ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Πρόγραμμα Προπτυχιακών Σπουδών

Ειδίκευση Δικτύων Υπολογιστών και Επικοινωνιών,

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

DISAGGREGATION IN DATA CENTERS

Βαλλάς Ευστάθιος
Α.Μ. 47240

Εισηγητής: Δρ Αντώνιος Μπόγρης, Καθηγητής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

DISAGGREGATION IN DATA CENTERS

Βαλλάς Ευστάθιος
Α.Μ. 47240

Εισηγητής: Δρ Αντώνιος Μπόγρης, Καθηγητής

Antonios Bogris
Digitally signed by Antonios Bogris
Date: 2023.05.03 21:51:18 +03'00'

Εξεταστική Επιτροπή: Δρ Αντώνιος Μπόγρης, Δρ Νίκος Ψαρράς και Δρ Παναγιώτης Καρκαζής

Panagiotis Karkazis
Digitally signed by Panagiotis Karkazis
Date: 2023.05.04 09:23:56 +03'00'

Digitally signed by Nikolaos Psarras
Date: 2023.05.03 23:47:36 +03'00'

Ημερομηνία εξέτασης: 7/10/2022

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Βαλλάς Ευστάθιος του Αναστασίου, με αριθμό μητρώου 47240 φοιτητής/τρια του Προγράμματος Προπτυχιακών Σπουδών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών της Σχολής Μηχανικών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:
«Είμαι συγγραφέας αυτής της προπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.
Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο/Η Δηλών/ούσα

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα εργασία, η οποία ολοκληρώθηκε μετά απο αρκετή έρευνα και επίμονη προσπάθεια, αποτελεί την διπλωματική εργασία του Προπτυχιακού Προγγράματος Σπουδών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών του Πανεπιστημίου Δυτικής Αττικής. Θα ήθελα να ευχαριστήσω εγκαρδίως τον επιβλέποντα καθηγητή κ. Αντώνιο Μπόγρη του οποίου η καθοδήγηση ήταν καταλυτική για την εκπόνηση της εργασίας αλλά και για την άμμεση ανταπόκριση που υπήρχε όταν προέκυπταν εκάστοτε απορίες. Ευχαριστώ επίσης οι οποιοί αποτελούν την εξεταστική επιτροπή. Τέλος θα ήθελα να ευχαριστήσω θερμά τους γόνεις μου Αναστάσιο και Ελένη, τον αδερφό μου Φρατζέσκο, καθώς και τον φίλο μου Νίκο για την πολύτιμη ψυχολογική υποστίρηξη που μου παρείχαν κατά την διάρκεια των σπουδών μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία ασχολείται με το Disaggregation (διαχωρισμό) στα Κέντρα δεδομένων. Αρχικά γίνεται εισαγωγή στις έννοιες του νέφους ώστε ο αναγνώστης να οικειοποιηθεί με κάποιες έννοιες αλλά και να κατανοήσει το νέφος στο σύνολο του. Έπειτα παρουσιάζονται, μελετούνται και αναλύονται δικτυακές τοπολογίες, συσκευές και αρχιτεκτονικές που υπάρχουν στα σύγχρονα Data Centers ενώ παρουσιάζεται και η δομή του εσωτερικού τους. Εν συνεχεία στο τρίτο πλέον κεφάλαιο παρουσιάζεται εις βάθος ο όρος του disaggregation. Αναλύεται και το πλήρες disaggregation αλλα και το μερικό ενώ δίνεται έμφαση στις απαιτήσεις όσον αφορά την CPU, storage και μνήμη. Στο τέταρτο κεφάλαιο αποτυπώνεται πως τα οπτικά δίκτυα είναι το κατάλληλο μέσο προς την σωστή κατεύθυνση για την υλοποίηση των πλήρως διαχωρισμένων κέντρων δεδομένων.Αρχικά παρατείθονται πληροφορίες σχετκά με τα οπτικά δύκτια και τις λειτουργίες τους.Έπειτα γίνεται ανάλυση για τις διάφορες τεχνολογίες και αρχιτεκτονικές που χρησιμοποιούνται, ενώ επιπλέον παρουσιάζονται κάποιες προτάσεις για την επίτευξη του πλήρους διαχωρισμού μέσα απο κάποιες μελέτες και έρευνες. Εν κατακλείδι στο πέμπτο και τελευταίο κεφάλαιο παρατίθονται τα συμπεράσματα της εργασίας και οι πιθανές μελλοντικές εξελίξεις στον τομέα των κέντρων δεδομένων.

ABSTRACT

This thesis deals with Disaggregation in Data Centers. Initially, the concepts of cloud are introduced so that the reader can become familiar with some concepts and also understand the cloud as a whole. Then network topologies, devices and architectures present in modern Data Centers are presented, studied and analyzed and the structure of their interior is presented. Next in the third chapter the term disaggregation is presented in depth. Both full and partial disaggregation are analyzed and the requirements in terms of CPU, storage and memory are emphasized. In the fourth chapter it is illustrated that the optical dumbbells are the appropriate means in the right direction for the realization of fully disaggregated data centers.First, information about optical dumbbells and their functions is presented.Then, an analysis of the different technologies and architectures used is made, and some proposals for achieving full disaggregation are presented through some studies and research. In conclusion, the fifth and last chapter presents the conclusions of the thesis and possible future developments in the field of data centers.

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ ΣΤΟ CLOUD COMPUTING (ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΕΦΟΥΣ)

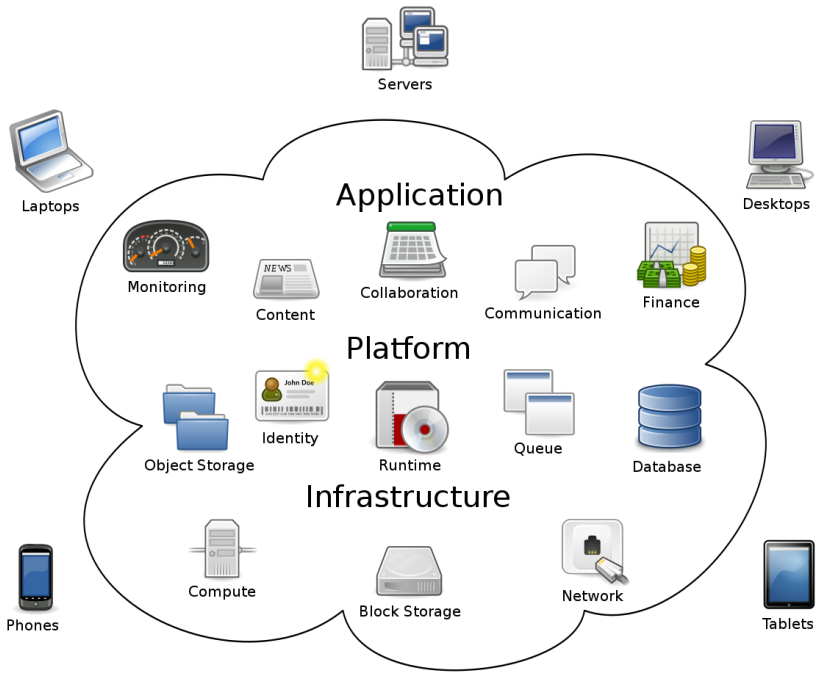
ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό γίνεται εισαγωγή στην υπολογιστική νέφους. Αρχικά παρατίθενται διάφοροί ορισμοί και έννοιες ώστε να γίνει κατανοητή η έννοια του νέφους. Έπειτα παρουσιάζονται τα βασικά χαρακτηριστικά του cloud computing, ενώ γίνεται ενδελεχής αναφορά στην εικονικοποιήση (virtualization). Αναφέρονται και αναλύονται οι υπηρεσίες του νέφους. Τέλος γίνεται αναφορά στις διανομές του νέφους και στον κλάδο του cloudonomics.

