξαμε το βιομηχανικό πρότυπο CISCO CATALYST 8500 SERIES 12, και συγκεκριμένα το μοντέλο C8500L-8S4X.

Εικόνα 14 : Δρομολογητής Cisco C8500L-8S4X

Οι δρομολογητές θα επιτρέπουν στους διακομιστές να επικοινωνούν με το Διαδίκτυο ή / και άλλα κέντρα δεδομένων μέσω VPN.

Για πλεονασμό, ο ελάχιστος αριθμός δρομολογητών είναι δύο, και ο καθένας θα συνδέεται επίσης με πολλαπλούς διακόπτες.

Καλώδια (Cables)

Για τα καλώδια επιλέξαμε τα καλώδια Dell Networking SFP28 έως SFP28 25GbE, Passive Copper Twinax Direct Attach.

Εικόνα 15 : Καλώδιο δικτύωσης Dell, SFP28 έως SFP28, 25GbE

Τα καλώδια είναι 25Gbit, συμβατά με τους διακόπτες μας, τριών μέτρων, τυπικά για καλωδίωση εντός των racks. Πρέπει να παρέχονται τουλάχιστον 480 καλώδια, όσος και ο συνολικός αριθμός θυρών στους διακόπτες.

Πρόσθετα κόστη

Ένα επιπλέον 15% κόστος έχει προστεθεί στο υποσύνολο δικτυακού εξοπλισμού για την κάλυψη του κόστους χάλκινης και οπτικής καλωδίωσης (ενεργό ή παθητικό) με απρόβλεπτο μήκος (καλωδιώσεις μεταξύ racks, καλωδίωση δρομολογητών κ.λπ.) και άλλο εξοπλισμό δικτύου όπως IP KVMs (Διακόπτες πληκτρολογίου, βίντεο και ποντικιού).

Κόστος υποδομής

Με βάση τα παραπάνω, το συνολικό κόστος της υποδομής (Compute και Network) συνοψίζεται στον ακόλουθο πίνακα:

	Τιμή ανά μονάδα	Μονάδες	Σύνολο
Διακομιστές Rack PowerEdge R7525 (Άδεια και υποστήριξη 5YR)	53'812,55 €	100	5'381'255,04 €
Διακόπτες (Switches)	8'500,00 €	10	85'000,00 €
Καλώδια (Cables)	55,00 €	480	26'401,54 €
Δρομολογητές (Routers)	18'000,00 €	2	36'000,00 €
Άλλος εξοπλισμός δικτύωσης (KVMs, οπτικά καλώδια κλπ.)		15%	22'110,23 €
Συνολικό κόστος δικτύου			169'511,77 €
Συνολικό κόστος υποδομής (εκτός από τα racks)			**5'720'278,57 €**

Πίνακας 5 : Συνολικό κόστος υπολογισμού και υποδομής δικτύου

Κόστος κέντρου δεδομένων

Οι υπολογισμοί μέχρι στιγμής δεν περιλαμβάνουν καμία από τις μόνιμες υποδομές κέντρων δεδομένων, όπως racks, ισχύ και ψύξη. Αυτό συμβαίνει επειδή, ο εξοπλισμός υπολογιστικής ισχύος και δικτύωσης έχει συνήθως μέγιστο κύκλο ζωής 5 ετών, ενώ ο υπόλοιπος εξοπλισμός έχει πολύ μεγαλύτερο κύκλο ζωής, με ελάχιστο τα 10 έτη. Καθώς η περίοδος απόσβεσης του ίδιου του Κέντρου δεδομένων είναι πολύ μεγαλύτερη από την υποδομή υπολογιστικής ισχύος και δικτύωσης, επιλέξαμε να παρουσιάσουμε δύο περιπτώσεις, μία που υποθέτει ότι ένα Κέντρο Δεδομένων υπάρχει διαθέσιμο σε λειτουργία, και μία που υποθέτει ότι ένα νέο Κέντρο Δεδομένων πρέπει να χτιστεί εξ αρχής.

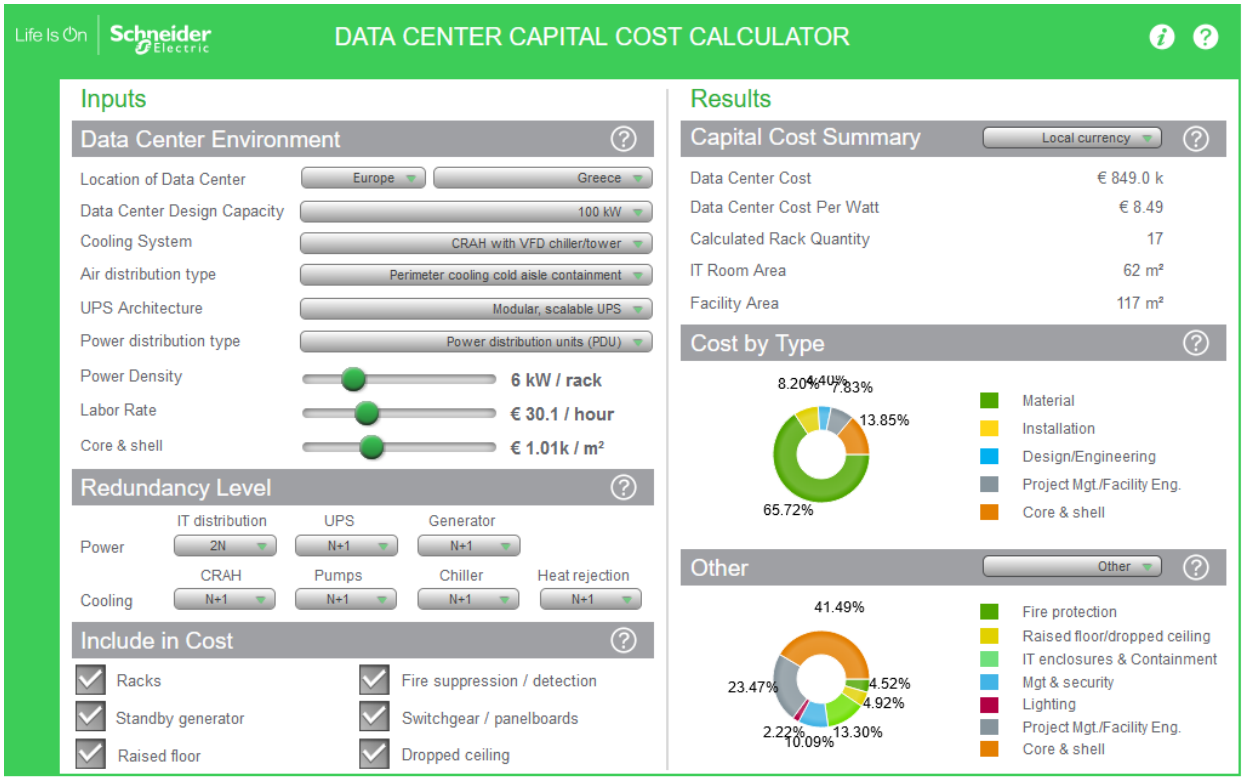
Ένας σημαντικός παράγοντας της ανάλυσης κόστους του κέντρου δεδομένων είναι η συνολική ισχύς που πρέπει να είναι σε θέση να υποστηρίξει. Σύμφωνα με την μέχρι τώρα ανάλυση των εξαρτημάτων μας, οι ανάγκες σε ισχύ μπορούν να υπολογιστούν σύμφωνα με τον παρακάτω πίνακα.

	Ισχύς ανά μονάδα	Μονάδες	Σύνολο
Διακομιστές	800 W	100	80'000 W
Διακόπτες (Switches)	500 W	10	5'000 W
Δρομολογητής	1'000 W	2	2'000 W
		Σύνολο	87'000 W

Πίνακας 6 : Απαιτήσεις ισχύος υποδομής

Οι συνολικές απαιτήσεις μέγιστης ισχύος για τον εξοπλισμό υπολογιστών και δικτύου μας φτάνουν τα 87kW. Επιλέξαμε όμως να αναλύσουμε το Data Center για χωρητικότητα σχεδιασμού 100kW, υπολογίζοντας ένα επιπλέον κόστος για μελλοντικές προσθήκες.

Η κατανομή του κόστους του κέντρου δεδομένων δεν εμπίπτει στο πεδίο αυτής της διπλωματικής και, ως εκ τούτου, χρησιμοποιήσαμε τον υπολογιστή κόστους κέντρων δεδομένων της Schneider Electric, όπως φαίνεται στο παρακάτω σχήμα.



Εικόνα 16 : Υπολογιστής Schneider Electric CAPEX

Όπως μπορεί να φανεί παραπάνω, επιλέξαμε σύγχρονες λύσεις ισχύος και ψύξης καθώς και πλεονασμό σε όλα τα επίπεδα:

- Χειριστής αέρα δωματίου υπολογιστή (CRAH) με μεταβλητή μονάδα συχνότητας (VFD) Chiller / Tower
- Περιμετρική ψύξη ψυχρού χώρου.
- Αρθρωτό και επεκτάσιμο UPS
- Μονάδες διανομής ισχύος
- 2N πλεονασμός στη διανομή ισχύος IT
- N + 1 πλεονασμός στο UPS
- N + 1 πλεονασμός στη Γεννήτρια
- N + 1 πλεονασμός στον εξοπλισμό CRAH, Pumps, Chiller και Heat Rejection

Ο Σχεδιασμός προϋποθέτει και περιλαμβάνει:

- 17 Racks (υποθέτοντας πυκνότητα ισχύος 6kW ανά rack σε κέντρο δεδομένων με δυνατότητα 100kW)
- Σχέδιο υπερυψωμένου δαπέδου
- Εξοπλισμός ανίχνευσης και καταστολής πυρκαγιάς
- Διακοπτικό υλικό και πίνακες
- Οροφή ασφαλείας.

Το κόστος περιλαμβάνει:

- Το κόστος του κτιρίου (πυρήνας και κέλυφος)
- Τον σχεδιασμό και το κόστος των μηχανικών
- Το εργατικό κόστος
- Το κόστος εγκατάστασης
- Το κόστος διαχείρισης έργου

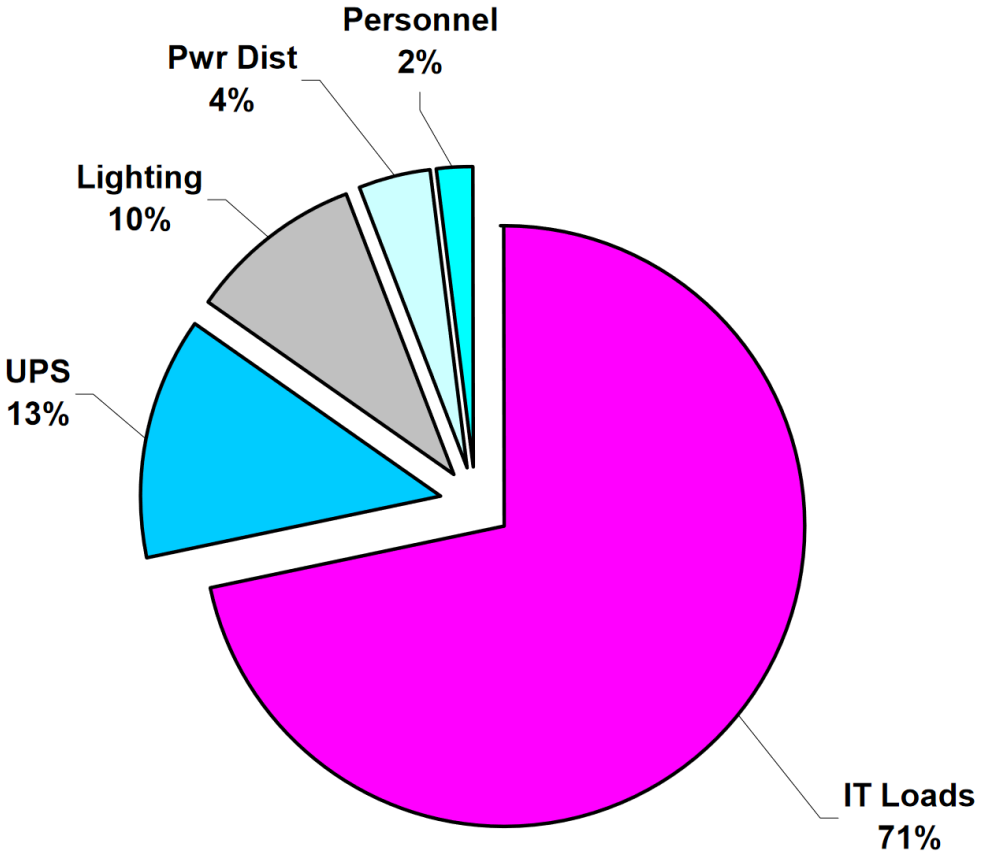
Το αναλυτικό κόστος μπορεί να βρεθεί στον ακόλουθο πίνακα, κατανεμημένο σε τρεις κύριες κατηγορίες (Ισχύς, Ψύξη και Άλλα, σύμφωνα με τον υπολογιστή κόστους της Schneider Electric), και αυτές χωρίζονται περαιτέρω σε υποκατηγορίες. Τα ποσοστά των υποκατηγοριών ανέρχονται στο 100% του συνόλου της κατηγορίας.

	Εξάρτημα	Τοις εκατό	Κόστος
Ισχύς		**39,0%**	**331'110,00 €**
εκ των οποίων	UPS	36,0%	119'199,60 €
	Γεννήτρια	27,0%	89'399,70 €
	Αλλαγή εξοπλισμού	21,0%	69'533,10 €
	Κατανομή κρίσιμης ισχύος	16,0%	52'977,60 €
Ψύξη		**28,0%**	**237'720,00 €**
εκ των οποίων	CRAH	13,0%	30'903,60 €
	Chiller	34,0%	80'824,80 €
	Πύργος ψύξης	13,0%	30'903,60 €
	Αντλίες CHW, σωληνώσεις, βαλβίδες	20,0%	47'544,00 €
	Αντλίες CW, σωληνώσεις, βαλβίδες	20,0%	47'544,00 €
Αλλα		**33,0%**	**280'170,00 €**
εκ των οποίων	Πυροπροστασία	4,5%	12'607,65 €
	Ανυψωμένο δάπεδο / οροφή με πτώση	4,9%	13'728,33 €
	Περιβλήματα IT + περιορισμός	13,3%	37'262,61 €
	Διαχείριση & Ασφάλεια	10,1%	28'297,17 €
	Φωτισμός	2,2%	6'163,74 €
	Project Mgmt & Facility Engineering	23,5%	65'839,95 €
	Πυρήνας & κέλυφος	41,5%	116'270,55 €
		Συνολικό κόστος:	**849'000,00 €**

Πίνακας 7 : Ανάλυση CAPEX Data Center

Κόστος ισχύος

Ο προηγούμενος υπολογισμός ισχύος των 87 kW ήταν η μέγιστη αναμενόμενη ισχύ από την υποδομή υπολογιστικής ισχύος και δικτύου. Φυσικά, ο εξοπλισμός δεν θα αντλεί συνεχώς το 100% της ονομαστικής ισχύος του, αλλά θα πλησιάζει το 30% κατά μέσο όρο, ενώ η συνολική κατανάλωση ενέργειας και η απαγωγή θερμότητας της υποδομής και διανομής ισχύος θα είναι 50% υψηλότερη από την κατανάλωση εξοπλισμού υπολογιστικής ισχύος και δικτύου, όπως φαίνεται στο παρακάτω σχήμα(Rasmussen, 2007) .



Εικόνα 17 : Σχετικές συνεισφορές στη συνολική θερμική απόδοση ενός τυπικού κέντρου δεδομένων

Έτσι, η συνολική απαγωγή θερμότητας και η κατανάλωση ισχύος του εξοπλισμού κέντρων δεδομένων (εξαιρουμένης της ψύξης) θα είναι (στρογγυλοποιημένη στο πρώτο δεκαδικό)

$$87kW \times 30\% = 26{,}1\ kW$$
$$26{,}1kW \times 150\% = 39{,}2\ kW$$

Εξίσωση 1 : Μέση κατανάλωση ισχύος

Ενώ η απαιτούμενη ισχύς ψύξης θα είναι 1,3 φορές η μέση κατανάλωση ισχύος:

$$39{,}2kW \times 130\% = 50{,}9\ kW$$

Εξίσωση 2 : Ψύξη Μέση κατανάλωση ισχύος

Και έτσι, η συνολική κατανάλωση και το κόστος για μια περίοδο 5 ετών φαίνεται στον παρακάτω πίνακα.

	Ισχύ	Ώρες σε 5YR	kWh σε 5YR	κόστος kWh	Μηνιαία τέλη	Κόστος 5YR
Υποδομή	39,2 kW	43'800	1'716'960 KWh	0,10	350,9968	192'755,81 €
Ψύξη	50,9 kW	43'800	2'229'420 KWh	0,10	455,7586	250'287,52 €
					Σύνολο	**443'043,32 €**

Πίνακας 8 : Κατανάλωση και κόστος ισχύος για περίοδο 5 ετών

Το κόστος ανά kWh καθώς και το μηνιαίο κόστος υπολογίστηκαν με βάση τον επίσημο τιμοκατάλογο του μεγαλύτερου τοπικού παρόχου και διανομέα ρεύματος κατά τη στιγμή της γραφής(DEI, 2020) .

Προσωπικό

Το ελάχιστο προσωπικό που θα απαιτείται για τη λειτουργία του κέντρου δεδομένων είναι το εξής:

- Μηχανικοί συστήματος / Διαχειριστές
- Διαχείριση εγκατάστασης:
    - Προσωπικό Ασφαλείας
    - Προσωπικό Υγιεινής

Για τους Μηχανικούς του Συστήματος και τους Διαχειριστές, οι απαιτήσεις αναφέρονται σε άτομα με υψηλή κατάρτιση και εμπειρία, με δύο άτομα να βρίσκονται στις εγκαταστάσεις κατά τη διάρκεια των ωρών εργασίας, τουλάχιστον ένα να βρίσκεται στις εγκαταστάσεις κατά τη διάρκεια εκτεταμένων ωρών εργασίας (ώρες εργασίας των φυσικών καταστημάτων όταν τα γραφεία δεν λειτουργούν - απογεύματα και Σαββατοκύριακα), και ένα κατά τις ώρες εκτός ωραρίου.

Έτσι, κατά τη διάρκεια μιας εβδομάδας θα χρειαστεί ο ακόλουθος αριθμός ανθρωποωρών:

$$2 \times 40 + 1 \times 40 + 1 \times 40 + 24 + 24 = 208\ MH$$

Εξίσωση 3 : Απαιτούμενες ανθρωποώρες σε μια εβδομάδα για SE / SA

Πρέπει επίσης να λάβουμε υπόψη τις ημέρες άδειας και τις κυλιόμενες βάρδιες, οπότε το ελάχιστο προσωπικό πλήρους ωραρίου που απαιτείται για τους Μηχανικούς / Διαχειριστές Συστήματος είναι έξι.

Για το προσωπικό διαχείρισης εγκαταστάσεων, χρειαζόμαστε τέσσερις υπαλλήλους ασφαλείας πλήρους απασχόλησης και ένα προσωπικό υγιεινής μερικής απασχόλησης.

Το μέσο κόστος ενός άρτια εκπαιδευμένου Μηχανικού Συστήματος / Διαχειριστή είναι 50'000 € ετησίως, λαμβάνοντας υπόψη τις κυλιόμενες βάρδιες και την εργασία κατά τη διάρκεια των Σαββατοκύριακων. Το μέσο κόστος για το προσωπικό διαχείρισης εγκαταστάσεων είναι 30'000 € ετησίως, λαμβάνοντας υπόψη και τις κυλιόμενες βάρδιες και την εργασία κατά τη διάρκεια των Σαββατοκύριακων.

Έτσι, το κόστος του προσωπικού, για μια περίοδο πέντε ετών φαίνεται στον παρακάτω πίνακα.

Προσωπικό	Αριθμός ατόμων	Κόστος / άτομο / έτος	Συνολικό κόστος 5 YR
Μηχανικοί συστημάτων	6	50'000,00 €	1'500'000,00 €
Διαχείριση εγκατάστασης	4,5	30'000,00 €	675'000,00 €
		Σύνολο	**2'175'000,00 €**

Πίνακας 9 : Κόστος προσωπικού για περίοδο 5 ετών

Συνολικό κόστος εγκατάστασης

Όπως αναφέρθηκε προηγουμένως, θα παρέχουμε ανάλυση και σύγκριση για δύο περιπτώσεις, με ή χωρίς το κόστος CAPEX του Data Center, και στην πρώτη περίπτωση, θα συμπεριλάβουμε περίοδο απόσβεσης 5 ετών και 10 ετών. Λαμβάνοντας υπόψη όλα τα παραπάνω, το συνολικό κόστος για το σενάριο εγκατάστασης σε ιδιόκτητο κέντρο δεδομένων φαίνεται στον παρακάτω πίνακα.

	Κόστος
Υποδομή	5'720'278,57 €
Κατανάλωση ενέργειας	443'043,32 €
Προσωπικό	2'175'000,00 €
Μερικό Σύνολο	**8'338'321,90 €**
DC Capex	849'000,00 €
Σύνολο με 50% Capex	**8'762'821,90 €**
Σύνολο με 100% Capex	**9'187'321,90 €**

Πίνακας 10 : Συνολικό κόστος εγκατάστασης για περίοδο 5 ετών

Το «υποσύνολο» στον πίνακα απεικονίζει το κόστος χωρίς το CAPEX κόστος του κέντρου δεδομένων, το "Σύνολο με 50% Capex" απεικονίζει μια περίοδο απόσβεσης 10 ετών για το Data Center, ενώ το "Σύνολο με 100% Capex" απεικονίζει 5ετή περίοδο απόσβεσης.

b. Υποδομή που παρέχεται από το Cloud

Στην περίπτωση της υποδομής που παρέχεται από το cloud, το κόστος αυξάνεται σχεδόν γραμμικά με το μέγεθος των VMs, οπότε δεν θα διερευνήσουμε διαφορετικά σενάρια μεγέθους, αλλά θα παρέχουμε το κόστος για 1000 VM με 4vCPU, 16 GB RAM και 500 GB αποθηκευτικού χώρου.

Η AWS έχει διαθέσει πολλά διαφορετικά μοντέλα τιμολόγησης, ανάλογα με τους τύπους και τις περιόδους δέσμευσης. Τα μοντέλα τιμολόγησης που θα διερευνήσουμε είναι:

- **AWS On Demand** : Πιο ευέλικτο και ακριβότερο, χωρίς περίοδο δέσμευσης.

- **AWS 1 έτος - Χωρίς προκαταβολή** : Ενοικίαση συγκεκριμένων τύπων δομών για 1 έτος χωρίς προκαταβολικό κόστος.
- **AWS 1 έτος - Πλήρης πληρωμή εκ των προτέρων** : Ενοικίαση συγκεκριμένων τύπων δομών για 1 έτος και προκαταβολή ολόκληρου το κόστους εκ των προτέρων.

- **AWS 3 ετών - Χωρίς προκαταβολή** : Ενοικίαση συγκεκριμένων τύπων δομών για 3 χρόνια χωρίς προκαταβολικό κόστος.
- **AWS 3 ετών - Πλήρης πληρωμή εκ των προτέρων** : Ενοικίαση συγκεκριμένων τύπων δομών για 3 έτη και προκαταβολή ολόκληρου το κόστους εκ των προτέρων.

- **AWS Compute Savings 3 ετών - Χωρίς προκαταβολή** : Το πρόγραμμα Compute Savings είναι ένα ευέλικτο μοντέλο τιμολόγησης που προσφέρει χαμηλές τιμές σε χρήση EC2, Lambda και Fargate, με αντάλλαγμα τη δέσμευση για ένα σταθερό ποσό χρήσης (μετρημένο σε $ / ώρα) για τριετή δέσμευση. Σε αυτήν την περίπτωση δεσμευόμαστε για μια τριετή δαπάνη ενός συγκεκριμένου ποσού υπολογιστικών πόρων, χωρίς κόστος εκ των προτέρων.
- **AWS Compute Savings 3 ετών - Πλήρης εκ των προτέρων** : Σε αυτήν την περίπτωση δεσμευόμαστε για μια τριετή δαπάνη ενός συγκεκριμένου ποσού υπολογιστικών πόρων, με πλήρες προκαταβολικό κόστος.

Οι δεσμευμένες δομές (Reserved Instances) είναι ανελαστικές καθώς μας περιορίζουν στον τύπο των δομών που μπορούμε να χρησιμοποιήσουμε και ενώ αυτό μπορεί να είναι αποδεκτό για σύντομο

χρονικό διάστημα (1 έτος) είναι μη αποδεκτό για μεγαλύτερες περιόδους, καθώς η ελαστικότητα είναι ένα κεντρικό προτέρημα του cloud, και τέτοιες δεσμεύσεις δεν έχουν νόημα μακροπρόθεσμα, αλλά παρέχουν την υψηλότερη εξοικονόμηση κόστους, επομένως θα συμπεριληφθούν και θα συγκριθούν με τα υπόλοιπα σχέδια εξοικονόμησης.

Το κόστος φαίνεται στον παρακάτω πίνακα, έχοντας μετατραπεί από Δολάρια ΗΠΑ χρησιμοποιώντας συναλλαγματική ισοτιμία USD σε EUR 0,84 USD ανά ευρώ, και έχει αναχθεί για περίοδο χρήσης 5 ετών.

Μοντέλο τιμολόγησης AWS	3YR σε USD	3YR σε EUR	5YR σε EUR	Εξοικονόμηση συγκριτικά με το «on-demand»
AWS κατ 'απαίτηση	$ 7'116'876,00	5'978'175,84 €	9'963'626,40 €	0%
AWS 1YR No Upfront	$ 5'253'624,00	4'413'044,16 €	7'355'073,60 €	26%
AWS 1YR Πλήρης εκ των προτέρων	$ 5'040'756,00	4'234'235,04 €	7'057'058,40 €	29%
AWS 3YR No upfront	$ 4'254'984,00	3'574'186,56 €	5'956'977,60 €	40%
AWS 3YR Πλήρης εκ των προτέρων	$ 3'973'788,00	3'337'981,92 €	5'563'303,20 €	44%
Compute Savings Plan 3YR No Upfront	$ 4'993'452,00	4'194'499,68 €	6'990'832,80 €	30%
Compute Savings Plan 3YR Full Upfront	$ 4'725'396,00	3'969'332,64 €	6'615'554,40 €	34%

Πίνακας 11 : Σχέδια και Κόστος Τιμολόγησης AWS

Όπως φαίνεται παραπάνω, ακόμη και η ελάχιστη δέσμευση ενός έτους μπορεί να προσφέρει σημαντική εξοικονόμηση κόστους (26%), ενώ η υψηλότερη εξοικονόμηση μπορεί να επιτευχθεί επιλέγοντας μια 3ετή δέσμευση σε reserved instances (40%). Το πρόγραμμα Compute Savings παρέχει μια καλή ισορροπία ευελιξίας και εξοικονόμησης (30%) και μπορούμε να δούμε ξεκάθαρα ότι η προπληρωμή ολόκληρης της κατανάλωσης δεν παρέχει σημαντικό όφελος (4% ή λιγότερο) και θα πρέπει να λαμβάνεται υπόψη μόνο σε περιπτώσεις όπου η πλήρης καταβολή μπορεί να μετατραπεί σε CAPEX αντί για OPEX, κάτι που θα παρουσιαστεί σε παρακάτω ενότητα.

c. Σύγκριση AWS με Ιδιόκτητες Εγκαταστάσεις

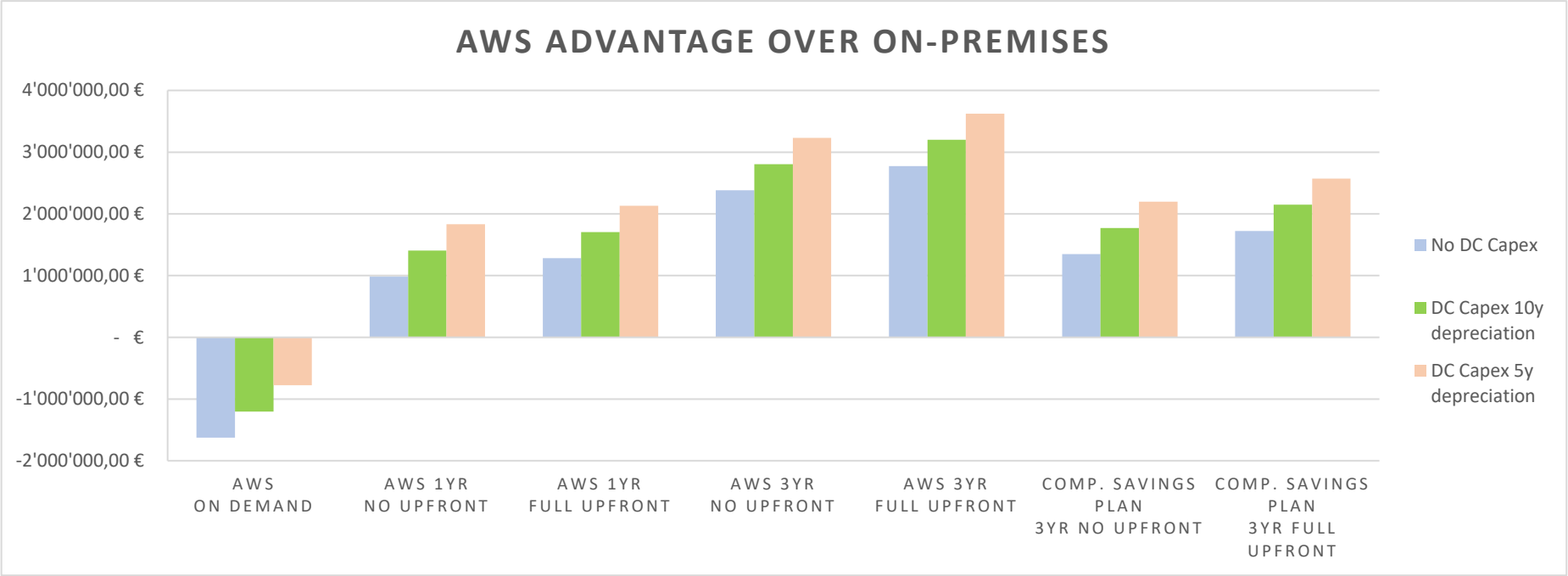
Στον ακόλουθο πίνακα και διάγραμμα, παρουσιάσαμε τη σύγκριση μεταξύ υποδομής που φιλοξενείται από AWS και ιδιόκτητου κέντρου δεδομένων σε μορφή "Πλεονέκτημα του AWS έναντι Ιδιόκτητων Εγκαταστάσεων", που σημαίνει ότι όλοι οι αρνητικοί αριθμοί αντιπροσωπεύουν το πλεονέκτημα του ιδιόκτητου κέντρου δεδομένων έναντι του AWS cloud, ενώ οι θετικοί αριθμοί αντιπροσωπεύουν το πλεονέκτημα του AWS cloud έναντι του ιδιόκτητου κέντρου δεδομένων, με το κόκκινο να είναι το πιο πλεονεκτικό για το On-Premises και το πράσινο το πιο πλεονεκτικό για το AWS.

	AWS On Demand	AWS 1YR No upfront	AWS 1YR Full upfront	AWS 3YR No upfront	AWS 3YR Full upfront	Comp. Savings Plan 3YR No upfront	Comp. Savings Plan 3YR Full upfront
Χωρίς DC Capex	-19,49%	11,79%	15,37%	28,56%	33,28%	16,16%	20,66%
DC Capex 10Y depreciation	-13,70%	16,07%	19,47%	32,02%	36,51%	20,22%	24,50%
DC Capex 5Y depreciation	-8,45%	19,94%	23,19%	35,16%	39,45%	23,91%	27,99%

Πίνακας 12 : Πλεονέκτημα του AWS έναντι των ιδιόκτητων (τοις εκατό)

		AWS On Demand	AWS 1YR No upfront	AWS 1YR Full upfront	AWS 3YR No upfront	AWS 3YR Full upfront	Comp. Savings Plan 3YR No upfront	Comp. Savings Plan 3YR Full Upfront
		9'963'626,40 €	7'355'073,60 €	7'057'058,40 €	5'956'977,60 €	5'563'303,20 €	6'990'832,80 €	6'615'554,40 €
Χωρίς DC Capex	8'338'321,90 €	-1'625'304,50 €	983'248,30 €	1'281'263,50 €	2'381'344,30 €	2'775'018,70 €	1'347'489,10 €	1'722'767,50 €
DC Capex 10Y depreciation	8'762'821,90 €	-1'200'804,50 €	1'407'748,30 €	1'705'763,50 €	2'805'844,30 €	3'199'518,70 €	1'771'989,10 €	2'147'267,50 €
DC Capex 5Y depreciation	9'187'321,90 €	-776'304,50 €	1'832'248,30 €	2'130'263,50 €	3'230'344,30 €	3'624'018,70 €	2'196'489,10 €	2'571'767,50 €

Πίνακας 13 : Πλεονέκτημα του AWS έναντι των ιδιόκτητων εγκαταστάσεων.



Εικόνα 18 : Πλεονέκτημα του AWS έναντι των ιδιόκτητων εγκαταστάσεων.

Όπως μπορούμε να δούμε ξεκάθαρα στους παραπάνω πίνακες, η υποδομή Cloud έχει ένα πλεονέκτημα έναντι της υποδομής ιδιόκτητων κέντρων δεδομένων, με εξαίρεση τις περιπτώσεις κατά παραγγελία (on-demand).

Οι «Κατά περίπτωση δομές» είναι υπηρεσίες pay-as-you-go που είναι κατάλληλες για βραχυπρόθεσμους απρόβλεπτους φόρτους εργασίας (Δοκιμές, Proof-of-concept (POC)) και σε καμία περίπτωση δεν πρέπει να χρησιμοποιούνται για προβλέψιμους, μακροπρόθεσμους φόρτους εργασίας. Για αυτά τα είδη φόρτου εργασίας, ενώ η αρχική εγκατάσταση θα χρησιμοποιεί δομές κατ' απαίτηση, αφότου ο φόρτος εργασίας έχει διευκρινιστεί και η ζήτηση έχει σταθεροποιηθεί, μπορούν να αγοραστούν «δεσμευμένες δομές», που θα αντιστοιχούν στην πραγματική ανάγκη ή «πλάνο εξοικονόμησης υπολογισμού» που θα παρέχει ένα μεγαλύτερο επίπεδο ευελιξίας.

Επομένως, συνιστάται ιδιαίτερα να εκτελούμε μια λεπτομερή διαδικασία βελτιστοποίησης κόστους αφού σταθεροποιηθούν οι εργασίες και να χρησιμοποιούμε εργαλεία αυτόματης κλιμάκωσης ή πρόβλεψης ζήτησης για να διατηρήσουμε το κόστος βελτιστοποιημένο.

Συγκρίναμε τη φιλοξενία VM (VM hosting), που είναι ο πιο βασικός και άμεσα συγκρίσιμος πόρος, μεταξύ του Cloud και του ιδιόκτητου κέντρου δεδομένων και έχουμε αποδείξει ότι το Cloud κατέχει ένα σαφές πλεονέκτημα. Αυτό το πλεονέκτημα γίνεται ακόμη πιο έντονο αν λάβουμε υπόψη εφαρμογές και υπηρεσίες χωρίς διακομιστές, που παρέχουν την υψηλότερη διακριτότητα κόστους.

Υπηρεσίες όπως Load Balancers, Functions, Gateways θα χρεώνονται ανά δευτερόλεπτο ή ανά όγκο μεταφοράς δεδομένων, θα κλιμακώνονται αυτόματα με διαφάνεια προς τον καταναλωτή, ανάλογα με τη ζήτηση και θα προσφέρουν υψηλή διαθεσιμότητα εξ’ ορισμού, κάτι που θα απαιτούσε μεγάλη προσπάθεια και κόστος σε ιδιόκτητο περιβάλλον.

Τέλος, σχεδόν όλες οι υπηρεσίες παρέχουν πρακτικά απεριόριστες επιλογές κλιμάκωσης, που είναι ουσιαστικά άμεσες, επίσης έχουν την δυνατότητα να διακοπούν όταν δεν χρησιμοποιούνται πλέον, ελαχιστοποιώντας το κόστος. Αντιθέτως, η ιδιόκτητη υποδομή πρέπει να σχεδιαστεί προβλέποντας την υψηλότερη ζήτηση και να διατηρείται σε λειτουργία πλήρους χωρητικότητας, είτε χρησιμοποιείται πλήρως είτε όχι. Αυτό γίνεται εμφανές κατά τη χρήση εφαρμογών Cloud Native, όπου η ελαστικότητα και η χρήση υπηρεσιών χωρίς διακομιστές (PaaS) αποτελούν μέρος του σχεδιασμού.

CAPEX και OPEX

Η τεχνολογία cloud επέφερε πρωτοποριακές αλλαγές στον τρόπο που επενδύουμε, επομένως και στον τον τρόπο δημιουργίας επιχειρηματικών στρατηγικών και προϋπολογισμών.

Αρχικά, ας καθορίσουμε την έννοια των CapEx και OpEx:

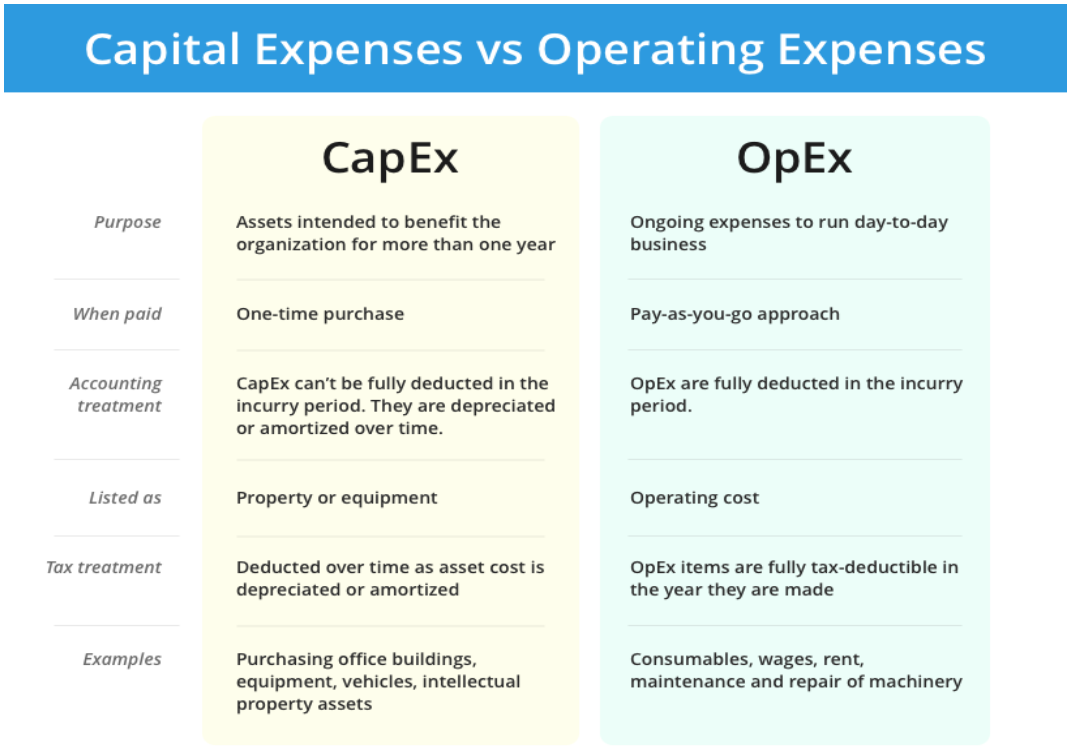
- **Οι κεφαλαιουχικές δαπάνες (CapEx)** είναι κεφάλαια που χρησιμοποιεί μια επιχείρηση για την απόκτηση, αναβάθμιση και συντήρηση φυσικών μακροπρόθεσμων περιουσιακών στοιχείων όπως γη, εγκαταστάσεις, κτίρια και τεχνολογία. Το CapEx χρησιμοποιείται συχνά από επιχειρήσεις για τη χρηματοδότηση νέων έργων ή επενδύσεων. Οι κεφαλαιουχικές δαπάνες για πάγια περιουσιακά στοιχεία μπορεί να περιλαμβάνουν την επισκευή στέγης, την αγορά εξοπλισμού ή την κατασκευή νέου εργοστασίου. Είναι ουσιαστικά - μια πληρωμή για αγαθά ή υπηρεσίες που έχουν συνήθως ωφέλιμη ζωή ενός έτους ή περισσότερο.

- **Λειτουργικές Δαπάνες (OpEx)** είναι τα κεφάλαια που χρησιμοποιούνται για την κάλυψη του κόστους που προκύπτει από μια επιχείρηση κατά τη διάρκεια συνήθων επιχειρηματικών λειτουργιών, όπως προμήθειες, ενοίκια, γενικά έξοδα γραφείου, μισθοί υπαλλήλων, κόστος αποθέματος, γενικά οτιδήποτε υποστηρίζει καθημερινές επιχειρηματικές λειτουργίες.