1.1 Εισαγωγή στο Νέφος

Σύμφωνα με το NIST (National Institute of Standards and Technology) το ‘cloud computing’ είναι ένα μοντέλο που προσφέρει την από οπουδήποτε, εύκολη (καθολική), κατ’ απαίτηση δικτυακή πρόσβαση σε μία κοινή «δεξαμενή» διαμορφώσιμων υπολογιστικών πόρων που μπορούν να παρασχεθούν γρήγορα και να απενεργοποιηθούν με ελάχιστη προσπάθεια διαχείρισης ή αλληλεπίδραση με τους παρόχους των υπηρεσιών. Επίσης το υπολογιστικό νέφος είναι η παροχή υπολογιστικών πόρων ως υπηρεσία και όχι ως προϊόν, με την οποία οι κοινόχρηστοι πόροι, το λογισμικό και οι πληροφορίες παρέχονται σε υπολογιστές και άλλες συσκευές ως χρησιμότητα μέσω δικτύου. Πρόκειται για έναν τύπο υπολογισμού που βασίζεται στον διαμοιρασμό υπολογιστικών πόρων και όχι στην ύπαρξη τοπικών διακομιστών ή προσωπικών συσκευών για τον χειρισμό εφαρμογών. Η έννοια του cloud computing χρονολογείται από τη δεκαετία του 1960.Το σύννεφο σαν όρος είναι μια μεταφορική έννοια για το διαδίκτυο και με τον τρόπο αυτό απεικονίζεται στα διαγράμματα δικτύων, είναι δηλαδή μια αφηρημένη έννοια για τις υποδομές που κρύβει μέσα του. Θεωρείται σημαντικό να διακρίνετε τον όρο του νέφους και το σύμβολο του από το Internet καθώς υπάρχουν πολλά νέφη τα οποία προσπελαύνονται μέσω του Internet. Ενώ το διαδίκτυο παρέχει ανοιχτή πρόσβαση σε πολλούς πόρους , το νέφος τυπικά είναι ιδιωτικό και προσφέρει μετρήσιμη πρόσβαση σε πόρους. Οι πόροι αυτοί μπορεί να είναι φυσικοί ή

εικονικοί και βασίζονται είτε σε κάποια φυσική συσκευή ή σε κάποιο πρόγραμμα λογισμικού αντιστοίχως. Στην εικόνα 1.1 απεικονίζεται ένα περιβάλλον νέφους, όπου φαίνεται ξεκάθαρα πως μπορεί να χρησιμοποιηθεί το σύμβολο του σύννεφου ώστε να οριοθετήσει τους πόρους οι οποίοι διατίθενται μέσω του νέφους καθώς και το ποιες συσκευές έχουν πρόσβαση στους πόρους αυτούς.



Εικόνα 1.1: Τοπολογία νέφους.
Πηγή:https://en.wikipedia.org/wiki/Cloud_computing

Βασικά Χαρακτηριστικά

Το νέφος διέπετέ εν πολλοίς από τα ακόλουθα βασικά χαρακτηριστικά:

- Πολυχρηστικότητα
- Εύκολη πρόσβαση
- Ευρεία κλιμάκωση
- Ελαστικότητα
- Πληρωμή ανάλογα με την χρήση

- Αυτοκαθορισμός των πόρων

Πολυχρηστικότητα: Οι παρεχόμενες υπηρεσίες πρέπει να υποστηρίζουν διαφορετικές εφαρμογές (πολλαπλότητα εφαρμογών), για διαφορετικούς χρήστες (πολλαπλότητα χρηστών), διάφανα και αποδοτικά. Οι εφαρμογές μπορούν να διαμοιράζονται πόρους αλλά πρέπει να εκτελούνται απομονωμένα και ανεξάρτητα για κάθε χρήστη. Η ανάπτυξη μιας νέας εφαρμογής θα πρέπει να απαιτεί μικρή προσπάθεια.

Εύκολη πρόσβαση: Οι παρεχόμενες υπηρεσίες και εφαρμογές πρέπει να είναι εύκολα προσβάσιμες απ' όλους (και σε όλα τα μέσα), με συνηθισμένους μηχανισμούς και ικανοποιητική ταχύτητα.

Ευρεία Κλιμάκωση: Εταιρείες και οργανισμοί θα μπορούσαν να διαθέτουν πιθανά έως εκατοντάδες ή και χιλιάδες υπολογιστικά συστήματα. Μέσω του cloud παρέχεται η δυνατότητα να ανέλθουν σε δεκάδες χιλιάδες κ.ο.κ. καθώς και η δυνατότητα να αυξηθεί μαζικά το εύρος ζώνης (bandwidth) και ο αποθηκευτικός χώρος.

Ελαστικότητα: Οι εφαρμογές πρέπει να είναι σε θέση να διαπραγματεύονται και να λαμβάνουν επιπλέον πόρους προκειμένου να καλύπτουν τις διαρκώς αυξανόμενες ανάγκες τους σε υπολογιστική ισχύ και αποθηκευτική δυνατότητα
Πληρωμή ανάλογα με την χρήση: Οι χρήστες πληρώνουν μόνο για τους πόρους που χρησιμοποιούν και μόνο για όσο χρόνο τους χρειάζονται.

Αυτοκαθορισμός των πόρων: Οι χρήστες προβλέπουν και ζητούν μόνοι τους τις ανάγκες που έχουν σε πόρους, όπως π.χ. επιπλέον συστήματα (επεξεργαστική ικανότητα, λογισμικό, αποθήκευση), πόρους δικτύου κ.α.

Πρότυπα στο Cloud Computing

Ο κλάδος του υπολογιστικού νέφους εργάζεται με αυτά τα αρχιτεκτονικά πρότυπα:

Εικονικοποιήση πόρων σε πλατφόρμες

Αρχιτεκτονική προσανατολισμένη στις υπηρεσίες

Πλαίσια εφαρμογών ιστού

Ανάπτυξη λογισμικού ανοικτού κώδικα

Τυποποιημένες υπηρεσίες ιστού

Αυτόνομα συστήματα

Υπολογιστική πλέγματος

Αυτά τα πρότυπα συμβάλλουν στην ενεργοποίηση διαφορετικών επιχειρηματικών μοντέλων που μπορούν να υποστηρίξουν οι πωλητές υπολογιστικού νέφους, κυρίως το λογισμικό ως υπηρεσία (SaaS), τις εφαρμογές Web 2.0 και το utility computing. Αυτές οι επιχειρήσεις απαιτούν ανοικτά πρότυπα, ώστε τα δεδομένα να είναι τόσο φορητά όσο και καθολικά προσβάσιμα

Το cloud computing θα γίνει αισθητό με τους ακόλουθους τρόπους τα επόμενα δέκα χρόνια:

Οι εφαρμογές στο νέφος θα αντικαταστήσουν τις εφαρμογές που είναι τοπικές στις συσκευές.

Οι πληροφορίες θα γίνουν φθηνότερες, πιο πανταχού παρούσες και ευκολότερα ανακτήσιμες, επειδή το νέφος καθιστά φθηνότερη την κλιμάκωση των εφαρμογών και των συνδέσεων σε δίκτυα που είναι πάντα σε λειτουργία.

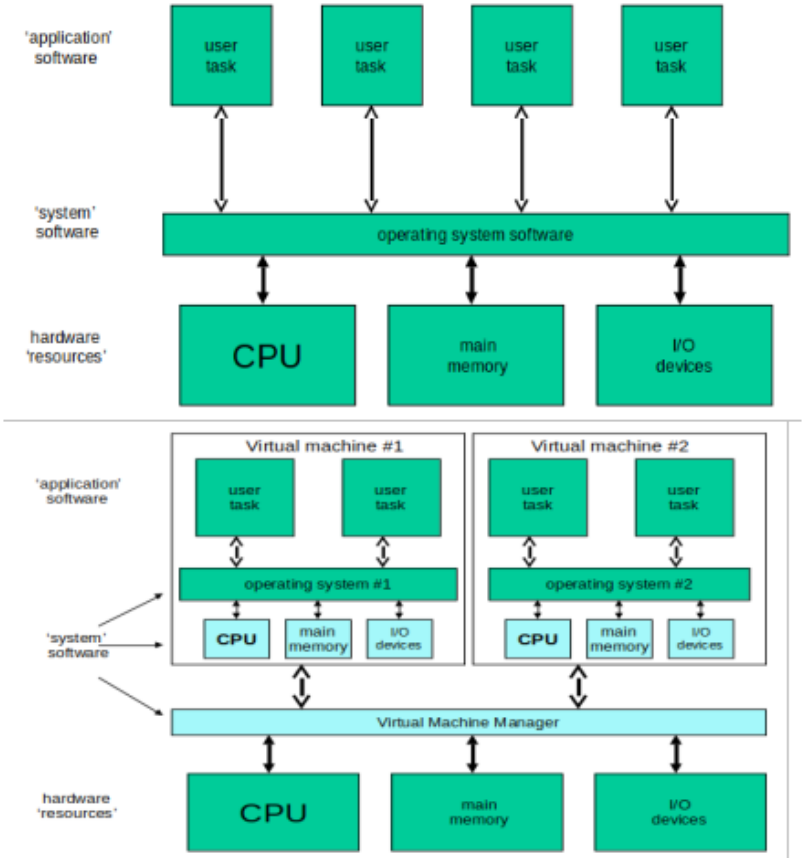
Το cloud θα επιτρέψει νέες κοινωνικές υπηρεσίες συνδέοντας τους χρήστες μέσω κοινωνικών δικτύων που κατασκευάζονται με τη χρήση πολλαπλών υπηρεσιών σύννεφου.

Οι νέες εφαρμογές θα είναι ευκολότερο να δημιουργηθούν.

Θα μειωθεί ο ρόλος που έχουν των λειτουργικών συστημάτων.

Θα υπάρχει σύνδεση μέσω του νέφους όπου κι αν βρίσκεστε και ανά πάσα στιγμή

1.2 Εικονικοποιήση (Virtualization)



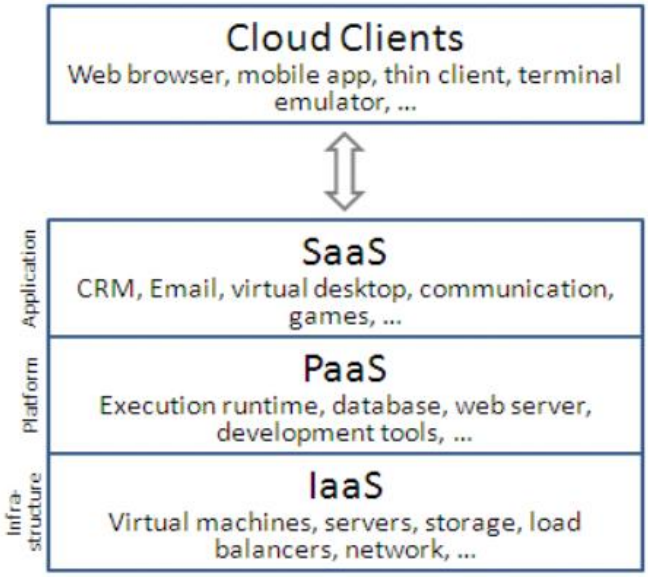
Εικόνα1.2:Τοπολογία πριν και μετά την εικονικοποιήση

Πηγή: Διαφάνειες Υπολογιστικής Νέφους και Υπηρεσίες ΠΑΔΑ

Μία από τις βασικότερες τεχνολογίες πίσω από το cloud computing είναι η 'εικονικοποιήση' (virtualization). Μέσω της εικονικοποιήσης είναι δυνατόν μια φυσική μηχανή να διαχωριστεί σε περισσότερες από μία 'εικονικές' μηχανές (VM – Virtual Machines) οι οποίες μπορούν να λειτουργήσουν εν συνεχεία ανεξάρτητα, με διαφορετικά λειτουργικά συστήματα, και να επιτελούν διαφορετικές εργασίες για διαφορετικούς χρήστες. Προσφέρει μεγαλύτερη αξιοποίηση των φυσικών μηχανών (utilization) αλλά ως αντίτιμο υπάρχει καθυστέρηση εκτέλεσης (overhead). Η εικονικοποιήση χωρίζεται σε δυο κατηγορίες στην πλήρη εικονικοποιήση και στην μερική εικονικοποιήση. Μερικά αντιπροσωπευτικά εργαλεία και υλοποιήσεις: VMware ESX / Workstation, Oracle VM / VirtualBox, Xen, KVM. Μερικά από

τα προτερήματα της εικονικοποιήση είναι η ασφάλεια, οι ευέλικτες λειτουργίες που προσφέρει καθώς εξαλείφει τα προβλήματα διατήρησης ή ανάκτησης χαμένων δεδομένων λόγω κατεστραμμένων ή συντριμμένων συσκευών και, ως εκ τούτου, προάγει την απόδοση της επένδυσης και εξοικονομεί χρόνο. Η ευέλικτη μεταφορά δεδομένων εφόσον δεν υπάρχει περιορισμός στη μεταφορά δεδομένων και μπορούν να μεταφερθούν σε μεγάλη απόσταση με ελάχιστες χρεώσεις. Η εξάλειψη των κινδύνων αποτυχίας του συστήματος αφού κατά την εκτέλεση οποιασδήποτε λειτουργίας, συχνά συμβαίνει το σύστημα να δυσλειτουργεί σε κρίσιμο χρόνο, έτσι ώστε η αποτυχία αυτή του συστήματος να είναι δυσμενής για τους πόρους μιας εταιρείας και να επιδεινώσει τη φήμη της, αυτή η αποτυχία του συστήματος μπορεί να προστατευθεί με την εικονικοποιήση, καθώς οι χρήστες θα μπορούσαν να εκτελούν την ίδια εργασία ταυτόχρονα σε πολλαπλές συσκευές και τα συσσωρευμένα δεδομένα μπορούν επίσης να ανακτηθούν ανά πάσα στιγμή με οποιαδήποτε συσκευή. Ενώ επιπλέον είναι οικονομική.

1.3 Υπηρεσίες Νέφους



Εικόνα 1.3: Μοντέλα Υπηρεσιών Νέφους.

Πηγή: Διαφάνειες Υπολογιστικής Νέφους και Υπηρεσίες ΠΑΔΑ

Το cloud computing προσφέρει τρία είδη υπηρεσιών. Τα διαθέσιμα μοντέλα του cloud computing είναι τα Software-as-as-a-Service, Platform-as-a-Service και Infrastructure-as-a-

Service. Ωστόσο δημιουργείται ένα νέο είδος υπηρεσίας το Software-plus-Services (Λογισμικό συν Υπηρεσίες).