Ιστορικά, οι τεχνολογικές επενδύσεις είχαν προτεραιότητα ως κεφαλαιουχικές δαπάνες και όχι ως λειτουργικές δαπάνες, καθώς οι CFOs θα μπορούσαν να αναβάλουν το κόστος αυτό για μεγάλο χρονικό διάστημα. Σήμερα, ένας αυξανόμενος αριθμός επιχειρήσεων μετατοπίζει τις επενδύσεις πληροφορικής του από το CapEx σε OpEx μεταφέροντας την υποδομή πληροφορικής τους στο cloud. Πρόσθετα οφέλη CapEx συσσωρεύονται ως αποτέλεσμα της εταιρικής στρατηγικής που δεν απαιτεί πλέον στατικές επενδύσεις σε υλικό, λογισμικό και πόρους. (Comindware, 2021)

Οι υπηρεσίες και άλλες ιδίου τύπου επιλογές αγοράζονται με βάση τις εκάστοτε ανάγκες, με κυμαινόμενο κόστος και σε αυτές τις περιπτώσεις το OpEx λειτουργεί καλύτερα. Από την άλλη πλευρά, το CapEx απεικονίζει τις επενδύσεις που έχουν πραγματοποιηθεί από την εταιρεία και ως αποτέλεσμα, έχει μεγάλη επιρροή στην επιχειρηματική εικόνα της εταιρείας, όσον αφορά τη βιωσιμότητα και την ανάπτυξη, επομένως πολλές εταιρείες ακολουθούν την προσέγγιση της προκαταβολής. Με αυτήν την επιλογή, οι εταιρείες μπορούν να δικαιολογήσουν το κόστος ενοικίασης της υποδομής, ως επένδυση και να αυξήσουν ανάλογα το CapEx, καθώς τα «αγαθά» που αγοράζουν έχουν περισσότερο από ένα χρόνο χρήσιμου κύκλου ζωής προκειμένου να θεωρηθούν ως επένδυση.

Πιο συγκεκριμένα, υπάρχουν δύο επιλογές για την ιδιοκτησία μακροπρόθεσμης υποδομής, όπως οι εικονικές μηχανές EC2 και αποκλειστικοί διακομιστές (dedicated host servers), τα οποία είναι συμβόλαια ενός έτους ή τριών ετών. Αυτή η προσέγγιση, όπως δείξαμε στο προηγούμενο κεφάλαιο, οδηγεί σε έκπτωση κόστους 5-10% κατ' ελάχιστο και είναι συνήθως ευεργετικό για βαριά φορτία εργασίας που πρέπει να είναι ενεργά 24/7, καθώς πληρώνουμε μόνο για τον χώρο αποθήκευσης και τις υπηρεσίες που εκτελούνται σε αυτά τα μηχανήματα, όχι τη χρήση του μηχανήματος. Καθώς περισσότερες επιχειρήσεις απομακρύνονται από την παραδοσιακή ιδιοκτησία υλικού (hardware) και λογισμικού (software) και επιλέγουν προϊόντα ως υπηρεσία (as-a-service), τα τμήματα πληροφορικής και οικονομικών πρέπει να συμφωνήσουν για τον καλύτερο τρόπο ταξινόμησης του κόστους cloud.



Εικόνα 19 : CapEx εναντίον OpEx (Comindware)

Σχέδιο μετεγκατάστασης (Migration Plan)

Ο προγραμματισμός μετεγκατάστασης είναι υψίστης σημασίας για έναν πάροχο υπηρεσιών, καθώς οποιαδήποτε διακοπή λειτουργίας, οποιαδήποτε στιγμή της ημέρας μπορεί να οδηγήσει σε απώλεια εσόδων. Ένας πάροχος υπηρεσιών πρέπει να σχεδιάσει προσεκτικά οποιαδήποτε μετεγκατάσταση για να επιτρέψει έναν ελάχιστο χρόνο διακοπής λειτουργίας, με ελάχιστο αντίκτυπο και να λάβει υπόψη αξιόπιστα και γρήγορα σενάρια επαναφοράς σε περίπτωση αποτυχίας.

Θα διαχωρίσουμε τον προγραμματισμό μετεγκατάστασης στα ακόλουθα στάδια:

1. **Αξιολόγηση Εφαρμογής / Μετεγκατάστασης Υπηρεσίας** : Κάθε υπηρεσία / εφαρμογή αξιολογείται προκειμένου να επιλεχθεί μια κατάλληλη στρατηγική.
2. **Προσδιορισμός εξαρτήσεων**: Θα παρουσιάσουμε μεθόδους για τον εντοπισμό των εξαρτήσεων μεταξύ υπηρεσιών και θα αναλύσουμε τον τρόπο με τον οποίο αυτές οι εξαρτήσεις επηρεάζουν το πρόγραμμα μετεγκατάστασης.
3. **Σχέδιο για ασφαλή μεταφορά δεδομένων** : Θα αναλύσουμε στρατηγικές για την ασφαλή και αξιόπιστη μετεγκατάσταση δεδομένων στο cloud.
4. **Σχεδιασμός για μηδενικό χρόνο διακοπής**: Θα διερευνήσουμε στρατηγικές που θα παρέχουν μηδενικό χρόνο διακοπής κατά τη μετεγκατάσταση των κρίσιμων υπηρεσιών (Mission Critical).
5. **Σχέδιο μετεγκατάστασης**: Θα συνοψίσουμε και θα αξιολογήσουμε το σχέδιο μετεγκατάστασης.
6. **Κόστος**: Θα απαριθμήσουμε τους παράγοντες που θα έχουν μικρό ή σημαντικό αντίκτυπο στο κόστος.

Οι στρατηγικές μετεγκατάστασης θα διαμορφωθούν γύρω από τα «6 R»(AWS, 2018) :

1. **Re-host (αναφέρεται και ως “lift and shift.”) :**

   Μεταφορά εφαρμογών χωρίς τροποποίηση. Οι οργανισμοί που υποβάλλονται σε μεγάλες μετακινήσεις παλαιού τύπου εφαρμογών πρέπει να κινηθούν γρήγορα προκειμένου να επιτύχουν τους επιχειρηματικούς τους στόχους. Οι περισσότερες από αυτές τις εφαρμογές φιλοξενούνται εκ νέου σε διακομιστές τρίτων. Στην προηγούμενη ανάλυση κόστους

αποδείξαμε ότι θα μπορούσαμε να εξοικονομήσουμε περίπου το 30% των δαπανών μέσω της επανεγκατάστασης.

Η πλειονότητα των εργασιών επανεγκατάστασης μπορεί να αυτοματοποιηθεί χρησιμοποιώντας εργαλεία (για παράδειγμα, AWS VM Import/Export). Ορισμένοι καταναλωτές προτιμούν να το κάνουν χειροκίνητα καθώς εξοικειώνονται με τη νέα πλατφόρμα cloud αλλά και με την προσαρμογή των παλαιών τους συστημάτων.

Όταν οι εφαρμογές έχουν μεταφερθεί στο cloud, είναι ευκολότερο να βελτιστοποιηθούν / επανασχεδιαστούν. Εν μέρει επειδή, ο οργανισμός θα έχει αποκτήσει τις απαραίτητες δεξιότητες και εν μέρει επειδή το δύσκολο μέρος - μετεγκατάσταση της εφαρμογής, δεδομένων και κυκλοφορίας - θα έχει ήδη ολοκληρωθεί.

2. **Re-platform (Αναφέρεται και ως "lift, tinker, and shift."):**

Με αυτή την τεχνική , πραγματοποιείται μια βελτιστοποιήση μερικών μόνο στοιχείων στο cloud, χωρίς όμως να αλλάξει η βασική αρχιτεκτονική της εφαρμογής. Τα οφέλη που αποκομίζονται σε αυτή την περίπτωση είναι, για παράδειγμα, η μείωση του χρόνου διαχείρισης δομών βάσης δεδομένων, μεταβαίνοντας σε μια πλατφόρμα βάσης δεδομένων ως υπηρεσία (DbaaS) όπως το Amazon Relational Database Service (Amazon RDS) ή μια πλήρως διαχειριζόμενη πλατφόρμα όπως το AWS Elastic Beanstalk (PaaS).

3. Re-factor / Re-architect :

Χρησιμοποιώντας δυνατότητες cloud native, επαναδημιουργούμε τον τρόπο με τον οποίο σχεδιάζεται και αναπτύσσεται η εφαρμογή. Αυτό οφείλεται σε μια επιτακτική ανάγκη για πρόσθεση χαρακτηριστικών κλιμάκωσης ή απόδοσης στην εφαρμογή που θα ήταν δύσκολο να επιτευχθεί στο τρέχον περιβάλλον της.

Ένα παράδειγμα είναι η μετάβαση από μια μονολιθική (monolithic architecture) σε μια αρχιτεκτονική προσανατολισμένη σε υπηρεσίες (Service Oriented Architecture - SOA) προκειμένου να αυξηθεί η ευελιξία και η επιχειρησιακή συνέχεια των εταιριών. Αυτή η στρατηγική είναι συνήθως η πιο ακριβή, αλλά μπορεί επίσης να είναι η πιο επωφελής αν το προϊόν και η αγορά μας είναι ισχυρά.

4. **Re-purchase:**

   Αγορά ενός μοντέλου λογισμικού ως υπηρεσία (SaaS) αντί των αδειών λογισμικού διάρκειας. Για παράδειγμα, η μετάβαση από ένα σύστημα διαχείρισης σχέσεων πελατών (CRM) στο Salesforce, από ένα σύστημα διαχείρισης ανθρώπινου δυναμικού (HR) στο Workday ή από ένα σύστημα διαχείρισης περιεχομένου (CMS) στο Drupal.

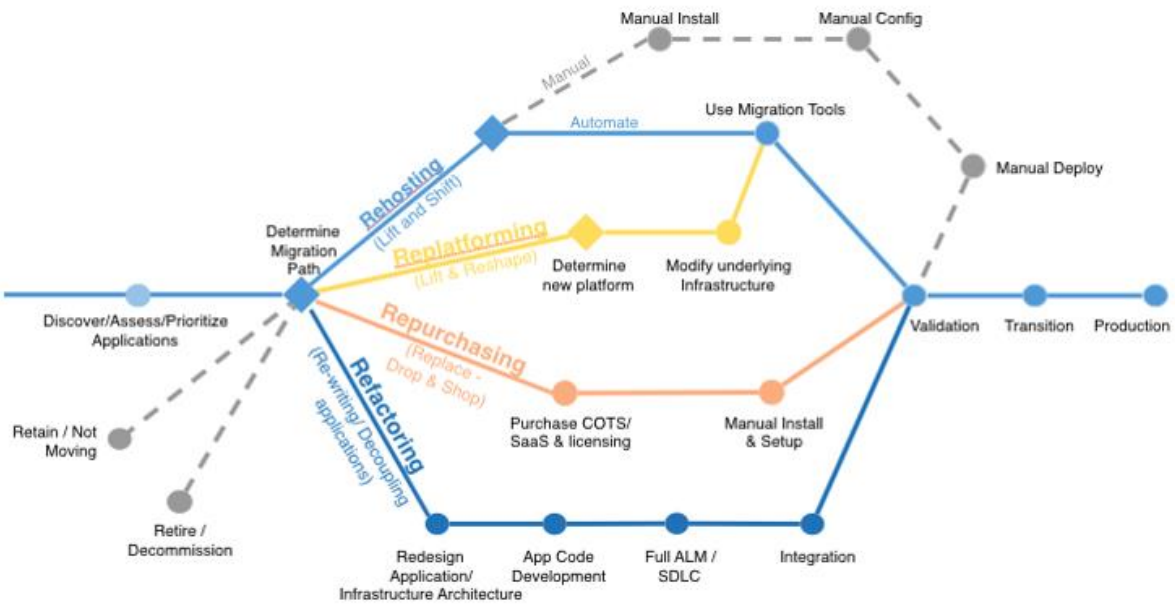
5. **Retire:**

   Εξάλειψη των μη χρησιμοποιούμενων εφαρμογών. Μόλις εντοπιστούν τα περιβάλλοντά μας, πρέπει να προσδιορίσουμε σε ποιόν ανήκει η κάθε εφαρμογή. Το 10% - 20% του χαρτοφυλακίου πληροφορικής μιας επιχείρησης δεν είναι πλέον απαραίτητο και μπορεί να παροπλιστεί. Αυτές οι εξοικονομήσεις μπορούν να μας βοηθήσουν να ενισχύσουμε την επιχειρηματική μας δραστηριότητα, να επαναπροσδιορίσουμε τις προσπάθειες της ομάδας σε σημαντικές εφαρμογές και να μειώσουμε τον αριθμό των εφαρμογών που πρέπει να ασφαλιστούν.

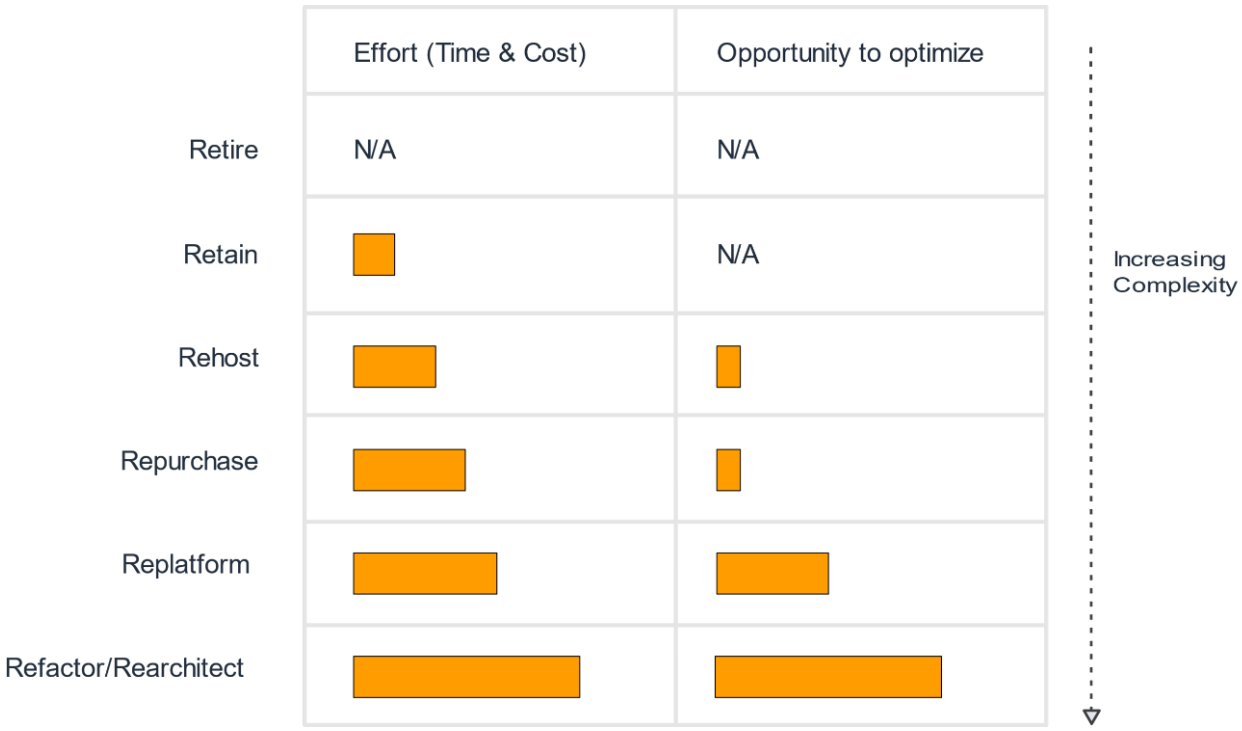
6. **Retain (Αναφέρεται και ως re-visit.):**

   Διατήρηση κρίσιμων εφαρμογών που απαιτούν εκτεταμένη αναδιαμόρφωση πριν μπορέσουν να μετεγκατασταθούν στο cloud. Μπορούμε να αναλύσουμε εκ νέου όλες τις εφαρμογές αυτής της κατηγορίας σε δεύτερο χρόνο. Αυτό περιλαμβάνει την κατηγορία που έχει συζητηθεί στην ενότητα «Υπηρεσίες που θα παραμείνουν στο ιδιόκτητο κέντρο δεδομένων».

Στα ακόλουθα σχήματα παρέχουμε ένα διάγραμμα υψηλού επιπέδου (HLD) των στρατηγικών καθώς και μια σύγκριση όσον αφορά το κόστος, την προσπάθεια και την πολυπλοκότητα τους.



Εικόνα 20 : The 6 R's - Διάγραμμα (AWS)



Εικόνα 21 : Σύγκριση των 6 στρατηγικών (AWS)

Στις ακόλουθες ενότητες υποθέτουμε ότι ο Οργανισμός γνωρίζει πλήρως τις υπηρεσίες του και παρόλο που θα προταθεί μια διαδικασία εντοπισμού εφαρμογών στην ενότητα «Αναγνώριση εξαρτήσεων», αναμένεται ότι η Εταιρεία διαθέτει ήδη ένα πλήρες χαρτοφυλάκιο Εφαρμογών και Υποδομών.

Φυσικά, υπάρχουν στρατηγικές εντοπισμού για την κάλυψη αυτής της ανάγκης, αλλά είναι εκτός του πεδίου αυτής της διπλωματικής.

Ο πρωταρχικός στόχος του σχεδίου μετεγκατάστασης είναι να κατευθύνει τη συνολική μετεγκατάσταση. Αυτό περιλαμβάνει τη διαχείριση της εμβέλειας του έργου, του σχεδίου πόρων, των προβλημάτων και των κινδύνων, καθώς και τον συντονισμό και την επικοινωνία με όλους τους εμπλεκόμενους στο έργο. Ο έγκαιρος προγραμματισμός μπορεί να βοηθήσει στην οργάνωση του έργου, ειδικά όταν πολλές ομάδες μετεγκαθιστούν πολλές εφαρμογές. Το πρόγραμμα μετεγκατάστασης λαμβάνει υπόψη κρίσιμους παράγοντες, όπως τη σειρά μετεγκατάστασης φόρτου εργασίας, το χρονοδιάγραμμα των απαιτήσεων για ανθρώπινους πόρους σε κάθε φάση του έργου και την παρακολούθηση της προόδου της μετεγκατάστασης. Συνιστούμε οι ομάδες να χρησιμοποιούν ευέλικτες μεθοδολογίες παράδοσης, βέλτιστες πρακτικές ελέγχου έργου, μια καλά καθορισμένη προσέγγιση παράδοσης και ένα ισχυρό επιχειρηματικό σχέδιο επικοινωνίας.

Μεταξύ των δραστηριοτήτων που προτείνονται για ένα σχέδιο μετεγκατάστασης είναι οι ακόλουθες(AWS, 2018) :

- Για τον εντοπισμό τυχόν κενών στις μεθόδους, τα εργαλεία και τις δυνατότητες διαχείρισης έργου, πραγματοποιείται μια ανασκόπηση των μεθόδων, των εργαλείων και των δυνατοτήτων.
- Ορίζουμε τις μεθόδους και τα εργαλεία για τη διαχείριση έργων που θα χρησιμοποιηθούν κατά τη μετεγκατάσταση.
- Ορίζουμε και αναπτύσσουμε τον Χάρτη (chapter)/Σχέδιο Επικοινωνίας (communication plan) του Προγράμματος μετεγκατάστασης, συμπεριλαμβανομένων διαδικασιών αναφοράς και κλιμάκωσης (escalation).
- Δημιουργούμε ένα σχέδιο έργου, ένα αρχείο καταγραφής κινδύνων και διαχείρισής τους και έναν πίνακα ρόλων και ευθυνών (π.χ. RACI chart) για να τους διαχειριστούμε τους και να τοποθετήσουμε τους υπεύθυνους για κάθε σχετικό πόρο.

- Προμηθευόμαστε και εφαρμόζουμε εργαλεία διαχείρισης έργων για να βοηθήσουμε στην παράδοση του έργου.

- Προσδιορίζουμε τους κρίσιμους πόρους και τις επαφές για καθεμία από τις ροές εργασίας μετεγκατάστασης που ορίζονται σε αυτήν την ενότητα.

- Διευκολύνουμε του συντονισμού και της εκτέλεσης των δραστηριοτήτων του σχεδίου.

- Περιγράφουμε τους πόρους, τα χρονοδιαγράμματα και το κόστος που σχετίζεται με τη μετεγκατάσταση του περιβάλλοντος προορισμού στο AWS

Αξιολόγηση μετεγκατάστασης εφαρμογής / υπηρεσίας

Η πιο σημαντική προϋπόθεση για μια επιτυχημένη μετεγκατάσταση Cloud είναι η πλήρης γνώση του χαρτοφυλακίου εφαρμογών. Ένα ολοκληρωμένο και πλήρες αποθετήριο (repository) εφαρμογών, συμπεριλαμβανομένων Εφαρμογών, Περιουσιακών στοιχείων (διακομιστών και υπηρεσιών) και διεπαφών / διασυνδέσεων, θα θέσει τον ακρογωνιαίο λίθο για το σχέδιο μετεγκατάστασης.

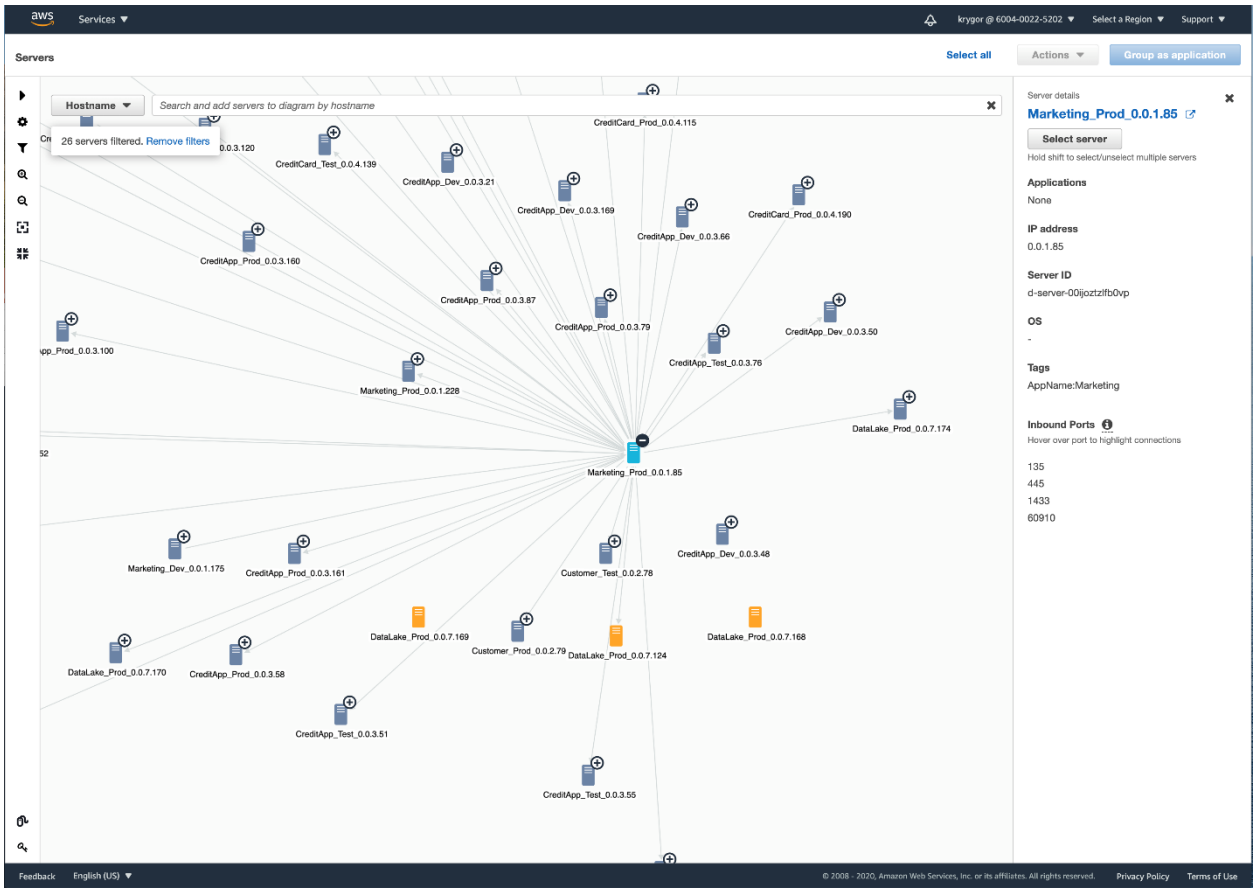
Για να δημιουργήσουμε ή να ενημερώσουμε το αποθετήριο, μπορούμε να χρησιμοποιήσουμε εργαλεία εντοπισμού που θα αυτοματοποιήσουν ολόκληρη τη διαδικασία. Τα εργαλεία μπορούν να χρησιμοποιηθούν για:

1. **Τον εντοπισμό διακομιστών και στοιχείων** : Οι συσκευές σάρωσης θα τοποθετηθούν στο δίκτυο, καλύπτοντας όλα τα τμήματα δικτύου, θα σαρώσουν όλες τις IP στο πρωτόκολλο ICMP και σε άλλες γνωστές / προσαρμοσμένες θύρες και θα ερωτήσουν τα στοιχεία που εντοπίστηκαν για περισσότερες πληροφορίες, όπως τύπος και έκδοση λειτουργικού (Operation System - OS). Αυτή θα είναι η βάση για το αποθετήριο και θα χρησιμοποιείται για την αντιστοίχιση στοιχείων σε Εφαρμογές / Υπηρεσίες, καθώς και για τον εντοπισμό και τεκμηρίωση άγνωστων στοιχείων. Οι σαρωτές μπορούν να παραμείνουν στο δίκτυο και να ελέγχουν περιοδικά για νέα στοιχεία ή άγνωστους διακομιστές (rogue servers).

2. **Τον εντοπισμό τμημάτων εφαρμογών και διασυνδέσεων**: Σε αυτό το βήμα, οι «agents» (λογισμικό που εγκαθίσταται σε διακομιστές και αναλύει τα περιεχόμενα τους) πρέπει να εγκατασταθούν σε όλους τους διακομιστές και θα σαρώσουν τις εφαρμογές και τα επιμέρους τμήματα τους, έπειτα θα τα τεκμηριώσουν στο αποθετήριο και θα παρακολουθήσουν τις συνδέσεις δικτύου για να προσδιορίσουν τη συνδεσιμότητα μεταξύ εφαρμογών, κάτι που θα διευκολύνει την ανακάλυψη των διεπαφών μεταξύ τους. Υπηρεσίες που δεν έχουν φυσικές δομές (για παράδειγμα SaaS υπηρεσίες που παρέχονται από third-parties) μπορούν επίσης να εντοπιστούν. Αυτό το βήμα πρέπει να εκτελεστεί για μεγάλο χρονικό διάστημα, προκειμένου να καταγραφούν συνδέσεις που σπάνια χρησιμοποιούνται - για παράδειγμα λογιστικές εφαρμογές που εκτελούνται μόνο στο τέλος του μήνα.

Ενώ υπάρχουν πολλά third-party εργαλεία και λύσεις για τον εντοπισμό, η AWS παρέχει το δικό της εργαλείο, την AWS Application Discovery υπηρεσία η οποία συλλέγει πληροφορίες για τις προδιαγραφές του διακομιστή (ονόματα κεντρικών υπολογιστών, διευθύνσεις IP, διευθύνσεις MAC, και την κατανομή των πόρων), τα δεδομένα απόδοσης (λεπτομέρειες αξιοποίησης των βασικών πόρων, συμπεριλαμβανομένης της CPU, του δικτύου, της μνήμης και του δίσκου), καθώς και λεπτομέρειες των διεργασιών που εκτελούνται και των συνδέσεων δικτύου.

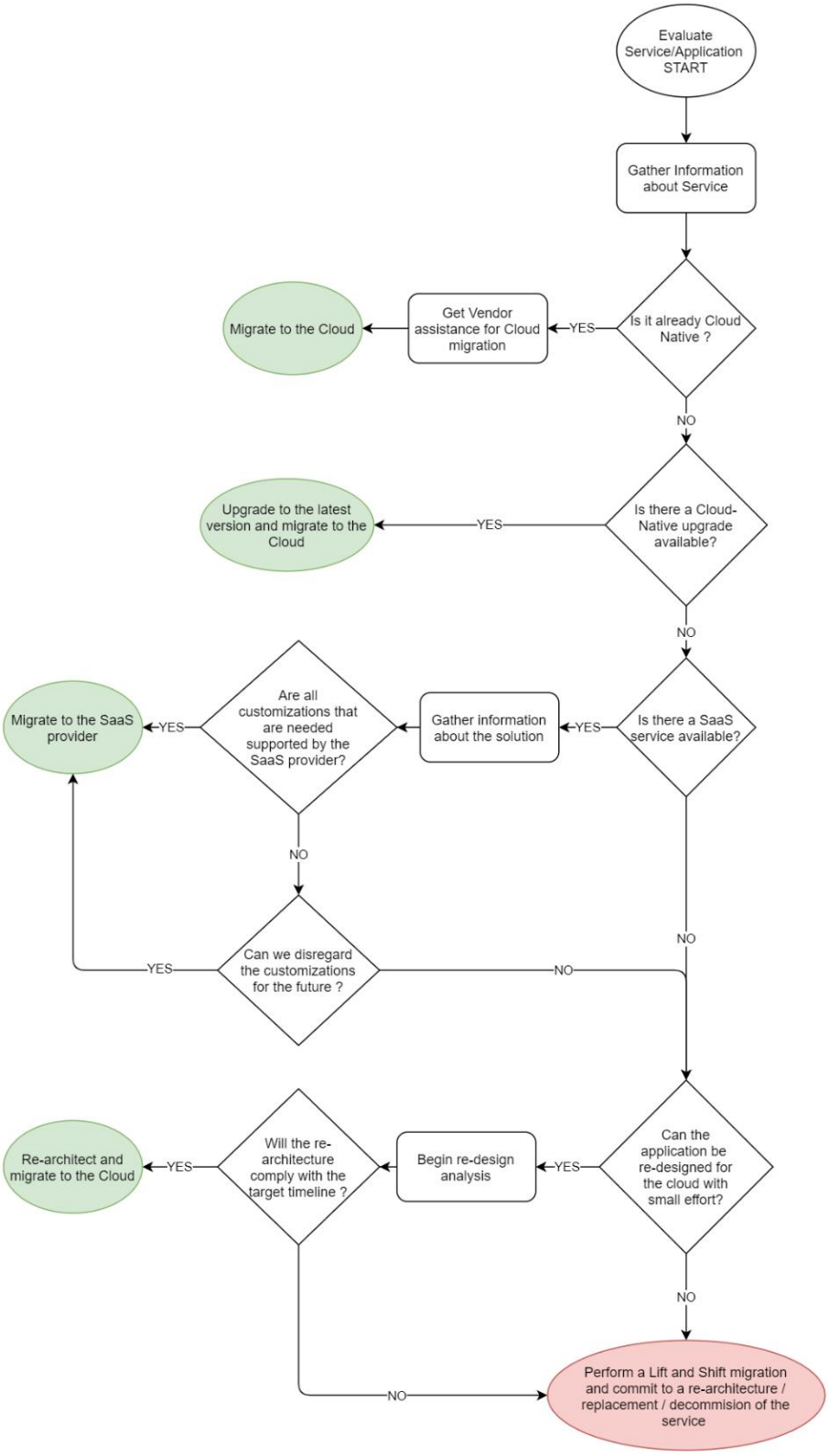
Τα δεδομένα από την υπηρεσία Application Discovery μπορούν να εισαχθούν στο AWS Migration Hub, όπου εκεί, τα στοιχεία μπορούν να ομαδοποιηθούν σε εφαρμογές και η ίδια η μετεγκατάσταση μπορεί να προγραμματιστεί και να παρακολουθηθεί. Το AWS migration hub παρέχει επίσης εργαλεία οπτικοποίησης όπως φαίνεται στην παρακάτω εικόνα.



Εικόνα 22 : Οπτικοποίηση HUB μετεγκατάστασης AWS (AWS)

Αφού ολοκληρωθεί η διαδικασία εντοπισμού, η Εταιρεία θα έχει ένα πλήρες χαρτοφυλάκιο Εφαρμογών / Υπηρεσιών και μπορεί να ξεκινήσει την αξιολόγηση των εφαρμογών. Πριν ξεκινήσει η αξιολόγηση των εφαρμογών προς μετεγκατάσταση, είναι πολύ σημαντικό να προσδιορισθούν όλες οι εφαρμογές που δεν χρησιμοποιούνται πλέον και μπορούν να αποσυρθούν. Αυτό θα μειώσει τον αριθμό των εφαρμογών που πρέπει να αξιολογηθούν και να μετεγκατασταθούν στο Cloud, και στη συνέχεια να μειώσει το κόστος και την προσπάθεια των επόμενων βημάτων.

Στο παρακάτω σχήμα, παρουσιάζουμε το διάγραμμα αποφάσεων που μπορεί να ακολουθήσει η εταιρεία για τη δημιουργία του σχεδίου μετεγκατάστασης.



Εικόνα 23 : Διάγραμμα απόφασης για μετεγκατάσταση υπηρεσίας / εφαρμογής

Το διάγραμμα απόφασης μπορεί να εξηγηθεί ως εξής:

- Εάν η εφαρμογή είναι ήδη Cloud Native, μπορεί να μεταφερθεί απευθείας στο Cloud.

- Εάν η εφαρμογή δεν είναι Cloud Native, αλλά υπάρχει διαθέσιμη μια αναβάθμιση σε Cloud Native, τότε η εφαρμογή μπορεί να αναβαθμιστεί και να μετεγκατασταθεί στο Cloud σε ένα βήμα (ανάπτυξη της πιο πρόσφατης έκδοσης στο cloud και, στη συνέχεια, μετεγκατάσταση των δεδομένων) ή δύο βήματα (αναβάθμιση τοπικά και μετά μετεγκατάσταση της εφαρμογής στο cloud).

- Εάν δεν υπάρχει διαθέσιμη αναβάθμιση Cloud Native ή δεν προβλέπεται αναβάθμιση σχετικά άμεσα διαθέσιμη στο μέλλον, τότε θα πρέπει να διερευνηθεί η διαθεσιμότητα ενός παρόχου λύσεων SaaS. Εάν υπάρχει μια τέτοια λύση, τότε πρέπει να αξιολογηθεί το επίπεδο προσαρμογών που έχουν εφαρμοστεί στην τρέχουσα λύση μας.

    - Εάν οι προσαρμογές είναι ελάχιστες ή / και μπορούν να υποστηριχθούν από τον πάροχο SaaS, τότε μπορούμε να μεταφέρουμε την εφαρμογή στον πάροχο SaaS.
    - Εάν οι προσαρμογές είναι εκτεταμένες και / ή δεν μπορούν να υποστηριχθούν από τον πάροχο SaaS, τότε η εφαρμογή δεν μπορεί να μετεγκατασταθεί στον πάροχο SaaS.

- Εάν μια λύση SaaS δεν μπορεί να χρησιμοποιηθεί, τότε θα πρέπει να αξιολογηθεί ο επανασχεδιασμός και μία καινούρια αρχιτεκτονική για την εφαρμογής. Η βασική απόφαση που πρέπει να ληφθεί σε αυτό το σημείο είναι εάν η προσπάθεια επανασχεδιασμού αξίζει τον χρόνο και τους πόρους και θα μπορέσει να υλοποιηθεί μέσα τις προθεσμίες. Εάν χρονικό και χρηματικό κόστος δικαιολογείται, τότε πρέπει να σχεδιαστεί ο επαναπροσδιορισμός (refactoring), να ξεκινήσει ένα έργο και να ολοκληρωθεί εντός των προθεσμιών. Εάν δεν μπορεί να δικαιολογηθεί, τότε έχουμε τρεις επιλογές:

    - Εκτελούμε ένα "Lift and Shift" που σημαίνει ότι μετεγκαθιστούμε την εφαρμογή "ως έχει" και προγραμματίζουμε έναν επανασχεδιασμό σε μεταγενέστερο χρόνο.
    - Αγοράζουμε μια νέα εφαρμογή για να αντικαταστήσουμε την παλιά, αφού βεβαιωθούμε ότι είναι Cloud Native και την αναπτύσουμε στο Cloud.
    - Επαναξιολογούμε την αναγκαιότητα της εφαρμογής και σχεδιάζουμε το μελλοντικό παροπλισμό της μετά την εκτέλεση του "Lift and Shift".

Προσδιορισμός εξαρτήσεων

Ο προσδιορισμός των εξαρτήσεων έχει μεγάλη σημασία για την επίτευξη μιας ομαλής και χωρίς προβλήματα μετεγκατάστασης. Είναι ζωτικής σημασίας να διεξαχθεί μια διεξοδική, βασισμένη σε δεδομένα αναλυτική διαδικασία που συγκεντρώνει όλα τα απαραίτητα δεδομένα για να κατανοήσει πραγματικά τις υποκείμενες εξαρτήσεις και πολυπλοκότητες. Στην πραγματικότητα, μια επιτυχημένη πορεία μετεγκατάστασης βασίζεται στην ομαδοποίηση αλληλοεξαρτώμενων εφαρμογών και στη μετεγκατάσταση της ομάδας, παρά σε μεμονωμένες μετεγκαταστάσεις εφαρμογών.

Η διαδικασία εντοπισμού και τεκμηρίωσης των εξωτερικών εξαρτήσεων μιας εφαρμογής λογισμικού, όπως διακομιστές, δίκτυα, αποθήκευση και άλλες εφαρμογές, ονομάζεται χαρτογράφηση εφαρμογών και εξαρτήσεων (discovery and dependency mapping). Αυτή η διαδικασία συνεπάγεται όχι μόνο την αναγνώριση όλων των στοιχείων του οικοσυστήματος του λογισμικού, αλλά και την κατανόηση του τρόπου με τον οποίο αλληλεπιδρούν και επηρεάζουν το ένα το άλλο.

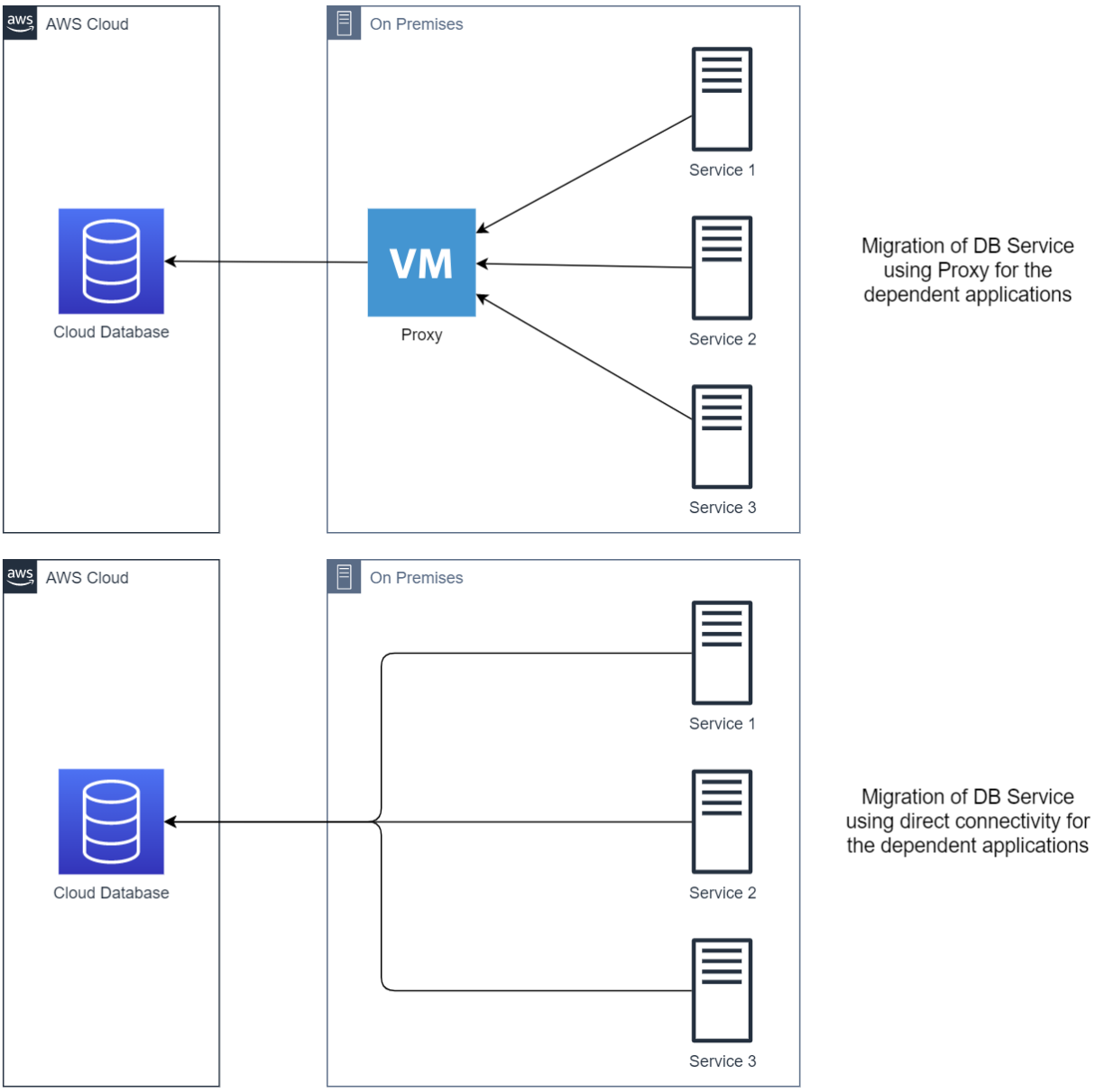
Η ύπαρξη ενημερωμένου χαρτοφυλακίου εφαρμογών θα βοηθήσει να εντοπιστούν η εφαρμογές που αλληλοεξαρτώνται και χρειάζονται ειδική μέριμνα. Τέτοιες εφαρμογές χρειάζονται ξεχωριστό προγραμματισμό, προκειμένου να ελαχιστοποιηθεί ο κίνδυνος διαχωρισμού στοιχείων που πρέπει να συνδεθούν, και θα προκαλούσε διακοπή λειτουργίας.

Πιο περίπλοκοι φόρτοι εργασίας όπως κεντρικές βάσεις δεδομένων ή αποθήκες δεδομένων, σίγουρα θα προκαλέσουν πολλές επιπλοκές. Σε τέτοιες περιπτώσεις έχουμε δύο επιλογές:

1. Μετεγκατάσταση της υπηρεσίας και ταυτόχρονη δημιουργία των απαιτούμενων ροών επικοινωνίας από τα εξαρτώμενα μέρη προς τη νέα υπηρεσία. Αυτή η προσέγγιση θα απαιτήσει μια σημαντική προσπάθεια για την αναδιάταξη όλων των εξαρτημένων εφαρμογών για τη χρήση των νέων ροών επικοινωνίας και η επαναφορά, σε περίπτωση που χρειαστεί, γίνεται πιο περίπλοκη.

2. Μετεγκατάσταση της υπηρεσίας και τοποθέτηση διακομιστή μεσολάβησης (proxy server) στη θέση των παλαιών στοιχείων. Αυτή η προσέγγιση θα απαιτήσει λιγότερη προσπάθεια, καθώς οι εξαρτημένες εφαρμογές δεν θα χρειαστούν αλλαγές, αλλά η προσθήκη των διακομιστών μεσολάβησης ενδέχεται να εισαγάγει καθυστερήσεις και μείωση απόδοσης. Όταν όλα τα εξαρτώμενα μέρη έχουν μεταφερθεί στο cloud, οι διακομιστές μεσολάβησης μπορούν να απενεργοποιηθούν με ασφάλεια.

Τα δύο διαφορετικά σενάρια παρουσιάζονται στο παρακάτω σχήμα.



Εικόνα 24 : Μετεγκατάσταση υπηρεσίας με εξαρτημένες εφαρμογές

Σχέδιο για ασφαλή μεταφορά δεδομένων

Η μετεγκατάσταση δεδομένων είναι η διαδικασία μεταφοράς δεδομένων από τη μία τοποθεσία στην άλλη. Αυτή η διαδικασία προσδιορίζει τα δεδομένα που πρόκειται να μεταφερθούν και στη συνέχεια τα προετοιμάζει για μετάδοση σε άλλη τοποθεσία αποθήκευσης. Είναι επίσης γνωστό ως μετεγκατάσταση αποθήκευσης συστήματος (system storage migration). Επιπλέον, οι υπηρεσίες μετεγκατάστασης δεδομένων μπορούν να βοηθήσουν στη μετεγκατάσταση ιδιόκτητης υποδομής σε cloud storage / εφαρμογές.

Τα ακόλουθα σημεία πρέπει να ληφθούν υπόψη κατά τον προγραμματισμό ή την εκτέλεση μετεγκατάστασης δεδομένων.

- **Αναγνώριση μορφής δεδομένων** - Πρέπει να προσδιορίσουμε τη μορφή των δεδομένων που θα μετεγκατασταθούν, πού είναι αποθηκευμένα, και τι μορφή θα πάρουν μετά τη μετεγκατάσταση. Κατά τη διάρκεια αυτού του σταδίου προ-προγραμματισμού, πρέπει να εντοπίσουμε όλα τα πιθανά σφάλματα και να καταγράψουμε τις προφυλάξεις ασφαλείας που πρέπει να ληφθούν κατά τη μεταφορά δεδομένων. Αυτό το βήμα θα συμβάλει στην προστασία των δεδομένων μας από κινδύνους.
- **Δημιουργία αντιγράφου ασφαλείας των δεδομένων** - Ένα αντίγραφο ασφαλείας των δεδομένων πρέπει πάντα να δημιουργείται πριν από την εκτέλεση μιας μεταφοράς δεδομένων. Εάν παρουσιαστεί πρόβλημα κατά τη διαδικασία μετεγκατάστασης, όπως καταστροφή δεδομένων, αρχεία που λείπουν ή αλλοίωση δεδομένων, θα έχουμε ήδη ένα αντίγραφο ασφαλείας των αρχικών μας δεδομένων αποθηκευμένο σε ασφαλή τοποθεσία. Μπορούμε να επιλύσουμε το πρόβλημα επαναφέροντάς τα στην αρχική τους κατάσταση.
- **Χρήση λογισμικού μετεγκατάστασης δεδομένων** - Όταν μετακινούνται μεγάλα αρχεία δεδομένων, η μεταφορά τους πρέπει γίνεται στρατηγικά. Η χειροκίνητη μεταφορά δεδομένων διαρκεί πολύ. Μπορούμε να χρησιμοποιήσουμε λογισμικό μετεγκατάστασης δεδομένων για να επισπεύσουμε τη διαδικασία μεταφοράς.
- **Διαδικασία μετεγκατάστασης δεδομένων** - Πρέπει να διασφαλίσουμε ότι τα δικαιώματα χρήστη/συστημάτων που εφαρμόζονται επαρκούν για την πλήρη εξαγωγή δεδομένων και μετεγκατάσταση στην επιθυμητή τοποθεσία, λαμβάνοντας υπόψη το σχέδιο. Τα δεδομένα που πρόκειται να μετεγκατασταθούν πρέπει να είναι απαλλαγμένα από σφάλματα / ιούς και να μετατραπούν στη σωστή μορφή. Στη συνέχεια, θα μεταφορτωθούν για μετεγκατάσταση. Τέλος, πρέπει να υπάρχει στενή παρακολούθηση της διαδικασίας, για τυχόν ζητήματα που ενδέχεται να προκύψουν στην πορεία.

- **Τελική δοκιμή** - Μετά τη μετεγκατάσταση, πρέπει να γίνει έλεγχος για τυχόν προβλήματα στο σύστημα προορισμού. Ο στόχος είναι η μετεγκατάσταση όλων των δεδομένων με ασφάλεια και χωρίς τροποποίηση. Οι δοκιμές πρέπει να διεξαχθούν στις εφαρμογές συστήματος, αποθήκευσης και ιστού, ώστε να επιβεβαιωθεί η επιτυχία της διαδικασίας.
- **Συντήρηση παρακολούθησης** - Επειδή μπορεί να προκύψουν σφάλματα κατά τη διάρκεια της διαδικασίας δοκιμής, πρέπει να πραγματοποιηθεί πλήρης επιθεώρηση του συστήματος προκειμένου να εντοπιστούν τυχόν πιθανά σφάλματα / ζημιές. Σε περίπτωση σφάλματος, τα δεδομένα μπορούν να αποκατασταθούν από το αντίγραφο ασφαλείας.

Στις ακόλουθες ενότητες θα παρουσιάσουμε τις διαθέσιμες λύσεις για τρεις διαφορετικούς τύπους ή / και στάδια μετεγκατάστασης δεδομένων:

- Online μεταφορά δεδομένων: Αυτές οι υπηρεσίες μπορούν να χρησιμοποιηθούν για τη μεταφορά δεδομένων μικρού έως μέτριου μεγέθους τα οποία μπορούν να μεταφερθούν μέσω του δικτύου σε εύλογο χρονικό διάστημα. Στον παρακάτω πίνακα μπορούμε να δούμε ότι, ανάλογα με την ταχύτητα μεταφοράς και το μέγεθος των δεδομένων, η διάρκεια της διαδικασίας μπορεί να είναι πολύ μεγάλη.

		Μέγεθος (GigaBytes)				
		10 GB	100 GB	1'000 GB	10'000 GB	100'000 GB
Ταχύτητα μεταφοράς (MegaBits / sec)	10 Mbit / s	0 ημέρες 02 ώρες 13 m 20 s	0 ημέρες 22 ώρες 13 m 20 s	9 ημέρες 06 h 13 m 20 s	92 ημέρες 14 ώρες 13 m 20 s	925 ημέρες 22 ώρες 13 m 20 s
	100 Mbit / s	0 ημέρες 00 ώρες 13 m 20 s	0 ημέρες 02 ώρες 13 m 20 s	0 ημέρες 22 ώρες 13 m 20 s	9 ημέρες 06 h 13 m 20 s	92 ημέρες 14 ώρες 13 m 20 s
	1'000 Mbit / s	0 ημέρες 00 ω 01 μ 20 δ	0 ημέρες 00 ώρες 13 m 20 s	0 ημέρες 02 ώρες 13 m 20 s	0 ημέρες 22 ώρες 13 m 20 s	9 ημέρες 06 h 13 m 20 s
	10'000 Mbit / s	0 ημέρες 00 ω 00 μ 08 δ	0 ημέρες 00 ω 01 μ 20 δ	0 ημέρες 00 ώρες 13 m 20 s	0 ημέρες 02 ώρες 13 m 20 s	0 ημέρες 22 ώρες 13 m 20 s

Πίνακας 14 : Χρόνος μεταφοράς

- Offline μεταφορά δεδομένων: Αυτές οι υπηρεσίες μπορούν να χρησιμοποιηθούν για τη μεταφορά μεγάλου όγκου δεδομένων εκτός σύνδεσης, με φυσικά μέσα. Αυτά είναι ιδανικά

για μεγάλες ποσότητες δεδομένων ή σε περιπτώσεις όπου μια σύνδεση δικτύου δεν είναι διαθέσιμη ή αξιόπιστη.

- Υβριδική αποθήκευση cloud: Αυτές οι υπηρεσίες μπορούν να χρησιμοποιηθούν για τη σύνδεση των ιδιόκτητων υπηρεσιών απευθείας με τις υπηρεσίες δεδομένων cloud. Με αυτόν τον τρόπο οι ιδιόκτητες υπηρεσίες θα μπορούν να χρησιμοποιούν απευθείας τα δεδομένα που έχουν μεταφερθεί.

Διαδικτυακή μεταφορά δεδομένων

- **AWS DataSync**

  Το AWS DataSync είναι μια υπηρεσία μεταφοράς δεδομένων που απλοποιεί τη διαδικασία αυτοματοποίησης της μεταφοράς δεδομένων μεταξύ ιδιόκτητου χώρου αποθήκευσης και Amazon S3 ή Amazon Elastic File System (Amazon EFS).

  Το DataSync αυτοματοποιεί έναν μεγάλο αριθμό εργασιών μεταφοράς δεδομένων που μπορούν να επιβραδύνουν τις μετακινήσεις ή να επιβαρύνουν τις λειτουργίες IT, όπως η εκτέλεση των δικών μας δομών, η διαχείριση της κρυπτογράφησης, η διαχείριση των scripts, η βελτιστοποίηση του δικτύου μας και η επικύρωση της ακεραιότητας των δεδομένων. Το DataSync μας επιτρέπει να μεταφέρουμε δεδομένα έως και δέκα φορές πιο γρήγορα από τα εργαλεία ανοιχτού κώδικα.

  Το DataSync συνδέεται με τα υπάρχοντα συστήματα αποθήκευσης ή αρχείων μέσω του πρωτοκόλλου Network File System (NFS), εξαλείφοντας την ανάγκη σύνταξης scripts ή τροποποίησης των εφαρμογών μας για χρήση AWS APIs. Το DataSync μας επιτρέπει να αντιγράφουμε δεδομένα από και προς το AWS χρησιμοποιώντας AWS Direct Connect ή συνδέσεις στο Διαδίκτυο.

  Η υπηρεσία επιτρέπει μεμονωμένες μετεγκαταστάσεις δεδομένων, αυτόματη αναπαραγωγή για προστασία και ανάκτηση δεδομένων και επαναλαμβανόμενες ροές εργασίας επεξεργασίας δεδομένων. Οι πράκτορες DataSync πρέπει να εγκατασταθούν εσωτερικά, να συνδεθούν σε ένα σύστημα αρχείων ή σε μια συστοιχία αποθήκευσης και να επιλέξουν είτε το Amazon EFS είτε το S3 ως τον στόχο αποθήκευσης AWS. Το κόστος υπολογίζεται ανάλογα με τα δεδομένα που μεταφέρονται.

- **AWS Transfer Family**

  Το AWS Transfer Family διαχειρίζεται όλες τις πτυχές των μεταφορών αρχείων από και προς το Amazon S3.

  Με υποστήριξη για το πρωτόκολλο ασφαλούς μεταφοράς αρχείων (Secure File Transfer Protocol - SFTP), το πρωτόκολλο μεταφοράς αρχείων μέσω SSL (FTPS) και το πρωτόκολλο μεταφοράς αρχείων (FTP), το AWS Transfer Family μας επιτρέπει να μεταφέρουμε τις ροές εργασίας μεταφοράς αρχείων στο AWS απρόσκοπτα, ενσωματώνοντας στα υπάρχοντα συστήματα ελέγχου ταυτότητας και παρέχοντας δρομολόγηση DNS μέσω του Amazon Route 53, διασφαλίζοντας ότι οι εφαρμογές μας παραμένουν ανεπηρέαστες. Είναι απλό να ξεκινήσετε με το AWS Transfer Family - δεν υπάρχει υποδομή για αγορά ή ρύθμιση.

Μεταφορά δεδομένων εκτός σύνδεσης

- **AWS Snowcone**

  Το AWS Snowcone είναι το μικρότερο μέλος της οικογένειας AWS Snow υπολογιστών, αποθήκευσης άκρης δικτύου και συσκευών μεταφοράς δεδομένων, βάρους 2,1 κιλών και διαθέτει 8 terabyte χρησιμοποιήσιμου χώρου αποθήκευσης. Το Snowcone είναι ανθεκτικό, ασφαλές και έχει σχεδιαστεί ειδικά για λειτουργία εκτός του παραδοσιακού κέντρου δεδομένων. Το μικρό του μέγεθος το καθιστά ιδανικό για στενούς χώρους ή καταστάσεις όπου απαιτείται φορητότητα, αλλά η συνδεσιμότητα δικτύου δεν είναι αξιόπιστη. Μπορείτε να εκτελέσετε υπολογιστικές εφαρμογές στην άκρη του δικτύου και έπειτα είτε να στείλετε τη συσκευή στο AWS για μεταφορά δεδομένων εκτός σύνδεσης είτε να χρησιμοποιήσετε το AWS DataSync για να μεταφέρετε δεδομένα μέσω διαδικτύου από άκρες δικτύου.

  Η συσκευή χρησιμοποιεί υλικό και λογισμικό για την παροχή ασφάλειας που πληροί ακόμη και τις πιο αυστηρές απαιτήσεις. Χρησιμοποιεί μονάδες Trusted Platform Modules (TPM) που βασίζονται στην αποθήκευση υλικών κλειδιών (device-specific keys) σε μια μη προσβάσιμη τοποθεσία από το λογισμικό, βοηθώντας έτσι στη διασφάλιση της ακεραιότητας της συσκευής. Τα δεδομένα του Snowcone κρυπτογραφούνται χρησιμοποιώντας δύο επίπεδα κρυπτογράφησης σε κατάσταση ηρεμίας, τα οποία βοηθούν στην προστασία των δεδομένων της συσκευής κατά την αποστολή. Η υπηρεσία AWS Key Management Service (KMS) διαχειρίζεται κλειδιά κρυπτογράφησης, τα οποία δεν αποθηκεύονται ποτέ στη συσκευή Snowcone.

Εικόνα: AWS SnowCone

- **AWS Snowball**

  Αυτή η συσκευή είναι μια λύση μεταφοράς δεδομένων κλίμακας petabyte που αξιοποιεί ασφαλείς συσκευές για τη μεταφορά μεγάλου όγκου δεδομένων εντός και εκτός του Amazon Web Services. Το Snowball ξεπερνά τα συνηθισμένα ζητήματα που σχετίζονται με τη μεταφορά δεδομένων μεγάλης κλίμακας, όπως το υψηλό κόστος δικτύου, οι μεγάλοι χρόνοι μεταφοράς και οι ανησυχίες σχετικά με την ασφάλεια. Η μεταφορά δεδομένων Snowball είναι απλή, γρήγορη και ασφαλής και μπορεί να κοστίσει μόλις το ένα πέμπτο της τιμής του Internet υψηλής ταχύτητας.

  Το Snowball εξαλείφει την ανάγκη να γράψουμε κώδικα ή να αγοράσουμε υλικό για τη μεταφορά των δεδομένων μας. Απλώς δημιουργούμε μια εργασία στο AWS Management Console και θα πραγματοποιηθεί αυτόματη αποστολή μιας συσκευής Snowball. Μόλις φτάσει η συσκευή, την συνδέουμε στο τοπικό μας δίκτυο και πραγματοποιούμε λήψη και εκτέλεση του προγράμματος-πελάτη Snowball και, στη συνέχεια, χρησιμοποιούμε τον πελάτη για να επιλέξουμε τους καταλόγους αρχείων που θα μεταφερθούν στη συσκευή. Στη συνέχεια, ο πελάτης θα κρυπτογραφήσει και θα μεταφέρει τα αρχεία με μεγάλη ταχύτητα στη συσκευή. Μόλις ολοκληρωθεί η μεταφορά και η συσκευή είναι έτοιμη για επιστροφή, η ετικέτα αποστολής E Ink θα ενημερωθεί αυτόματα έτσι ώστε να μπορούμε να παρακολουθούμε την κατάσταση της εργασίας μέσω της υπηρεσίας απλής ειδοποίησης Amazon (SNS), μηνυμάτων κειμένου ή απευθείας στην κονσόλα.

Το Snowball προστατεύει τα δεδομένα μας με πολλαπλά επίπεδα ασφάλειας, συμπεριλαμβανομένων περιβλημάτων ανθεκτικών σε παραβιάσεις, κρυπτογράφησης 256-bit και μιας μονάδας Trusted Platform Module (TPM) βιομηχανικού προτύπου που διασφαλίζει τόσο την ασφάλεια όσο και την πλήρη αλυσίδα φύλαξης. Η AWS εκτελεί μια διαγραφή λογισμικού της συσκευής Snowball μετά την επεξεργασία και την επαλήθευση της εργασίας μεταφοράς δεδομένων.

Εικόνα 25 : AWS SnowBall

- **AWS Snowmobile**

Το AWS Snowmobile είναι μια υπηρεσία μεταφοράς δεδομένων κλίμακας petabyte που επιτρέπει τη μεταφορά τεράστιων ποσοτήτων δεδομένων στο AWS. Μπορούμε να μεταφέρουμε έως και 100 PB ανά Snowmobile, ένα ανθεκτικό κιβώτιο αποστολής 15 μέτρων που τραβιέται από ένα φορτηγό ημιρυμουλκούμενου. Το Snowmobile καθιστά απλή τη μετεγκατάσταση μεγάλων ποσοτήτων δεδομένων στο cloud, όπως βιβλιοθήκες βίντεο, αποθετήρια εικόνων ή ακόμη και ολόκληρο το κέντρο δεδομένων. Η μεταφορά δεδομένων Snowmobile είναι ασφαλής, γρήγορη και οικονομική.

Μετά από μια αρχική αξιολόγηση, ένα Snowmobile θα παραδοθεί στο κέντρο δεδομένων μας και θα διαμορφωθεί από το προσωπικό της AWS έτσι ώστε να μπορεί να έχει πρόσβαση ως στόχος αποθήκευσης δικτύου. Όταν το Snowmobile φτάσει επί τόπου, το προσωπικό της AWS θα συνεργαστεί με την ομάδα μας για να το συνδέσει στο τοπικό μας δίκτυο μέσω ενός αφαιρούμενου διακόπτη δικτύου υψηλής ταχύτητας. Στη συνέχεια, μπορούμε να ξεκινήσουμε τη μεταφορά δεδομένων υψηλής ταχύτητας στο Snowmobile από οποιονδήποτε αριθμό πηγών στο κέντρο δεδομένων μας. Μετά τη φόρτωση των δεδομένων μας, το Snowmobile επιστρέφει στο AWS, όπου εισάγεται στο Amazon S3 ή S3 Glacier.

Το AWS Snowmobile προστατεύει τα δεδομένα μας με πολλαπλά επίπεδα ασφάλειας, συμπεριλαμβανομένων αποκλειστικού προσωπικού ασφαλείας, παρακολούθησης GPS, παρακολούθησης συναγερμών, παρακολούθησης βίντεο 24/7 και προαιρετικού οχήματος ασφαλείας συνοδείας κατά τη μεταφορά. Όλα τα δεδομένα κρυπτογραφούνται με χρήση κλειδιών κρυπτογράφησης 256-bit AWS KMS, τα οποία έχουν σχεδιαστεί για να διασφαλίζουν την ασφάλεια και την πλήρη αλυσίδα φύλαξης των δεδομένων μας.

Εικόνα 26 : AWS SnowMobile

Υβριδική αποθήκευση cloud

- **AWS Storage Gateway**

Το AWS Storage Gateway μπορεί να χρησιμοποιηθεί για την εύκολη ανάπτυξη του AWS Storage στο ιδιόκτητο δίκτυο και επιτρέπει τη δυνατότητα σύνδεσης και επέκτασης των ιδιόκτητων εφαρμογών μας στο AWS Storage με απρόσκοπτο τρόπο. Για παράδειγμα

μετεγκατάστασητων βιβλιοθηκών ταινιών στο cloud, παρέχοντας υποστήριξη αποθήκευσης cloud κοινή χρήση αρχείων και δημιουργία προσωρινής μνήμης χαμηλής καθυστέρησης για εφαρμογές ιδιόκτητων εγκαταστάσεων για πρόσβαση σε δεδομένα στο AWS.

Η υπηρεσία υποστηρίζει τρεις διαφορετικούς τύπους πύλης: πύλες αρχείων (file gateways), πύλες κασέτας (tape gateways) και πύλες όγκου (volume gateways).

- Το File Gateway αποθηκεύει δεδομένα αρχείων στο Amazon S3 ως ανθεκτικά αντικείμενα ή σε πλήρως διαχειριζόμενες κοινοποιήσεις αρχείων μέσω του Amazon FSx File Gateway.

- Η διαμόρφωση της εικονικής βιβλιοθήκης ταινιών Tape Gateway (VTL) ενσωματώνεται με ένα υπάρχον λογισμικό δημιουργίας αντιγράφων ασφαλείας, επιτρέποντας την οικονομική αντικατάσταση ταινιών στο Amazon S3 και μακροπρόθεσμη αρχειοθέτηση στο S3 Glacier και το S3 Glacier Deep Archive.

- Η τοπική αποθήκευση ή η προσωρινή αποθήκευση των block volumes παρέχεται από το Volume Gateway, μαζί με αντίγραφα ασφαλείας point-in-time με τη μορφή στιγμιότυπων EBS. Η ανάκτηση μέσω cloud είναι δυνατή για αυτά τα στιγμιότυπα.

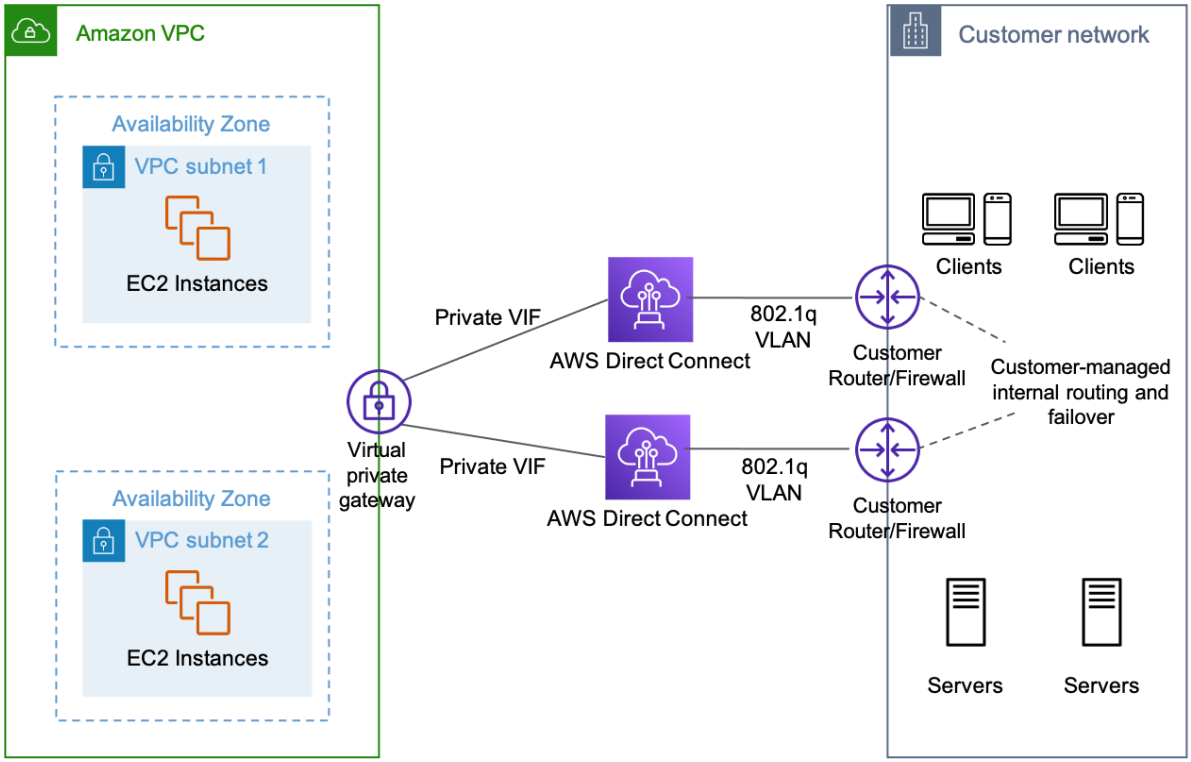
Εικόνα 27 : AWS Storage Gateway (AWS)

- **AWS Direct Connect**

Το AWS Direct Connect είναι μια ειδική φυσική σύνδεση σε κέντρα δεδομένων AWS, η οποία μπορεί να χρησιμοποιηθεί για την επιτάχυνση των μεταφορών δικτύου μεταξύ των ιδιόκτητων κέντρων δεδομένων και των κέντρων δεδομένων AWS. Καθώς είναι μια φυσική σύνδεση με εγγυημένο εύρος ζώνης και διαθεσιμότητα, μπορεί να χρησιμοποιηθεί για τη σύνδεση υπηρεσιών και τη μεταφορά δεδομένων αξιόπιστα και με συνέπεια, παρακάμπτοντας το δημόσιο Διαδίκτυο.

Το AWS Direct Connect μας επιτρέπει να συνδέσουμε το δίκτυό μας σε μία από τις θέσεις AWS Direct Connect μέσω μιας αποκλειστικής σύνδεσης δικτύου. Αυτή η αποκλειστική σύνδεση μπορεί να χωριστεί σε πολλαπλές εικονικές διεπαφές χρησιμοποιώντας βιομηχανικά πρότυπα 802.1q VLAN, τα οποία μας επιτρέπουν να χρησιμοποιούμε την ίδια σύνδεση για πρόσβαση σε δημόσιους πόρους, όπως αντικείμενα που είναι αποθηκευμένα στο Amazon S3 και ιδιωτικούς πόρους, όπως δομές Amazon EC2 που εκτελούνται εντός ένός Amazon Virtual Private Cloud (VPC), διατηρώντας παράλληλα τον διαχωρισμό δικτύου μεταξύ δημόσιου και ιδιωτικού περιβάλλοντος. Οι εικονικές διεπαφές μπορούν να

αναδιαμορφωθούν ανά πάσα στιγμή για να ικανοποιήσουν τις μεταβαλλόμενες επιχειρηματικές απαιτήσεις.



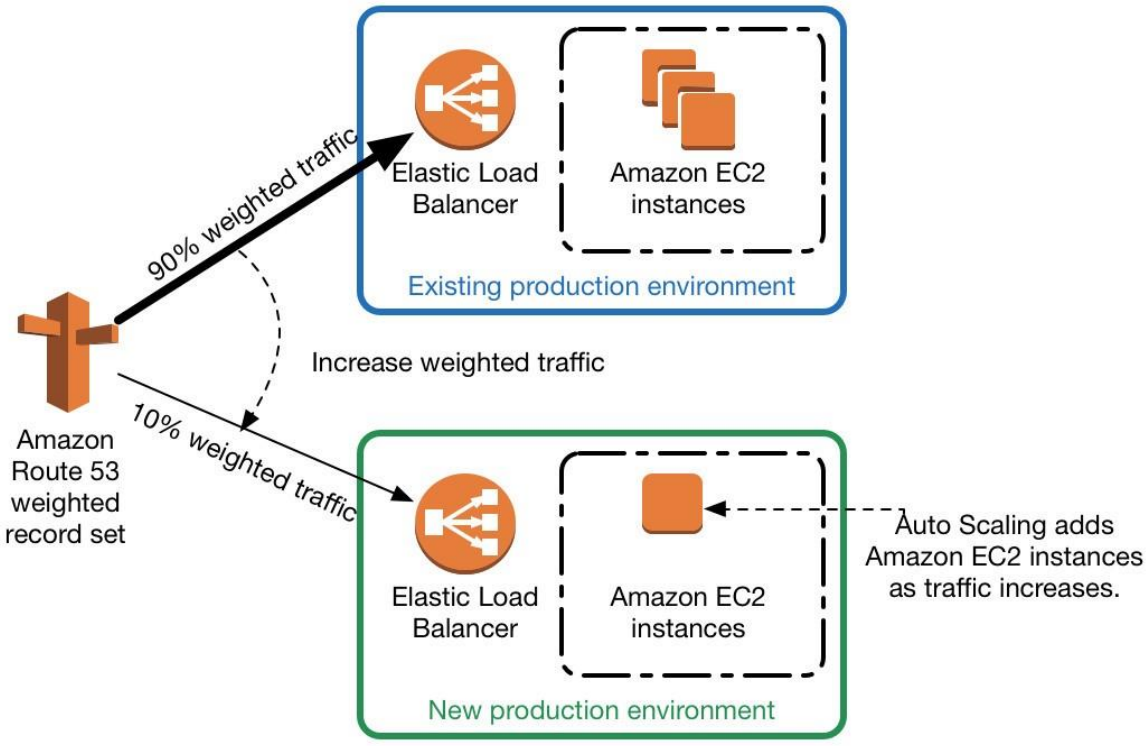
Εικόνα 28 : AWS redundant Direct Connect

Προγραμματισμός για μηδενικό χρόνο διακοπής

Για κρίσιμες εφαρμογές που χρειάζονται μηδενικό χρόνο διακοπής λειτουργίας ή σχεδόν μηδενική διακοπή λειτουργίας, μπορεί να χρησιμοποιηθεί μια στρατηγική ανάπτυξης blue/green αφότου η εφαρμογή και τα δεδομένα της έχουν δημιουργηθεί εκ νέου ή αντιγραφεί στο cloud. Έχοντας δύο σχεδόν πανομοιότυπες εγκαταστάσεις, μπορούμε να μεταφέρουμε σταδιακά την εφαρμογή από τη μία εγκατάσταση στην άλλη.

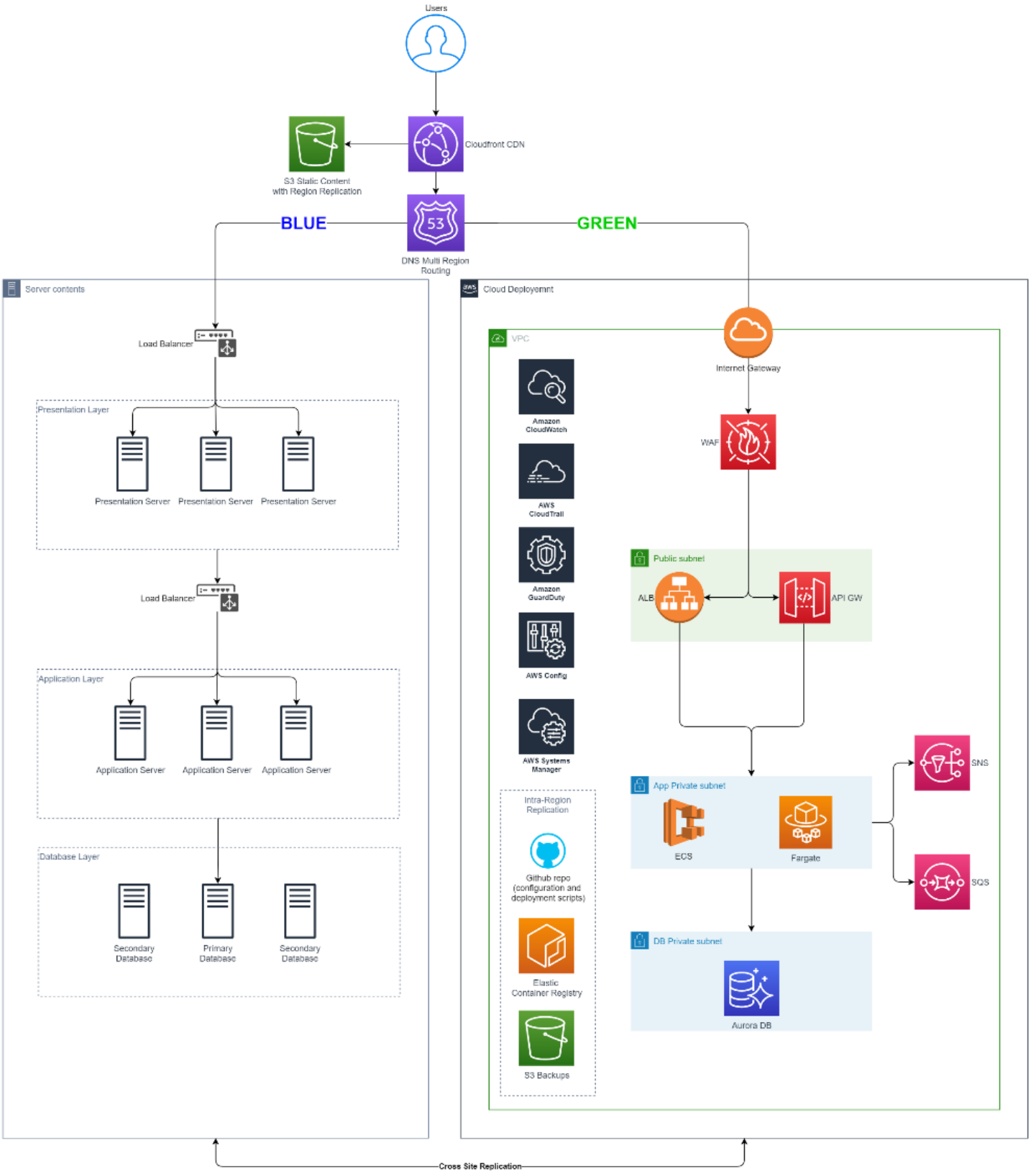
Οι blue/green αναπτύξεις καθιστούν δυνατή την ανάπτυξη και επαναφορά με σχεδόν μηδενικό χρόνο διακοπής λειτουργίας. Η βασική ιδέα της ανάπτυξης blue/green είναι η δρομολόγηση της κίνησης μεταξύ δύο πανομοιότυπων περιβαλλόντων, που το καθένα εκτελεί μια διαφορετική έκδοση της εφαρμογής μας. Το μπλε περιβάλλον είναι η τρέχουσα έκδοση της εφαρμογής που εξυπηρετεί την παραγωγή. Ταυτόχρονα, δημιουργείται και διαμορφώνεται ένα πράσινο περιβάλλον για να εκτελεί μια διαφορετική έκδοση της εφαρμογής μας. Μόλις το πράσινο περιβάλλον έχει προετοιμαστεί και δοκιμαστεί, η κίνηση της παραγωγής ανακατευθύνεται από την μπλε στην πράσινη. Εάν εντοπιστούν προβλήματα, μπορούμε να επαναφέρουμε την κίνηση στο μπλε περιβάλλον.

Αφότου το πράσινο περιβάλλον (η εφαρμογή που φιλοξενείται στο cloud) έχει ελεγχθεί διεξοδικά, χρησιμοποιώντας την υπηρεσία DNS του παρόχου cloud (στην περίπτωσή μας AWS Route 53), η κίνηση μπορεί να κατανεμηθεί μεταξύ των μπλε και των πράσινων εφαρμογών, σταδιακά, προβλέψιμα και με δυνατότητα εύκολης και γρήγορης επαναφοράς.



Εικόνα 29 : AWS Route 53 green/blue σταθμισμένη δρομολόγηση (AWS)

Παρακάτω μπορούμε να δούμε ένα παράδειγμα ανάπτυξης / μετεγκατάστασης blue/green με βάση τα σχέδια που παρουσιάστηκαν προηγουμένως.



Εικόνα 30: Green/Blue σχέδιο

Ζώνες Προσγείωσης (Landing Zones)

Ο όρος "Landing Zones" αναφέρεται σε ένα διαμορφωμένο περιβάλλον που περιλαμβάνει ένα τυπικό σύνολο ασφαλών υποδομών cloud, πολιτικών, βέλτιστων πρακτικών, οδηγιών και υπηρεσιών με κεντρική διαχείριση. Αυτό επιτρέπει στους πελάτες να αναπτύξουν γρήγορα ένα ασφαλές περιβάλλον πολλαπλών λογαριασμών Cloud που συμμορφώνεται με τα καθιερωμένα πρότυπα. Έχοντας διαθέσιμη μια πληθώρα επιλογών σχεδίασης, η διαμόρφωση ενός περιβάλλοντος πολλαπλών λογαριασμών μπορεί να έχει υψηλό κόστος χρόνου και προσπάθειας, καθώς περιλαμβάνει τη διαμόρφωση πολλών λογαριασμών και υπηρεσιών, κάτι που απαιτεί πλήρη κατανόηση των υπηρεσιών παροχής cloud.

Αυτή η λύση μπορεί να μας βοηθήσει να εξοικονομήσουμε χρόνο αυτοματοποιώντας τη διαδικασία δημιουργίας περιβαλλόντων για την εκτέλεση ασφαλών και επεκτάσιμων φορτίων εργασίας, ενώ παράλληλα δημιουργούμε μια αρχική βάση ασφαλείας μέσω της δημιουργίας βασικών κοινών λογαριασμών και πόρων.

Κατά τη δημιουργία μιας ζώνης προσγείωσης (landing zone), πρέπει να ληφθούν υπόψη τα ακόλουθα σημεία.

1. **Συμμόρφωση και ασφάλεια:** Μια ζώνη προσγείωσης μας επιτρέπει να θεσπίζουμε ελέγχους ασφαλείας σε επίπεδο οργανισμού και λογαριασμού και να καθορίσουμε μια βάση ασφαλείας με προληπτικά και αναγνωριστικά στοιχεία ελέγχου. Με τις ζώνες προσγείωσης, μπορούμε να εφαρμόσουμε πολιτικές συμμόρφωσης και διαχείρισης δεδομένων σε ολόκληρο τον οργανισμό. Η αρχιτεκτονική της υποδομής αναπτύσσεται με σταθερότητα, ως μέρος αυτής της διαδικασίας, για την αντιμετώπιση ζητημάτων όπως ασφάλεια δικτυακών άκρων, διαχείριση απειλών, διαχείριση των vulnerabilities και ασφάλεια δικτυακής μετάδοσης.

2. **Τυποποιημένη μίσθωση (Standardized Tenancy):** Μια ζώνη προσγείωσης παρέχει ένα πλαίσιο για τη δημιουργία και τη διαχείριση ενός περιβάλλοντος πολλαπλών λογαριασμών. Ένα αυτοματοποιημένο βασικό περιβάλλον για πολλαπλούς λογαριασμούς συμβάλλει στη μείωση του χρόνου εγκατάστασης, ενώ παράλληλα δημιουργεί μια αρχική γραμμή ασφαλείας για οποιοδήποτε ψηφιακό περιβάλλον χρησιμοποιούμε. Οι απαιτήσεις ασφάλειας, ελέγχου και κοινής χρήσης διαχειρίζονται από την αυτοματοποιημένη δομή πολλαπλών λογαριασμών. Σημαντική είναι και η εφαρμογή πολιτικών χρήσης ετικετών (tags)

σε πολλαπλούς λογαριασμούς cloud καθώς και η παροχή τυποποιημένων λογαριασμών για διαφορετικά προφίλ ασφαλείας (dev / staging / prod).

3. **Διαχείριση ταυτότητας και πρόσβασης:** Ο καθορισμός των ρόλων και των πολιτικών πρόσβασης γίνεται σύμφωνα με την αρχή του λιγότερο προνομίου (principle of least privilege). Με τον καθορισμό ρόλων και πολιτικών πρόσβασης, μπορούμε να εφαρμόσουμε την αρχή του ελάχιστου δικαιώματος. Θα πρέπει επίσης να εφαρμοστεί το SSO για σύνδεση στο cloud.

4. **Δικτύωση:** Ο σχεδιασμός και η υλοποίηση της δικτύωσης στο Cloud αποτελούν κρίσιμα στοιχεία των προσπαθειών υιοθέτησης του Cloud. Η δικτύωση αποτελείται από μια ποικιλία προϊόντων και υπηρεσιών που το καθένα προσφέρει ένα ξεχωριστό σύνολο δικτυακών δυνατοτήτων. Πρέπει να ληφθούν μέτρα κατά την υλοποίηση του δικτύου για να διασφαλιστεί ότι το δίκτυο έχει υψηλή διαθεσιμότητα, ανθεκτικότητα και επεκτασιμότητα.

5. **Λειτουργίες Διαχείρισης (Operations):** Χρήση κεντρικού logging και συλλογή των αρχείων καταγραφής από πολλούς λογαριασμούς που χρησιμοποιούν διάφορες υπηρεσίες παροχής cloud. Χρησιμοποιώντας μια ποικιλία εγγενών εργαλείων cloud, μπορούμε να διαμορφώσουμε την αυτόματη δημιουργία αντιγράφων ασφαλείας και την αποκατάσταση καταστροφών, την διαχείριση της παρακολούθησης και προειδοποίησης για την διαχείριση κόστους, την αντιδραστική επεκτασιμότητα (reactive scalability) και την αξιοπιστία. Η προσθήκη διορθώσεων λογισμικού στους διακομιστές γίνεται αυτόματα και προγραμματισμένα.

Όλα τα παραπάνω θα πρέπει να αντικατοπτρίζονται στο σχεδιασμό και το σχεδιάγραμμα της ζώνης προσγείωσης (μέσω κώδικα IaC) μαζί με τις βέλτιστες πρακτικές που έχουμε παρουσιάσει σε ένα προηγούμενο κεφάλαιο και παρότι είναι πιο στοχευμένα για το AWS, πρέπει να λάβουμε υπόψη, ότι όλοι οι μεγάλοι πάροχοι δημόσιου cloud υποστηρίζουν παρόμοιες υπηρεσίες.

Σχέδιο μετεγκατάστασης (Migration Plan)

Το τελικό σχέδιο μετεγκατάστασης θα είναι περίπλοκο, αλλά μπορεί να συνοψιστεί στα ακόλουθα βήματα.

1. Εντοπίζουμε εφαρμογές, υπηρεσίες και διακομιστές ή άλλα στοιχεία.
2. Αποσύρουμε οποιαδήποτε υπηρεσία δεν χρειάζεται πλέον ή μπορεί να αντικατασταθεί από κάποια άλλη υπάρχουσα υπηρεσία.