Το Software-as-a-Service βασίζεται στη λογική της υπενοικίασης λογισμικού από έναν πάροχο υπηρεσιών, αντί της αγοράς της άδειας χρήσης. Το λογισμικό λειτουργεί σε ένα κεντρικοποιημένο δίκτυο servers προκειμένου να διατίθεται ως υπηρεσία από το web ή το διαδίκτυο. Επίσης καλείται και ως «software on demand» και αποτελεί τον πλέον γνωστό τύπο cloud computing λόγω της μεγάλης ευελιξίας, ποιότητας υπηρεσιών, υψηλής σταθερότητας και της ελάχιστης δυνατής συντήρησης που απαιτεί. Προσφέρεται στον πελάτη η δυνατότητα χρήσης (και πιθανά παραμετροποίησης) ολοκληρωμένων εφαρμογών λογισμικού, μέσω μόνο ενός browser. Ο παροχέας είναι εδώ αυτός που διαχειρίζεται τόσο την απαιτούμενη υποστηρικτική υποδομή (εξοπλισμό) όσο και το σύνολο των εργαλείων λογισμικού (πλατφόρμα) που απαιτούνται για τη λειτουργία των διατιθέμενων εφαρμογών. Το SaaS μοντέλο είναι πολύ αποτελεσματικό στη μείωση του κόστους αφού παρέχεται στην επιχείρηση ως μηνιαίο λειτουργικό κόστος το οποίο συνήθως είναι κατά πολύ οικονομικότερο από την αγορά των αντίστοιχων αδειών χρήσης και υποδομής. Στο SaaS μοντέλο δεν απαιτείται καμία συντήρηση ή αναβάθμιση, αφού ο τελικός αποδέκτης δε χρειάζεται να μεριμνήσει για τη διαθεσιμότητα, την κλιμάκωση, τη χωρητικότητα και το SLA της υποδομής, της πλατφόρμας και της υπηρεσίας. Μερικές SaaS υπηρεσίες είναι: Exchange Online (ηλεκτρονικό ταχυδρομείο), SharePoint Online (Σύστημα διαχείρισης κειμένων και περιεχομένου) CRM Online, Office Live Meeting(ηλεκτρονικός χώρος συναντήσεων).

Το Platform-as-a-Service παρέχει μια cloud πλατφόρμα εφαρμογών για εταιρείες ή ιδιώτες που κατασκευάζουν λογισμικό είτε για ιδία χρήση είτε για τρίτους. Προσφέρεται δηλαδή στον πελάτη η δυνατότητα χρήσης μίας πλατφόρμας (συνήθως αποτελούμενης από λειτουργικό σύστημα, περιβάλλον χρήσης γλώσσας προγραμματισμού, σύστημα διαχείρισης βάσεων δεδομένων, web-εξυπηρετητή, κ.α.) με δυνατότητες/εργαλεία ανάπτυξης ολοκληρωμένων εφαρμογών. Ενώ και σε αυτήν την κατηγορία ο παροχέας διαχειρίζεται την απαιτούμενη υποδομή στην οποία λειτουργεί η πλατφόρμα και την απαιτούμενη διασύνδεση. Το μοντέλο αυτό παρέχει τις κατάλληλες υπηρεσίες προκειμένου κάποιος να μπορέσει να αναπτύξει, να δοκιμάσει, να διαθέσει και να συντηρήσει εφαρμογές και υπηρεσίες μέσα ένα ενιαίο περιβάλλον πλατφόρμας το οποίο είναι εγγενώς υψηλά διαθέσιμο, ελαστικό και ευέλικτο, με δυνατότητες πλήρης αυτοδιαχείρισης, αυτό-συντήρησης και αυτό-κλιμάκωσης της υποδομής, του λειτουργικού συστήματος και της πλατφόρμας εφαρμογών. Το PaaS

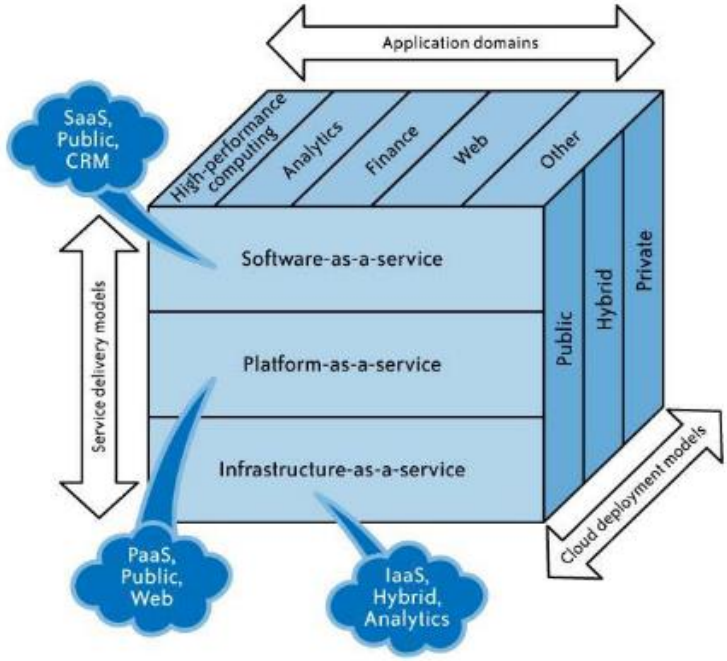
βασίζεται στο μοντέλο «Pay-per-use» με τέτοιο τρόπο έτσι ώστε να επιτυγχάνεται η πλήρης αξιοποίηση των υπολογιστικών πόρων που χρησιμοποιούνται σε σχέση με το κόστος χρήσης. Αν συνδυαστεί με το χαρακτηριστικό της αυτόκλιμάκωσης μπορούμε να πετύχουμε τη διάθεση υπηρεσιών που να μπορούν να ανταποκρίνονται σε οποιαδήποτε ραγδαία ή αναμενόμενη μεταβολή χωρητικότητας (ισχύς, μνήμη, αποθηκευτικό χώρο, δίκτυο) που θα απαιτηθεί ανά πάσα χρονική στιγμή χωρίς να έχω δεσμευτεί εκ των προτέρων είτε με αγορά υποδομής, λογισμικού πλατφόρμας, δικτυακή γραμμή υψηλής χωρητικότητας κλπ. είτε με ένα συμβόλαιο παροχής υπηρεσιών φιλοξενίας υποδομής και πλατφόρμας συγκεκριμένης χωρητικότητας και χρονικής διάρκειας. Παραδείγματα υπηρεσιών PaaS είναι: Google App Engine, Microsoft Azure, Amazon AWS.

Το τρίτο μοντέλο είναι το Infrastructure-as-a-Service το οποίο είναι ηπαροχή υπολογιστικών και δικτυακών υποδομών ως μια πλήρως outsourced υπηρεσία. Η εταιρεία ή ο ιδιώτης μπορεί να υπενοικιάσει υποδομή (όχι όμως και πλατφόρμα όπως στο PaaS) ανάλογα με τις απαιτήσεις εκείνης της χρονικής στιγμής με λογική, όπως και στο PaaS, «Pay as you go» αντί να προβεί στην αγορά εξοπλισμού (υπολογιστικού, δικτυακού, κλπ) ή στη σύναψη συμβολαίου παροχής υπηρεσιών φιλοξενίας υποδομής για συγκεκριμένο χρονικό διάστημα. Σημαντικό πλεονέκτημα του IaaS είναι επίσης η δυνατότητα μεταφοράς εικονικών μηχανών από το ιδιόκτητο περιβάλλον της εταιρείας ή του ιδιώτη στο cloud, με συνοπτικές διαδικασίες. Προσφέρεται στον πελάτη η δυνατότητα χρήσης της υπολογιστικής υποδομής (υπολογιστικού εξοπλισμού) του παροχέα (συνήθως virtual machines). Ένας άλλος όρος είναι ‘Environment as a Service’ (EaaS). Εκτός από υπολογιστική ισχύ, και μνήμη προσφέρονται συνήθως και άλλα συστατικά υποδομής, όπως π.χ.: storage (διαφόρων μορφών), networking, firewalls, IP addresses, VLANs. Μερικές IaaS υπηρεσίες είναι: Amazon (Elastic Compute Cloud - EC2), GoGrid, GCE, RAX, IBM, MS Azure, Alibaba, Okeanos (cyclades).