3. Προσδιορίζουμε τις εξαρτήσεις μεταξύ των υπηρεσιών
4. Δίνουμε προτεραιότητα στη μετεγκατάσταση υπηρεσιών σύμφωνα με
    a. Εξαρτήσεις
    b. Κρισιμότητα
    c. Ετοιμότητα για το Cloud (Cloud Readiness)
5. Αναλύουμε και σχεδιάζουμε την αρχιτεκτονική "to be", και κάνουμε μία εκτίμηση για το κόστος.
6. Εξασφαλίζουμε την διαθεσιμότητα του προϋπολογισμού.
7. Προετοιμάζουμε το περιβάλλον με ζώνες προσγείωσης.
8. Σχεδιάζουμε και υλοποιούμε την συνδεσιμότητα δικτύου και δοκιμάζουμε τις δικτυακές ροές σε όλα τα περιβάλλοντα.
9. Μεταφέρουμε τα δοκιμαστικά δεδομένα και εκτελούμε System Integration Tests (SITs) / User Acceptance Tests (UATs) στα περιβάλλοντα δοκιμής.
10. Μεταφέρουμε τα πραγματικά δεδομένα και ρυθμίζουμε την on-line συνεχή αντιγραφή τους (replication) , εάν χρειάζεται.
11. Εκτελούμε μια προσέγγιση μετεγκατάστασης με μηδενικό χρόνο διακοπής για τις κρίσιμες εφαρμογές ή προγραμματίζουμε ένα διάστημα με διακοπή λειτουργίας και εκτελούμε την μεταφορά εκτός σύνδεσης για τις υπόλοιπες εφαρμογές.
12. Δοκιμάζουμε το νέο παραγωγικό περιβάλλον, διατηρώντας το παλιό, σε περίπτωση που απαιτηθεί επαναφορά.
13. Εάν όλα λειτουργήσουν όπως αναμενόταν, δημιουργούμε αντίγραφα ασφαλείας, εάν χρειάζεται, και παροπλίζουμε το παλιό περιβάλλον.

Τα βήματα προετοιμασίας του περιβάλλοντος στο Cloud, που πρέπει να εκτελεσθούν παράλληλα και να ολοκληρωθούν πριν από το βήμα 6, έχουν ως εξής.

1. Δημιουργούμε τη δομή οργάνωσης AWS Organizations με τουλάχιστον έναν λογαριασμό διαχείρισης και διάφορους λογαριασμούς κοινών υπηρεσιών.

2. Σχεδιάζουμε και εφαρμόζουμε τις βασικές πολιτικές ασφαλείας, όπως:
    a. Συνδέουμε τον πάροχο ταυτότητας (Identity provider) της εταιρείας με το Cloud.
    b. Συνδέουμε το λογισμικό καταγραφής, παρακολούθησης και ασφάλειας της εταιρείας με το Cloud.
    c. Ορίζουμε και εφαρμόζουμε πολιτικές ελέγχου πρόσβασης σύμφωνα με την αρχή των "ελάχιστων δικαιωμάτων". Αντιστοιχούμε υπάρχουσες ομάδες ασφαλείας σε AWS ρόλους.
    d. Ορίζουμε και εφαρμόζουμε κανόνες συμμόρφωσης (Compliance) με τις πολιτικές.

e. Δημιουργούμε ασφαλή πρότυπα VM που θα έχουν προεγκατεστημένους όλους τους απαιτούμενους παράγοντες και τα στοιχεία ελέγχου και ασφαλείας.

3. Προετοιμάζουμε την διασύνδεση:
    a. Δημιουργούμε έναν λογαριασμό "δικτυακού κόμβου" με πολλαπλή συνδεσιμότητα προς στα τοπικά κέντρα δεδομένων.
    b. Δημιουργούμε ζώνες DNS για χρήση στο Cloud και τις συνδέουμε με τις εσωτερικές τοπικές υπηρεσίες DNS.
    c. Προετοιμάζουμε ζώνες IP που μπορούν να χρησιμοποιηθούν στο Cloud αλλά είναι δρομολογήσιμες / προσβάσιμες από τα τοπικά κέντρα δεδομένων. Αυτές θα χρησιμοποιηθούν μόνο στον λογαριασμό του δικτυακού κόμβου.
    d. Εφαρμόζουμε πολιτικές που θα επιβάλλουν την αρχή δικτύωσης «εξαρχής απόρριψη όλων των συνδέσεων».

4. Δημιουργούμε πρότυπα υποδομής (infrastructure templates) που θα περιλαμβάνουν τις βασικές αρχές ασφαλείας και δικτυακής υποδομής και οι οποίες θα χρησιμοποιηθούν για τη δημιουργία των ζωνών προσγείωσης.

Κόστος

Το κόστος μετεγκατάστασης μπορεί να χωριστεί στις ακόλουθες τέσσερις κατηγορίες:

1. **Κόστος υποδομής:** Αυτό περιλαμβάνει το κόστος της υποδομής του νέου περιβάλλοντος στο Cloud, κατά τη διάρκεια που κατασκευάζεται και αξιολογείται, και έως ότου μεταφερθεί η υπηρεσία και ολοκληρωθεί ο παροπλισμός των παλαιών συστημάτων.

2. **Κόστος μεταφοράς δεδομένων:** Αυτό περιλαμβάνει το εύρος ζώνης (bandwidth), την κυκλοφορία δεδομένων (traffic) και άλλες δαπάνες υπηρεσιών μεταφοράς δεδομένων, online ή εκτός σύνδεσης, οι οποίες καταναλώνονται κατά τη μεταφορά δεδομένων.

3. **Κόστος Υπηρεσιών μετεγκατάστασης:** Εδώ περιλαμβάνονται τυχόν Υπηρεσίες Μετεγκατάστασης του παρόχου Cloud που έχουν χρησιμοποιηθεί κατά τη διάρκεια των διαδικασιών εντοπισμού και μετεγκατάστασης. Η περισσότερες εξ' αυτών παρέχονται δωρεάν από τους παρόχους ως ένα μέσο διευκόλυνσης της διαδικασίας μετεγκατάστασης των υπηρεσιών στο cloud τους.

4. **Κόστος παροχής επαγγελματικών υπηρεσιών (Professional Services):** Αυτό περιλαμβάνει οποιεσδήποτε επαγγελματικές υπηρεσίες θα χρειαστεί να παρέχουν οι προμηθευτές κατά τη διαδικασία μετεγκατάστασης εφαρμογών, όπως:

a. Την προετοιμασία της εφαρμογής
b. Τις αναβαθμίσεις της εφαρμογής
c. Την προετοιμασία των δεδομένων της εφαρμογής
d. Την προετοιμασία του νέου περιβάλλοντος
e. Τις δοκιμές μετά τη μεταφορά, τις δοκιμές ενοποίησης συστήματος (SITs) / Δοκιμές αποδοχής χρήστη (UATs)

Συμπεράσματα

Ο «ψηφιακός μετασχηματισμός» έχει γίνει θέμα συζήτησης τα τελευταία χρόνια, καθώς τόσο οι μικρές και μεσαίες όσο και οι μεγάλες επιχειρήσεις έχουν μεταφέρει ένα μεγάλο μέρος των εφαρμογών τους στο cloud και χρησιμοποιούν τα εργαλεία παραγωγικότητας και συνεργασίας στο διαδίκτυο. (Aggarwal, 2021)

Η πανδημία Covid-19 άλλαξε ριζικά τον τρόπο που εργαζόμαστε, επιβάλλοντας πολλαπλά lockdowns και μακρόχρονη απομακρυσμένη εργασία. Αυτό χρησίμευσε ως μία αναπάντεχη αφορμή για επανεκτίμηση του καθεστώτος λειτουργίας για οργανισμούς όλων των μεγεθών σε όλο τον κόσμο. Οι περιορισμοί της πανδημίας ανέτρεψαν την καθημερινότητα, με τις βαριές βιομηχανίες μας όπως οι ταξιδιωτικές και η φιλοξενία να καταρρέουν και πολλές επιχειρήσεις να αντιμετωπίζουν προκλήσεις επιβίωσης.

Λόγω της αβεβαιότητας των καιρών και του «νέου φυσιολογικού», ένας αυξανόμενος αριθμός εταιριών έχει ήδη κάνει βραχυπρόθεσμα ή μακροπρόθεσμα σχέδια για τον ψηφιακό τους μετασχηματισμό, τα οποία περιλαμβάνουν, σε μεγάλο βαθμό, την υιοθέτηση του Cloud ή το "Cloudification".

Ο Διευθύνων Σύμβουλος της Microsoft Satya Nadella δήλωσε λίγους μόλις μήνες μετά την αρχή της πανδημίας ότι η εταιρεία είχε πραγματοποιήσει έναν ψηφιακό μετασχηματισμό διετίας μέσα σε δύο μήνες, ως αποτέλεσμα της υιοθέτησης των λύσεων cloud από τους πελάτες της. (Spataro, 2020)

Αυτό κατέστησε σαφές, ότι το Cloudification δεν είναι μια εφήμερη τάση, αλλά μια καθιερωμένη και μεγάλης κλίμακας μετάβαση της βιομηχανίας.

Σε αυτήν τη διπλωματική έχουμε δείξει ότι το Cloud Migration σε επίπεδο Επιχείρησης, ενώ μπορεί να είναι ένα μακροπρόθεσμο σχέδιο, δεν είναι αδύνατο να ολοκληρωθεί εντός ενός λογικού χρονοδιαγράμματος, και σίγουρα αξίζει τον προσπάθεια.

Τα πλεονεκτήματα του Cloud έναντι λύσεων βασισμένων σε κλασσικό Data Center, μπορούν να συνοψιστούν ως εξής:

1. **Ευελιξία / Ελαστικότητα:** Δεν υπάρχει ανάγκη εκτίμησης του ακριβούς μεγέθους της εγκατάστασης ούτε και πρόβλεψης μελλοντικών αναγκών. Καθώς το cloud είναι μια λύση pay-as-you-go, μπορούμε να αυξήσουμε ή να μειώσουμε τα μεγέθη των υπηρεσιών ή να προσθέσουμε και να αφαιρέσουμε πόρους όπως απαιτείται, ενώ έχουμε τη δυνατότητα να χρησιμοποιήσουμε δεσμευμένους υπολογιστικούς πόρους, για να απολαμβάνουμε υψηλότερη εξοικονόμηση κόστους στους πιο προβλέψιμους φόρτους εργασίας μας. Μπορούμε να αναπτύξουμε δοκιμαστικά περιβάλλοντα μέσα σε λίγα λεπτά και μετά να τα καταστρέψουμε όταν δεν τα χρειαζόμαστε πλέον. Οι υπηρεσίες, όπως η αποθήκευση δεδομένων, μπορούν, συνήθως, να κλιμακωθούν αυτόματα σε μεγέθη που υπερβαίνουν πολύ τις ανάγκες των περισσότερων επιχειρήσεων - Petabytes ή ακόμα και Exabytes. Για παράδειγμα, το AWS S3 δεν έχει όριο στο πόσα δεδομένα μπορούμε να αποθηκεύσουμε.

2. **Εξοικονόμηση κόστους**: Όπως έχουμε δείξει σε προηγούμενο κεφάλαιο, το cloud μπορεί να είναι σημαντικά οικονομικότερο από τις λύσεις εσωτερικής εγκατάστασης, χρησιμοποιώντας προγράμματα εξοικονόμησης κόστους. Η εξοικονόμηση κόστους, εκτός από τη δέσμευση για ελάχιστη χρήση υπολογιστικών πόρων, μπορεί να επιτευχθεί με τη χρήση ελαστικής κλιμάκωσης και υπηρεσιών PaaS cloud. Ένας άλλος παράγοντας εξοικονόμησης κόστους είναι οι οικονομίες κλίμακας που περιλαμβάνουν την εξοικονόμηση κόστους που σχετίζεται με την εργασία με μεγαλύτερες ποσότητες ενός προϊόντος. Όταν οι όγκοι των προϊόντων είναι υψηλότεροι, το ποσοστό του σταθερού κόστους ανά μονάδας παραγωγής είναι χαμηλότερο, καθώς το κόστος κατανέμεται σε μεγαλύτερο αριθμό προϊόντων.

3. **Ασφάλεια:** Αναθέτοντας την ευθύνη της φυσικής ασφάλειας στον δημόσιο πάροχο cloud, μπορούμε να απολαύσουμε ένα επίπεδο ασφάλειας που είναι πολύ δύσκολο να επιτευχθεί σε μεσαίες επιχειρήσεις. Ακόμα και ένας μεμονωμένος τελικός χρήστης μπορεί να αποκτήσει ένα διακομιστή που φιλοξενείται σε ένα κέντρο δεδομένων με φυσική ασφάλεια πολλαπλών επιπέδων και βιομετρικούς σαρωτές, εάν επιλέξει έναν μεγάλο δημόσιο πάροχο cloud (Public Cloud Provider - PCP).
   Η ασφάλεια των υπηρεσιών μας μπορεί επίσης να ενισχυθεί χρησιμοποιώντας υπηρεσίες ασφαλείας που παρέχονται από τον PCP και είναι καλά ενσωματωμένες και βελτιστοποιημένες για το συγκεκριμένο περιβάλλον. Αυτά τα στοιχεία ελέγχου μπορούν επίσης να αυτοματοποιηθούν για νέες αναπτύξεις και να επιτρέψουν σε όλες τις υπηρεσίες να "Ασφαλιστούν από το σχεδιασμό" (Secured by Design).

4. **Αξιοπιστία:** Σχεδόν όλες οι υπηρεσίες PCP φιλοξενούνται και αναπαράγονται σε πολλά Κέντρα Δεδομένων (Availability Zones) σε μια περιοχή, και μερικά μπορούν ακόμη και να αναπαραχθούν σε πολλές γεωγραφικές περιοχές (Regions). Η ανθεκτικότητα ορισμένων υπηρεσιών, όπως το AWS S3, μπορεί να φτάσει τα "11 9's" (99,99999999%), κάτι που είναι σχεδόν αδύνατο να επιτευχθεί στις εγκαταστάσεις. Η δημιουργία διαθέσιμων και ανθεκτικών

στις καταστροφές υπηρεσιών μπορεί να γίνει πολύ εύκολη στο cloud, όπως έχει αποδειχθεί στα σχέδια των κρίσιμων υπηρεσιών των προηγούμενων κεφαλαίων.

5. **Διαχείριση / Λειτουργίες / Διακυβέρνηση:** Στο cloud πολλές από τις εργασίες διαχείρισης, λειτουργίας και διακυβέρνησης μπορούν να αυτοματοποιηθούν πλήρως. Χρησιμοποιώντας υπηρεσίες PCP που είναι στενά συνδεδεμένες με την υποδομή, εργασίες όπως διαχείριση δομών, συμμόρφωση, παρακολούθηση, ειδοποίηση ακόμη και διορθώσεις (Patching) μπορούν να αυτοματοποιηθούν. Αυτό σημαίνει ότι οι, ανθρώπινοι και μη, πόροι μπορούν να διατεθούν σε άλλους, πιο παραγωγικούς τομείς.

Οι Υπηρεσίες Cloud έχουν απλοποιήσει σε μεγάλο βαθμό την έννοια «Απομακρυσμένη εργασία», καθιστώντας τις καθημερινές υπηρεσίες προσβάσιμες από παντού. Η ανάγκη να είμαστε φυσικά παρόντες στις εγκαταστάσεις μιας εταιρείας για πρόσβαση στις υπηρεσίες της, έχει μειωθεί σημαντικά, η ευελιξία της επιλογής ενός φορητού υπολογιστή ή ενός σταθμού εργασίας πέρα από την εξαιρετικά ασφαλή και πολύ ακριβή συσκευή που προσφέρει η εταιρία, είναι ένα από τα πλεονεκτήματα που μας προσφέρει το DaaS ενώ πολλές άλλες υπηρεσίες SaaS έχουν καταστήσει δυνατή την εργασία σε οποιοδήποτε σημείο υπάρχει ένα σύγχρονο πρόγραμμα περιήγησης.

Είναι αδιαμφισβήτητο ότι οι υπηρεσίες cloud έχουν διαδραματίσει σημαντικό ρόλο κατά τη διάρκεια της πανδημίας και προβλέπεται ότι αυτός ο ρόλος θα συνεχίσει να αναπτύσσεται για αρκετά χρόνια ακόμα. Η υιοθέτηση cloud σε όλες τις επιχειρηματικές περιοχές μιας επιχείρησης θα οδηγήσει τον κλάδο στο μέλλον, καθώς τα λειτουργικά μοντέλα θα συνεχίζουν να αλλάζουν προς πιο ευέλικτες προσεγγίσεις.

Το Cloud διαθέτει πολλά πλεονεκτήματα, και όπως αποδείξαμε είναι επίσης πιο οικονομικό.

Λίστα εικόνων

Λίστα πινάκων

Βιβλιογραφία

Aggarwal, G. (2021, 01 15). *How The Pandemic Has Accelerated Cloud Adoption*. Retrieved from Forbes: https://www.forbes.com/sites/forbestechcouncil/2021/01/15/how-the-pandemic-has-accelerated-cloud-adoption/

AWS. (2018). *AWS Migration Whitepaper.* Retrieved from AWS: https://docs.aws.amazon.com/whitepapers/latest/aws-migration-whitepaper/welcome.html

AWS. (2020, July). *AWS Well-Architected Framework.* Retrieved from AWS Documentation: https://docs.aws.amazon.com/wellarchitected/latest/framework/welcome.html

AWS. (2021). *AWS Single Sign-On*. Retrieved from AWS Documentation: https://aws.amazon.com/single-sign-on/

AWS. (2021). *Directory Service*. Retrieved from AWS Documentation: https://aws.amazon.com/directoryservice/

AWS. (2021). *General Data Protection Regulation (GDPR) Center*. Retrieved from AWS Documentation: https://aws.amazon.com/compliance/gdpr-center/

Azeem, S. A., & Sharma, S. K. (2017). Study of Converged Infrastructure & Hyper Converge Infrastructre As Future of Data Centre. *International Journal of Advanced Research in Computer Science, 8*(5), 900-903. Retrieved 6 6, 2021, from http://ijarcs.info/index.php/ijarcs/article/view/3476

Citrix. (2021, 06). *VDI and DaaS*. Retrieved from Citrix Web Site: https://www.citrix.com/solutions/vdi-and-daas/what-is-desktop-as-a-service-daas.html

Cloudflare. (2021). *What is SSO.* Retrieved from Cloudflare website: https://www.cloudflare.com/learning/access-management/what-is-sso/

Comindware. (2021, 06 21). *What is CapEx and OpEx*. Retrieved from Comindware: https://www.comindware.com/blog-what-is-capex-and-opex/

DEI. (2020, March). *DEI Pricelist.* Retrieved from dei.gr: https://www.dei.gr/Documents2/TIMOLOGIA/TIM-MARTIOS-2020/TIMOK-MT-2020-BG-MARCH20full.pdf

European Union. (2021). *What are the GDPR Fines?* Retrieved from European Union GDPR site: https://gdpr.eu/fines/

Forbes. (2020, October 26). *The History And The Future Of Cloud Office Suites*. Retrieved from Forbes: https://www.forbes.com/sites/forbestechcouncil/2020/10/26/the-history-and-the-future-of-cloud-office-suites/

RackSpace. (2020, 08 26). *What is Cloud Native*. Retrieved from RackSpace: https://www.rackspace.com/blog/what-is-cloud-native

Rasmussen, N. (2007, February). *Calculating Total Cooling Requirementsfor Data Centers.* Retrieved from APC: https://www.apcdistributors.com/white-papers/Cooling/WP-25%20Calculating%20Total%20Cooling%20Requirements%20for%20Data%20Centers.pdf

RedHat. (2020). *What is CI/CD*. Retrieved from RedHat: https://www.redhat.com/en/topics/devops/what-is-ci-cd

SDX Central. (2016, 05 18). *What Is Cloud Native? Definition*. Retrieved from SDX Central: https://www.sdxcentral.com/cloud/cloud-native/definitions/what-is-cloud-native-definition/

Spataro, J. (2020, 04 30). *2 years of digital transformation in 2 months* . Retrieved from Microsoft: https://www.microsoft.com/en-us/microsoft-365/blog/2020/04/30/2-years-digital-transformation-2-months/

Wikipedia. (2021, 3 12). *Hyper-converged infrastructure*. Retrieved from wikipedia.org: https://en.wikipedia.org/wiki/Hyper-converged_infrastructure

English Version

Below we present the English version of the Master Thesis.

Table of Contents

Introduction

During the recent COVID-19 pandemic, it has become clear that Enterprises of any size need to re-evaluate their operating model in order to become able to adapt to the new challenges. The pandemic has abruptly introduced the need for flexibility in both the workload of the IT systems, as well as the working conditions of the employees and the consumer behavior. And while the old, traditional operating model was invalidated in a matter of weeks, the new operating model presents us with many challenges as well as opportunities, but most of all, reminds us that many of these changes were long due.

The cloud transformation has been going on for the past decade, but has really taken off during 2020, when the need for digital services and remote working rose internationally and exponentially. Customers require more of the, traditionally physical, services to be offered via the internet and employees were required to perform their duties without having access to the office premises.

The target of this thesis it to analyze and plan the cloudification of an Enterprise in a core sector that was affected in both the service load as well as the working conditions of its employees, a Telecommunications Company (TelCo).

We will

A. Analyze the current state of the company.
    a. Analyze the company structure.
    b. Analyze the IT services.

B. Design a Cloud based solution for every type of IT services.
    a. Define Cloud native services and best practices.
    b. Translate the current services to Cloud Ready / Native services
    c. Perform Cost analysis and comparison to on-premise Data Center costs.

C. Provide a migration plan.
    a. Analyze Strategies.
    b. Analyze Cost Vectors.

We will not dive deep into the organizational structure of the Company, hence major groups of IT services such as Contact Center will be analyzed, and minor services such as legal services will be integrated into larger groups. We will also not have in scope the Telecommunication Network Support services, for both Fixed and Mobile, as they are engineering services and are required to be as near as possible to the physical network.

Finally, we choose AWS as a Cloud service provider, based on its high availability and vast portfolio of cloud services.

“As is” analysis

Enterprise structure

As the entire enterprise structure is out of the scope of this thesis, we have concentrated the analysis on the IT related organizational units and have based our results on data from real enterprises. Since these data are of strategic nature, and cannot be publicized, special care has been given to sanitization and while the numbers that follow are a very close approximation, they cannot be directly attributed to any single Telecommunication Company (TelCo).

In any TelCo we can identify two major Service Categories: Information Technology (IT) services and Network (NW) Services. Network Services comprise of all the services that support the network operation for both fixed and mobile services.

Fixed services include DSLAM management services, Customer-Premises Equipment (CPE) management services, Connectivity Monitoring services, Physical Network (Copper / Fiber) mapping, planning and operation services and internet connectivity services (backbone connectivity).

Mobile services include Tower operations services (connectivity monitoring, tower security, tower power, tower access), network health monitoring, network capacity monitoring and consolidation services, 5G services, Text Routing (SMS services).

These services need to remain close to the physical network, and unless there is a Public Cloud Provider data center in the same region as the physical network, it is inefficient to transfer them to the Cloud. At the time of the writing no Public Cloud Provider has physical presence in Greece, so these services will remain out of scope of this thesis.

The IT services are decoupled from the network and interact with it only through other network services, and as such do not require close proximity to the physical network. We will focus our analysis on these services in the following section.

IT Services

The following units were identified, with granularity up to 1% of the TCO, with any smaller units being consolidated into larger ones.

Consolidated List	
Department	**% of TCO**
Customer Management	23,00%
Billing Services	14,98%
Business Intelligence	14,29%
Service Order Management	10,45%
Contact Center	8,01%
Backend Support	6,97%
Frontend Applications	6,97%
Middleware Services	6,97%
ERP	5,92%
Incident Management	1,39%
Content Management	1,05%
Total	100,00%

Table 1: Consolidated list of Departments

As we can see in Table 1, the Organizational Units shorted for cost, are:

1. **Customer Management**, including Customer Relationship Management (CRM) such as Siebel and Business Relationship Management (BRM), which are the backbone of any service provider type of enterprise and as such, consume by far the greatest portion of the TCO.

   A CRM software aids businesses in staying in touch with customers, streamlining procedures, and increasing profits. BRM, on the other hand, is a strategic business software focusing on account management and collaboration in the commercial process. It integrates with CRM systems to provide for a two-way flow of pertinent data (push and pull).

   These are usually huge monolithic applications with large amounts of historical data and also large amounts of customizations.

2. **Billing Services**, including billing, accounting and other financial services like rating, invoicing and pdf creation for customers.
   These usually include localized applications that follow local legislature and might lack cloud-compatible versions.

3. **Business Intelligence,** including all Big Data related services for collecting, processing, and storing Big Data, as well as Business Intelligence (BI) services for analyzing such data to extract business usable information. BI services such as predictive analysis, always ON marketing, etc. are also included in this category.

   These are usually applications that can be easily migrated to the cloud, as they have been developed recently, they have cloud-compatible versions, as well as cloud-optimized versions.

4. **Service Order Management,** including all front-end and automation services for processing service orders from customers.
   These are typical applications that can be directly migrated to the cloud as they do not have any specific requirements.

5. **Contact Center,** including:

   Computer Telephony Integration (CTI) - software that enables you to have all the capabilities of a call center (Call/response recording, Routing etc...) without the use of a typical phone device. It also integrates with customer-related software, providing specific customer data and history directly to the agent's workstation at the beginning of the inbound call, in order to maximize efficiency and improve service quality.

   Intelligent Workload Distribution (IWD) system - a solution that establishes a centrally managed and prioritized task list. It enables work to be presented to the appropriate resource at the appropriate time and location. It collects non-real-time work (tasks) from multiple source systems, prioritizes or reprioritizes the tasks based on business rules, and then distributes the tasks to the most appropriate resource/agent, reducing time and cost.

   Interactive Voice Response (IVR) - a telephone system technology that enables incoming callers to access information through a voice response system of pre-recorded messages without speaking with an actual agent.

   This includes also other systems, like Intelligent Text Chat systems and Social Media integrations.

Special care should be given to these services, due to their latency sensitivity. In most cases 250ms is the maximum acceptable latency while ITU-T G.114 recommends a maximum of a 150ms one-way latency. For our case, the data centers of the public cloud providers in Germany, Italy and France provide low enough latency, as we can see in the following table 2.

AWS	
Region description	**Latency (ms)**
Milan	57
Frankfurt	62
Paris	67
London	78
Stockholm	83
Ireland	83
Azure	
Region description	**Latency (ms)**
Paris	81
Frankfurt	85
Zurich	87
London	93
Cardiff	94
Ireland	97
Netherlands	101
Norway	112
GCP	
Region description	**Latency (ms)**
Frankfurt	62
Eemshaven (NL)	71
St. Ghislain (BE)	73
London	77

Table 2: Latency of Public Cloud Provider Regions from Greece

6. **Backend Support,** including Security Services, Identity and Access Management, Internal Information and Collaboration portals, HR Services and Legal Services.

   Some of these (for example, Identity and access management and Security Services) must remain partially on-premises, as they are used also by the Network Services. For these services, a hybrid cloud approach will be followed. The rest of the services can be safely migrated, as there is no dependency on latency or proximity.

7. **Frontend Applications,** including Customer facing portals (Company's website, Youth Portal) and campaign portals.

These are ideal candidates for cloud migration, as web applications of this kind are latency tolerant and will benefit greatly from cloud services such as Content Delivery Network (CDN), auto-scaling, etc.

8. **Middleware Services,** including Enterprise service bus (ESB) and any API / File transfer abstraction layer services.

   The applications that depend on these Middleware Services, will be migrated to the cloud, but some might also remain on premise – such as the Network services – we will follow a hybrid approach for these too.

9. **ERP,** including any Enterprise Resource Planning application such as SAP.

   These are also usually large monolithic applications, but with less customization than the CRM and BRM, so it will be much easier to migrate. Additionally, all major ERP providers already have cloud-hosted solutions that can be consumed “as a Service” (SaaS).

10. **Incident Management,** including all internal ticketing systems.

    These are typical applications that can be directly migrated to the cloud as they do not have any specific requirements.

11. **Content Management,** including all customer facing content management systems such as Internet TV.

    These are heavily bandwidth dependent, so they must remain on-premises, at least partly. The content itself should remain on-premises, close to the local network, to leverage capacity and free local content delivery, while the control plane can be transferred to the cloud and create a hybrid cloud installation.

“To Be” Design.

Cloud Services

The term "cloud services" refers to a broad range of on-demand services supplied over the internet to businesses and consumers. These services are intended to provide simple, cost-effective access to applications and resources, without requiring internal infrastructure or hardware.

Cloud computing vendors and service providers manage all aspects of cloud services. They are made available to customers via the providers' servers, which eliminates the need for a business to host the applications on-premises.

Cloud computing has fundamentally altered the way consumers and businesses operate, by providing significant benefits over the more traditional on-premises datacenter options, such as:

1. **Agility**: The ability to have access to a diverse set of technologies that enables us to innovate more quickly and build nearly anything we can imagine. We can rapidly provision resources as needed–from infrastructure services such as compute, storage, and databases to Internet of Things, machine learning, data lakes, and analytics, among other services. Deployment of any type of service or IT resource occurs in a matter of minutes.

2. **Elasticity**: With cloud computing, we avoid the upfront cost of over-provisioning resources to handle future spikes in business activity. Rather than that, we provision only the resources that we require. We can instantly scale up or down these resources based on our current business requirements, in most cases - in a matter of minutes.

3. **Cost savings**: The cloud makes it feasible to trade Capital Expenses (CAPEX) for Operational Expenses (OPEX) by eliminating the costs of owning and operate data centers and physical servers and only pay for the resources that we consume. Additionally, the variable costs are significantly lower than what we would pay to do it on our own, due to economies of scale.

4. **Global Deployment**: In minutes, our applications can be globally deployed to a new geographic region. By deploying our services in multiple physical locations, we are able to maintain a close proximity to the end users (customers), thereby reducing latency and improving the overall customer experience.

There are three primary types of cloud computing, and each one offers a different level of control, flexibility, and management to meet the unique needs of each customer.

- **Infrastructure as a Service (IaaS)**: IaaS provides the fundamental building blocks for cloud computing. It typically provides access to networking capabilities, computers (virtual or dedicated), and storage space for data. IaaS enables us to have the greatest degree of control and flexibility over our IT resources.

- **Platform as a Service (PaaS):** PaaS eliminates the need to manage underlying infrastructure (typically hardware and operating systems) and makes more time available to focus on application deployment and management. This increases our efficiency because we are no longer responsible for resource acquisition, capacity planning, software maintenance, patching, or any of the other heavy lifting associated with running the applications. This type of service is ideal for developers that want to build, test and deploy their code without the need of provisioning and managing the underlying infrastructure.

- **Software as a Service (SaaS):** SaaS is a fully functional product that is managed and supported by the service provider. Typically, when people refer to SaaS, they are referring to end-user applications (such as web-based email). With a SaaS offering, customers are not responsible for the service's maintenance or management of the underlying infrastructure, they simply need to consider how they will use the software. This type is the most recognized type of cloud service with a variety of services, such as file storage, backup, web-based email, and project management tools.

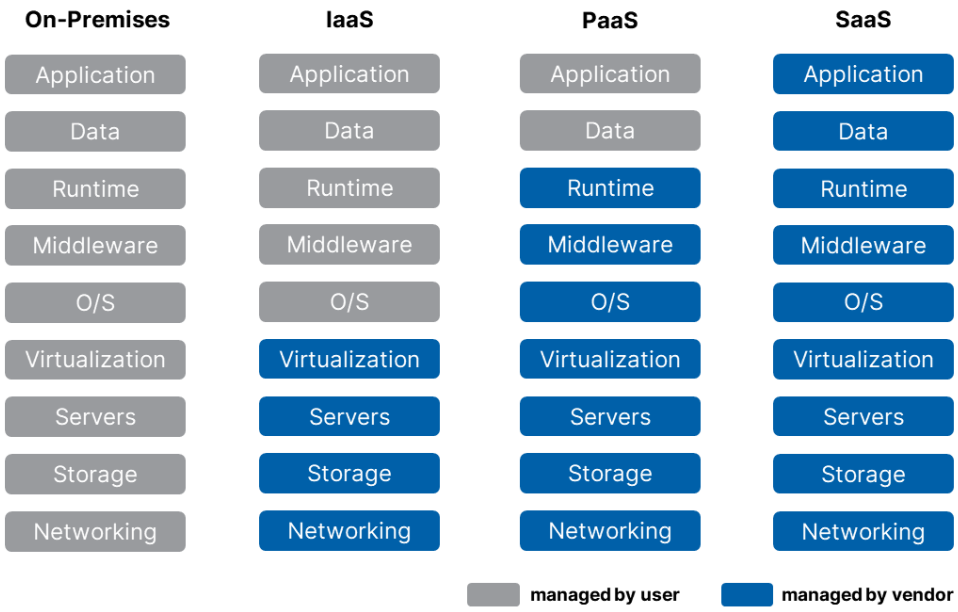


Figure 1: On-Prem, IaaS, PaaS, SaaS (PentaSecurity)

Cloud Native

As we already now, we live in a fast-paced technological era, in which applications have become increasingly complex, as users have increased their expectations. Users expect instantaneous response, cutting-edge features, and zero downtime. Performance issues, recurrent errors, and inability to move quickly, are no longer acceptable. These requirements led to the development and use of technologies and methodologies that provide faster deployments, agility, reliability, increased customer satisfaction and cost efficiency. (RedHat, 2020) (RackSpace, 2020) (SDX Central, 2016)

Cloud-native refers to the approach used to build cloud-based applications. Cloud-native applications and services differ from their legacy counterparts because they are specifically designed for the cloud from ground up. They can be deployed and repaired more quickly, have a more fluid architecture, and can be placed and moved easily in a variety of environments.

Cloud-native technologies enable businesses to develop and run scalable applications in modern, dynamic environments such as public, private, and hybrid clouds. They enable engineers to make high-impact changes frequently and predictably, with minimal effort when combined with robust automation.

The applications and services that meet the aforementioned requirements are called Cloud-Native.

Cloud Native Applications have a few common elements such as:

- **Containerized:** Containers are logical packaging software that contain all of the necessary elements for an application to run, such as application code, runtime, configuration files and dependencies.

- **Microservices:** They are an architectural approach for application development and provision. Prior to this technology, applications or systems were composed of one piece by using a single code base for all their functionalities and services. This approach is called monolithic architecture, and these applications have a major disadvantage when it comes to deploying code. The whole systems/services have to be redone from scratch, leading to service unavailability, extra effort and major delays. Microservices are deployed in containers and apply the "divide and conquer" method to the whole service or application, by isolating each application's function as its own service, inside its own container connecting them via APIs. This method offers loosely coupled applications that enable engineers to build and deploy their code without the risk of breaking or shutting down the whole application.

- **Serverless computing:** This is the technology that enables engineers and developers to run their processes and operations without the need of provisioning the underlying platforms. Serverless does not mean that no servers are involved in computing processes. It means that there is no longer a need to maintain the servers required to keep the operations running. All the heavy lifting of operations management takes place outside of the business. This leads to delivering greater efficiency, stronger security, and bigger cost savings.

- **Continuous Integration/Continuous Delivery:** Also called CI/CD Pipeline is a methodology, visualized in a form of a pipelined workflow that enables frequent application delivery to customers by introducing continuous automation and monitoring into all stages of application development lifecycle - from integration and testing phases to deployment and monitoring. This approach offers an automated, reliable, and faster way to constantly fix bugs, test and deploy code, without the risk of breaking down or affect the application.

  To implement and facilitate this process, a need for collaboration and combined effort of different engineering teams is required. Both software development (plan, code, build, test) and IT operations (release, deploy, operate, monitor) teams are combined into one consolidated cross-functional group called DevOps.

DevOps also refers to a work environment, set of practices and culture that this combined team needs to have, to constantly communicate and collaborate with the goal of achieving software and infrastructure service provisioning rapidly and frequently via automated and standardized processes.

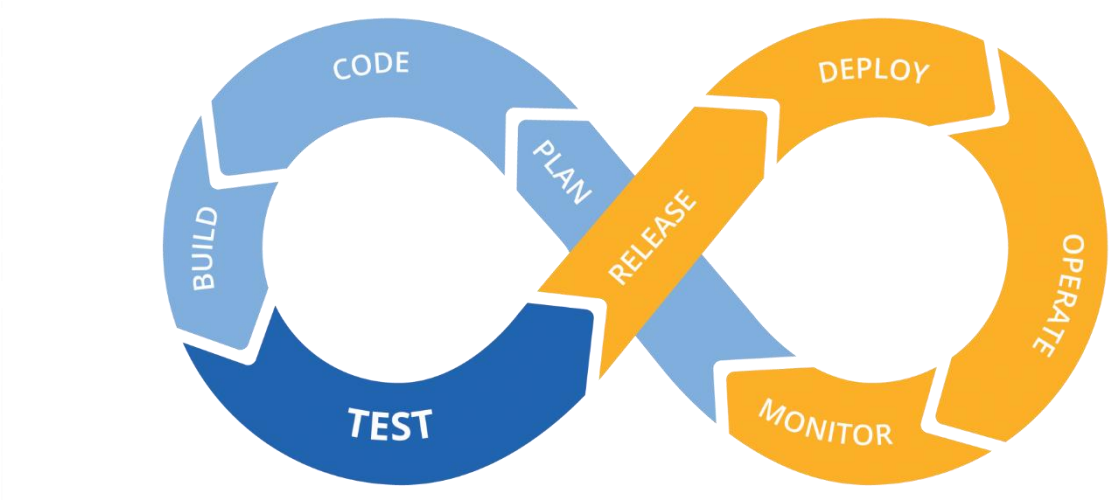


Figure 2: CI/CD Pipeline

Security by Design

Nowadays, data can influence everything, from market trends, political elections, to starting or ending wars. For these reasons, data have become more valuable than gold and proper use and safety of them, has never been more important. The need of physical and logical security of the hosting infrastructure, services and customer data is constantly rising. Enterprises are trying to find more secure ways of accessing their applications, migrate and store their data and protect their on-premises datacenters, while mitigating their costs and efforts for provisioning and monitoring. Below, we underly some of the key services and practices for better security of the infrastructure, accounts, and services, not only regarding a TelCo company, but for any medium and large-sized enterprise.

One of the main services regarding user access control and accounts provision to the AWS services and resources, is the AWS Identity and Access Management (IAM) service. AWS Identity and Access Management (IAM) enables you to define individual user accounts with permissions across AWS resources.

For privileged accounts, AWS Multi-Factor Authentication (MFA) is available, with options for software- and hardware-based authenticators. This feature offers an extra layer of authentication of the user, by providing security tokens or passwords via a mobile application or a hardware device connection. IAM can be also used to grant federated access to the AWS Management Console and AWS service APIs to the employees and applications via the existing identity management systems, such as Microsoft Active Directory.

Identity federation is a trust relationship between two parties that enables the authentication of users and the transmission of information necessary to authorize their access to resources. Federation is a widely used approach for developing access control systems, that centrally manage users within a single IdP (Identity Provider) and manage their access to multiple applications and services acting as Service Provider (SP).

A powerful IAM source tool is AWS Directory Service, also known as AWS Managed Microsoft Active Directory (AD), which enables the use of managed Active Directory (AD) in AWS for directory-aware workloads and AWS resources. AWS Managed Microsoft AD is built on top of an existing Microsoft Active Directory and does not require synchronization or replication of data from the on-premises Active Directory to the cloud. We can use standard Active Directory administration tools and leverage built-in Active Directory features, such as Group Policy and single sign-on, and in the same time reducing administrative overhead of keeping data across multiple directories.

Single Sign-On (SSO) is a technology that combines multiple application login screens into a single one. SSO requires a user to enter their login credentials (username, password, etc.) once, on a single page, in order to access all of their SaaS applications. AWS SSO offers the ability to manage access of multiple AWS accounts and applications, centrally by configuring and maintaining all of the necessary permissions for the company's accounts automatically, without additional setup. In addition, it offers user permissions based on job roles (Role Based Access Control - RBAC) and customization of these permissions in terms of alignment with specific security requirements. The most common and best-practice rule of creating and managing permissions to user or group accounts is called "Least Privilege" and it means that individuals can have a level of access and permission sufficient for their purpose or role only. (AWS, 2021) (AWS, 2021) (Cloudflare, 2021)

Segregated Accounts

Rather than replicating the organization's reporting structure, it is recommended that workloads are organized in separate accounts and these accounts are grouped based on function, compliance requirements, or a common set of controls. Accounts are a hard boundary in AWS and separating production workloads from development and test workloads is strongly recommended. (AWS, 2020)

- **Utilize accounts to segment workloads**: Begin with security and infrastructure in mind to enable the organization to establish standardized safeguards as the workloads grow. This approach establishes boundaries and maintains control over workloads. Separating workloads at the account level is strongly recommended for isolating production environments from development and test environments, as well as for establishing a strong logical boundary between workloads that process data with varying degrees of sensitivity, as defined by external compliance requirements (such as PCI-DSS or GDPR), and workloads that do not.

- **Secure all AWS accounts**: Securing an AWS account entails a number of steps, including securing and not using the root user, as well as maintaining current contact information. AWS Organizations enables us to centrally manage and govern the accounts as the workloads grow and scale in AWS. AWS Organizations allows users to manage accounts, configure security settings, and configure services across multiple accounts.

- **Centrally manage accounts**: AWS Organizations automates the creation and management of AWS accounts, as well as the control of those accounts once they are created. Organizations enables us to organize accounts into organizational units (OUs), as will be discussed in a later section, which can represent distinct environments depending on the requirements and purpose of the workload.

- **Centralize control**: Control what the AWS accounts can do by restricting access to specific services, regions, and service actions. AWS Organizations enables us to apply permission guardrails at the organization, organizational unit, or account level that apply to all AWS Identity and Access Management (IAM) users and roles via service control policies (SCPs). For instance, we can implement a SCP that prevents users from launching resources in regions that we have not expressly permitted. AWS Control Tower simplifies the process of setting up and managing multiple accounts. It automates account creation in our AWS Organization, automates provisioning, applies guardrails (including prevention and detection), and provides a dashboard for visibility.