Η υπηρεσία Software+Services (Λογισμικό συν υπηρεσίες), βρίσκεται ακόμα σε πρώιμο στάδιο. Αυτό που προσφέρει ο πάροχος είναι οι επιπρόσθετες υπηρεσίες σε διασύνδεση με το λογισμικό του χρήστη μέσω του cloud. Ο χρήστης ουσιασικά χρησιμοποιεί τοπικά το δικό του λογισμικό ή εφαρμογή και χρησιμοποιεί από τον παροχέα πρόσθετες υπηρεσίες. Οι λόγοι που ένα άτομο ή μια εταιρία μπορεί να επιλέξει αυτού του είδους την υπηρεσία είναι η καλύτερη απόδοση που προσφέρει , ή για απλή συμπληρωματικότητα, παρέχει επιπλέον απόλυτη μυστικότητα. Κάποιες S+S εφαρμογές είναι οι Microsoft strategy (κυρίως στα

πλαίσια των cloud λύσεων ERP, CRM αλλά και γενικότερα) Google APIs (libraries για Java, Python) Google Android, iPhone SDK.

1.4 Μοντελα Διανομης Cloud



Εικόνα 1.4: Συσχέτιση υπηρεσιών και μοντέλων του νέφους.
Πηγή: Διαφάνιες Υπολογιστικής Νέφους και Υπηρεσίες ΠΑΔΑ

Public Cloud (Δημόσιο Σύννεφο)

Ο όρος δημόσιο cloud περιγράφει το cloud computing υπό μία γενική τάση, όπου οι πόροι παρέχονται δυναμικά, σε μια αυτόεξυπηρετούμενη βάση μέσω του διαδικτύου, μέσω web εφαρμογών ή web υπηρεσιών, από έναν εξωτερικό τρίτο πάροχο που μοιράζει τους πόρους και χρεώνει σύμφωνα με την χρήση. Ο τρόπος με τον οποίο λειτουργούν τα public clouds είναι με διαχωρισμό των τμημάτων τους για την αποκλειστική χρήση ενός μόνο πελάτη,δημιουργώντας ένα εικονικό ιδιωτικό κέντρο δεδομένων Έτσί οι χρήστες έχουν μια ολοκληρωμένη εντύπωση στην υποδομή του με αποτέλεσμα να μπορούν να ελέγξουν όχι μόνο τις εικόνες της εικονικής μηχανής, αλλά και τους διακομιστές, τα αποθηκευτικά συστήματα, τις συσκευές δικτύου και την τοπολογία του δικτύου. Επιπλέον η δημιουργία ενός εικονικού ιδιωτικού κέντρου δεδομένων βοηθά να στην ελάττωση της τοπικότητας των

δεδομένων, καθώς υπάρχει άφθονο και δωρεάν εύρος ζώνης έφοσον η σύνδεση γίνεται ανάμεσα στους πόρους της ίδιας μονάδας.

Private Cloud (Ιδιωτικό Σύννεφο)

Ο όρος ιδιωτικό cloud είναι όρος που χρησιμοποιείται για να περιγράψει προϊόντα και υπηρεσίες που έχουν σαν σκοπό να εξομοιώσουν το cloud computing στα ιδιωτικά δίκτυα. Τα private clouds κατασκευάζονται για την αποκλειστική χρήση από έναν πελάτη, παρέχοντας τον μέγιστο έλεγχο των δεδομένων, την ασφάλεια και την ποιότητα των υπηρεσιών. Ως επι το πλείστον τα private clouds μπορούν να αναπτυχθούν σε ένα κέντρο δεδομένων μιας επιχείρησης και επίσης μπορούν να αναπτυχθούν στις ίδιες κτηριακές εγκαταστάσεις. Ωστόσο υπάρχουν κάποιες κατηγορίες ιδιωτικού cloud στις οποίες διαφοροποιούνται οι προδιαγραφές όσον αφορά τις εγκαταστάσεις.

Αφιερωμένο(dedicated)

Το ιδιωτικό cloud φιλοξενείται σε data center που ανήκει στον χρήστη και διαχειρίζεται από τα εσωτερικά τμήματα τεχνολογίας πληροφορικής.

Κοινότητας(Community)

Το ιδιωτικό cloud βρίσκεται στις εγκαταστάσεις ενός τρίτου, ανήκει, διαχειρίζεται και λειτουργεί από έναν πωλητή, ο οποίος μπορεί να διαθέτει αντίστοιχες υπηρεσίες και σε άλλους πελάτες. Ο πωλητής περιορίζεται από τον πελάτη με συμφωνητικά για το επίπεδο των υπηρεσιών (SLAs)[1].

Διαχειριζόμενο(managed)

Η υποδομή του ιδιωτικού cloud ανήκει στον πελάτη και διαχειρίζεται από έναν πωλητή.

Hybrid Cloud (Υβριδικό Σύννεφο)

Το Hybrid cloud ουσιαστικά συνδυάζει τόσο τα public όσο και τα private μοντέλα σύννεφων και μπορεί να αποτελείται από πολλαπλούς εσωτερικούς ή εξωτερικούς παρόχους. Για παράδειγμα ένας οργανισμός ο οποίος χρησιμοποιεί υβριδικό cloud μπορεί να τρέχει μη κρίσιμες εφαρμογές σε ένα δημόσιο cloud (public cloud), ενώ να διατηρούν τις κρίσιμες εφαρμογές και τα ευαίσθητα δεδομένα εντός του οργανισμού σε ένα ιδιόκτητο δίκτυο (private cloud).

Community Cloud (Κοινοτικό Σύννεφο)

Σε αυτήν τη διανομή, η υποδομή του cloud μοιράζεται μεταξύ πολλών οργανισμών και υποστηρίζει μια συγκεκριμένη κοινότητα που έχει κοινές ανησυχίες (π.χ. απαιτήσεις ασφαλείας, πολιτική και θέματα συμμόρφωσης). Η διαχείρισή της μπορεί να γίνεται από έναν εκ των οργανισμών ή από τρίτους και μπορεί να βρίσκεται εντός ή εκτός των εγκαταστάσεων των εκάστοτε οργανισμών.Ένα κοινοτικό νέφος βρίσκεται μεταξύ δημόσιου και ιδιωτικού νέφους όσον αφορά το δίκτυο και το σύνολο των καταναλωτών-στόχων. Είναι κάπως παρόμοιο με ένα ιδιωτικό νέφος, αλλά η υποδομή και οι υπολογιστικοί πόροι ανήκουν αποκλειστικά σε δύο ή περισσότερους οργανισμούς που έχουν κοινά ζητήματα προστασίας της ιδιωτικής ζωής, ασφάλειας και κανονιστικών ρυθμίσεων. Επίσης προσφέρει λιγότερο κόστος έναντι των κοινών πόρων.

1.5 Cloudonomics (Νεφονομία)

Η νεφονομία είναι ένας νεοσύστατος όρος και κλάδος που ιδρύθηκε από τον συγγραφέα (Weinman, 2008). Αυτό που επιδιώκει είναι να παρέχει μια αυστηρή βάση βασισμένη στον λογισμό, τη στατιστική, την τριγωνομετρία, τη δυναμική των συστημάτων, τα οικονομικά και τη θεωρία της υπολογιστικής πολυπλοκότητας, η οποία μπορεί να χρησιμοποιηθεί για την ερμηνεία των εμπειρικών αποτελεσμάτων. Ουσιαστικά ορίζει το νέφος απο οικονομική άποψη ώστε να γίνει αντιληπτό στις επιχειρήσεις. Υπάρχουν δέκα βασικοί κανόνες οι διέπουν τον κλάδο αυτο.

1. Οι υπηρεσίες κοινής ωφέλειας κοστίζουν λιγότερο, παρόλο που κοστίζουν περισσότερο.

2. Η ζήτηση ξεπερνά την πρόβλεψη.

3. Η κορυφή του αθροίσματος δεν είναι ποτέ μεγαλύτερη από το άθροισμα των κορυφών.

4. Η συνολική ζήτηση είναι πιο ομαλή από την ατομική.

5. Το μέσο μοναδιαίο κόστος μειώνεται με την κατανομή του σταθερού κόστους σε περισσότερες μονάδες παραγωγής.

6. Η αριθμητική υπεροχή είναι ο σημαντικότερος παράγοντας για το αποτέλεσμα μιας μάχης.

7. Ο χωροχρόνος είναι ένα συνεχές.

8. Η διασπορά είναι το αντίστροφο τετράγωνο της καθυστέρησης.

9. Μην βάζετε όλα τα αυγά σας σε ένα καλάθι.

10. Ένα αντικείμενο σε ηρεμία τείνει να παραμείνει σε ηρεμία.

ΚΕΦΑΛΑΙΟ 2

ΔΙΚΤΥΑ ΣΕ DATACENTERS

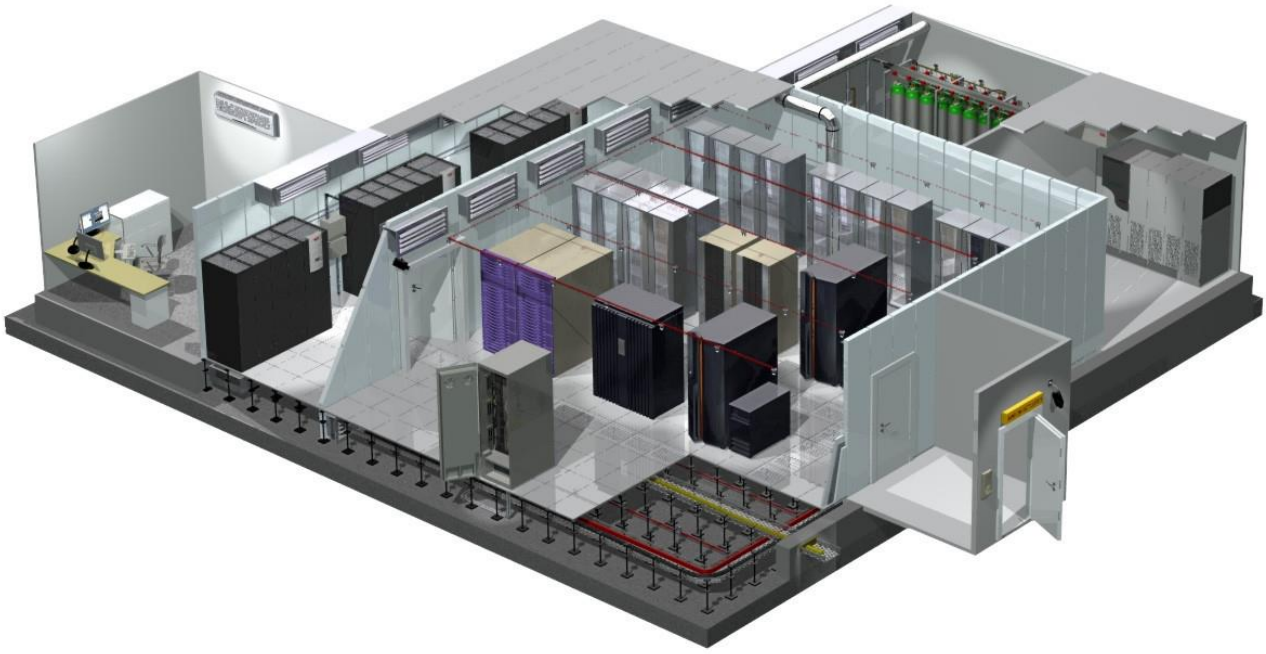
ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο παρουσιάζονται, μελετούνται και αναλύονται δικτυακές τοπολογίες, συσκευές και αρχιτεκτονικές που υπάρχουν στα σύγχρονα Data Centers. Αρχικά γίνεται μια ιστορική αναδρομή για το πως δημιουργήθηκαν τα πρώτα Data Centers και περιγράφεται η δομή τους. Έπειτα γίνεται αναφορά στα δομικά στοιχεία των servers και στα switches μαζί με την χρήση του 400G Ethernet. Τέλος περιγράφονται και αναλύονται οι διάφορες τοπολογίες διασύνδεσης.

2.1 Τα πρώτα Data Centers

Η εξάπλωση των υπολογιστικών και πληροφοριακών συστημάτων στις αρχές του 2000, οδήγησε πολλές επιχειρήσεις στην μεταφορά τον υπηρεσιών τους στο Internet μάζι με την δημιουργία ιστοσελίδων. Όμως με την πάροδο του χρόνου και την εδραίωση συναλλαγών μέσω διαδικτύου οι ανάγκες των επιχειρήσεων όσον αφορά τους υπολογιστικούς πόρους αυξήθηκαν. Μια εταιρία παραγωγής, για παράδειγμα ενώ αρχικά μπορούσε να ανταπεξέλθει στις υπολογιστικές ανάγκες της με την εγκατάσταση ενός ή μερικών server και ενός σχετικά μικρού IT τμήματος, αργότερα αντιμετώπιζε πρόβλημα καθώς χρειαζόταν πολλούς παραπάνω servers και ουσιαστικά ένα μεγαλύτερο κατα πολύ τμήμα IT. Η υλοποιήση αυτών των απαιτήσεων είναι εξαιρετικά δαπανηρή για μια επιχείρηση/εταιρία. Επομένως καθώς υπήρξε η ανάγκη για τεράστιους υπολογιστικούς πόρους δημιουργήθηκαν τα Data Centers. Σύμφωνα με την Cisco “Στην απλούστερη μορφή του, ένα κέντρο δεδομένων είναι μια φυσική εγκατάσταση που χρησιμοποιούν οι οργανισμοί και επειχηρήσεις για να στεγάσουν τις κρίσιμες εφαρμογές και τα δεδομένα τους. Ο σχεδιασμός ενός κέντρου δεδομένων βασίζεται σε ένα δίκτυο υπολογιστικών και αποθηκευτικών πόρων που επιτρέπουν την παροχή κοινών εφαρμογών και δεδομένων. Τα βασικά στοιχεία του σχεδιασμού ενός κέντρου δεδομένων περιλαμβάνουν δρομολογητές, μεταγωγείς, τείχη προστασίας, συστήματα αποθήκευσης, διακομιστές και ελεγκτές παροχής εφαρμογών.” Στην Εικόνα 2.1

φαίνεται πως μοιάζει ενα σχετικά μικρό Data Center ώστοσο πρέπει να σημειωθεί πως οι εγκαταστάσεις που φιλοξενούν τα Dcs πρέπει να είναι κατάλληλα διαμορφωμένες. Ουσιαστικά πρέπει παρέχουν έναν τεχνικό χώρο προετοιμασμένο με υπερυψωμένο δάπεδο κάτω από το οποίο εγκαθίστανται ηλεκτρικές πρίζες για τη σύνδεση των racks. Έτσι, όλα πρέπει να ελέγχονται και να εξασφαλίζονται προσεκτικά. Για το σκοπό αυτό, τα κέντρα δεδομένων μεγάλης κλίμακας εφαρμόζουν πολυάριθμους μηχανισμούς ασφαλείας: εφεδρικά συστήματα παροχής ηλεκτρικής ενέργειας, εφεδρικές γεννήτριες ντίζελ, εφεδρικά και πολύ αποδοτικά συστήματα ψύξης, ανίχνευση και κατάσβεση πυρκαγιάς, ανιχνευτές διαρροής νερού και ελέγχους ασφαλείας.