- **Centrally manage services and resources**: AWS Organizations enables us to centrally manage AWS services that apply to all of our accounts. For instance, we can use AWS CloudTrail to centrally log all actions performed across our organization and prevent member accounts from disabling logging. Additionally, we can centrally aggregate data for rules defined with AWS Config, which enables us to audit the workloads for compliance and respond quickly to

changes. Infrastructure as Code (IaC) automation enables us to provision a new account automatically to meet our security requirements.

Secure Network

Users, both internal and external to our organization, can be located anywhere. We must abandon traditional models of trusting anyone and everything with network access. When we adhere to the principle of security at all layers, we are implementing a Zero Trust strategy. Zero Trust security is a model in which application components or microservices are considered distinct from one another and no component or microservice trusts any other component or microservice. (AWS, 2020)

Cloud Security best practices include:

- **Creating network layers**: To segment components such as Compute instances, database clusters, and functions that share reachability requirements, subnets can be used to create layers. For instance, a database cluster that does not require internet access should be placed in subnets that do not have any route to or from the internet. These subnets should also be configured with a “least access” design principle in mind, which means that communication between subnets and assets should be permitted on a case by case and directionally, so as not to allow subnet level access – and thus an attacker that has gained access on a subnet, compromise an entire other subnet. This layered approach to network access control configuration mitigates the impact of a single layer misconfiguration which would allow unintended access.

- **Control traffic at all layers**: When designing our network's topology, we should consider each component's connectivity requirements. For instance, whether a component requires internet connectivity (both inbound and outbound), connectivity to virtual private clouds (VPCs), edge services, or external data centers. Multiple controls should be applied in a defense-in-depth fashion to both inbound and outbound traffic, including the use of security groups (stateful inspection firewalls) on an instance level, network access control lists (ACLs) on a subnet level, subnets, and route tables. Each subnet may have its own route table, which defines the routing rules for managing the paths taken by traffic within the subnet.

  When an instance, database, or other service is launched, each network interface has its own security group. This firewall operates on a separate layer from the operating system and can be used to define rules for allowed inbound and outbound traffic. Additionally, relationships between security groups can be defined. For example, instances within a database tier

security group accept traffic only from instances within the application tier, based on the security groups assigned to the corresponding instances. Moreover, a subnet can be associated with a network ACL, which acts as a stateless firewall. The Network ACL should be configured to restrict the type of traffic allowed between layers.

Connectivity to the internet should also be controlled and layered. No instance should be allowed to freely access the internet. The requirement should be documented, whitelisted using an outgoing proxy, which should be the only asset that can directly access the internet, and connectivity must be logged. Any unauthorized request must raise a security alert and be investigated.

- **Conduct inspections and provide protection**: At each layer, we should inspect and filter the traffic. A web application firewall (WAF) can help protect components that communicate via HTTP-based protocols from common attacks. By filtering and monitoring HTTP traffic between a web application and the Internet, a WAF helps protect web applications. It typically safeguards web applications against cross-site forgery, cross-site scripting (XSS), file inclusion, and SQL injection, among other attacks. A WAF is a layer 7 (in the Open Systems Interconnection model) protocol defense, that is not intended to defend against all types of attacks. This method of attack mitigation is typically used in conjunction with a suite of tools (Layer 4 Firewalls, Intrusion Prevention Systems) that together provide a comprehensive defense against a variety of attack vectors.

  When a WAF is deployed in front of a web application, it creates a barrier between the web application and the Internet. While a proxy server protects the identity of a client machine through the use of an intermediary, a WAF acts as a reverse proxy, shielding the server from exposure by requiring clients to pass through the WAF prior to reaching the server.

  A WAF is governed by a set of rules, frequently referred to as policies. These policies are designed to guard against application vulnerabilities by filtering out malicious traffic. The value of a WAF is partly due to the speed and ease with which policies can be modified, allowing for quicker response to varying attack vectors. For example, rate limiting can be quickly implemented during DDoS attacks by modifying WAF policies.

  An intrusion prevention system (IPS) is a type of network security that monitors for and prevents the occurrence of identified threats. Intrusion prevention systems continuously monitor the network, looking for and logging possible malicious incidents. The IPS notifies system administrators of these events and takes preventative action, such as shutting down access points and configuring firewalls, to ward off future attacks. Additionally, IPS solutions

can be used to detect violations of corporate security policies, deterring employees and network guests from violating the rules contained in these policies.

- **Automate network protection**: On the cloud the previous services can work together to alert and mitigate security risks without delays and human intervention and are usually offered as a Service. Using threat intelligence and anomaly detection, intrusion detection and prevention tools can adapt to and mitigate the impact of current threats. A web application firewall is an example of how we can automate network protection: by utilizing the AWS WAF Security Automations solution to automatically block requests originating from known threat actor IP addresses.

For communication between accounts, it is recommended to use an isolation layer, like AWS Private link. This is a service that creates a link between a network load balancer (NLB) and a VPC endpoint, using the cloud provider's backbone network, which is hidden and inaccessible to all consumers. This allows us to maintain all accounts network-independent, without the need to have non-overlapping IP ranges across all accounts, keep an IP management system, nor carefully plan routing. The VPC Endpoint resides and is addressed inside the network of the source system, while the NLB resides and is addressed in the network of the target system. This provides, apart from network-independence, also an additional layer of security, as communication needs to be explicitly established by using predefined IaC templates.

Data Protection/ GDPR

In April of 2016, a new European privacy law known by its abbreviation "GDPR" (General Data Protection Regulation) has been legislated, in order to enforce and harmonize certain data protections laws throughout the European Union and its state members accordingly. The GDPR applies to all organizations that process personal data in the EU, whether they have an establishment in the EU or they process personal data of EU residents when offering goods or services to individuals in the EU, or when monitoring their behavior in the EU. Any information relating to an identified or identifiable natural person is considered personal data. Violation or non-compliance of this privacy law could result in a fine of up to 20 million euros or 4% of the company's worldwide annual revenue from the preceding financial year, whichever amount is higher. (European Union, 2021)

As we have already underlined, security of data is of paramount importance, therefore encryption techniques need to be applied, both in data **at rest** and in data **in transit**.

For regulatory compliance and data protection, encrypted data at rest is critical. It contributes to the security of sensitive data stored on disks by ensuring that no user or application can read it without a valid key. AWS offers a variety of options for encryption at rest and key management. For example, we can encrypt data before it is written to a non-volatile storage, like encryption of AWS EBS volumes or configure AWS S3 bucket for Server-Side Encryption (SSE) by using AES-256 Encryption, or even Client-Side encryption which offers the ability to encrypt data before sending them to a S3 bucket. Encryption of data at EC2 Instance level is feasible too, by having disk-level or file system-level encryption.

On the other hand, AWS has also put a lot of effort to secure and encrypt data in transit from one system to another, including resources within or outside of AWS. By creating an AWS account, a logically isolated section of the AWS Cloud called Amazon Virtual Private Cloud (VPC) is provisioned to it. There, we can launch AWS resources within a defined virtual network. We have complete control over the virtual networking environment, including the ability to configure our own IP address range, subnets, route tables, and network gateways.

Additionally, we can establish a hardware Virtual Private Network (VPN) connection between our corporate datacenter and our Amazon VPC, enabling us to leverage the AWS Cloud as an extension of our corporate datacenter. Regarding protected communication between Amazon VPC and corporate on-premises datacenters, AWS has a variety of VPN connectivity options according to the business and technical needs of the company. The AWS Client VPN enables secure access to AWS resources via client-based VPN services. On the AWS Marketplace, we can purchase a third-party software VPN appliance that we can install on an Amazon EC2 instance in our Amazon VPC.

Alternatively, we can establish an IPsec VPN connection between our VPC and our remote network to secure communication. We can use AWS Direct Connect to establish a dedicated private connection

from a remote network to our Amazon VPC. This connection can be combined with an AWS Site-to-Site VPN to create an encrypted private connection using IPsec tunnels. (AWS, 2021)

Disaster Recovery and Business Continuity Plans

A business continuity plan is a comprehensive plan that outlines how a business will continue to operate in the event of a disaster. This plan is comprehensive in scope but drills down to specific scenarios that could result in operational risks. The goal of business continuity planning is to maintain critical operations so that our business can continue to operate normally even in the face of unusual circumstances.

When implemented properly, a business continuity plan should enable customers to continue receiving services with minimal disruption during or immediately after a disaster. A thorough plan should also consider the requirements of business partners and vendors.

The continuity plan should exist as a written document outlining the critical functions of the business. This is likely to include a list of critical supplies and business functions, as well as copies of critical records and contact information for key employees. The information contained in the plan should enable the business to resume normal operations as quickly as possible following a disruptive event.

A disaster or data recovery plan is a more specialized component of a larger business continuity plan. Occasionally, the scope of a disaster recovery plan is condensed to focus exclusively on a business's data and information systems. In the simplest terms, a disaster recovery plan is used to save data in the event of a disaster with the sole purpose of quickly recovering it. With this objective in mind, disaster recovery plans are typically developed to address the specific requirements of the information technology department for resuming operations—which ultimately affects the entire business.

Depending on the nature of the disaster, the plan may include everything from recovering a single file to recovering an entire datacenter. Because the majority of businesses rely heavily on information technology, a disaster recovery plan is a critical component of successful business continuity planning.

In some instances, disaster recovery planning may also refer to protocols that exist outside of the information technology department. For instance, disaster recovery plans could include procedures for recovery personnel to locate a backup business location in order to resume critical operations. This could be advantageous in the event of a natural disaster, such as flooding, rendering the existing business premises unusable. Additionally, the plan may include instructions on how to re-establish communication between emergency personnel if the usual lines of communication are unavailable. If

our IT department is developing an IT-focused plan, it is critical to incorporate all non-IT recovery protocols into the broader business continuity plan.

To summarize, disaster recovery is the process of restoring data, servers, files, software applications, and operating systems following a destructive event. By contrast, business continuity refers to how a business continues to operate in the event of a technological failure or outage. In other words, a disaster or data recovery plan specifies how a business should respond to a disaster, whereas a business continuity plan specifies how a business can continue operations in the event of a disaster.

The core services, that are absolutely critical for the continuous operation of the enterprise, should be part of the Business Continuity (BC) Plan and have strict Disaster Recovery (DR) plans, while non-critical services can have less strict DR plans, if any.

RTO/RPO and MTPoD

RTO stands for Recovery Time Objective and is a term that refers to the amount of time an application can be unavailable without causing significant business disruption. Certain applications can be unavailable for days without causing significant damage. Certain critical applications can be unavailable for a few seconds without causing employee annoyance, customer anger, or lost business.

RTO is not synonymous with the time period between loss and recovery. Additionally, the objective takes into account the steps required by IT to restore the application and its data. If information technology has invested in failover services for critical applications, they can safely express RTO in seconds. (IT is still responsible for restoring the on-premises environment. However, because the application is being processed in the cloud, IT can take as much time as necessary.)

RPO stands for Recovery Point Objective and refers to the organization's loss tolerance: the amount of data that can be lost without causing significant harm to the organization. The objective is specified in terms of the time interval between the loss event and the most recent preceding backup.

If we back up all or most of our data in regularly scheduled 24-hour increments, we will lose 24 hours' worth of data in the worst-case scenario. This is acceptable for some applications. For others, it is categorically not the case.

The maximum tolerable period of disruption (MTPoD) is the time period following a disaster during which an organization's viability will deteriorate irreversibly if production is not resumed. When

conducting a business impact analysis (BIA) and developing a disaster recovery/business continuity plan, MTPoD is critical.

The above three metrics are very important to determine the criticality of each service, and make plans for the Backup Strategies, Disaster Recovery and Business Continuity.

Services that have MTPoD and RTOs in days (1 or more), usually only have some default backup strategies. The restore process should be documented, and the restoration duration should fall within the RTO.

Services that have MTPoD and RTOs measured in hours, usually have simple disaster recovery plans that account for recreating the infrastructure on a secondary site when a disaster strikes. Thus, the backup strategy should account for immediate availability of backups on the secondary site.

Services that have MTPoD and RTOs measured in minutes, usually have far stricter disaster recovery plans, and are part of the Business Continuity Plan too. They usually have Active or Passive Standbys in another region (more than 300 KMs apart), make use of online replication of data, and their recovery time can range from near real-time to a few minutes.

All Plans should be part of periodical Disaster Tests, in order to test and prove readiness and effectiveness.

Backup Strategies

Backup strategies should also be well documented, implemented and periodically tested by doing restores and data integrity tests. They are designed according to the required RPO and the criticality of the service.

So, an RPO of 24 hours, means that the Business can accept the loss of a maximum of 24 hours' worth of data and an adequate backup strategy would be taking full backups every day at 00:00 (or any other low-load time).

An RPO of 1 hour would force us to create a more complex backup strategy, for example full backups on low-load times and incremental backups every 1-hour otherwise, but this is highly depended on the type of the data that are being backed up. For example, Containers are immutable, so keeping a backup whenever they change is enough, as well as with configuration files that should be updated through a code repository, while databases are always changing so the narrowest backup intervals should be used.

An RPO of less than 1 hour, or even zero means that practically online data replication should be used, so that almost no loss of data will be experienced.

The backup strategies should also account for backup availability, so backups should not reside close to the original data (off-site backups), they should be directly copied to secondary sites in case of mission critical services, and the backup medium should make sure that the data will be available within the RTO period. For example, a Tape Backup that is send to an off-site 500 KMs away without a tape reader, or with very slow connectivity is not a valid solution for a 1 hour RTO.

Cross-Region Secondary Sites

In case of mission critical applications, where a strict DR plan is needed, having secondary sites is of paramount importance. This is extremely easy to accomplice on the cloud, by using the already existing Infrastructure as Code that was used for the deployment of the primary, to create a secondary, and setting up cross-region replication of data.

We can identify the following scenarios:

1. Cold Standbys: These are replicas of the primary sites that are basically powered off, and only data are being replicated. These have the lowest cost, while being able to become fully operational, either manually or automatically, within several minutes.

2. Hot Standbys: These are replicas of the primary sites, that are half powered off, meaning that they do not operate at their full capacity. They are more expensive than the cold standbys, while keeping much of the functionality available in matter of seconds and reaching full capacity in a matter of minutes.

3. Active secondary sites: These are fully functional replicas of the primary sites. They are always-on with the same capacity as the primary, and are available within seconds, usually transparently to the end user. This is the most expensive solution, effectively doubling the cost of a single primary site with the addition of the data replication costs, but it provides near real-time failover.
   These sites allow for Active-Active load balancing configuration in case the applications support it.

Organizational Design

AWS Organizations service allows for a company to implement an organization tree structure in the cloud, which helps with central management, governance, and cost management. It provides a way to quickly scale workloads, while keeping them separated and secure, making possible to centrally audit all child accounts, delegate permissions and access controls, take advantage of quantity discounts, while still having a central yet granular cost management.

The structure is very similar to the Active Directory Forest, with a Master Account at the root, children accounts or Organizational units at the stems. An organizational unit (OU) can hold one or more child accounts. Access rights, permissions and policies can be applied on an Account or OU level.

With that in mind we will target to create a structure where the enterprise units have corresponding OUs in the cloud, tracking their own budget and actual consumption, and allowing for the staff that operated their services on-premises to continue operate them on the cloud without ever having access to other infrastructure, thus also adhering to the “least privilege” rule.

A diagram depicting the basic Organizations structure can be viewed in the following figure, also showing shared services, and environments which will be analyzed in the next section.

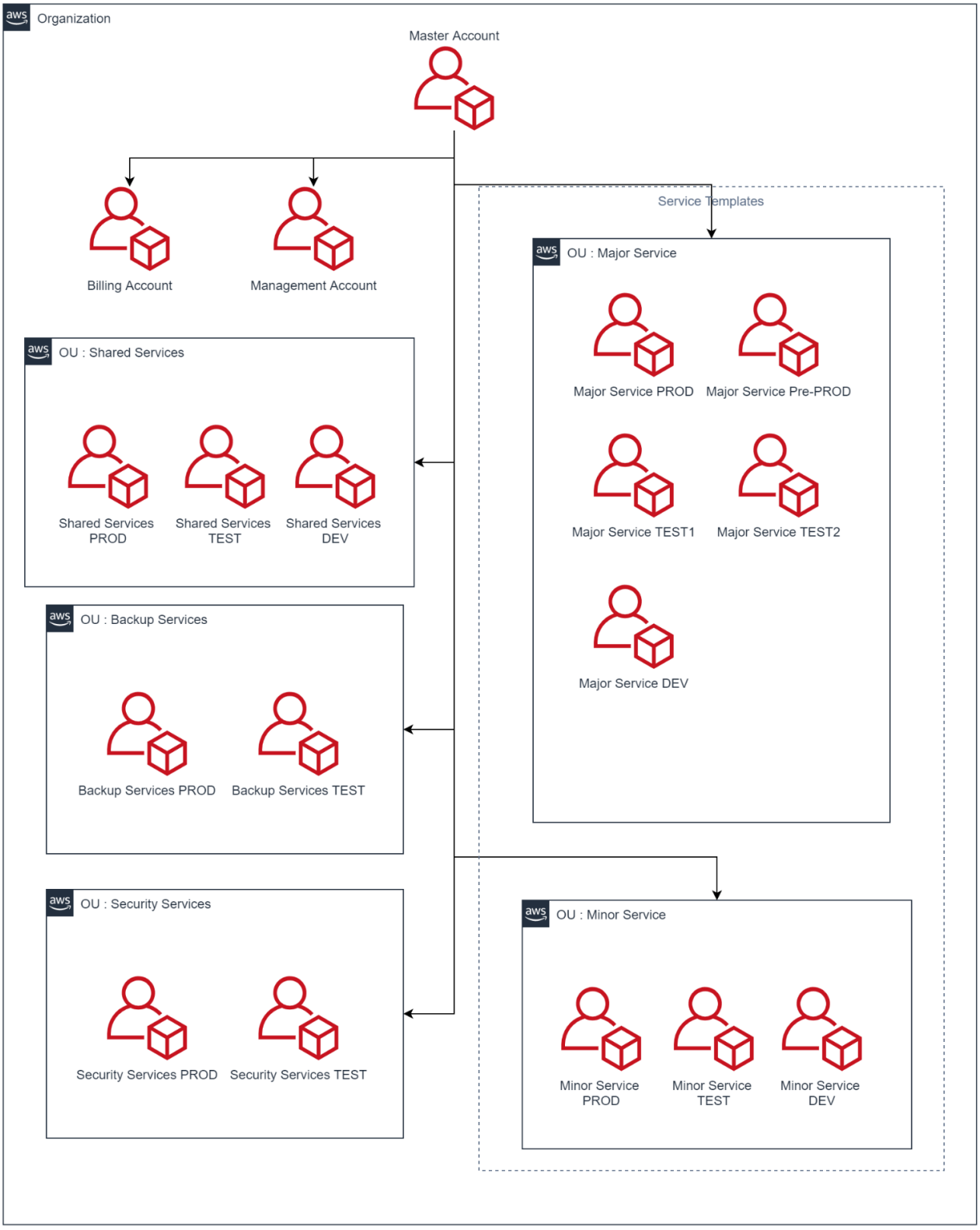


Figure 3: Organization Diagram

IT Services Design

IT services can be categorized in two axes:

1. Depending on the size of the service to Major and Minor
2. Depending on the business impact to Mission Critical, Standard and Basic

Major Services typically have large infrastructure requirements and multiple test and development environments while minor services typically have simpler infrastructure requirements and require one of each development, test and production environment.

Mission Critical services are services whose unavailability will have a significant impact to the company and will brake business continuity. This would result in a large revenue loss and any risks need to be immediately identified and remediated. These services require geographical redundancy and RPO/RTO times measured in minutes. Examples of such services are CRM, ERP and CTI (Contact Center Services).

For these services we will consider using regional redundancy and design the solution in such a way, that the services will be remain operational, with minimal, if any, unavailability in case of a regional disaster. Hot standbys will be placed in a second region, databases and file storages will be replicated to the second region, and global services will be used to automatically perform the failover in case of disaster.

Standard services are services whose unavailability will have a medium impact to the company and will not break business continuity. Any revenue loss resulting from such unavailability is not significant enough to consider the additional costs of geographical redundancy, and risks are identified, assessed, and accepted or remediated based on the medium business impact. RPO/RTO is measured in hours.

For these services we will consider using multiple availability zones only, in order to not be impacted by the provider's data center outages. Backups will be copied to a second region, and a disaster recovery process will be implemented where the entire production environment will be re-deployed in the second region and backups will be restored, in case of extended regional unavailability, according to required RTO. This is feasible using infrastructure as a code, and cloud native applications, and can even be automated if needed.

Basic services are services whose unavailability will have minimal impact to the company and no impact to the Business Continuity.

For these services, the designs are the same as with Standard services, but there is no need for high availability, thus they will reside on a single data center (Availability zone) of a single region and their backups will reside on a single region to. As the designs are the same, we will not provide additional designs for the basic services.

1. Cloud Native – Minor – Standard Service

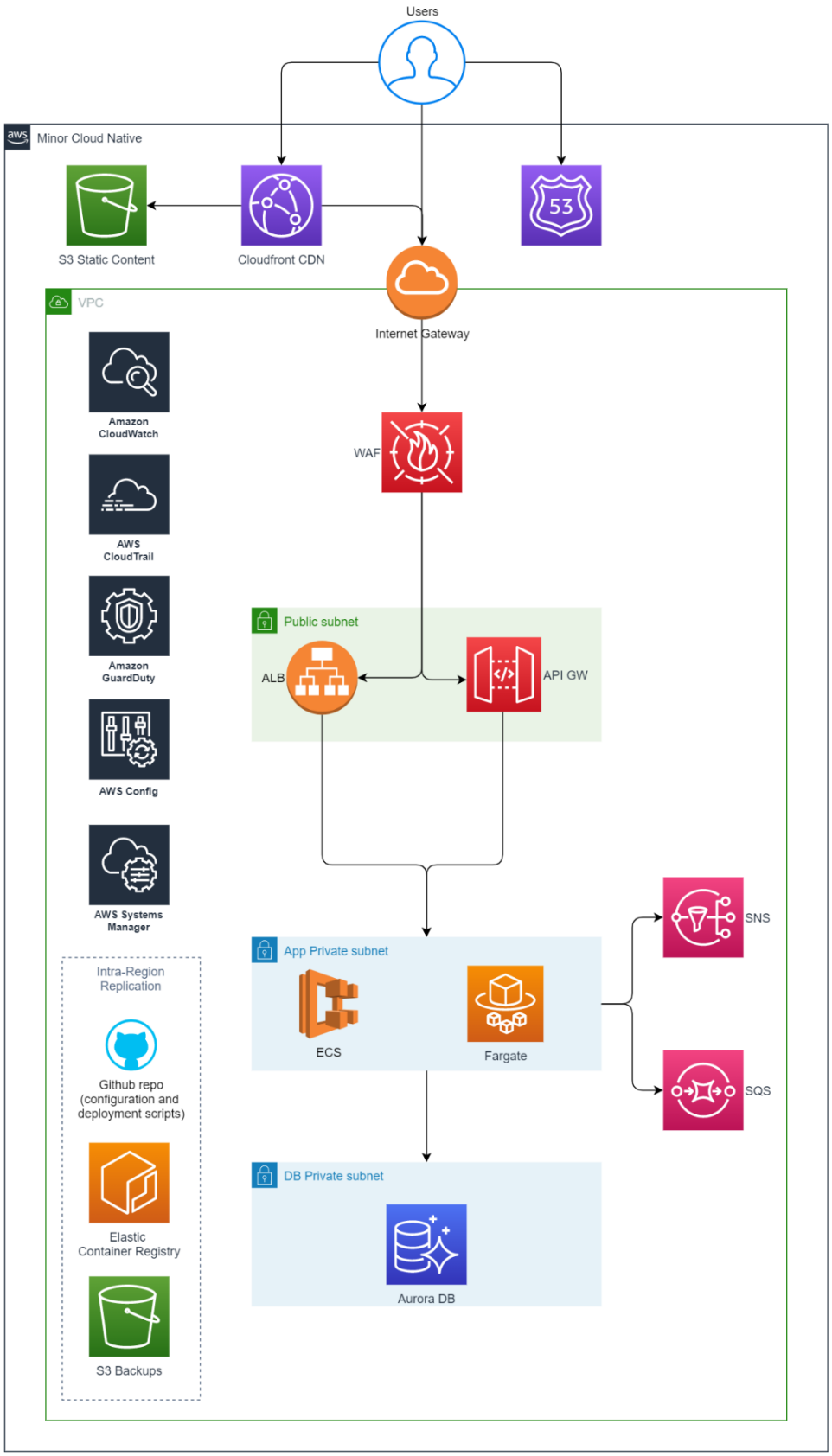


Figure 4: Cloud Native minor standard Service

In the above diagram we can see a typical architecture for a Cloud Native minor standard service, which is a service that does not have high infrastructure requirements and does not need cross-region level redundancy. We should point out that all services in the diagram are highly available within the region by spanning automatically across 3 Availability Zones.

The services in use are the following:

- **Route 53**: Global Domain Name System (DNS) services, with almost 100% availability, as they span across all AWS Regions, and also provide traffic management based on geo-location, filtering based on geo-location, and load balancing / failover across regions using health checks.

- **Cloudfront**: A global Content Delivery Network (CDN) with close integration to the AWS services and many edge locations (including Greece), with the additional capability of providing SSL termination, executing lambda functions for transforming requests and responses, and managing traffic based on rules.

- **S3**: Object Storage with high availability and reliability. Objects are stored and retrieved with HTTP/HTTPS requests or through AWS APIs. In our architecture S3 is used to store static objects (S3 Static Content) for delivery through Cloudfront, as well as in the back-end for storing backups, logs, etc.

- **Internet Gateway**: A serverless gateway service for connecting to and from the internet.

- **WAF**: A Web application firewall protecting from most common attacks like cross-site scripting and SQL injection. It can use managed rules that are provided by AWS or 3rd party providers that are automatically updated.

- **Elastic Load Balancer**: A managed load balancer service providing either Application Load Balancers (ALB) working on layer 7 or Network Load Balancers (NLB) working on layer 4. Both provide load balancing capabilities with health checks, SSL termination, with ALB providing additionally URL based routing, and TLS offloading.

- **API Gateway**: A managed service for publishing, maintaining, monitoring, and securing APIs. It offloads tasks like traffic management, Cross-origin resource sharing (CORS) support, authorization, access control, throttling, monitoring and API version management from the servers, containers, or functions behind it.

- **ECS**: Elastic Container Service is a managed container orchestration platform provided by AWS with close integration with the rest of the AWS services. Not Kubernetes compatible.

- **Fargate**: A compute engine for running containers on-demand in a serverless manner. Highly available and scalable, can be used by ECS directly to orchestrate small to medium workloads.

- **Aurora DB**: A scalable Relational Database based on AWS engines that are MySQL and PostgreSQL compatible, but provide higher performance and closer integration with AWS services.

- **Simple Notification Service** (SNS): A managed notification service for Application to Application (A2A) communication using pub/sub topology and Application to Person (A2P) communication using common channels like email, mobile push, and SMS. The pub/sub topology allows for "topics" with many "subscribers" that will all receive the message/notification "published" by one or more "publishers". It is a one or many to many broadcast service.

- **Simple Queue Service** (SQS): A managed queuing service for queuing and delivering messages with guaranteed delivery. Messages that are added to the queue can be retrieved (pulled) by other services, applications on VMs or containers, and Lambda functions, and will remain hidden for a predetermined period of time (while they are being processed), and either deleted by the retriever after processing or be restored in the queue if the timeout period is exceeded. This implies that if something went wrong with the processing, the messages will become available to be picked up by another instance. Such failed messages can then be added to failed queue for further inspection. This mechanism provides a decoupling layer between services, application components and microservices.

- **CloudWatch**: A managed monitoring and observability service, which collects monitoring and operational data in the form of logs, metrics, and events, and provides visualizations, alarms, and insights for the infrastructure and applications.

- **CloudTrail**: A managed service that provides event history of the AWS account activity, including actions taken through the AWS Management Console, AWS SDKs, command line tools, and other AWS services. It can forward logs to CloudWatch.

- **GuardDuty**: A managed threat detection service. It monitors continuously for malicious activity and unauthorized behavior, using AI and Machine learning on all logs to detect anomalies. It can also use Lambda to automatically respond to threats.

- **Config**: A managed service that allows to assess, audit, and evaluate the configuration of the resources in the account. It is rule based and uses these rules to continuously evaluate the configurations and relationships between resources, and detect any non-compliances or changes. This is the easiest method to detect unprotected S3 buckets or databases with public endpoints that so far have led to numerous security leaks.

- **Systems manager**: A managed operations hub which provides operational tasks centralization and automation by using runbooks. It includes an incident manager, an application manager with an AppConfig feature that automates application configuration deployment and management, a parameter store which can be used for storing application parameters, a patch manager for patching automation within maintenance windows and Session Manager that allows for safe and secure infrastructure access using the AWS console.

- **Elastic Container Registry**: A managed container registry for storing, managing, sharing, and deploying container images and artifacts. As a fully managed service (SaaS) it can be consumed without any management overhead and is fully scalable.

- **Github / AWS CodeCommit**: Github is a 3rd party source control service that is also provided as a service. It integrates well with the AWS ecosystem and is proposed due to developer's familiarity with the service. A great alternative is AWS's CodeCommit that is a fully managed service too, is based on GIT but of course provides the highest level of integration with AWS.

2. Cloud Native – Minor – Mission Critical Service

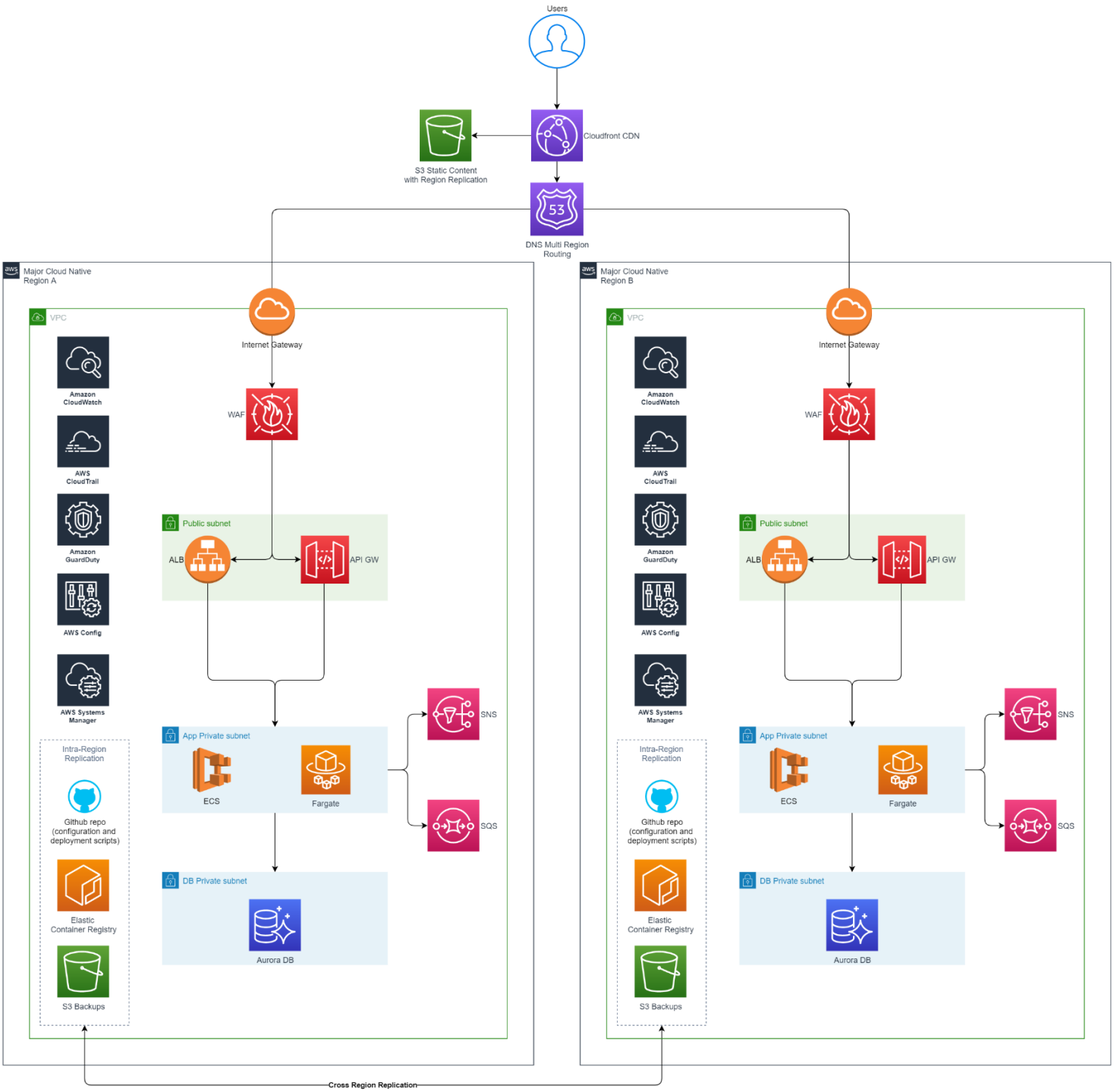


Figure 5: Cloud Native minor Mission Critical Service

In the above diagram we can see a typical architecture for a Cloud Native minor Mission Critical service, which is a service that does not have high infrastructure requirements but needs cross-region level redundancy. The high level of redundancy is achieved by spanning to two different regions which are physically apart by more than 100km and thus provide a high level of disaster isolation. The traffic is distributed across the regions by Route53 and can be dynamically allocated between the Regions based on the following policies:

- **Failover routing policy** – This is the simplest routing policy, Route53 performs health checks and routes traffic from primary to standby in case of a region failure.
- **Geolocation routing policy** – This routing policy distributes traffic based on the location of the user according to predefined rules set by the admin. Using this policy, you can select which geographic locations will be served by which AWS region.
- **Geoproximity routing policy** – This routing policy distributes traffic based on the location of the resources and can, optionally, shift traffic from resources in one region to resources in another.
- **Latency routing policy** – This routing policy distributes traffic based on the latency from the user to the location of the resources. Using this policy, we can route traffic to the Region that provides the best latency with less round-trip time.
- **Multivalue answer routing policy** – This routing policy distributes traffic randomly among up to eight healthy DNS records. This is the most basic form of load balancing (round-robin) but with the additional feature of health checking the destinations.
- **Weighted routing policy** – This routing policy distributes traffic among resources based on a weight we assign to each resource. Using this routing policy we can route traffic to different versions of an application for testing, or for migrations.

We should point out that all services in the diagram are highly available within each region by spanning automatically across three Availability Zones.

The services in use are the following:

- **Route 53**: Global DNS services, with almost 100% availability, as they span across all AWS Regions, and also provide traffic management based on geo-location, filtering based on geo-location, and load balancing / failover across regions using health checks. In this setup it will be used for routing traffic among the regions.

- **Cloudfront**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **S3**: Object Storage with high availability and reliability. Objects are stored and retrieved with HTTP/HTTPS requests or through AWS API. In our architecture S3 is used to store static objects (S3 Static Content) for delivery through Cloudfront, as well as in the back-end for storing backups, logs, etc. AWS S3 allows for cross-region replication, which will be used in order to keep the object-storage automatically up-to-date on both regions.

- **Internet Gateway**: Same as in "1. Cloud Native – Minor – Standard Service".

- **WAF**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Elastic Load Balancer**: Same as in "1. Cloud Native – Minor – Standard Service".

- **API Gateway**: Same as in "1. Cloud Native – Minor – Standard Service".

- **ECS**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Fargate**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Aurora DB**: A scalable Relational Database based on AWS engines that are MySQL and PostgreSQL compatible, but provider higher performance and closer integration with AWS services. In this case we will use Aurora Global Database which will span across the 2 regions, with one primary and one secondary. Only one region can handle both writes while both regions can handle reads.

- **Simple Notification Service** (SNS): Same as in "1. Cloud Native – Minor – Standard Service".

- **Simple Queue Service** (SQS): Same as in "1. Cloud Native – Minor – Standard Service".

- **CloudWatch**: Same as in "1. Cloud Native – Minor – Standard Service".

- **CloudTrail**: Same as in "1. Cloud Native – Minor – Standard Service".

- **GuardDuty**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Config**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Systems manager**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Elastic Container Registry**: A managed container registry for storing, managing, sharing, and deploying container images and artifacts. As a fully managed service (SaaS) it can be consumed without any management overhead and is fully scalable. ECR can automatically replicate the registry to multiple regions, a feature that will be used in this case.

- **Github / AWS CodeCommit**: Github is a 3rd party source control service that is also provided as a service. It integrates well with the AWS ecosystem and is proposed due to developer's familiarity with the service. A great alternative is AWS's CodeCommit that is a fully managed service too, is based on GIT but of course provides the highest level of integration with AWS. Both services are available across multiple regions, but CodeCommit needs manual (scripted) repository replication.

3. Cloud Native – Major – Standard Service

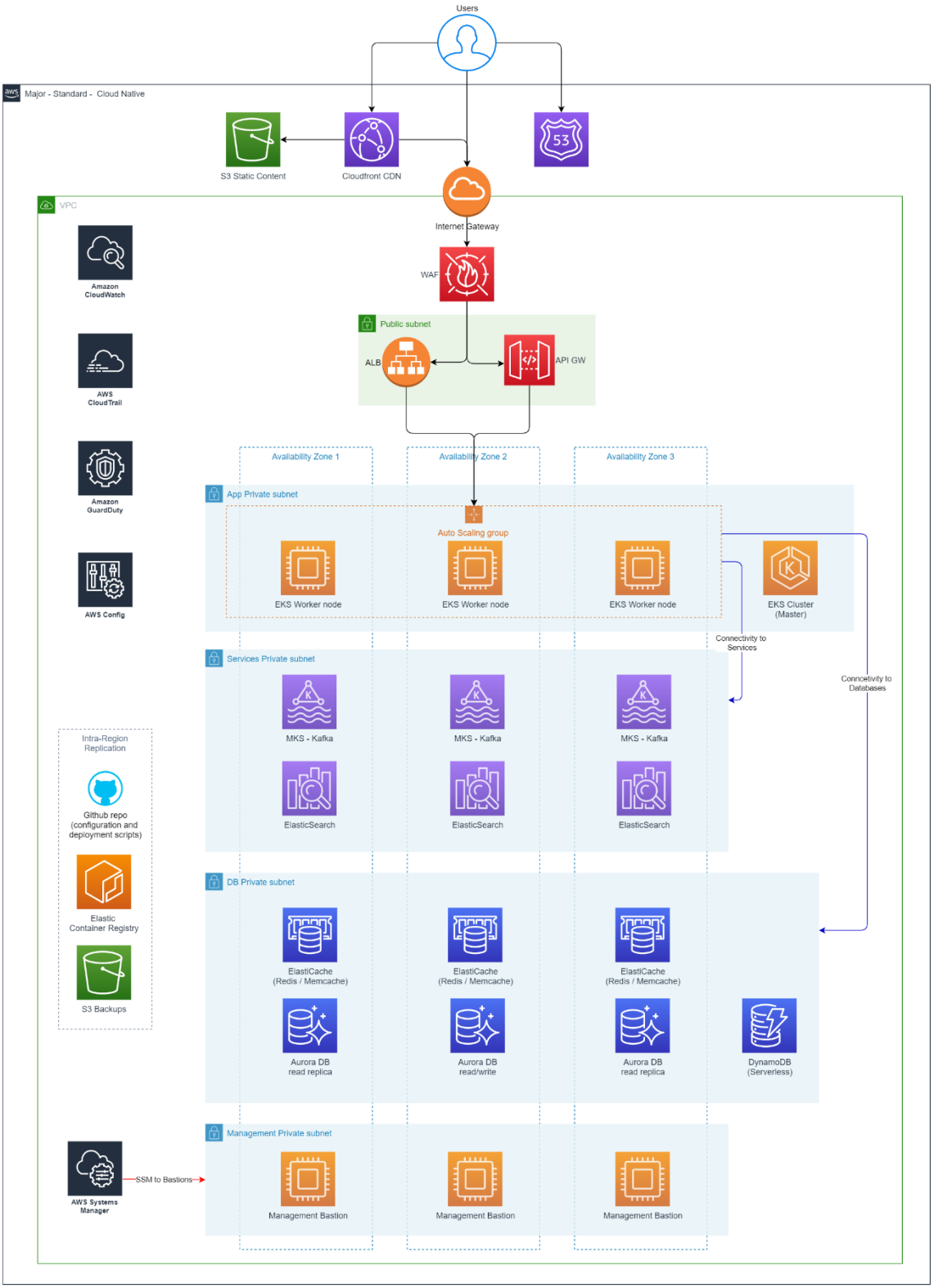


Figure 6:Cloud Native – Major – Standard Service

In the above diagram we can see a typical architecture for a Cloud Native major standard service, which is a service that does have high infrastructure requirements but does not need cross-region level redundancy. We have split the services in the three Availability Zones and placed resources in each of them to provide high availability. The high performance/capacity needs of the major applications are more fitted to be hosted on non-serverless infrastructure, which, while increase somewhat the operational overhead, provide provisioned capacity at great scales.

The applications are still hosted on containers but this time the orchestrator is the industry standard Kubernetes, in an AWS managed flavor, the Elastic Kubernetes Service (EKS). Messaging is handled by a managed Kafka service and search and indexing needs are covered by a managed Elasticsearch service.

Database needs are available in three flavors, ElasticCache for caching, session storage, and any other temporary DB storage needs, Aurora DB for relational DB needs, and DynamoDB for NoSQL needs.