Εικόνα 2.1: Data Center
Πηγή:https://www.signalscontrol.com/dcm.html

Μεταξύ του 2003 και του 2010, τα virtual DC’s δημιουργήθηκαν καθώς η επανάσταση της virtual τεχνολογίας κατέστησε δυνατή τη συγκέντρωση των πόρων υπολογιστών, δικτύου και αποθήκευσης από διάφορα πρώην απομονωμένα κέντρα δεδομένων για τη δημιουργία ενός κεντρικού, πιο ευέλικτου πόρου που θα μπορούσε να ανακατανεμηθεί ανάλογα με τις ανάγκες. Ένα virtual data center εξαλείφει πολλές από τις προκλήσεις των παραδοσιακών κέντρων δεδομένων, καθώς αυξάνει την ευελιξία του IT και μειώνει την πολυπλοκότητα και το κόστος. Είναι επίσης γνωστά ως ιδιωτικά νέφη, κέντρα δεδομένων νέφους ή κέντρα

δεδομένων καθορισμένα από λογισμικό (SSDC). Μια φυσική τοπολογία Data Center μπορεί να φιλοξενήσει πολλά virtual DC’s με το καθένα απο αυτά να είναι ανεξάρτητο εξασφαλίζοντας μέγιστη ασφάλεια, υψηλή διαθεσιμότητα και ευελιξία.

2.2 Data Center Components

Τα δομικά στοιχεία στα σύγχρονα DC’s ειναι τα τρία ακόλουθα: servers, switches και routers. Αυτά βρίσκονται εγκατεστημένα σε ειδικές καμπίνες και συνεργατικά διαχειρίζονται τον απαιτούμενο όγκο δεδομένων.

2.2.1 Servers

Υπάρχουν δυο κατηγορίες στην δομή των servers στα data centers, οι rack servers και server blades. Η κύρια διαφορά μεταξύ τους είναι ότι ένας rack server είναι ανεξάρτητος, ενώ οι blade servers πρέπει να συνεργάζονται μεταξύ τους.

Rack Server

Ουσιαστικά πρόκειται για servers γενικής χρήσης οι οποίοι μπορούν να διαμορφωθούν κατάλληλα ώστε να υποστιρίζουν ένα ευρύ φάσμα απαιτήσεων. Στα DC’s οι rack servers βρίσκονται μεσα στα server racks όπως φαίνεται στην εικόνα 2.2. Στην δομή του διαφέρει απο έναν παραδοσιάκο server καθώς είναι πιο φαρδύς. Προσφέρει αυτονομία, καθώς διαθέτει τα απαραίτητα εξαρτήματα, ενώ μπορεί να υποστιρίξει και μεγάλο αριθμό RAM. Επίσης έχει εύκολη τοποθέτηση και ψύξη.

Εικόνα 2.2: Server Rack

Πηγή: https://www.turbosquid.com/3d-models/server-rack-data-center-3d-model-1366707

Blade Serves

Ο blade server είναι ουσιαστικά modular server που επιτρέπει τη στέγαση πολλαπλών server σε μικρότερο χώρο. Η φυσική τους σχεδίαση είναι πιο λεπτή και συνήθως διαθέτουν μόνο CPU, μνήμη, ενσωματωμένους ελεγκτές δικτύου και μερικές φορές δίσκους αποθήκευσης. Αντί για τα server racks ενσωματώνονται στα blade carriers όπως απεικονίζεται στην εικόνα 2.3.Λόγω της ικανότητάς τους να χωρούν τόσους πολλούς διακομιστές σε ένα μόνο ράφι παρέχουν υψηλή επεξεργαστική ισχύ. Βασικό χαρακτηριστικό τους είναι πως μειώνουν την συνολική κατανάλωση καθώς ένα Server Blade τροφοδωτεί πολλούς servers. Επίσης δεν απαιτείται χρήση πολλών καλωδίων, καθώς οι blade servers μπορούν να έχουν ένα καλώδιο συχνά οπτικών ινών. Επιπλέον οι Blade Servers καταλαμβάνουν ελάχιστο χώρο ενώ μπορούν να αντικατασταθούν εν θερμώ, έτσι ώστε εάν ένα blade έχει πρόβλημα, να μπορεί να αφαιρεθεί και να αντικατασταθεί πολύ πιο εύκολα.

Εικόνα 2.3: Blade Server

Πηγή:https://www.turbosquid.com/3d-models/3d-photoreal-blade-server-computer-1169298

2.2.2 Switch

Το switch που χρησιμοποιείται στα DC’s είναι απλά ένα switch υψηλών επιδόσεων κυρίως για μεγάλες επιχειρήσεις και παρόχους cloud που βασίζονται σε μεγάλο βαθμό στην εικονικοποίηση. Μπορεί να αναπτυχθεί σε όλο το Data Center ή να αγκυροβολήσει τοπικά

με μια αρχιτεκτονική επίπεδου πλέγματος ή υφάσματος δύο επιπέδων (leaf-spine). Αυτού του είδους τα switches διέπονται απο τα εξής χαρακτηριστικά.

Μπορούν να χειριστούν τόσο north-south ροές κυκλοφορίας όσο και east-west.

Είναι συμβατοί με και με τις δύο αρχιτεκτονικές που χρησιμοποιούνται (top-of-rack (ToR) και end-of-row (EoR)).

Υποστηρίζουν διασυνδέσεις υψηλού εύρους ζώνης χρησιμοποιώντας τόσο το πρωτόκολλο LAN Ethernet όσο και τα πρωτόκολλα SAN.

Διαθέτουν εκτεταμένα συστήματα υψηλής διαθεσιμότητας και ανοχής σφαλμάτων στο υλικό και στο λογισμικό. Με αποτέλεσμα να παρέχουν καλύτερο χρόνο διαθεσιμότητας για κρίσιμες εφαρμογές.

Είναι επίσης φιλικό προς τον χρήστη καθώς όλα τα στοιχεία ενός κατανεμημένου switch μπορούν να διαχειριστούν από ένα ενιαίο περιβάλλον διαχείρισης.

2.3 Δικτύωση οριζόμενη απο λογισμικό SDN

Το Software-Defined Networking (SDN) είναι ένας θεμελιώδης τρόπος προγραμματισμού των switches που χρησιμοποιούνται στα σύγχρονα κέντρα δεδομένων. Ουσιαστικά πρόκειται για μια δυναμική αρχιτεκτονική που είναι, διαχειρίσιμη, οικονομικά αποδοτική και προσαρμόσιμη, καθιστώντας έτσι κατάλληλη για τη δυναμική φύση των σημερινών εφαρμογών που απαιτούν υψηλό εύρος ζώνης. Αυτή η αρχιτεκτονική αποσυνδέει τις λειτουργίες ελέγχου και προώθησης επιτρέποντας στον έλεγχο του δικτύου να γίνει άμεσα προγραμματιζόμενος. Το πρωτόκολλο OpenFlow® αποτελεί θεμελιώδες στοιχείο για την οικοδόμηση λύσεων SDN. Τα οφέλη του SDN ποικίλουν και καταδεικνύουν την χρησιμότητα του:

<u>Άμμεσα προγραμματιζόμενο</u>

Ο έλεγχος του δικτύου είναι άμεσα προγραμματιζόμενος επειδή είναι αποσυνδεδεμένος από τις λειτουργίες προώθησης.