The services in use are the following:

- **Route 53**: Same as “in 1. Cloud Native – Minor – Standard Service”.
- **Cloudfront**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **S3**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **Internet Gateway**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **WAF**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **Elastic Load Balancer**: A managed load balancer service providing either Application Load Balancers (ALB) working on layer 7 or Network Load Balancers (NLB) working on layer 4. Both provide load balancing capabilities with health checks, SSL termination, with ALB providing additionally URL based routing, and TLS offloading. In this case ALBs and NLBs can be created automatically by EKS when creating an “Ingress” or “Load Balance” service.
- **API Gateway**: Same as in “1. Cloud Native – Minor – Standard Service”
- **Elastic Kubernetes Service (EKS)**: A managed container orchestration platform provided by AWS, based on upstream Kubernetes with close integration with the rest of the AWS services. Most of the tools designed for Kubernetes will work on EKS, with managed availability and scalability for the control plane nodes and the pods can run on either EC2 (VMs – as in our case) or Fargate.

- **Elastic Cloud Compute (EC2)**: A compute engine for running VMs on the cloud. There are many types and sizes available, compute, memory or network optimized, with both x86 and ARM flavors. In our case, they are used to host the worker nodes of the EKS cluster, but also internally for the MKS, ElasticSearch and ElastiCache nodes.

- **Aurora DB**: Same as in "1. Cloud Native – Minor – Standard Service".

- **DynamoDB**: A managed key-value and document database, multi-region, multi-active with high durability and built-in security, backup and restore as well as in-memory caching. It is a serverless service and can auto-scale out to "virtually unlimited throughput and storage".

- **ElastiCache**: A managed in-memory data store, compatible with either Redis or MemCached. These data stores can be used to store frequently accessed data such as session data and tracking data. As the entire database is stored in-memory it can provide sub-millisecond response times, but with much lower durability, as the data are not stored in a permanent storage, by default.

- **ElasticSearch**: A managed ElasticSearch service. Elasticsearch is a distributed, search and analytics engine for all types of data, including textual, numerical, geospatial, structured, and unstructured. It can be used to index and provide search capabilities to a wide array of applications, from website applications to Business Analytics. AWS ElasticSearch, based on forked open-source ElasticSearch code is compatible with other often used Elastic products like Kibana and Logstash.

- **Managed Streaming for Apache Kafka (MKS)**: A managed Kafka service. Apache Kafka is an open-source platform for building real-time streaming data pipelines and applications and is used to publish (write) and subscribe to (read) streams of events, including continuous import/export of data from other systems, to store streams of events durably and reliably for as long as needed and to process streams of events as they occur or retrospectively. MSK automatically provisions and runs the clusters, monitors cluster health, and automatically replaces unhealthy nodes with no downtime and secures the cluster by encrypting data at rest.

- **CloudWatch**: Same as in "1. Cloud Native – Minor – Standard Service".

- **CloudTrail**: Same as in "1. Cloud Native – Minor – Standard Service".

- **GuardDuty**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Config**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **Systems manager**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **Elastic Container Registry**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **Github / AWS CodeCommit**: Same as in “1. Cloud Native – Minor – Standard Service”.

4. Cloud Native – Major – Mission Critical Service

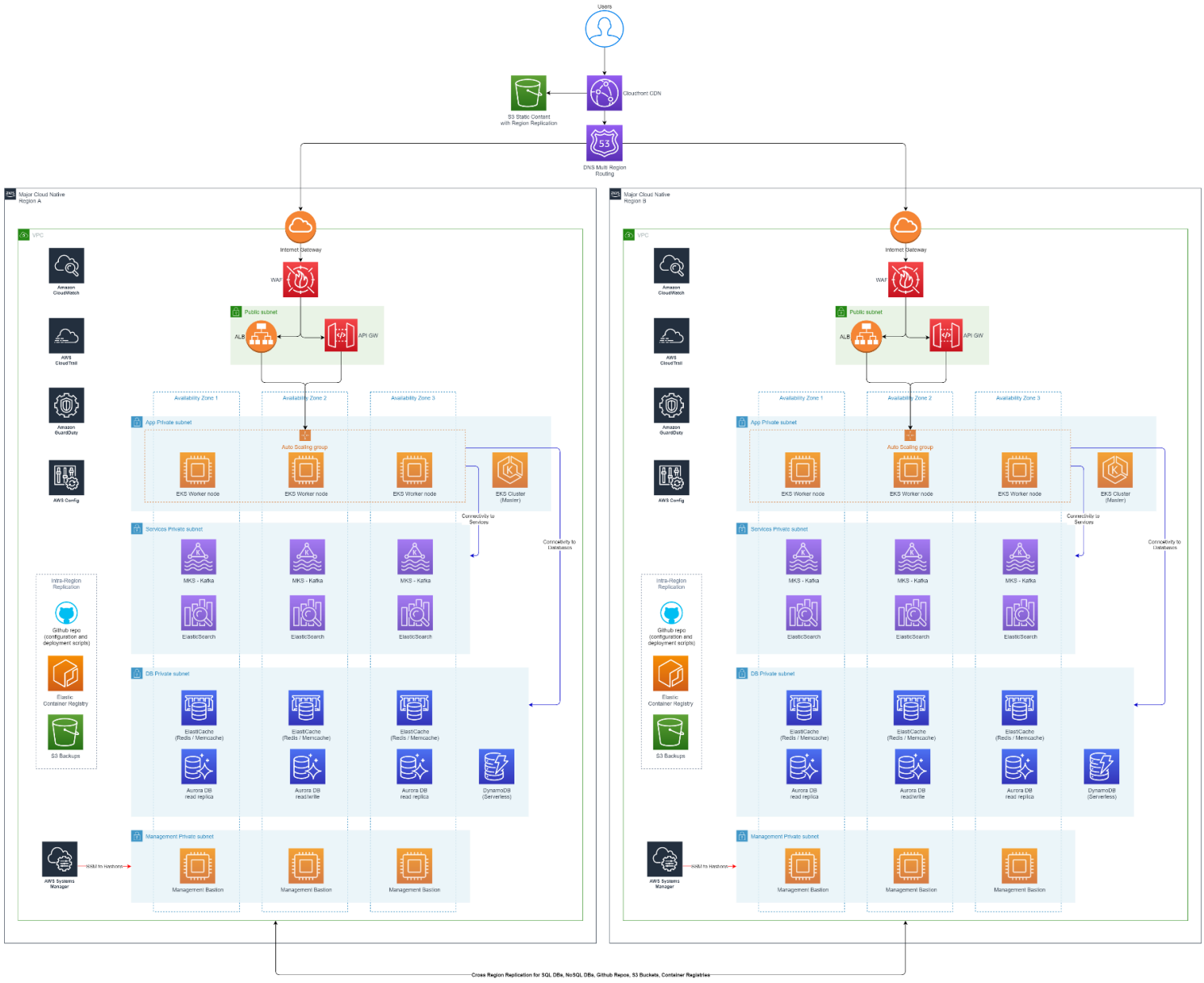


Figure 7:Cloud Native – Major – Mission Critical Service

In the above diagram we can see a typical architecture for a Cloud Native major Mission Critical service, which is a service that does have high infrastructure requirements and needs cross-region level redundancy. As in “3.” we have split the services in the three Availability Zones, and placed resources in each of them to provide high availability. As in “3.” the applications are still hosted on containers orchestrated by EKS. Messaging is handled by a managed Kafka service and search and indexing needs are covered by a managed ElasticSearch search.

Database needs are available in three flavors, ElasticCache for caching, session storage, and any other temporary DB storage needs, Aurora DB for relational DB needs, and DynamoDB for NoSQL needs.

As in “2. Cloud Native – Minor – Mission Critical Service”, the traffic is distributed across the regions by Route53 and can be dynamically allocated between the Regions based on the previously mentioned policies.

The services in use are the following:

- **Route 53**: Same as in “2. Cloud Native – Minor – Mission Critical Service”.
- **Cloudfront**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **S3**: Same as in “2. Cloud Native – Minor – Mission Critical Service”.
- **Internet Gateway**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **WAF**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **Elastic Load Balancer**: Same as in “3. Cloud Native – Major – Standard Service”.
- **API Gateway**: Same as in “1. Cloud Native – Minor – Standard Service”.
- **Elastic Kubernetes Service (EKS)**: Same as in “3. Cloud Native – Major – Standard Service”.
- **Elastic Cloud Compute (EC2)**: Same as in “3. Cloud Native – Major – Standard Service”.
- **Aurora DB**: Same as in “2. Cloud Native – Minor – Mission Critical Service”.
- **DynamoDB**: Same as in “3. Cloud Native – Major – Standard Service”.
- **ElastiCache**: Same as in “3. Cloud Native – Major – Standard Service”.
- **ElasticSearch**: Same as in “3. Cloud Native – Major – Standard Service”.

- **Managed Streaming for Apache Kafka (MKS)**: Same as in “3. Cloud Native – Major – Standard Service”.

- **CloudWatch**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **CloudTrail**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **GuardDuty**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Config**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Systems manager**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Elastic Container Registry**: Same as in “2. Cloud Native – Minor – Mission Critical Service”.

- **Github / AWS CodeCommit**: Same as in “2. Cloud Native – Minor – Mission Critical Service”.

5. Legacy – Minor/Major – Standard Service

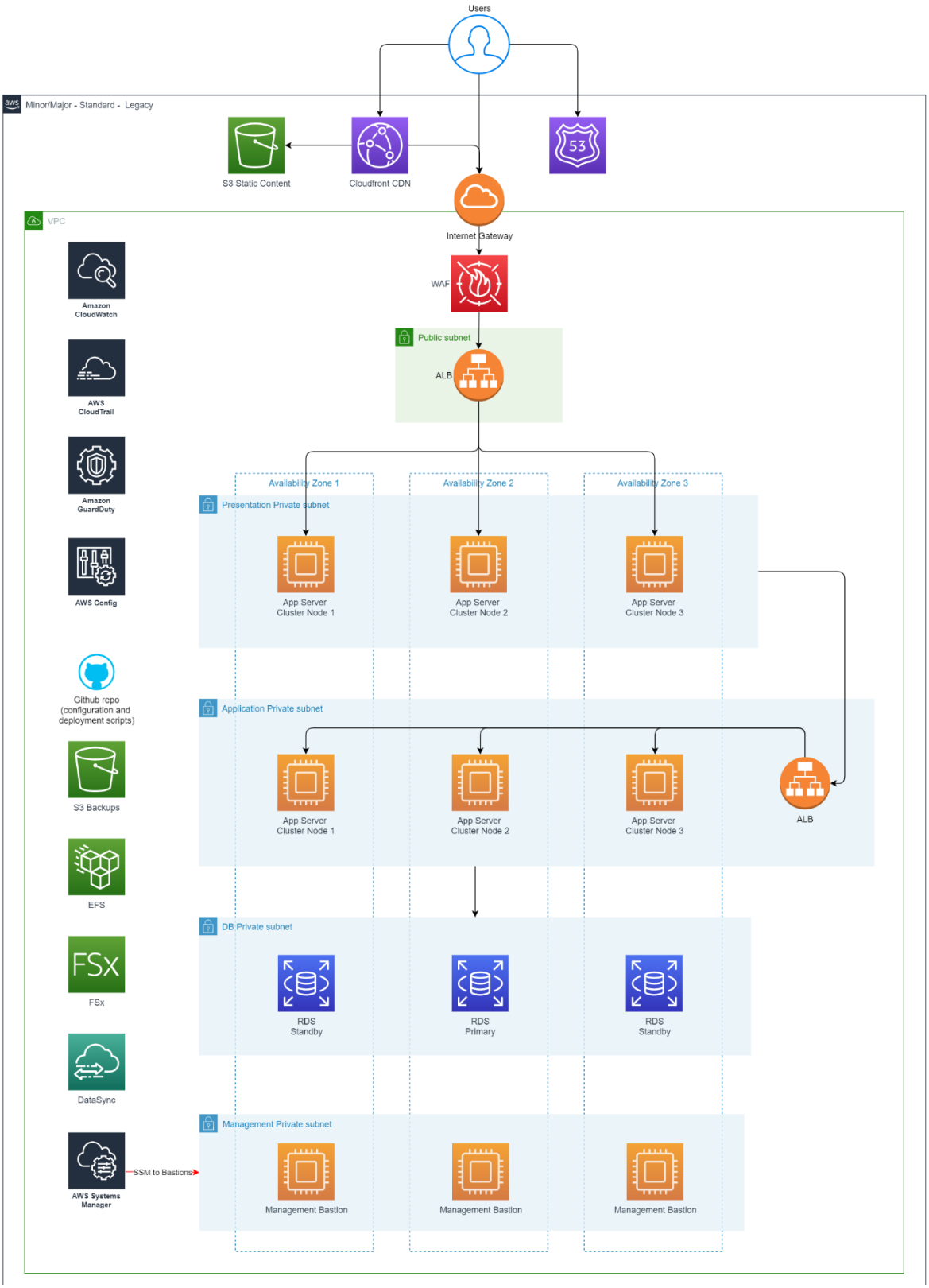


Figure 8: Legacy – Minor/Major – Standard Service

In the above diagram we can see a typical architecture for a Legacy standard service, which is a service that does have components that are not Cloud Native, which means that neither containers, nor serverless functions can be used. In this case a more classic VM architecture is in use, while still taking advantage of cloud services like managed storage, managed databases and managed Load Balancers. As the application only supports VMs, there is no distinction between major or minor services, the difference only being in the size of the VMs, that is not depicted in the diagram. So, the same architecture is applied in both cases.

In most cases of legacy applications any kind of elasticity cannot be used, so auto-scaling groups are missing from the diagram. Unless the application does not even support multiple instances, the VMs can span across availability zones for high availability and configure the ALB to produce alarm events in case any one instance becomes unhealthy. In case the application can only function in an "active-passive" manner, then an additional server is required to act as a "watchdog", that will monitor the health of the application and change the active instance using scripts, in the event of a failure.

As a last resort, a cluster environment can be replicated on the cloud, using existing clustering technologies, like Microsoft Windows Server Failover Cluster, Veritas Cluster for Unix/Linux, or "keepalived" software, which will not take advantage of cloud technologies and incur additional costs.

Database needs should be covered by Amazon RDS managed databases, which support MS SQL, Oracle, MySQL, PostgreSQL and MariaDB engines, and should cover most of the use cases. A popular exception is IBM DB2, that will need to be installed on EC2 instances, and be entirely self-managed.

The services in use are the following:

- **Route 53**: Same as in "1. Cloud Native – Minor – Standard Service."
- **Cloudfront**: Same as in "1. Cloud Native – Minor – Standard Service".
- **S3**: Same as in "1. Cloud Native – Minor – Standard Service".
- **Internet Gateway**: Same as in "1. Cloud Native – Minor – Standard Service".
- **WAF**: Same as in "1. Cloud Native – Minor – Standard Service".
- **Elastic Load Balancer**: Depending on the case, the Load Balancer in this scenario will perform Load balancing among the instances or direct the traffic to a single instance with manual failover using a third health-check server but will never use its autoscaling features. It can be either an ALB when Layer 7 LB is supported, and SSL/HTTPS termination can be performed on the ALB, or Network Load Balancer (NLB) if the packets need to arrive unaltered to the application.

- **Elastic Cloud Compute (EC2)**: A compute engine for running VMs on the cloud. There are many types and sizes available, compute, memory or network optimized, with both x86 and ARM flavors. In our case, they are used to host the applications and depending on the case, an independent health-check server or even the custom databases. EC2 instances are supported by Elastic Block Storage (EBS) volumes which is a Storage Attached Network (SAN) block storage system and Elastic File Storage Shares (EFS) which is a NFS type Network Attached Storage (NAS) for unix/linux servers or Amazon FSx for Windows File Server which is a SMB type NAS for Windows servers.

- **EBS**: A fully managed, easy to use, high-performance, block-storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads at any scale. EBS volumes are replicated within a single availability zone, and scale up to Petabytes. EBS snapshots with automated lifecycle policies are available to use for backups and can be replicated across regions, or even used to create new instances, identical to the original.

- **EFS**: A fully managed, simple, serverless, set-and-forget, NFSv4 compatible file system that allows us to share file data without provisioning or managing storage. EFS provides automatically scalable storage and performance. In the base model the performance is dependent on the storage (the performance scales with the storage space that is being used), but performance can be also provisioned on a monthly basis if not enough storage is in use.

- **Amazon FSx for Windows File Server**: A fully managed, highly reliable, and scalable file storage that is accessible over the industry-standard Server Message Block (SMB) protocol which is the default file sharing protocol for Windows. It is built on Windows Server, delivering a wide range of administrative features such as user quotas, end-user file restore, and Microsoft Active Directory (AD) integration. It offers single-AZ and multi-AZ deployment options, fully managed backups, and encryption of data at rest and in transit. It can also be used in Unix/Linux but NFS has a slight performance advantage and native integration.

- **Amazon RDS**: A fully managed Relational Database Service that provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching and backups. Standby Replicas or Read Replicas can be used across Availability Zones.

- **CloudWatch**: A managed monitoring and observability service, which collects monitoring and operational data in the form of logs, metrics and events, and provides visualizations, alarms, and insights for the infrastructure and applications. In this case, Cloudwatch agent will need to be installed on the VMs that host the applications, in order to gather logs and metrics.

- **CloudTrail**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **GuardDuty**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Config**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Systems manager**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Github / AWS CodeCommit**: Github is a 3rd party source control service that is also provided as a service. It integrates well with the AWS ecosystem and is proposed due to developer’s familiarity with the service. A great alternative is AWS’s CodeCommit that is a fully managed service too, is based on GIT but of course provides the highest level of integration with AWS. Both services are available across multiple regions, but CodeCommit needs manual (scripted) repository replication. Although the Applications are not Cloud Native, it is still a good practice to keep the configuration files in a repository and manually deploy changes, or manually integrate the application to a CI/CD pipeline to automate the process as much as possible.

6. Legacy – Minor/Major – Mission Critical Service

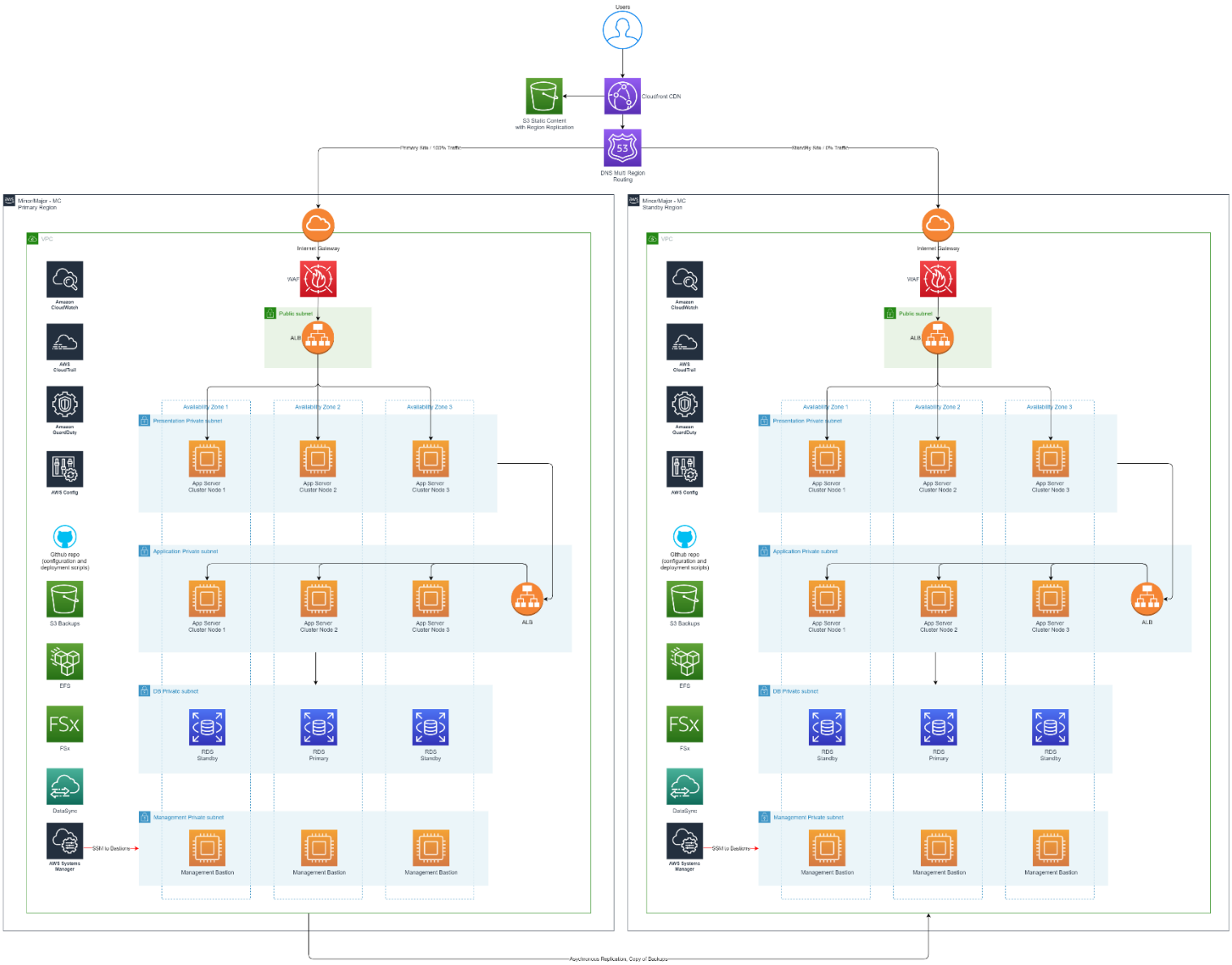


Figure 9: Legacy – Minor/Major – Mission Critical Service

In the above diagram we can see a typical architecture for a Legacy Mission Critical service, which is a service that does have components that are not Cloud Native, which means that neither containers, nor serverless functions can be used, but still is critical for business continuity and must suffer minimal downtime. In this case the architecture is identical to “5. Legacy – Minor/Major – Standard Service”, but it spans across two regions.

In contrast to the Cloud Native Mission Critical solutions, Route 53 can only be used with either Failover routing policy (this is the simplest routing policy, Route53 performs health checks and routes traffic from primary to standby in case of region failure) or Weighted routing policy (this routing policy distributes traffic among resources based on a weight we assign to each resource and can be used to perform Green/Blue type application upgrades).

Database needs should be covered by Amazon RDS managed databases, which support MS SQL, Oracle, MySQL, PostgreSQL and MariaDB engines, and should cover most of the use cases, and read replicas will be creates across regions. A popular exception is IBM DB2, that will need to be installed on EC2 instances, and be entirely self-managed including cross region replication. If replication is not supported, then a semi-manual backup/copy/restore procedure must be implemented or a third-party tool must be used, with additional cost.

The services in use are the following:

- **Route 53**: Global DNS services, with almost 100% availability, as they span across all AWS Regions, and also provide traffic management. In this setup it will be used for routing traffic based on Failover or Weighted Routing policy .

- **Cloudfront**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **S3**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Internet Gateway**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **WAF**: Same as in “1. Cloud Native – Minor – Standard Service”.

- **Elastic Load Balancer**: Same as in “5. Legacy – Minor/Major – Standard Service”.

- **Elastic Cloud Compute (EC2)**: Same as in “5. Legacy – Minor/Major – Standard Service”.

- **EBS**: A fully managed, easy to use, high-performance, block-storage service designed for use with Amazon Elastic Compute Cloud (EC2) for both throughput and transaction intensive workloads at any scale. EBS volume are replicated within a single availability zone and scale up to Petabytes. EBS snapshots with automated lifecycle policies are available to use for backups and can be replicated across regions, or even used to create new instances, identical

to the original. In this case the EBS snapshots will need to be replicated to the other region using a replicated S3 bucket, as snapshots are automatically saved in S3.

- **EFS**: A fully managed, simple, serverless, set-and-forget, NFSv4 compatible file system that allows us to share file data without provisioning or managing storage. Cross region replication can be achieved using AWS Datasync, another AWS service that allows transferring across regions or accounts.

- **Amazon FSx for Windows File Server**: A fully managed, highly reliable, and scalable file storage that is accessible over the industry-standard Server Message Block (SMB) protocol which is the default file sharing protocol for Windows. Just like EFS, cross region replication can be achieved using AWS Datasync.

- **Amazon RDS**: A fully managed Relational Database Service that provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching and backups. Standby Replicas or Read Replicas can be used across Availability Zones, and in our case across Regions.

- **CloudWatch**: Same as in "5. Legacy – Minor/Major – Standard Service".

- **CloudTrail**: Same as in "1. Cloud Native – Minor – Standard Service".

- **GuardDuty**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Config**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Systems manager**: Same as in "1. Cloud Native – Minor – Standard Service".

- **Github / AWS CodeCommit**: Same as in "5. Legacy – Minor/Major – Standard Service".

7. SaaS type Services

There are services that are already provided as SaaS, where the company is not involved in the hosting or the operations at all. These services are mostly Office IT services, with the most common being the email and office software (Word Processor, Spreadsheet, Presentation, etc.). The most common providers are Microsoft with Microsoft 365 and Google with Google Workspace taking almost 100% of the market. (Forbes, 2020)

These suites provide everything that is needed to perform common day to day tasks and collaborate with colleagues, share, and manage the daily workload. They are hosted on the provider's infrastructure and the customer is only responsible for managing access and installing any software that might be needed on end user devices – if any. Google's solution is entirely web based, while Microsoft's requires the installation of the standalone software suite to take advantage of the full functionality.

Many software vendors have already started to provide their solutions as SaaS, so there is the possibility to consume CRM, BRM, ERP, accounting, asset management and other types of software as a service.

These services need minimal effort, as only the authentication / authorization integrations need to be migrated, while they will continue to be provided as before.

Various SaaS applications enable end users to manage their SSO configurations independently of the vendor. To avoid losing access in these cases, it is recommended creating a test instance and include a method for authenticating as an administrator natively within the application. The test instance allows to validate the migration without affecting any existing users. This will allow that the SaaS service will work using the Identity services that have been migrated to the cloud.

In this category we can also add Corporate end-user services, like Desktop-as-a-Service, or office IT self-service applications.

Desktop-as-a-Service, or DaaS, is a cloud-based service that securely delivers virtual apps and desktops to any device or location. Secure SaaS and legacy applications, as well as full Windows-based virtual desktops, are provisioned and delivered to our workforce via this desktop virtualization solution. DaaS provides a simple and predictable pay-as-you-go subscription model that enables on-demand scaling up or down. This turnkey service is simple to manage, as it eliminates many of the IT administration tasks associated with desktop solutions. (Citrix, 2021)

The advantages of DaaS are as follow:

1. **Improved accessibility:** The service's anywhere, anytime, anyhow nature enables users to remotely access their desktops via PC, laptop, tablet, or even smartphone, providing them with unprecedented freedom and flexibility. They can work from anywhere as long as they have a client device and a reliable internet connection.

2. **Capital expenditures are reduced:** With its traditional subscription model, Desktop as a Service (DaaS) breaks the cycle of investment in desktop hardware, servers, and licensing, freeing up funds previously spent on depreciating assets for higher-value initiatives. There is no longer a need for expensive, enterprise level end-user hardware, as almost any off-the-self notebook or PC can fulfill the role of the DaaS client. Users can also use their own devices, if the company provided one is not to their liking.

3. **Lower operating costs:** DaaS alleviates much of the heavy housekeeping burden, allowing for resource consolidation or redeployment, while also reducing space, power, and cooling requirements. Support contracts for enterprise level end-user devices can also be eliminated.

4. **Greater agility and responsiveness:** The dynamic nature of DaaS, with its rapid provisioning and inherent scalability, combined with its low-cost Opex model, makes it an excellent fit for organizations that need to expand quickly or respond to opportunity, whether it is migrating hosted desktops online or pushing new applications across a hosted desktop estate.

5. **Enhanced security:** DaaS shifts the security burden away from individual devices and places it within a secure data center infrastructure. Data is no longer exposed on a local device but is instead stored – and regularly backed up – in a secure hosted environment. It is also encrypted and accessible only via multi-factor authentication protocols, adhering to the strict company security guidelines regardless of the access method.

6. **Closer alignment with business requirements:** Businesses frequently scale up to meet peak resource requirements but then find themselves over-provisioned, which is inherently wasteful. DaaS can scale and adapt to changing needs, ensuring that we only pay for what we use. Role-based desktops can provide an additional level of alignment and cost efficiency.

7. **Increased resiliency and dependability:** Users can rely on consistent access and performance when desktops are delivered by solution providers that typically guarantee 99.99 percent uptime, based on a compelling mix of high-grade infrastructure, security, and support.

8. **Enhancement of business continuity:** DaaS is an attractive alternative to the traditional disaster recovery option of a stand-by secondary site, due to the ubiquitous availability of a desktop and centralized data backup. Additionally, it addresses the growing issue of weather/travel-related disruptions to Business as Usual, by allowing workers to work remotely or from home.

9. **Increased financial predictability:** Monthly fixed-fee pricing provides cost certainty and simplifies budgeting and forecasting. This can have a beneficial effect on cash flow and assist

with strategic planning. Different user groups may have different sizing needs, but the cost will remain predictable for each group.

10. **Increased consistency and readiness for the future:** Staff across an organization benefit from a standardized core desktop build (while still allowing for customization), the same versions of applications, regular refresh, and ongoing back-end investment, all of which contribute to a unified and uniform environment and optimal user experience.

11. **Productivity enhancements:** Increased uptime, improved performance, and increased collaboration opportunities can also increase productivity. The ability to work from anywhere, on hardware that can be customized to meet the individuals' needs, makes work even more productive.

12. **A more environmentally friendly profile:** Reduced demand for new equipment, reduced energy consumption, extended PC life, and more flexible working, all contribute to a company's carbon footprint reduction and compliance with environmental targets.

Services that can be offered in a SaaS model and can help reduce the end-user management overhead by transferring every-day tasks to the end-user include:

- **Password Reset:** Resetting a forgotten or expired password is one of the most requested service desk activities and it can be easily automated. By using multiple authentication methods – phone, SMS, email, authenticator app – most identity providers have developed solutions for secure password reset by the end users themselves.

- **Software ordering and installation:** By using predefined packages and an automated approval and delivery process, the users can "order" the software they need inside an internal software "shop" and the software can be installed automatically by the same mechanism that delivers update packages.

- **Access rights management:** End-Users can request access to services or assets using an automated access management platform, that will request management approval and then automatically add the user to the corresponding Identity Provider (for example Active Directory) security groups. Additionally, the platform can also request for the implementation of network connectivity (firewall rules), if required, by either raising tickets for manual implementation or with direct integration to a SDN (Software defined network). In case of DaaS, this process can be fully automated, by also automatically adding the user to the corresponding network security groups.

- **Remote, live-classroom or on-demand, Education:** Employees can choose their own pace, but still get high quality education from the safety of their homes. Live seminars can be easily recorded for later playback/review, questions can be asked electronically without, if not needed, interrupting the presenter, and viewers can return to their daily tasks within minutes.

Services that will remain on-premises

As was previously mentioned, some services will remain on-premises. These include:

- Services that need to be very close to the physical fixed network, such as DSLAM management, performance measure and monitor services, IPTV services that will incur substantial costs if they were provided over the public network, etc.

- Services that need to be close to the mobile network, such as tower equipment management software.

- Services that support the physical installations, such as Access management, Generator / UPS management, Alarms and CCTV recording and management.

- Services that are needed as an extension of the Cloud Migrated Services, such as storage connectors for backup, AD Connectors for Active Directory, security connectors for security scanners, malware scanners etc.

These services will be operated in a hybrid-cloud environment and take advantage of a close integration with the Cloud Services.

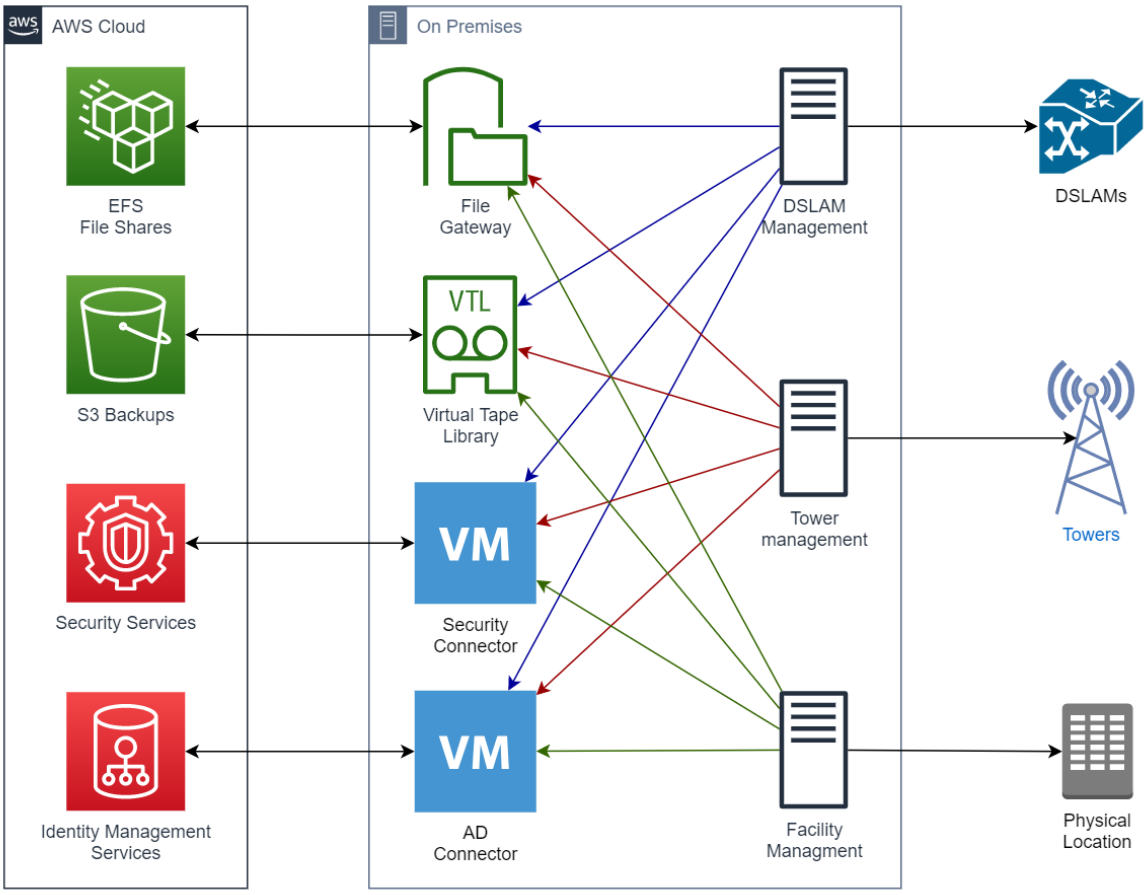


Figure 10: Hybrid Architecture design

Cost Analysis and Comparison

In order to be able to compare the cost of infrastructure housed in a self-managed data center to the cost of consuming infrastructure on a public cloud provider data center, we will be analyzing the costs of a similar set of assets on-premises and on the cloud.

For this exercise we assume that the on-premises infrastructure is based on the latest Hyper-converged architecture. Hyper-converged infrastructure (HCI) is a software-defined IT infrastructure that virtualizes all the elements of conventional "hardware-defined" systems. HCI includes, at a minimum, virtualized computing (a hypervisor), software-defined storage, and virtualized networking (software-defined networking). HCI typically runs on commercial off-the-shelf (COTS) servers. (Wikipedia, 2021)

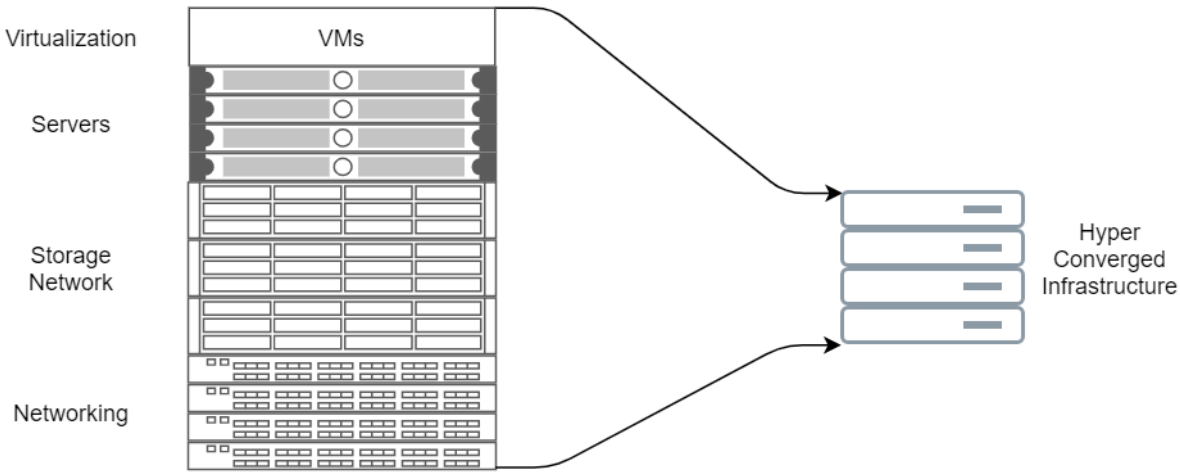


Figure 11: Hyper Convergence

Hyperconvergence is a transition from discrete, hardware-defined systems that are connected and bundled together to a purely software-defined environment in which all functional aspects run on commercial, off-the-shelf (COTS) servers, with hypervisor-assisted convergence. Server systems with Direct-Attached Storage (DAS) are typically used in HCI environments, as opposed to traditional Storage Attached Networks (SAN) and Network Attached Storages (NAS). HCI offers the ability to directly integrate into a pool of similar systems in a data center. For both hardware and software layers, all physical data-center resources are managed through a single administration platform. Traditional data-center inefficiencies are eliminated, and the total cost of ownership (TCO) for data centers is reduced, thanks to the consolidation of all functional elements at the hypervisor level, as well as federated administration.

Hyper-Converged Architecture eliminates the need for discreet storage, complicated network topologies, explicit high availability designs or multiple management consoles.

IT executives are increasingly considering HCI as an efficient way to increase their business agility, cite increased business agility, IT staff productivity, operational efficiency, and faster time to value. HCI also provides greater simplicity, reduces operations expenditures (OPEX) and risk, and increases business agility. (Azeem & Sharma, 2017)

For the purpose of this comparison, we chose virtual infrastructure of 1'000 virtual servers with the average specifications in the following table:

	Per Server	Total (for 1000 Servers)
vCores	4	4'000
Memory (GB)	16 GB	16'000 GB
Storage (GB)	500 GB	500'000 GB

Table 3:Virtual Server Specifications

a. On-Premises Infrastructure

Servers

For the Servers we chose mainstream rack servers from one of the top infrastructure providers, DELL EMC. The model we chose is the "PowerEdge R7525 Rack Server" with a configuration of two CPUs with 24 cores each totaling 48 cores, 256GB of memory and 6'553.6 GB of redundant high-performance storage on board. The server also includes redundant power supply, Quad 25Gbps network adapter, service module for remote control, ProSupport (next day on site) for 5 years and licenses for VMWare hyper-converged solution for 5 years.

Figure 12 : Dell PowerEdge R7525 Rack Server

When calculating the number of servers required, we need to consider the following factors:

1. Hardware failures: In case of hardware failures, the rest of the servers will need to be able to handle the added workload until the failed servers are restored.

2. Usage increase/spikes: In case of periodical (for example holiday season) or permanent increase of usage, the servers should be able to handle the added load (increase of specifications of VMs or introduction of additional servers) until new infrastructure can be purchased, configured, and installed. This process, based on the company processes, can take anywhere from a few days to a few months (Cost Estimation, Securing the Budget, Purchase order, Delivery).

3. New services / upgrades of old services: In the case of introducing new services or upgrading old services, the number of servers need to be able to handle the additional load until, same as above, new infrastructure can be purchased, configured and installed.

The servers will need to have a maximum utilization limit which will provide the required threshold of the above conditions, and thus need to be over-provisioned. An overprovisioning of 120% and therefor a utilization of 83,3%, is a safe amount and will be used in our case.

Base on the above the number of the required servers will be the highest calculated from all requirements (Cores, Memory, Storage):

	Total (for 1000 Servers)	Total (Overprovisioned)	Per Physical Server	Servers Required
vCores	4'000	4'800	48	**100**
Memory (GB)	16'000 GB	19'200 GB	256 GB	75
Storage (GB)	500'000 GB	600'000 GB	6'553 GB	92

Table 4: Number of servers required.

Switches

For the Switches we chose the Dell EMC Networking S5248F-ON from the same Vendor, which are 48 port 25 Gbit Managed switches with additional 4 x 100 Gigabit QSFP28 + 2 x 200 Gigabit QSFP28-DD ports.

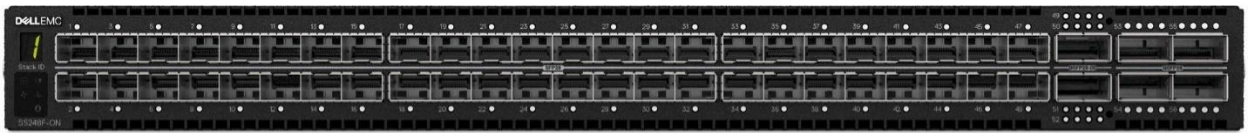


Figure 13: Dell S-Series S5248F-ON Switch

The servers have been configured with quad-25Gpbs network cards, as Hyperconverged infrastructure relies heavily on the underlying network for most of its core functionalities, such as storage replication across nodes, VM relocation for load balancing and fault tolerance, as well as the VM communication needs, and a high-performance and redundant network is of paramount importance for the operation.

Each of the quad ports of the servers will connect to a different switch for redundancy, and the switches will be independently interconnected using their 100Gbit and 200Gbit ports, and as such, the network can survive the failure of as many as 3 switches at the same time, although with degraded performance.

For calculating the number of switches required, we compute the total number of server ports and divide them by the number of ports per switch, which will provide the absolute minimum number of switches required:

- 100 Servers * 4 ports per server = 400 ports total
- 400 ports / 48 ports per switch = 8,33 → round up to next integer = 9

At this point we should consider the best practice of having free ports on every switch for future expansion or in case of single port malfunctions and add an additional switch (10 switches in total) which will be bring the port utilization down to about 400 / (48 * 10) = 83,3%, which lines up with our 120% overprovision server policy.

Routers

For the routers we chose the industry standard CISCO CATALYST 8500 SERIES 12, and specifically the model C8500L-8S4X.

Figure 14:Cisco C8500L-8S4X Router

The routers will allow the servers to communicate with the internet and/or other data centers over VPN.

For redundancy, the minimum number of routers is two, each also connecting to multiple switches.

Cables

For the Cables we chose the Dell Networking SFP28 to SFP28 25GbE, Passive Copper Twinax Direct Attach cables.

Figure 15: Dell Networking Cable, SFP28 to SFP28, 25GbE

The cables are 25Gbit, compatible with our switches, three meter cables, typical for cabling inside a rack. A minimum of 480 cables should be provisioned, same as the total number of ports on the switches.

Additional Costs

An additional 15% cost has been added to the sub-total to cover for the cost of copper and optical cabling (active or passive) with unpredictable length (inter-rack cabling, router cabling, etc.), and other network equipment like IP KVMs (Keyboard, Video and Mouse Switches).

Infrastructure Costs

Based on the above the total cost of the infrastructure (Compute and Network) is summarized on the following table:

	Price per Unit	Units	Total
PowerEdge R7525 Rack Server (+ 5YR License and Support)	53'812,55 €	100	5'381'255,04 €
Switches	8'500,00 €	10	85'000,00 €
Cables	55,00 €	480	26'401,54 €
Router	18'000,00 €	2	36'000,00 €
Other Networking equipment (KVMs, Optical cables, etc.)		15%	22'110,23 €
Network Total			169'511,77 €
Infra Total (excluding Racks)			**5'720'278,57 €**

Table 5: Total Compute and Network Infrastructure Costs

Data Center Costs

The calculations so far did not include any of the permanent data center infrastructure, such as racks, power and cooling. This is the case because, the Compute and Networking equipment has typically a lifecycle of 5 years maximum, whereas the rest of the equipment has a much longer lifecycle, 10 years minimum. As the depreciation period of the Data Center itself is much longer than the Compute and Network infrastructure, we chose to present two cases, one that assumes that a Data Center is in place and operational, and one that assumes that a new Data Center needs to be build.

A major factor of the data center cost analysis is the total power that it needs to be able to support. According to our analysis of the components so far, the power needs can be calculated as per the following table.

	Power per Unit	Units	Total
Servers	800 W	100	80'000 W
Switches	500 W	10	5'000 W
Router	1'000 W	2	2'000 W
		Total	87'000 W

Table 6: Infrastructure Power Requirements

The total Maximum power requirements for our compute and network equipment is summed up to 87kW, and allowing an overhead for future additions, we chose to analyze the Data Center for a 100kW design capacity.

Breaking down the data center costs is out of the scope of this thesis, and therefore we have used the data center cost calculator of Schneider Electric as can be seen in the following figure.

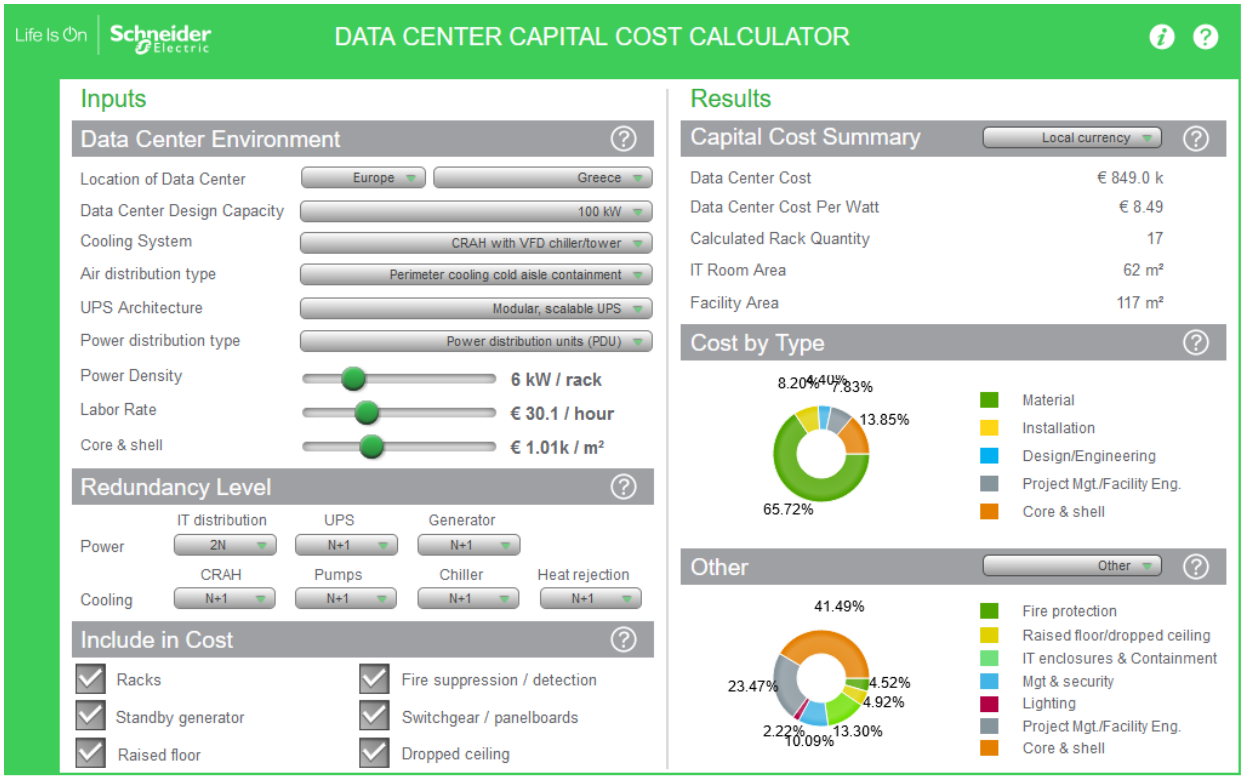


Figure 16: Schneider Electric CAPEX calculator

As can be seen above, we chose modern power and cooling solutions as well as redundancy in all levels:

- Computer Room Air Handler (CRAH) with Variable Frequency Drive (VFD) Chiller/Tower
- Perimeter cooling cold aisle containment.
- Modular and scalable UPS
- Power Distribution Units
- 2N redundancy in IT power distribution
- N+1 redundancy in the UPS
- N+1 redundancy in the Generator
- N+1 redundancy in the CRAH, Pumps, Chiller, and Heat Rejection equipment

The Design assumes and includes:

- 17 Racks (assuming a power density of 6kW per rack on a 100kW capable data center)
- Raised Floor design
- Fire detection and suppression equipment
- Switchgear and panelboards
- Dropped ceiling.

The cost includes:

- Building (core and shell) costs
- Design and engineering
- Labor costs
- Installation costs
- Project Management costs

The detailed cost can be found in the following table, broken down in three major categories (Power, Cooling and Other, as per the Schneider Electric Calculator), and these further divided to sub-categories. The percentages of the sub-categories sum up to 100% of the category total.

	Component	Percent	Cost
Power		**39,0%**	**331'110,00 €**
of which	UPS	36,0%	119'199,60 €
	Generator	27,0%	89'399,70 €
	Switch Gear	21,0%	69'533,10 €
	Critical Power Distribution	16,0%	52'977,60 €
Cooling		**28,0%**	**237'720,00 €**
of which	CRAH	13,0%	30'903,60 €
	Chiller	34,0%	80'824,80 €
	Cooling Tower	13,0%	30'903,60 €
	CHW pumps, piping, valves	20,0%	47'544,00 €
	CW pumps, piping, valves	20,0%	47'544,00 €
Other		**33,0%**	**280'170,00 €**
of which	Fire Protection	4,5%	12'607,65 €
	Raised Floor/Dropped Ceiling	4,9%	13'728,33 €
	IT Enclosures + Containment	13,3%	37'262,61 €
	Management & Security	10,1%	28'297,17 €
	Lighting	2,2%	6'163,74 €
	Project Mgmt & Facility Engineering	23,5%	65'839,95 €
	Core & Shell	41,5%	116'270,55 €
		Total Cost:	**849'000,00 €**

Table 7: Data Center CAPEX analysis

Power Costs

The previous power calculation of 87 kW was the maximum power draw expected of the Compute and Network infrastructure. Of course, the equipment will not draw continuously 100% of its rated power, but will be closer to 30% on average, while the total power consumption and heat dissipation of the infra and power distribution will be 50% higher than the compute and network equipment consumption as can be seen in the following figure (Rasmussen, 2007).

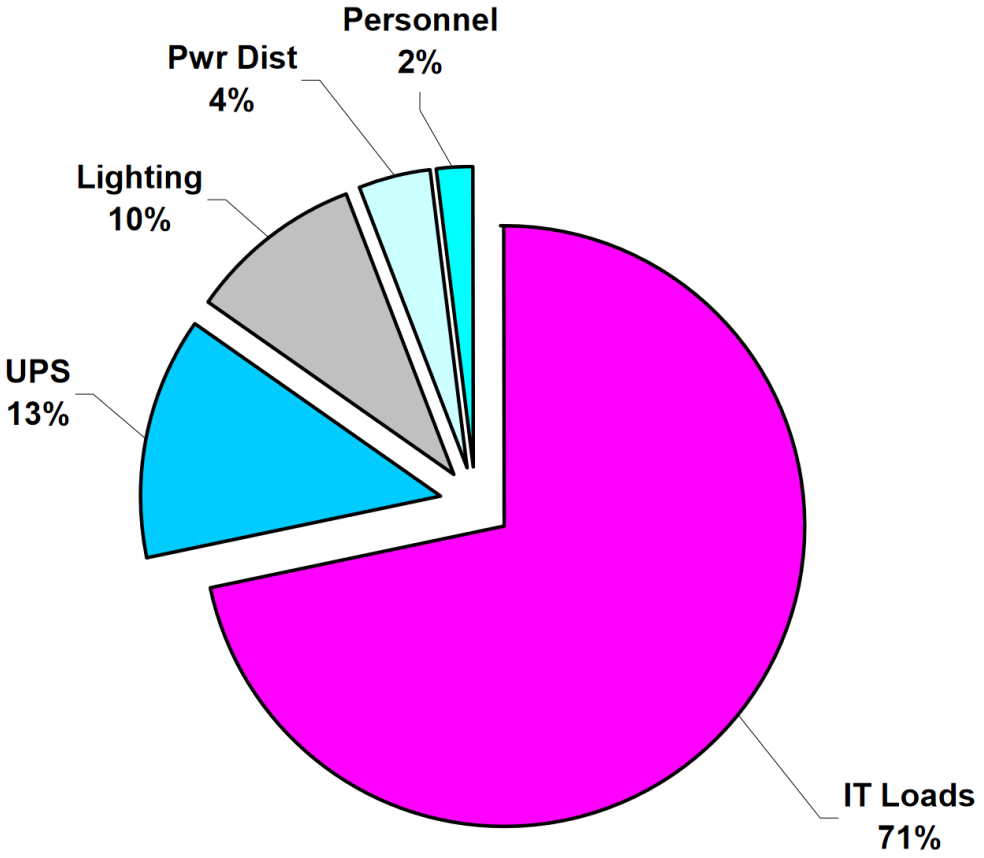


Figure 17:Relative contributions to the total thermal output of a typical data center

Thus, the total heat dissipation and power consumption of the datacenter equipment (excluding cooling) will be (rounded up to first decimal):

$$87kW \times 30\% = 26{,}1\ kW$$
$$26{,}1kW \times 150\% = 39{,}2\ kW$$

Equation 1: Average Power Consumption

While the cooling power required will be 1.3 times the average power consumption:

$$39{,}2kW \times 130\% = 50{,}9\ kW$$

Equation 2: Cooling Average power consumption

And thus, the total consumption and cost over a 5-year period can be seen in the following table.

	Power	Hours in 5YR	kWh in 5YR	kWh cost	Monthly Fees	5YR Cost
Infra	39,2 kW	43'800	1'716'960 KWh	0,10	350,9968	192'755,81 €
Cooling	50,9 kW	43'800	2'229'420 KWh	0,10	455,7586	250'287,52 €
					Total	**443'043,32 €**

Table 8:Power Consumption and Cost over a 5 year period

The cost per kWh as well as the monthly costs have been calculated based on the official pricelist of the major local power provider and distributor at the time of the writing (DEI, 2020).

Personnel

The minimum personnel that will be required for the operation of the data center is as follows:

- System Engineers / Administrators
- Facility Management
    - Security
    - Sanitation

For the System Engineers and Administrators, the requirements are for highly trained and experienced people, with two people being on premises during working hours, at least one being on premises during extended working hours (working hours of the physical shops when the offices are not operational – afternoons and weekends), and one during the off-hours periods.

Thus, during the course of 1 week the following number of man-hours are going to be needed :

$$2 \times 40 + 1 \times 40 + 1 \times 40 + 24 + 24 = 208\ MH$$

Equation 3: Man-Hours needed in a week for SEs/SAs

We need to also account for leave days and rolling shifts, so the minimum amount of full time System Engineers / Administrators that are needed is six.

For the facility management personnel, we need four full time security officers, and one part time sanitation personnel.

The average cost of a highly trained System Engineer/Administrator is 50'000€ per annum, taking into account the rolling shifts and working during off-days. The average cost for the facility management

personnel is 30'000€ per annum, taking again into account the rolling shifts and working during off-days.

Thus the personnel cost, over a period of five years can be seen in the following table.

Personel	People	Cost / Person / Year	5 YR Total Cost
System Engineers	6	50'000,00 €	1'500'000,00 €
Facility management	4,5	30'000,00 €	675'000,00 €
		Total	**2'175'000,00 €**

Table 9:Personnel cost over a 5 year period

Total On-Premises Cost

As mentioned previously, we will provide analysis and comparison for two cases, with or without the CAPEX cost of the Data Center, and in the first case, we will include 5-year and 10-year depreciation period. Taking all of the above into account, the total cost for the on-premises scenario can be seen in the following table.

	Cost
Infrastructure	5'720'278,57 €
Power Consumption	443'043,32 €
Personnel	2'175'000,00 €
Sub-Total	**8'338'321,90 €**
DC Capex	849'000,00 €
Total with 50% Capex	**8'762'821,90 €**
Total with 100% Capex	**9'187'321,90 €**

Table 10: Total On-Premises cost for a 5-Year period

The Sub-Total in the table depicts the cost without the Date Center CAPEX cost, the “Total with 50% Capex” depicts a 10-Year depreciation period for the Data Center, while the “Total with 100% Capex” depicts a 5-Year depreciation period.

b. Cloud Provided Infrastructure

In the case of the cloud provided infrastructure, the cost scales almost linearly with the size of the VM, so we will not explore different size scenarios, but will provide the cost for 1000 VMs with 4vCPUs, 16GB RAM and 500GB of storage.

AWS has made available many different pricing models, depending on commitment types and periods. The pricing models we will explore are:

- **AWS On Demand**: Most flexible and most expensive, no commitment period.

- **AWS 1 year – No upfront**: Reserve specific instance types for 1-year with no upfront costs.
- **AWS 1 year – Full upfront**: Reserve specific instance types for 1-year and pre-pay the entire compute costs upfront.

- **AWS 3 year – No upfront**: Reserve specific instance types for 3-years with no upfront costs.
- **AWS 3 year – Full upfront**: Reserve specific instance types for 3-years and pre-pay the entire compute costs upfront.

- **Compute Savings Plan 3 year – No upfront**: Compute Savings Plans are a flexible pricing model that offer low prices on EC2, Lambda, and Fargate usage, in exchange for a commitment to a consistent amount of usage (measured in $/hour) for a 3-year term. In this case we commit to a 3-year spending of a specific amount of compute resources, with no upfront costs.
- **Compute Savings Plan 3 year – Full upfront**: In this case we commit to a 3-year spending of a specific amount of compute resources, with full upfront costs.

Reserved instances are inflexible as the lock us down in the type of instances we can use, and while that might be acceptable for a short period of time (1 year) it is unacceptable for longer periods, as elasticity is a central part of the cloud promise, and such commitments do not make sense in the long run, but they do provide the highest cost savings, so they will be included and compared to the rest of the saving plans.

The costs can be seen in the following table, converted from US Dollars using a USD to EUR currency exchange rate of 0.84 USD per Euro, and also calculated for a 5-year period.

AWS Pricing Model	3YR in USD	3YR in EUR	5YR in EUR	Savings vs On-demand
AWS On-demand	$ 7'116'876,00	5'978'175,84 €	9'963'626,40 €	0%
AWS 1YR No Upfront	$ 5'253'624,00	4'413'044,16 €	7'355'073,60 €	26%
AWS 1YR Full upfront	$ 5'040'756,00	4'234'235,04 €	7'057'058,40 €	29%
AWS 3YR No upfront	$ 4'254'984,00	3'574'186,56 €	5'956'977,60 €	40%
AWS 3YR Full upfront	$ 3'973'788,00	3'337'981,92 €	5'563'303,20 €	44%
Compute Savings Plan 3YR No Upfront	$ 4'993'452,00	4'194'499,68 €	6'990'832,80 €	30%
Compute Savings Plan 3YR Full Upfront	$ 4'725'396,00	3'969'332,64 €	6'615'554,40 €	34%

Table 11: AWS Pricing Plans and Costs

As can be seen above, even the minimum commitment of 1-year can provide significant cost savings (26%), while the highest savings can be achieved by selecting a 3-year commitment on reserved instances (40%). The compute savings plan provides a good balance of flexibility and savings (30%) and we can clearly see that pre-paying the entire consumption doesn't provide a significant benefit (4% or less) and should only be consider in cases where full-upfront payment can be transformed into CAPEX instead of OPEX, which will be presented in the following section.

c. Comparison of AWS and On-Premises

In the following table and chart, we have presented the comparison between AWS hosted infrastructure and on-premises datacenter in an "Advantage of AWS over On-Premises" form, meaning that all negative numbers represent the advantage of On-Premises over AWS, while positive numbers represent the advantage of AWS over On-Premises, with red being the most advantageous for On-Premises and green the most advantageous for AWS.

	AWS On demand	AWS 1YR No Upfront	AWS 1YR Full upfront	AWS 3YR No upfront	AWS 3YR Full upfront	Comp. Savings Plan 3YR No Upfront	Comp. Savings Plan 3YR Full Upfront
No DC Capex	-19,49%	11,79%	15,37%	28,56%	33,28%	16,16%	20,66%
DC Capex 10y depreciation	-13,70%	16,07%	19,47%	32,02%	36,51%	20,22%	24,50%
DC Capex 5y depreciation	-8,45%	19,94%	23,19%	35,16%	39,45%	23,91%	27,99%

Table 12: Advantage of AWS over On-Premises (percent)

		AWS On demand	AWS 1YR No Upfront	AWS 1YR Full upfront	AWS 3YR No upfront	AWS 3YR Full upfront	Comp. Savings Plan 3YR No Upfront	Comp. Savings Plan 3YR Full Upfront
		9'963'626,40 €	7'355'073,60 €	7'057'058,40 €	5'956'977,60 €	5'563'303,20 €	6'990'832,80 €	6'615'554,40 €
No DC Capex	8'338'321,90 €	-1'625'304,50 €	983'248,30 €	1'281'263,50 €	2'381'344,30 €	2'775'018,70 €	1'347'489,10 €	1'722'767,50 €
DC Capex 10y depreciation	8'762'821,90 €	-1'200'804,50 €	1'407'748,30 €	1'705'763,50 €	2'805'844,30 €	3'199'518,70 €	1'771'989,10 €	2'147'267,50 €
DC Capex 5y depreciation	9'187'321,90 €	-776'304,50 €	1'832'248,30 €	2'130'263,50 €	3'230'344,30 €	3'624'018,70 €	2'196'489,10 €	2'571'767,50 €

Table 13: Advantage of AWS over On-Premises

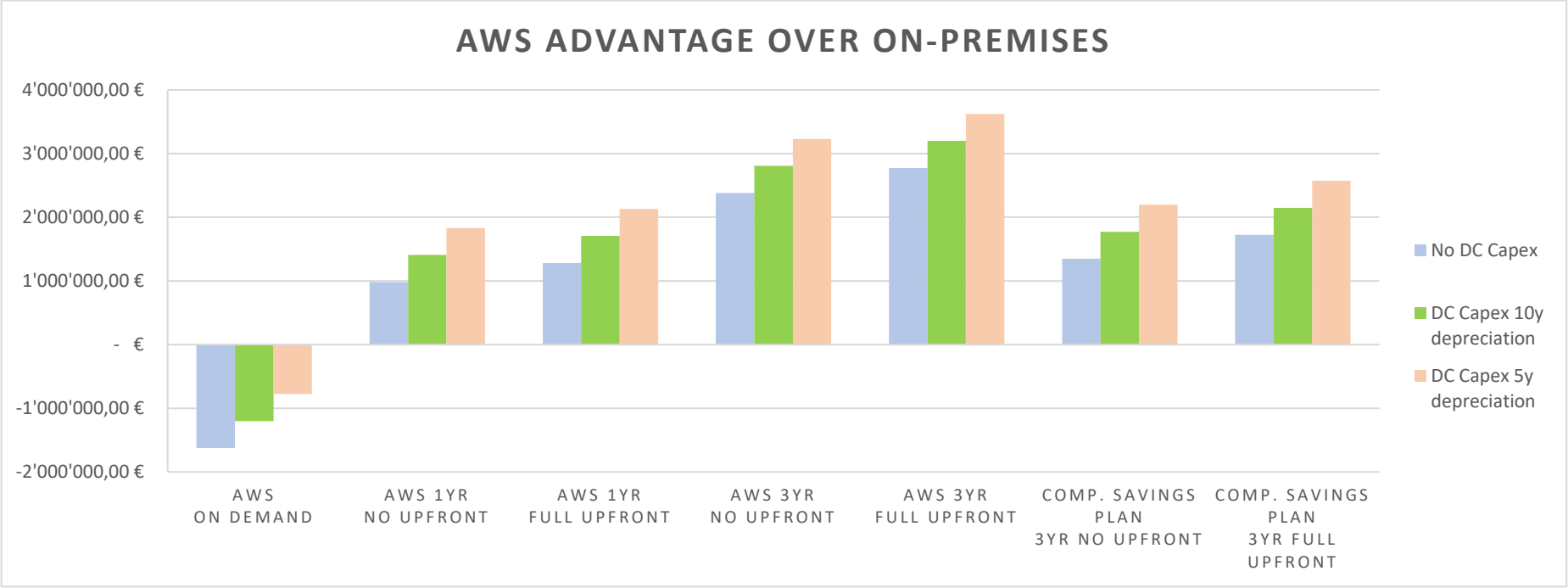


Figure 18: Advantage of AWS over On-Premises

As we can clearly see above, the Cloud infrastructure holds an advantage over on-premises infrastructure, with the exception of On-Demand instances.

On Demand instances are pay-as-you-go services that are suitable for short-term unpredictable workloads (Testing, Proof of Concepts), and under no circumstances should be used for predictable, long-term workloads. For these kinds of workloads, while the initial setup will use On-Demand instances, after the workload has gone live and the demand has stabilized, reserved-instances can be purchased, that will correspond to the actual need, or Compute savings that will provide a great level of flexibility.

So, it is highly recommended to perform a detailed Cost Optimization process after the workloads have stabilized and use Auto-Scaling or Demand Prediction tools to keep the cost optimized.

We have compared VM hosting, which is the most basic and directly comparable resource, between the Cloud and On-Premises and have demonstrated that the Cloud holds a clear advantage. This advantage becomes even more pronounced if we take into consideration serverless applications and services, that provide the highest cost granularity.

Services like Load Balancers, Functions, Gateways will be charged by second or per data transferred, will auto-scale transparently according to demand, and will be highly available by default, which is something that will take great effort and cost in an On-Premises environment.

Finally, almost all of the services provide virtually infinite scale out options, that are nearly immediate, can scale in when not used, and can be shut down minimizing the costs. On the contrary, On-Premises infrastructure must be provisioned by predicting the highest demand, and be kept operational for its full capacity, whether it is fully used or not. This becomes apparent when using Cloud Native applications, where elasticity and use of serverless (PaaS) services are part of the design.

CAPEX vs OPEX

Cloud technology brought ground-breaking changes in the way we invest therefore the way of creating budgets and curve business strategies.

Initially, let us define the concept of CapEx and OpEx:

- **Capital Expenditure (CapEx)** are funds that a business uses to acquire, upgrade, and maintain physical long-term assets such as land, plants, buildings, and technology. CapEx is frequently used by businesses to fund new projects or investments. Capital expenditures on fixed assets may include the repair of a roof, the purchase of equipment, or the construction of a new factory. It is essentially – a payment for goods or services that typically have a useful life of one year or more.

- **Operational Expenditure (OpEx)** are the funds that are used for covering cost incurred by a business during normal business operations, like supplies, rent, office overhead, salaries of employees, inventory costs, generally anything that supports day-to-day business.

Historically, technology investments were prioritized as capital expenditures rather than operating expenditures, as CFOs could defer these costs for an extended period of time. Nowadays, an increasing number of businesses are shifting their IT investments from CapEx to OpEx, and they have a compelling reason for doing so – moving their IT infrastructure to the cloud. Additional CapEx benefits accumulate as a result of the company no longer requiring static investments in hardware, software, and resources. (Comindware, 2021)

Services and options are purchased on a need-to-know basis, costs fluctuate, and OpEx works better for these types of expenses and provides the necessary scalability. On the other hand, CapEx depicts the investments that have been occurred by the company and as a result, it has a great influence on the business image of the company in terms of sustainability and growth, therefore many companies follow the approach of upfront payment. With this option, companies can justify the costs of renting the infrastructure, as investment and increase CapEx accordingly, since the "goods" that they are buying have more than one year useful lifecycle in order to be considered as investment.

More specifically, there are two options for long term infrastructure owning such as virtual machines EC2 and dedicated host servers like Bare Metal machines and that is – one-year or three-year contracts. This approach leads to 5-10% cost discount at minimum and it is usually beneficial for heavy workloads that need to be live 24/7 since we only pay for the storage allocated and the services running on these machines, not the use of the machine. As more businesses migrate away from traditional hardware and software ownership toward as-a-service models, IT and finance departments must agree on the best way to classify cloud costs.

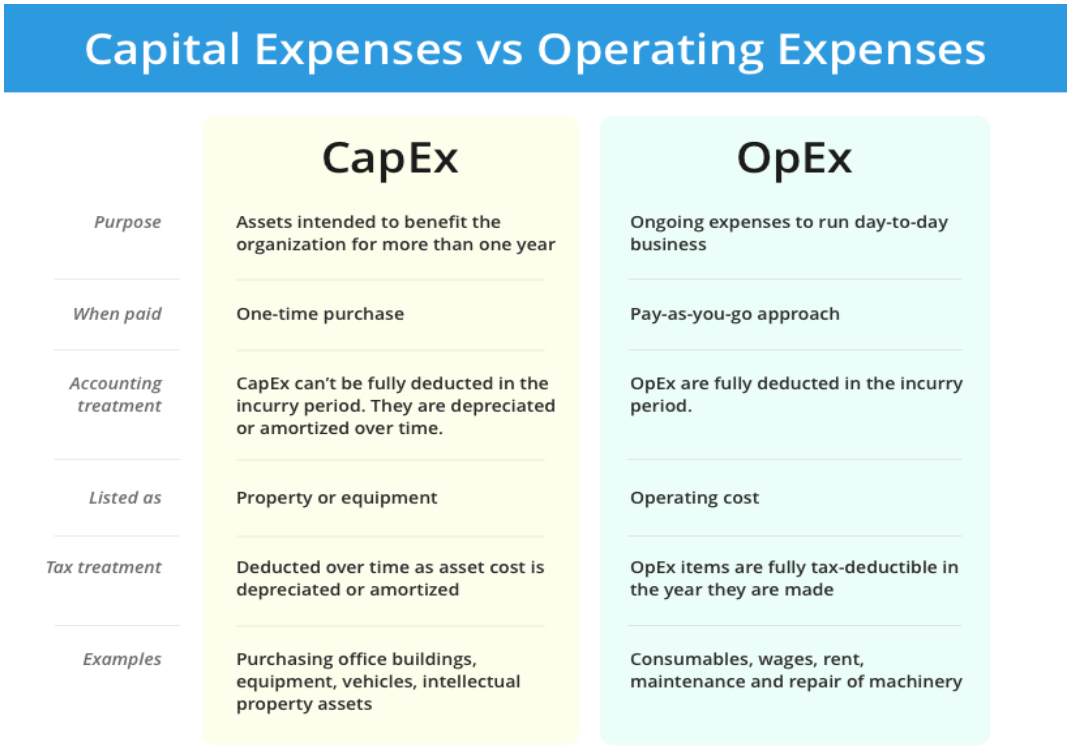


Figure 19: CapEx vs OpEx (Comindware)

Migration Plan

The migration planning is of paramount importance for a Service Provider, as any outage, at any time of day may -and most probably will- result in revenue loss. A Service provider needs to plan carefully any migration to allow for minimal downtime, with minimal impact and account for reliable and fast roll back scenarios in case of failure.

We will split the migration planning in the following stages:

1. **Application / Service Migration Evaluation**: Each service/application is evaluated in order to choose an appropriate Strategy.
2. **Identifying Dependencies**: We will provide methods for identifying inter-service dependencies and analyze how these dependencies affect the migration plan.
3. **Plan for secure transfer of data**: We will analyze strategies for migrating securely and reliably data to the cloud.
4. **Plan for zero downtime**: We will explore strategies that will provide zero-downtime during the migration of Mission Critical services.
5. **Migration Plan**: We will summarize and evaluate the migration plan.
6. **Cost**: We will try to enumerate the factors that will have a minor or major cost impact.

The migration strategies will be architected around “the 6 R’s” (AWS, 2018) :

1. **Re-host (Referred to as a “lift and shift.”)**:

   Transfer applications without modifying them. Organizations undergoing large-scale legacy migrations need to move quickly in order to meet business objectives. Most of these applications are re-hosted on third-party servers. We have demonstrated in our previous cost analysis that we could save roughly 30% of the costs by re-hosting.

   The majority of re-hosting tasks can be automated using tools (for example, AWS VM Import/Export). Certain consumers prefer to do this manually as they become familiar with the new cloud platform and how to adapt their legacy systems.

Once applications are running in the cloud, they are easier to optimize/re-architect. Partly because the organization will have acquired the necessary skills, and partly because the difficult part — migrating the application, data, and traffic — will have already been completed.

2. **Re-platform (Referred to as "lift, tinker, and shift.")**:

   Optimize a few cloud components to realize a tangible benefit. The Organization will not alter the application's core architecture. Reduce the time spent managing database instances, for example, by migrating to a database-as-a-service platform like Amazon Relational Database Service (Amazon RDS) or a fully managed platform like AWS Elastic Beanstalk.

3. **Re-factor / Re-architect**:

   Using cloud-native features, reimagine how the application is architected and developed. This is motivated by a compelling business need to add features, scale, or performance to the application that would be difficult to achieve in its current environment.

   An example is migrating from a monolithic to a service-oriented (or server-less) architecture in order to increase agility and business continuity. This strategy is typically the most expensive, but it can also be the most beneficial if our product and market fit are strong.

4. **Re-purchase**:

   Adopt a software-as-a-service (SaaS) model in place of perpetual licenses. For instance, migrate from one customer relationship management (CRM) system to Salesforce.com, from one human resource management (HR) system to Workday, or from one content management system (CMS) to Drupal.

5. **Retire**:

   Eliminate unused applications. Once our environments have been discovered, we need to determine who owns each application. As much as 10% – 20% of an enterprise's IT portfolio is no longer necessary and can be decommissioned. These savings can help us strengthen our business case, refocus our team's efforts on the applications that people use, and reduce the number of applications that need to be secured.

6. **Retain (Referred to as re-visit.)**

   Maintain mission-critical applications that require extensive refactoring before they can be migrated. We can re-visit all applications in this category at a later date. This includes the “remain” category that has been discussed in the “Services that will remain on Premise” section.

In the following figures we provide a high-level diagram (HLD) of the strategies as well as a high-level comparison in terms of cost, effort and complexity of the strategies.

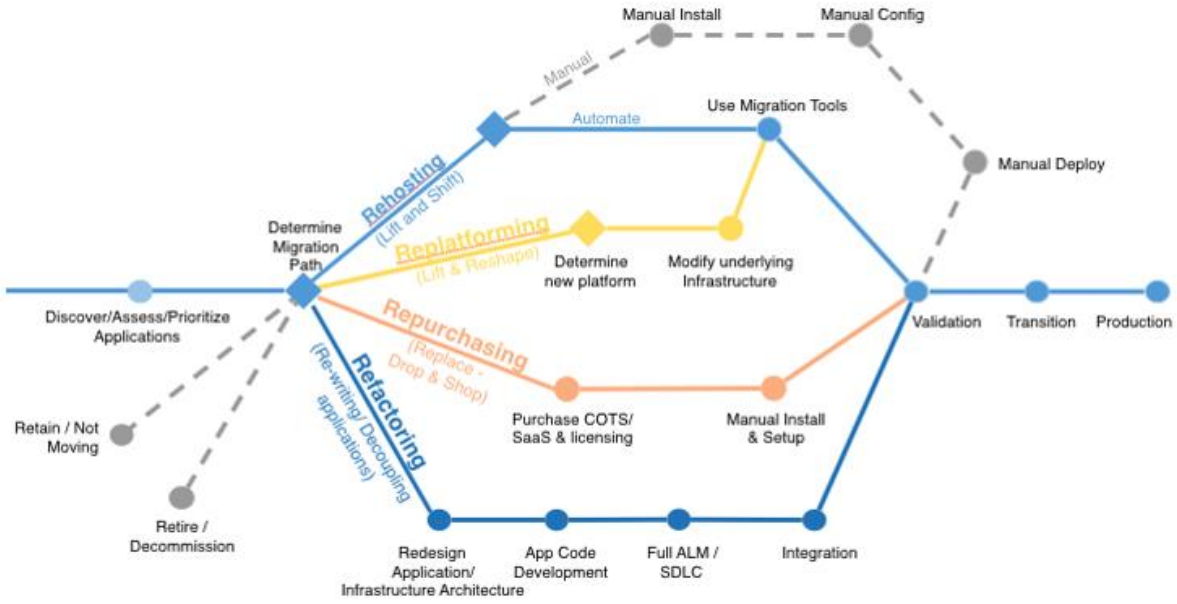


Figure 20: The 6 R's - Diagram (AWS)

Figure 21: High Level Comparison of the 6 strategies (AWS)

In the following sections we assume that the Organization is fully aware of its services, and although an Application discovery will be proposed in “Identifying Dependencies” section, it is expected that the Company already possess a complete Application and Infrastructure portfolio.

Of course, discovery strategies do exist to cover this need, but they are out of the scope of this thesis.

The migration plan's primary objective is to direct the overall migration effort. This includes managing the project's scope, schedule, resource plan, issues, and risks, as well as coordinating with and communicating with all stakeholders. Early planning can help organize the project, especially when multiple teams migrate multiple applications. The migration plan takes into account critical factors such as the migration order of workloads, the timing of resource requirements, and tracking the migration's progress. We recommend that the teams employ agile delivery methodologies, project control best practices, a well-defined delivery approach, and a robust business communication plan.

Among the activities recommended for a migration plan are the following (AWS, 2018):

- To identify any gaps in project management methods, tools, and capabilities, a review of the methods, tools, and capabilities is conducted.
- Define the methods and tools for project management that will be used during the migration.
- Define and develop the Migration Project Charter/Communication Plan, including procedures for reporting and escalation.
- Create a project plan, risk/mitigation log, and roles and responsibilities matrix (e.g., RACI) to manage project risks and assign ownership to each resource involved.
- Procure and implement project management tools to assist in the project's delivery.
- Determine the critical resources and contacts for each of the migration work streams defined in this section.
- Facilitate the coordination and execution of the plan's activities.
- Outline the resources, timelines, and costs associated with migrating the target environment to AWS.

Application / Service Migration Evaluation

The most important prerequisite for a successful Cloud Migration is to have full awareness of the application portfolio. A comprehensive and complete application repository, including Applications, Assets (servers and services) and interfaces / interconnections, will set the corner stone for the migration plan.

In order to create or update the repository, discovery tools can be used that will automate the entire process. The tools can be used to:

1. **Discover servers and assets**: Scanner appliances will be placed in the network, covering all network segments, will scan all IPs on ICMP and well known / custom ports, and will query the discovered assets for more information such as OS Type and Version. This will be the basis for the repository and should be used to map assets to Applications/Services as well as discover and document unknown assets. The scanners can remain on the network, and periodically scan for new assets or rogue servers.

2. **Discover Application components, interfaces, inter-connections**: In this step, agents need to be installed in all of the servers that will scan the applications and their components and document them in the repository and monitor network connections to identify inter-application connectivity that will make interface discovery easier. Services that do not have physical assets (for example SaaS services provided by 3rd parties) can also be discovered here. This step needs to run for long periods of time in order to be able to discover connectivities that are rarely used – for example accounting applications that only run at the end of the month.

While there are plenty of 3rd party tools and solutions for the discovery, in our case, AWS provides their own tool, AWS Application Discovery Service which collects server specification information (hostnames, IP addresses, MAC addresses, and resource allocation), performance data (utilization details of key resources including CPU, network, memory, and disk), and details of running processes and network connections.

The data from the Application Discovery service can be imported into the AWS Migration Hub, where the assets can be grouped into applications, and the migration itself can be planned and tracked. AWS migration hub also provides visualization tools as can be seen in the figure below.

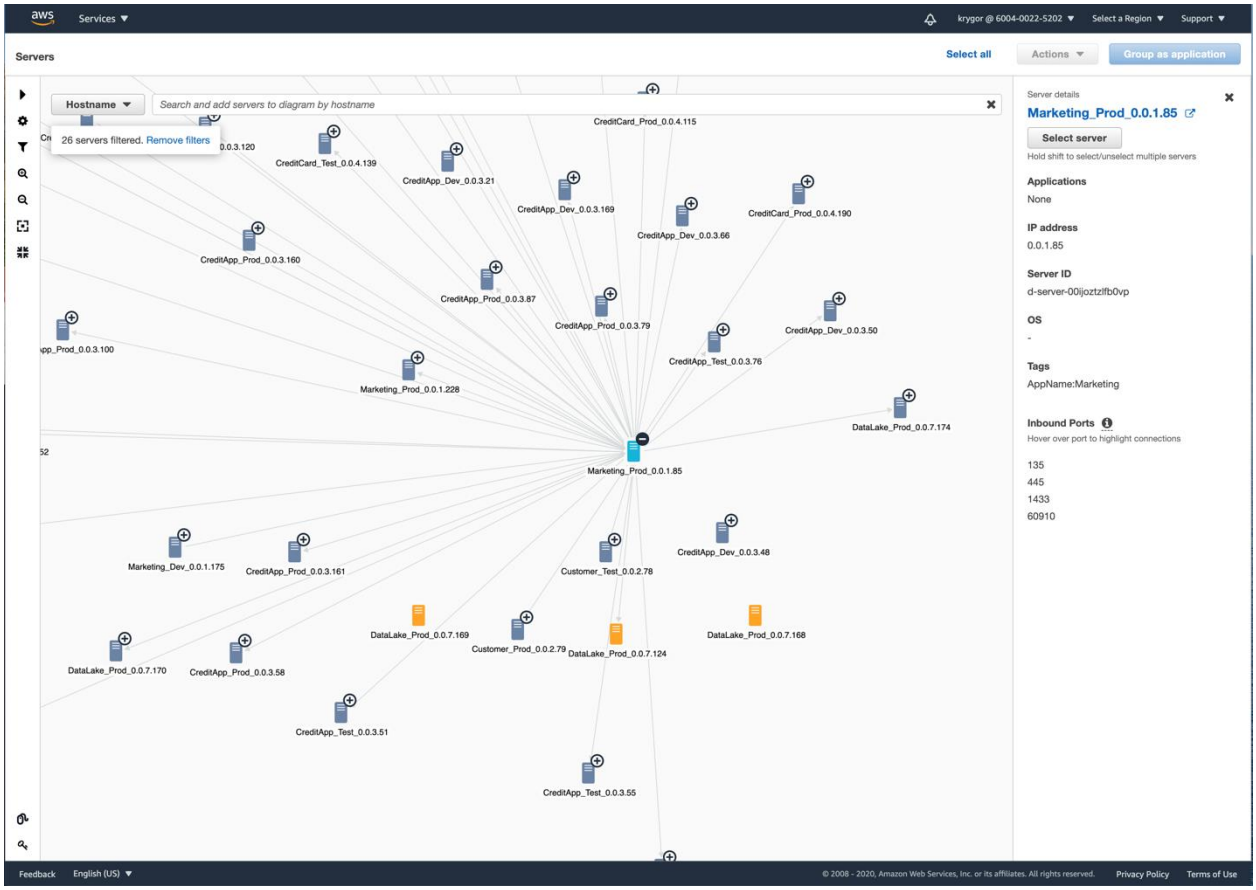


Figure 22: AWS migration HUB visualization (AWS)

After the discovery has finished, the Company will have a complete Applications/Services portfolio and can start the application evaluation. Before starting to evaluate the applications for migration it is very important to identify all applications that are no longer in use and can be retired. This will reduce the amount of applications that need to be evaluated and migrated, and subsequently reduce the cost and effort of the next steps.

In the following figure, we have presented the decision diagram that the company can follow to build up the Migration plan.

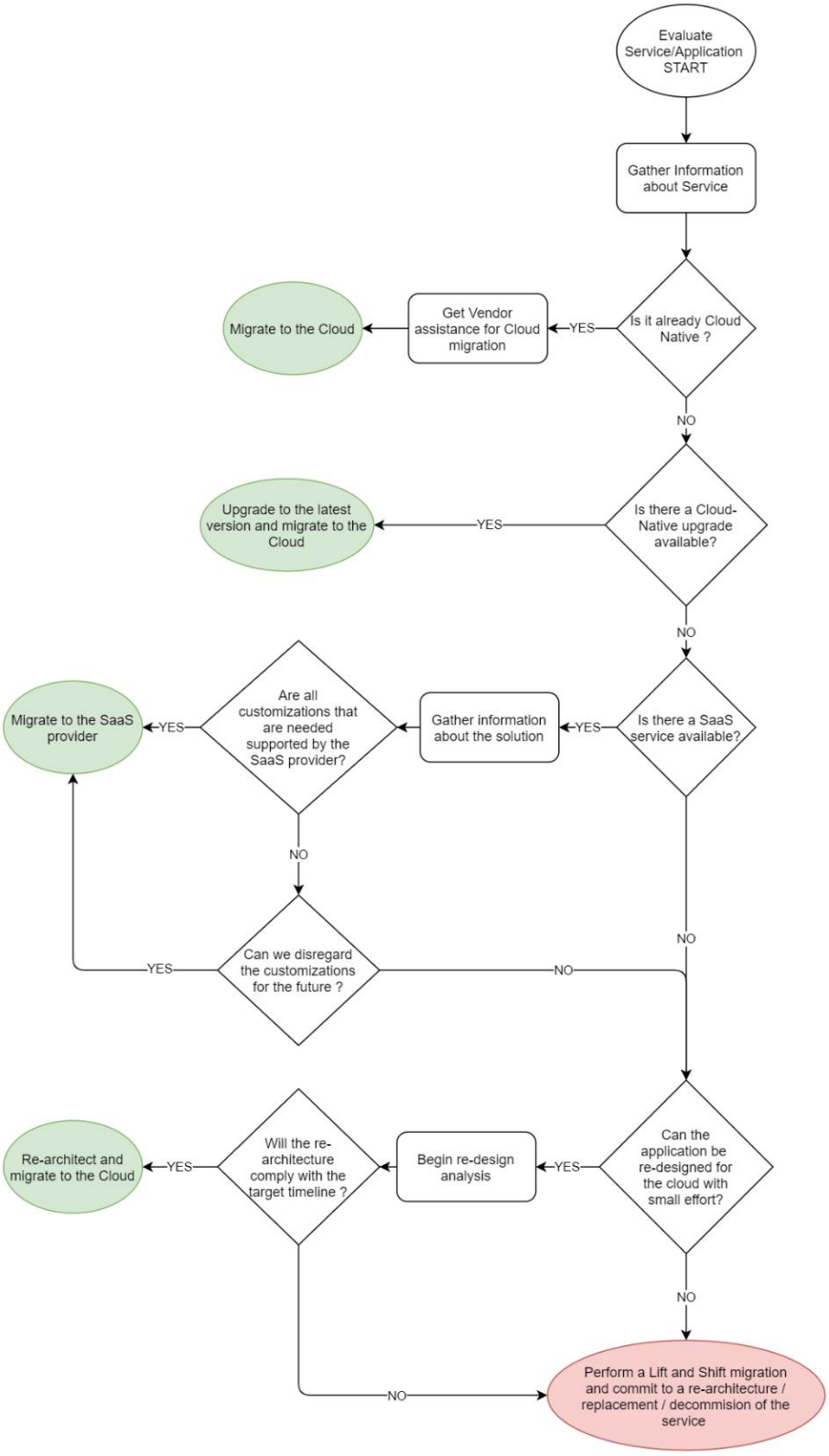


Figure 23:Decision diagram for Service/App migration

The Decision Diagram can be explained as follows:

- If the application is already cloud native, it can be directly migrated to the Cloud.

- If the application is not Cloud Native, but a Cloud Native upgrade is available, then the application can be upgraded and migrated to the Cloud in one (deploy the latest version on the cloud and then migrate the data) or two steps (upgrade in place, and then migrate the application to the cloud).

- If there is no Cloud Native upgrade (or near-in-the-future upgrade) available, then the availability of a SaaS solution provider should be explored. If such a solution is available, then the level of customizations of our current solution should be evaluated.

    - If the customizations are minimal and/or can be supported by the SaaS provider then we can migrate the application to the SaaS provider.
    - If the customizations are extensive and/or cannot be supported by the SaaS provider, then the application cannot be migrated to the SaaS provider.

- If a SaaS solution is not viable, then the re-design / re-factoring / re-architecture of the application should be evaluated. The key decision to be made here is whether the re-factoring effort is worth the time and resources and can meet the deadlines. If the effort can be justified, then a re-factoring plan needs to be made, a project initiated, and completed within the deadlines. If it cannot be justified, then we have three options:

    - Perform a “Lift and Shift” which means that we migrate the application “as is” and plan for a re-factoring at a later time.
    - Purchase a new application to replace the old one, making sure that it is Cloud Native and deploy it to the Cloud.
    - Re-Evaluate the necessity of the application and plan for a future decommission of the application after performing a “Lift and Shift”.

Identifying dependencies

Identifying dependencies is of great importance in order to achieve a smooth and trouble-free migration. It is critical to conduct a thorough, data-driven analytical process that gathers all necessary data to truly comprehend underlying dependencies and complexities. As a matter of fact, a successful migration path relies on grouping interdependent applications and migrating the group, rather than a single application.

The process of identifying and documenting a software application's external dependencies, such as servers, networks, storage, and applications, is called application discovery and dependency mapping. This process entails not only recognizing all of the software ecosystem's components, but also comprehending how they interact and affect one another.

Having an up-to-date Application portfolio will help to identify the applications that are interdependent and need special care. Such applications need special planning, in order to minimize the risk of separating components that have to be connected and thus causing outages.

More complicated workloads like centralized databases, or data warehouses, will certainly introduce numerous complications. In such cases we have two options:

1. Migrating the service while also creating the required communication flows from the dependents to the new service. This approach will require a certain effort to reconfigure all of the dependent applications to using the new communication flows and the rollback, in case of migration trouble is more complicated.

2. Migrating the service and placing a proxy / proxies in the place of the old assets. This approach will require less effort, as the dependent applications will not need any changes, but the addition of the proxies might introduce latencies and performance degradation. When all dependents have been migrated to the cloud, the proxies can be safely decommissioned.

The two different scenarios are presented in the following figure.

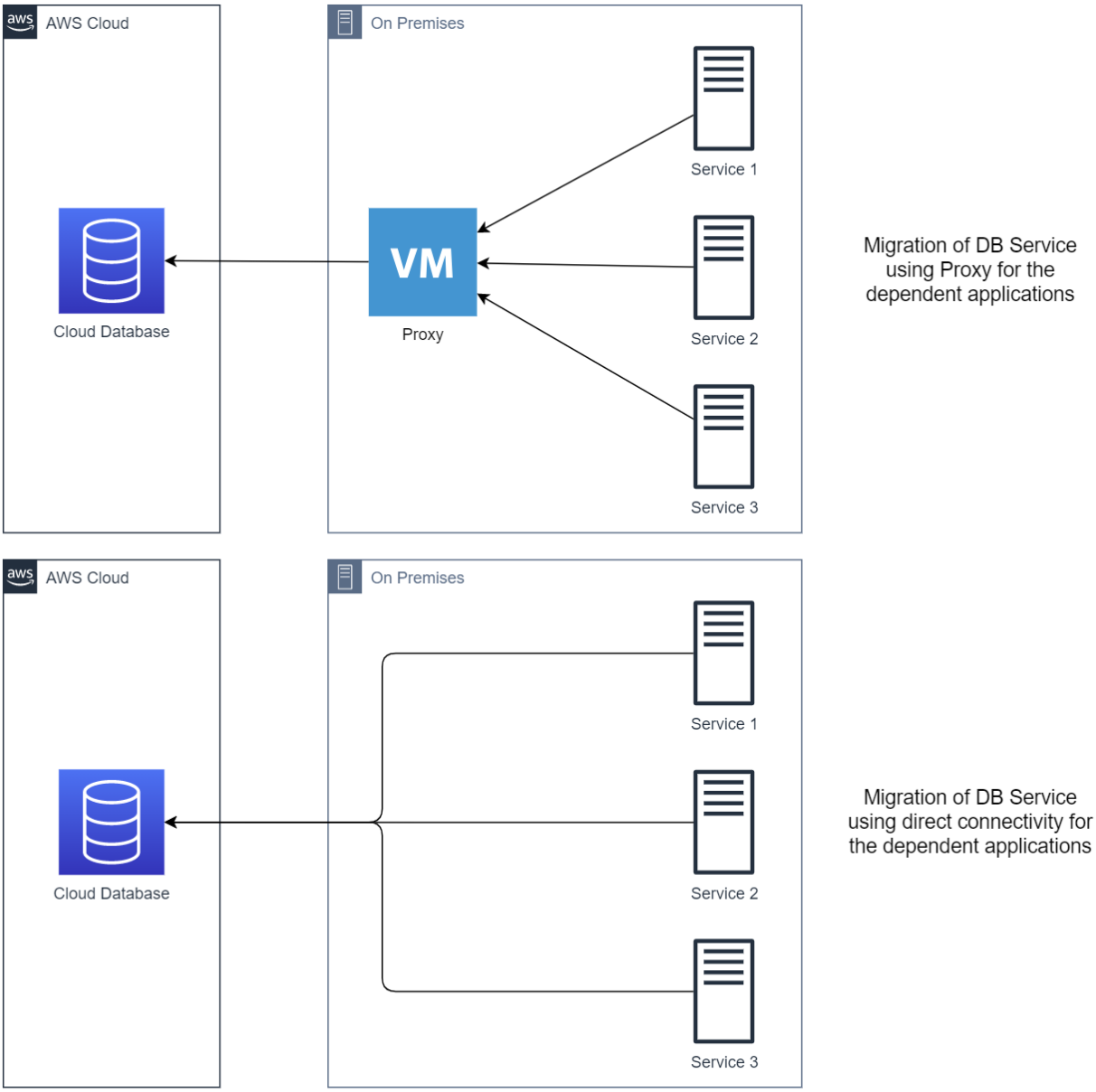


Figure 24: Migration of Service with Dependent applications

Plan for secure transfer of data

Data migration is the process of transferring data from one location to another. This process identifies the data to be transferred and then prepares it for transmission to another system storage location. It is also known as "system storage migration." Additionally, data migration services can assist with the migration of on-premises infrastructure to cloud storage/applications.

The following points need to be taken into account when planning or performing a data migration.

- **Data format identification** – We need to determine the format of the data to be migrated. Where it is stored, and the format into which it will be transferred following migration. During this pre-planning stage, we need to identify all possible errors and security precautions that must be taken when transferring data. This step can help safeguard our data from harm.
- **Make a backup of the data** – A backup of the data must always be created prior to performing a data transfer. If a problem occurs during the migration process, such as data corruption, missing files, or data alteration, we will already have a backup of our original files stored in a secure location. We can resolve the issue by restoring them to their original state.
- **Utilize data migration software** – When large data files are moved, data transfer becomes a tactical task. Manual data movement takes a long time. We can use data migration software to expedite the data transfer process.
- **Procedure for data migration** – We must ensure that permissions are successfully applied for complete data extraction and migration to the desired location, keeping the plan in mind. The data that are about to be migrated must be free of bugs/viruses and transformed into the correct format. Following that, they will be uploaded for migration. Finally, close monitoring of the process must be in place, for any issues that may arise along the way.
- **Final testing** – Following migration, checking for any problems to the target system must be done. The objective is to migrate all data safely and without modifying it. Testing must be conducted on the system, volume, web, and batch applications to verify this.
- **Follow-up maintenance** – Because errors can occur during the testing process, a full system inspection must be conducted in order to identify any possible errors/damage. In the event of an error, the data can be restored from the backup.

In the following sections we will present the available solutions for three different types and/or stages of data migration:

- Online Data Transfer: These services can be used for transferring data that are of small to moderate size and can be transferred over the network in a reasonable time. In the table below we can see that depending on the speed and size the transfer time can take exorbitant amounts of time.

		Size (GigaBytes)				
		10 GB	100 GB	1'000 GB	10'000 GB	100'000 GB
Transfer Speed (MegaBits/sec)	10 Mbit/s	0 days 02 h 13 m 20 s	0 days 22 h 13 m 20 s	9 days 06 h 13 m 20 s	92 days 14 h 13 m 20 s	925 days 22 h 13 m 20 s
	100 Mbit/s	0 days 00 h 13 m 20 s	0 days 02 h 13 m 20 s	0 days 22 h 13 m 20 s	9 days 06 h 13 m 20 s	92 days 14 h 13 m 20 s
	1'000 Mbit/s	0 days 00 h 01 m 20 s	0 days 00 h 13 m 20 s	0 days 02 h 13 m 20 s	0 days 22 h 13 m 20 s	9 days 06 h 13 m 20 s
	10'000 Mbit/s	0 days 00 h 00 m 08 s	0 days 00 h 01 m 20 s	0 days 00 h 13 m 20 s	0 days 02 h 13 m 20 s	0 days 22 h 13 m 20 s

Table 14: Transfer Times

- Offline Data Transfer: These services can be used to transfer large amounts of data offline, by physical means. These are ideal for large amounts of Data or in cases where a network connection is not available or reliable.

- Hybrid cloud storage: These services can be used to connect the on-premises services directly to the cloud data services. This way the on-premises services will be able to use the migrated data directly.

Online data transfer

- **AWS DataSync**

  AWS DataSync is a data transfer service that simplifies the process of automating data transfers between on-premises storage and Amazon S3 or Amazon Elastic File System (Amazon EFS).

DataSync automates a large number of data transfer tasks that can slow down migrations or burden our IT operations, such as running our own instances, managing encryption, managing scripts, optimizing our network, and validating data integrity. DataSync enables us to transfer data up to ten times faster than open-source tools.

DataSync connects to our existing storage or file systems via the Network File System (NFS) protocol, eliminating the need to write scripts or modify our applications to use AWS APIs. DataSync enables us to copy data to and from AWS using AWS Direct Connect or internet connections.

The service enables one-time data migrations, automated replication for data protection and recovery, and recurring data processing workflows. DataSync agents need to be installed on-premises, connected to a file system or storage array, and select either Amazon EFS or S3 as the target AWS storage. Cost is calculated depending on the data transferred.

- **AWS Transfer Family**

The AWS Transfer Family manages all aspects of file transfers into and out of Amazon S3.

With support for Secure File Transfer Protocol (SFTP), File Transfer Protocol over SSL (FTPS), and File Transfer Protocol (FTP), the AWS Transfer Family enables us to migrate our file transfer workflows to AWS seamlessly by integrating with existing authentication systems and providing DNS routing through Amazon Route 53, ensuring that our applications remain unaffected. It is simple to get started with the AWS Transfer Family - there is no infrastructure to purchase or set up.

Offline data transfer

- **AWS Snowcone**

The AWS Snowcone is the smallest member of the AWS Snow Family of edge computing, edge storage, and data transfer devices, weighing 2,1 kg and featuring 8 terabytes of usable storage. Snowcone is ruggedized, secure, and designed specifically for operation outside of a traditional data center. Its compact size makes it ideal for tight spaces or situations where portability is required but network connectivity is unreliable. You can run compute applications at the edge and then either ship the device to AWS for offline data transfer or use AWS DataSync to transfer data online from edge locations.

The device employs both hardware and software to provide security that meets even the most strict requirements. It makes use of hardware-based Trusted Platform Modules (TPM) to store device-specific keys in an inaccessible location to software, thereby assisting in ensuring the device's integrity. The Snowcone's data is encrypted using two layers of at-rest encryption, which helps protect the device's data during shipment. AWS Key Management Service (KMS) manages encryption keys, which are never stored on the Snowcone device.

Figure: AWS SnowCone

- **AWS Snowball**

This device is a petabyte-scale data transport solution that leverages secure appliances to move large amounts of data into and out of Amazon Web Services. Snowball overcomes common issues associated with large-scale data transfers, such as high network costs, lengthy transfer times, and security concerns. Snowball data transfer is simple, quick, and secure, and can cost as little as one-fifth the price of high-speed Internet.

Snowball eliminates the need for us to write code or purchase hardware to transfer our data. Simply create a job in the AWS Management Console, and an automatic shipment of a Snowball appliance will occur. Once the appliance arrives, connect it to our local network, download and run the Snowball client, and then use the client to select the file directories to transfer to the appliance. The client will then encrypt and transfer the files at high speed to the appliance. Once the transfer is complete and the appliance is ready to be returned, the E

Ink shipping label will update automatically, and we can monitor the job's status via the Amazon Simple Notification Service (SNS), text messages, or directly in the console.

Snowball protects our data with multiple layers of security, including tamper-resistant enclosures, 256-bit encryption, and an industry-standard Trusted Platform Module (TPM) that ensures both security and complete chain of custody. AWS performs a software erasure of the Snowball appliance once the data transfer job has been processed and verified.

Figure 25: AWS SnowBall

- **AWS Snowmobile**

AWS Snowmobile is a petabyte-scale data transfer service that enables the transfer of massive amounts of data to AWS. We can transfer up to 100 PB per Snowmobile, a ruggedized 15m shipping container pulled by a semi-trailer truck. Snowmobile makes it simple to migrate large amounts of data to the cloud, such as video libraries, image repositories, or even an entire data center. Snowmobile data transfer is secure, quick, and economical.

Following an initial assessment, a Snowmobile will be delivered to our data center and configured by AWS personnel so that it can be accessed as a network storage target. When our Snowmobile arrives on-site, AWS personnel will work with our team to connect it to our local network via a removable, high-speed network switch. Then we can begin transferring high-speed data to the Snowmobile from any number of sources within our data center. After loading our data, the Snowmobile returns to AWS, where it is imported into Amazon S3 or S3 Glacier.

AWS Snowmobile protects our data with multiple layers of security, including dedicated security personnel, GPS tracking, alarm monitoring, 24/7 video surveillance, and an optional escort security vehicle while in transit. All data is encrypted using AWS KMS-managed 256-bit encryption keys, which are designed to ensure the security and complete chain of custody of our data.

Figure 26: AWS SnowMobile

Hybrid cloud storage

- AWS Storage Gateway

AWS Storage Gateway can be used to easily deploy AWS Storage on-premises and enables the capability of connecting and extending our on-premises applications to AWS Storage in a seamless manner, for example migrate from tape libraries to cloud storage, provide cloud storage-backed file shares, and create a low-latency cache for on-premises applications to access data in AWS.

The service supports three distinct gateway types: file gateways, tape gateways, and volume gateways.

- The File Gateway stores file data in Amazon S3 as durable objects or in fully managed file shares via Amazon FSx File Gateway.
- Tape Gateway's virtual tape library (VTL) configuration integrates seamlessly with an existing backup software, allowing for cost-effective tape replacement in Amazon S3 and long-term archiving in S3 Glacier and S3 Glacier Deep Archive.
- Local storage or caching of block volumes is provided by Volume Gateway, along with point-in-time backups in the form of EBS snapshots. Cloud-based recovery is possible for these snapshots.

Figure 27: AWS Storage Gateway (AWS)

- **AWS Direct Connect**

AWS Direct Connect is a dedicated physical connection to AWS data centers, which can be used to accelerate network transfers between their datacenters and AWS datacenters. As it is a physical connection with guaranteed bandwidth and availability it can be used to connect services and transfer data reliably and consistently, bypassing the public internet.

AWS Direct Connect enables us to connect our network to one of the AWS Direct Connect locations via a dedicated network connection. This dedicated connection can be partitioned into multiple virtual interfaces using industry-standard 802.1q VLANs, which enables us to use the same connection to access public resources, such as objects stored in Amazon S3, and private resources, such as Amazon EC2 instances running within an Amazon Virtual Private Cloud (VPC), while maintaining network separation between the public and private environments. Virtual interfaces can be reconfigured at any time to accommodate changing business requirements.

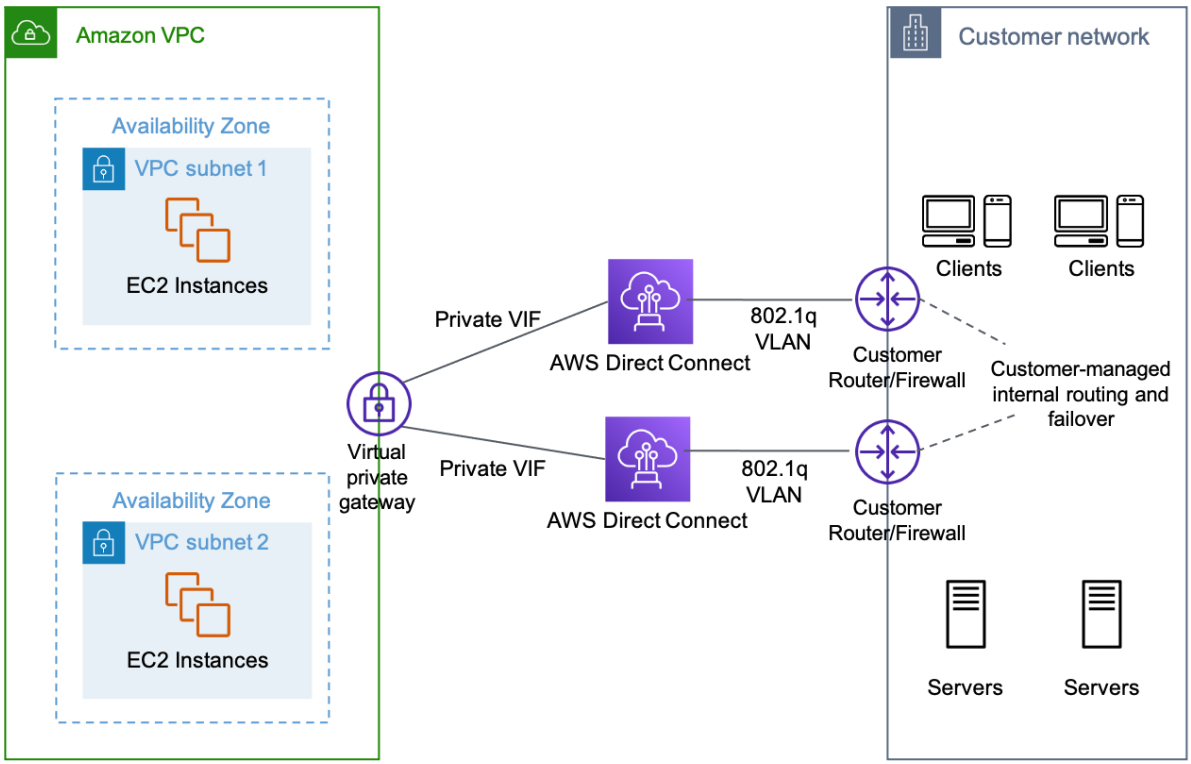


Figure 28: AWS redundant Direct Connect (AWS)

Plan for zero downtime

For mission critical applications that need zero downtime or near-zero downtime a blue/green deployment strategy can be employed after the application and its data has been re-created or copied to the cloud. By having two near identical installations, we can gradually transfer the workload from one installation to the other.

Blue/green deployments enable near-zero-downtime deployments and rollbacks. The underlying concept of blue/green deployment is to route traffic between two identical environments that each run a different version of our application. The blue environment is the current version of the application that is serving production traffic. Simultaneously, a green environment is created and configured to run a different version of our application. Once the green environment has been prepared and tested, production traffic is rerouted from blue to green. If any issues are discovered, we can revert traffic to the blue environment.

After the green deployment (cloud hosted application) has been thoroughly tested, using the cloud provider's DNS service (in our case AWS Route 53), the traffic can be distributed between the blue and the green application sets, gradually, predictably, and with easy and fast rollback.

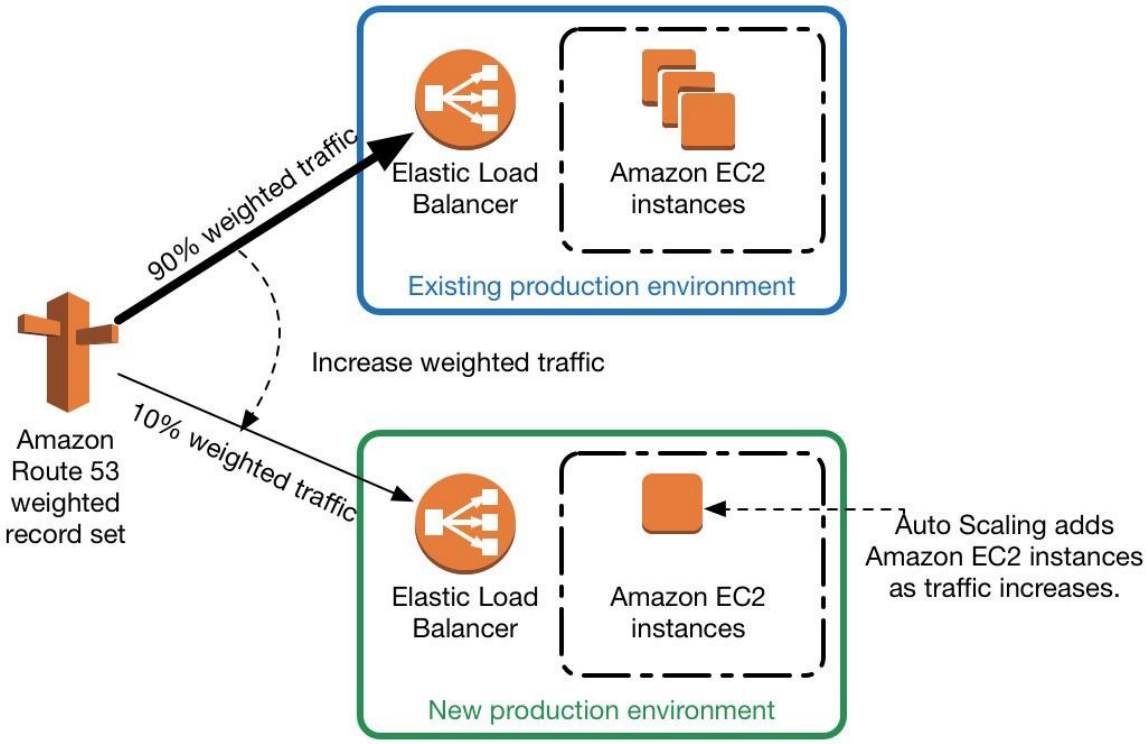


Figure 29: AWS Route 53 green/blue weighted routing (AWS)

Below we can see an example of a blue/green deployment/migration based on the previously presented designs.

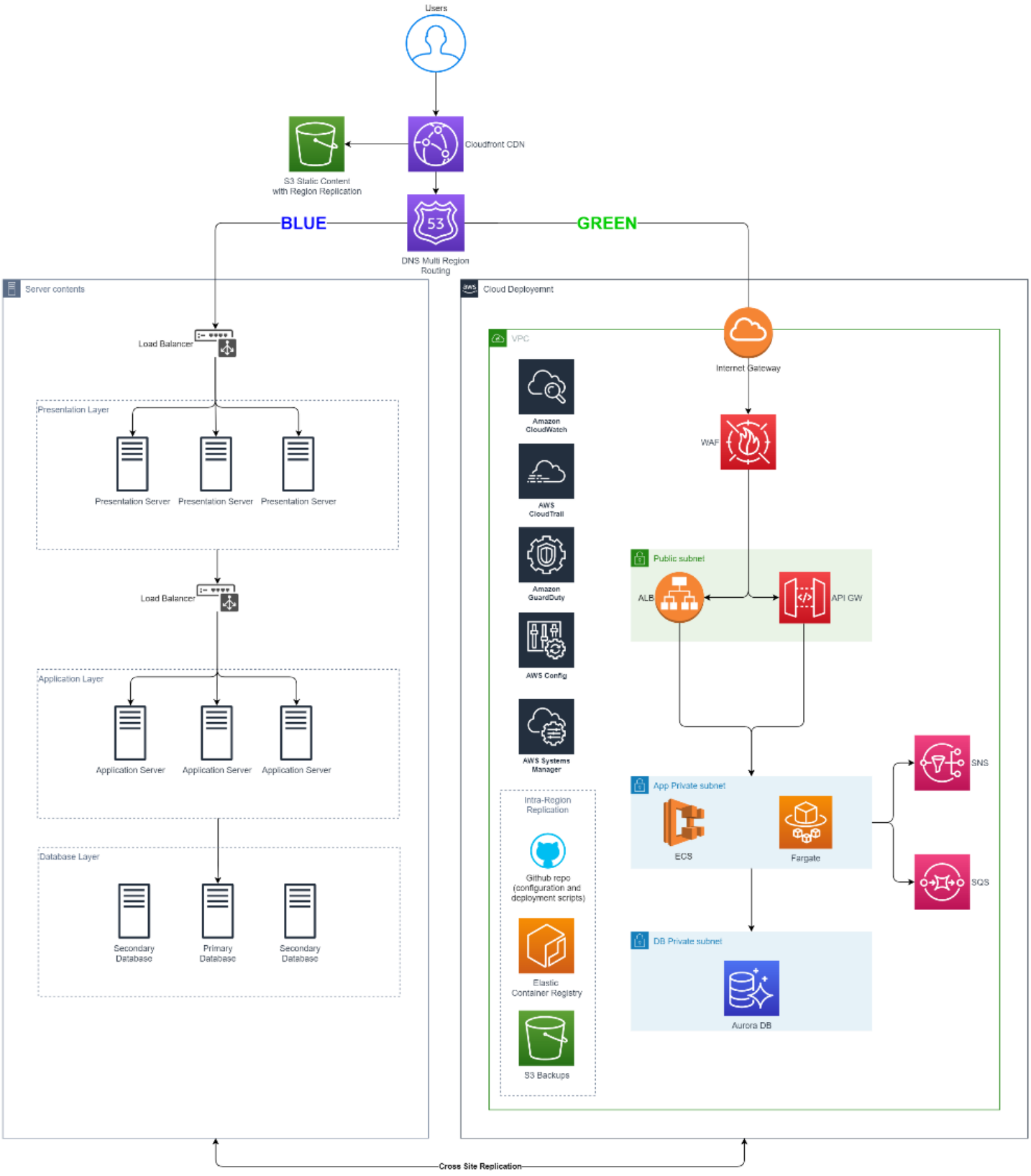


Figure 30: Green/Blue Design

Landing Zones

The term "Landing Zone" refers to a configured environment that includes a standard set of secured cloud infrastructure, policies, best practices, guidelines, and centrally managed services. This enables customers to quickly deploy a secure, multi-account Cloud environment that adheres to industry standards. With a plethora of design options, configuring a multi-account environment can take considerable time, as it involves configuring multiple accounts and services, which requires a thorough understanding of cloud provider services.

This solution can help save time by automating the process of setting up an environment for running secure and scalable workloads while also establishing an initial security baseline through the creation of core accounts and resources.

When creating a landing zone, the following points need to be considered.

1. **Compliance and Security:** A landing zone enables us to enact global and account-level security controls. Defining a security baseline with proactive and detective controls. With landing zones, we can implement company-wide compliance and data residency policies. Consistent architecture is deployed as part of this process to address issues such as edge security, threat management, vulnerability management, and transmission security.

2. **Standardized Tenancy:** A landing Zone provides a framework for establishing and managing a multi-account environment. Automated environment for multi-accounts helps reduce setup time while also establishing an initial security baseline for any digital environment we use. Security, auditing, and shared service requirements are all addressed by the automated multi-account structure. Implementing tagging policies across multiple cloud tenants and provide standardized tenants for various security profiles (dev/staging/prod).

3. **Management of Identity and Access:** Defining roles and access policies in accordance with the principle of least privilege. By defining roles and access policies, we can apply the least privilege principle. SSO for cloud logins should be also implemented.

4. **Networking:** Cloud networking capability design and implementation are critical components of our cloud adoption efforts. Networking is made up of a variety of products and services that each offer a unique set of networking capabilities. Measures should be taken during network implementation to ensure that the network is highly available, resilient, and scalable.

5. **Operations:** Logging centrally from multiple accounts that utilize various cloud provider services. Using a variety of cloud native tools, configure automated backup and disaster recovery. Monitoring and alerting configuration for cost management, reactive scalability, and reliability. Server patching is automated and scheduled.

All of the above should be reflected in the landing zone design and blueprint (IaC implementation) along with the best practices we have presented in a previous chapter, and are more specific to AWS, keeping in mind that all major public cloud providers, support similar services.

Migration Plan

The final Migration Plan will be complex, but it can be summed up to the following steps.

1. Discover Applications, Services and Servers or other assets.
2. Retire any service that is no longer needed or can be replaced by an existing service.
3. Identify Dependencies between the services
4. Prioritize Service Migration according to
    a. Dependencies
    b. Service Criticality
    c. Cloud Readiness
5. Analyze and design the “to be” architecture, estimate cost.
6. Secure Budget.
7. Prepare the target environment on the Landing Zones.
8. Implement network connectivity and test network flows on all environments.
9. Migrate test data and perform SITs/UATs on the test environments.
10. Migrate the real data and set up on-line replication if needed.
11. Perform a zero-downtime migration approach for Mission Critical Applications or schedule a downtime and perform the switch offline for standard applications.
12. Test the new Prod environment while keeping the old environment in case a rollback is needed.
13. In case everything works as expected, backup if needed, and decommission the old environment.

The Cloud preparation steps that should be run in parallel, and be completed before step 6 are as follows.

1. Create the AWS Organizational structure with, at least, a management account and Shared services accounts.

2. Design and implement the security baselines, including:
    a. Integration with the company's Identity provider.
    b. Integration with the company's Logging, monitoring and security software.
    c. Define and implement policies for least privilege access controls. Map existing security groups to AWS Roles.
    d. Define and implement Compliance rules.
    e. Create hardened VM images that will have preinstalled all required agents and controls.

3. Prepare connectivity:
    a. Create a network hub account with redundant connectivity to the local data centers.
    b. Create DNS zones for Cloud use and connect on-premises with Cloud DNS services.
    c. Prepare IP-ranges that can be used on the Cloud but are routable/reachable from the local data centers. These will be used on the network hub only.
    d. Apply policies that will enforce the "deny all by default" networking principle.

4. Create infrastructure templates that will include the security and network baselines and will be used to create the landing zones.

Cost

The migration costs can be split in the following four categories:

1. **Infrastructure costs:** These include the costs of the new environment's infrastructure, while it is being built and evaluated, and until the workload is transferred and the old systems have been decommissioned.

2. **Data Transfer costs:** These include the bandwidth, traffic, and other data transfer services costs, online or offline, that are consumed during the Data migration.

3. **Migration Services costs:** These include any Public Cloud Provider Migration Services that have been used during the Discovery and Migration processes. Most are provided free of

charge by the providers as a means to ease the process of migrating workloads to their cloud services.

4. **Professional Services costs:** these include any professional services, vendors will provide during the applications migration process:

    a. Application preparation
    b. Application upgrades
    c. Application data preparation
    d. New environment preparation
    e. Post migration tests, System Integration Tests (SITs) / User Acceptance Tests (UATs)

Conclusion

"Digital transformation" has become a talking point in recent years, as small and medium-sized businesses as well as large enterprises alike have moved operations to the cloud and embraced online productivity and collaboration tools. (Aggarwal, 2021)

The Covid-19 pandemic changed radically the way people worked by enforcing multiple lockdowns and long-time remote working. This has served as an unexpected reality check for organizations of all sizes all around the world. The pandemic's restrictions turned the world upside down, with industries such as travel and hospitality collapsing and many businesses facing survival challenges.

Due to the uncertainty of the times and the likely realities of the "new normal," an increasing number of organizations have been making short- or long-term plans for their digital transformation which include, in a large part, Cloud adoption or “Cloudification”.

Microsoft CEO Satya Nadella stated just a few months into the pandemic that the company had experienced a two-year digital transformation in two months as a result of its customers' adoption of cloud solutions. (Spataro, 2020)

This has made clear, that the Cloudification is not an ephemeral trend, but a well-established Industry transition on a large scale.

In this thesis we have showed that Cloud Migration on an Enterprise level, while it can be a long-term plan, is not impossible to complete within a reasonable timeline, and it is certainly worth the effort.

The advantages of the Cloud over on-premises and legacy solutions can be summed up as follows:

1. **Flexibility / Elasticity:** There is no need to precisely size and plan ahead. As the cloud is a pay-as-you-go solution, we can increase or decrease sizes or add and remove resources as needed, and still use reserved compute resources to enjoy higher cost savings on our more predictable workloads. We can deploy test environments in a matter of minutes and then destroy them when they are not needed anymore. Services, like storage, can usually be automatically scaled up to sizes that are far beyond the need of most enterprises – Petabytes or even Exabytes. For example, AWS S3 has no limit on how much data you can store.

2. **Cost Savings**: As we have demonstrated in a previous chapter, cloud can be significantly less expensive than on-premises solutions by employing cost saving plans. Cost savings, apart from committing to a minimum compute usage, can be achieved by using elastic scaling and public cloud provider PaaS services. Another beneficial factor of cost savings is Economies of scale which include the cost savings associated with working with larger quantities of a product. When product volumes are higher, the percentage of fixed costs per unit of output is lower, as the costs are spread across a greater number of products.

3. **Security:** By delegating the physical security responsibility to the public cloud provider, we can enjoy a level of security that is very difficult to achieve in medium enterprises. Even a single end user can have a server hosted in a Data Center with multi-layer physical security and biometric scanners, if they choose a large public cloud provider (PCP).
   The security of our services can also be enhanced by using security services that are provided by the PCP and are tightly integrated and optimized for the specific environment. These controls can also be automated for new deployments and allow all services to be "Secured by Design".

4. **Reliability:** Almost all PCP services are hosted and replicated across multiple Data Centers (Availability Zones) within a region, and some can even be replicated across multiple Regions. The durability of some services, like AWS S3, can reach 11 9's (99,99999999%) which is nearly impossible to achieve on premise. Building highly available and disaster resilient services can become very easy on the cloud, as has been demonstrated in the mission critical designs of the previous chapters.

5. **Management / Operations / Governance:** On the cloud many of management, operations and governance tasks can be fully automated. By using PCP services that are closely integrated to the infrastructure, tasks like asset management, compliance, monitoring, alerting and even remediations can be automated. This means that resources can be allocated to other, more productive areas.

Cloud Services have greatly simplified the "Remote Working" concept, by making everyday services accessible from everywhere. The need to be physically present in a company's facilities to access company's services has been greatly reduced, the flexibility of choosing a notebook or workstation other than an enterprise grade, highly secured and very expensive one has been enabled by DaaS, and many other SaaS services have made it possible to work basically wherever there is a modern browser available.

It is indisputable that cloud services have played a role of significant importance during the pandemic, and it is projected that this role will continue for the years to come. Cloud adoption in all Business Areas of an enterprise will drive the industry to the future, while the operating models continue to change towards more Agile and flexible approaches.

The Cloud holds many advantages, and as we have demonstrated is more cost-efficient too.

List of Figures

List of Tables

Bibliography

Aggarwal, G. (2021, 01 15). *How The Pandemic Has Accelerated Cloud Adoption*. Retrieved from Forbes: https://www.forbes.com/sites/forbestechcouncil/2021/01/15/how-the-pandemic-has-accelerated-cloud-adoption/

AWS. (2018). *AWS Migration Whitepaper.* Retrieved from AWS: https://docs.aws.amazon.com/whitepapers/latest/aws-migration-whitepaper/welcome.html

AWS. (2020, July). *AWS Well-Architected Framework.* Retrieved from AWS Documentation: https://docs.aws.amazon.com/wellarchitected/latest/framework/welcome.html

AWS. (2021). *AWS Single Sign-On*. Retrieved from AWS Documentation: https://aws.amazon.com/single-sign-on/

AWS. (2021). *Directory Service*. Retrieved from AWS Documentation: https://aws.amazon.com/directoryservice/

AWS. (2021). *General Data Protection Regulation (GDPR) Center*. Retrieved from AWS Documentation: https://aws.amazon.com/compliance/gdpr-center/

Azeem, S. A., & Sharma, S. K. (2017). Study of Converged Infrastructure & Hyper Converge Infrastructre As Future of Data Centre. *International Journal of Advanced Research in Computer Science, 8*(5), 900-903. Retrieved 6 6, 2021, from http://ijarcs.info/index.php/ijarcs/article/view/3476

Citrix. (2021, 06). *VDI and DaaS*. Retrieved from Citrix Web Site: https://www.citrix.com/solutions/vdi-and-daas/what-is-desktop-as-a-service-daas.html

Cloudflare. (2021). *What is SSO.* Retrieved from Cloudflare website: https://www.cloudflare.com/learning/access-management/what-is-sso/

Comindware. (2021, 06 21). *What is CapEx and OpEx*. Retrieved from Comindware: https://www.comindware.com/blog-what-is-capex-and-opex/

DEI. (2020, March). *DEI Pricelist.* Retrieved from dei.gr: https://www.dei.gr/Documents2/TIMOLOGIA/TIM-MARTIOS-2020/TIMOK-MT-2020-BG-MARCH20full.pdf

European Union. (2021). *What are the GDPR Fines?* Retrieved from European Union GDPR site: https://gdpr.eu/fines/

Forbes. (2020, October 26). *The History And The Future Of Cloud Office Suites*. Retrieved from Forbes: https://www.forbes.com/sites/forbestechcouncil/2020/10/26/the-history-and-the-future-of-cloud-office-suites/

RackSpace. (2020, 08 26). *What is Cloud Native*. Retrieved from RackSpace: https://www.rackspace.com/blog/what-is-cloud-native

Rasmussen, N. (2007, February). *Calculating Total Cooling Requirementsfor Data Centers.* Retrieved from APC: https://www.apcdistributors.com/white-papers/Cooling/WP-25%20Calculating%20Total%20Cooling%20Requirements%20for%20Data%20Centers.pdf

RedHat. (2020). *What is CI/CD*. Retrieved from RedHat: https://www.redhat.com/en/topics/devops/what-is-ci-cd

SDX Central. (2016, 05 18). *What Is Cloud Native? Definition*. Retrieved from SDX Central: https://www.sdxcentral.com/cloud/cloud-native/definitions/what-is-cloud-native-definition/

Spataro, J. (2020, 04 30). *2 years of digital transformation in 2 months* . Retrieved from Microsoft: https://www.microsoft.com/en-us/microsoft-365/blog/2020/04/30/2-years-digital-transformation-2-months/

Wikipedia. (2021, 3 12). *Hyper-converged infrastructure*. Retrieved from wikipedia.org: https://en.wikipedia.org/wiki/Hyper-converged_infrastructure