<u>Κεντρικά διαχειριζόμενο</u>

Η ευφυΐα του δικτύου είναι συγκεντρωμένη σε ελεγκτές SDN βασισμένους σε λογισμικό οι οποίοι διατηρούν μια συνολική εικόνα του δικτύου, η οποία εμφανίζεται στις εφαρμογές και τις πολιτικές ως ένα ενιαίο, λογικό switch.

<u>Βασισμένο σε ανοιχτά πρότυπα</u>

Εφόσον υλοποιείται μέσω ανοικτών προτύπων, το SDN απλοποιεί το σχεδιασμό και τη λειτουργία του δικτύου, καθώς οι οδηγίες παρέχονται από τους ελεγκτές SDN αντί για πολλαπλές συσκευές και πρωτόκολλα συγκεκριμένων κατασκευαστών.

Διαμορφωμένο προγραμματιστικά

Το SDN επιτρέπει στους διαχειριστές δικτύου να διαμορφώνουν, να διαχειρίζονται, να διασφαλίζουν και να βελτιστοποιούν τους πόρους του δικτύου πολύ γρήγορα μέσω δυναμικών, αυτοματοποιημένων προγραμμάτων SDN, τα οποία μπορούν να δημιουργήσουν οι ίδιοι, καθως τα προγράμματα δεν εξαρτώνται από ιδιόκτητο λογισμικό.

Δυναμικό

Η αφαίρεση του ελέγχου από την προώθηση επιτρέπει στους διαχειριστές να προσαρμόζουν δυναμικά τη ροή της κυκλοφορίας σε όλο το δίκτυο ώστε να ανταποκρίνονται στις μεταβαλλόμενες ανάγκες.

2.4 400G Ethernet

Στα περισσότερα σύγχρονα κέντρα δεδομένων χρησιμοποιείται 100G Ethernet, ωστόσο με την άνοδο εφαρμογών υψηλού ρυθμού δεδομένων, όπως το 5G και το cloud computing, το 200G και το 400G Ethernet γίνονται οι πλέον βέλτιστες επιλογές. Είναι σημαντικό όμως να σημειωθεί πως λόγω των αναγκών της αγοράς, το πρότυπο 400G είναι πιο ολοκληρωμένο από αυτό του 200G. Μέχρι στιγμής υπάρχουν δύο τύποι 400G. Πομποδέκτες με βάση το PAM4 που δημιουργήθηκαν βάσει προδιαγραφών IEEE 802.3bs και αποτελούν καλύτερη επιλογή για το εσωτερικό των DC's. Το IEEE 802.3bs προσφέρει ένα ευρύ φάσμα προδιαγραφών για διαφορετικές απαιτήσεις Ethernet:

400GBASE-SR16, η οποία καλύπτει τουλάχιστον 100 μέτρα σε πολύτροπη ίνα μέσω 16 ινών εκπομπής και άλλων 16 ινών λήψης, καθεμία από τις οποίες εκπέμπει με ταχύτητα 25 Gbps.

400GBASE-DR4, για τουλάχιστον 500 m μέσω μονότροπης ίνας με χρήση τεσσάρων παράλληλων ινών σε κάθε κατεύθυνση με μετάδοση 100 Gbps σε κάθε ίνα.

400GBASE-FR8, το οποίο χρησιμοποιεί WDM οκτώ μηκών κύματος για την αντιμετώπιση αποστάσεων τουλάχιστον 2 km πάνω από μια μονότροπη ίνα σε κάθε κατεύθυνση.

400GBASE-LR8, το οποίο είναι παρόμοιο με το -FR8, εκτός από το ότι η εμβέλεια επεκτείνεται σε τουλάχιστον 10 km πάνω από μονότροπη ίνα.

Συσκευές 400ZR με δυνατότητα συνοχής που έχουν κατασκευαστεί σύμφωνα με το OIF Implementation Agreement οι οποίες είναι κατάλληλες για διασυνδέσεις ανάμεσα σε DC's.Οι εν λόγω μονάδες ενσωματώνουν την τεχνολογία συνεκτικής μετάδοσης προς τα κέντρα δεδομένων. Η αλλαγή σε 400G επίσης συμφέρει όσον αφορά το κόστος και την ενέργεια, καθώς μια θύρα 400G σε έναν δρομολογητή, μαζί με τα οπτικά, θα κοστίζει λιγότερο από τέσσερις μεμονωμένες θύρες 100G με το δικό τους σετ οπτικών. Το ίδιο ισχύει και για την ισχύ εφόσον μια μεμονωμένη θύρα 400G καταναλώνει λιγότερη ενέργεια από τη συνολική ενέργεια που καταναλώνουν τέσσερις μεμονωμένες θύρες 100G.Επιπλέον, οι ταχύτητες 400Gb/s επιτρέπουν scale-up και scale-out αρχιτεκτονικές, για αυξημένη ανθεκτικότητα.

2.5 Data Center Interconnect (DCI)

Η τεχνολογία Data Center Interconnect (DCI) συνδέει δύο ή περισσότερα κέντρα δεδομένων μεταξύ τους σε μικρές, μεσαίες ή μεγάλες αποστάσεις. Η πιο αποτελεσματική τεχνολογία διασύνδεσης για DCI είναι οι συνεκτικές οπτικές ίνες νέας γενιάς που παρέχουν μεγάλο εύρος ζώνης και ταχύτητες έως και 800 Gb/s. Ωστόσο, χρησιμοποιείται κυρίως για point-to-point συνδέσεις, ενώ η διασύνδεση εντός του DataCenter Network (DCN) εξακολουθεί να βασίζεται σε ηλεκτρικές δομές καλωδίωσης, τα οποία όμως έχουν υψηλή κατανάλωση ενέργειας και περιορισμένη χωρητικότητα εύρους ζώνης. Επί του παρόντος, η ισχύς που καταναλώνουν τα δίκτυα κέντρων δεδομένων αντιπροσωπεύει το 23% της συνολικής κατανάλωσης ισχύος IT παγκοσμίως. Τα κέντρα δεδομένων πρέπει να επικοινωνούν μεταξύ τους για να μοιράζονται δεδομένα, περιεχόμενο και να παρέχουν εφεδρικά αντίγραφα ασφαλείας. Η επικοινωνία έχει μεγάλες απαιτήσεις καθώς υπάρχουν τέραστιες αποστάσεις που πρέπει να καλυφθούν ενώ υπάρχουν και ορισμένες εφαρμογές DCI που χρειάζονται το υψηλότερο επίπεδο χωρητικότητας και επεκτασιμότητας μαζί με περισσότερο έλεγχο λογισμικού και αυτοματισμό και άλλες που μπορούν να ανταλλάξουν κάποιες επιδόσεις για να ικανοποιήσουν ένα συγκεκριμένο προφίλ ισχύος και παράγοντα μορφής, απαιτούνται λύσεις DCI βελτιστοποιημένων επιδόσεων για να ξεπεραστούν οι περιορισμοί κλιμάκωσης και απόστασης για την παροχή συνδεσιμότητας υψηλότερης χωρητικότητας. Για να αντιμετωπιστούν αυτές οι αυξανόμενες απαιτήσεις εύρους ζώνης, ο εξοπλισμός δικτύωσης πρέπει να παρέχει αξιόπιστες συνδέσεις υψηλής χωρητικότητας που κλιμακώνονται απλά και γρήγορα. Οι οπτικές ίνες ανοίγουν το δρόμο για την κίνηση δεδομένων με ρυθμούς έως και 800 Gb/s σε ένα μόνο μήκος κύματος, αυξάνοντας τη

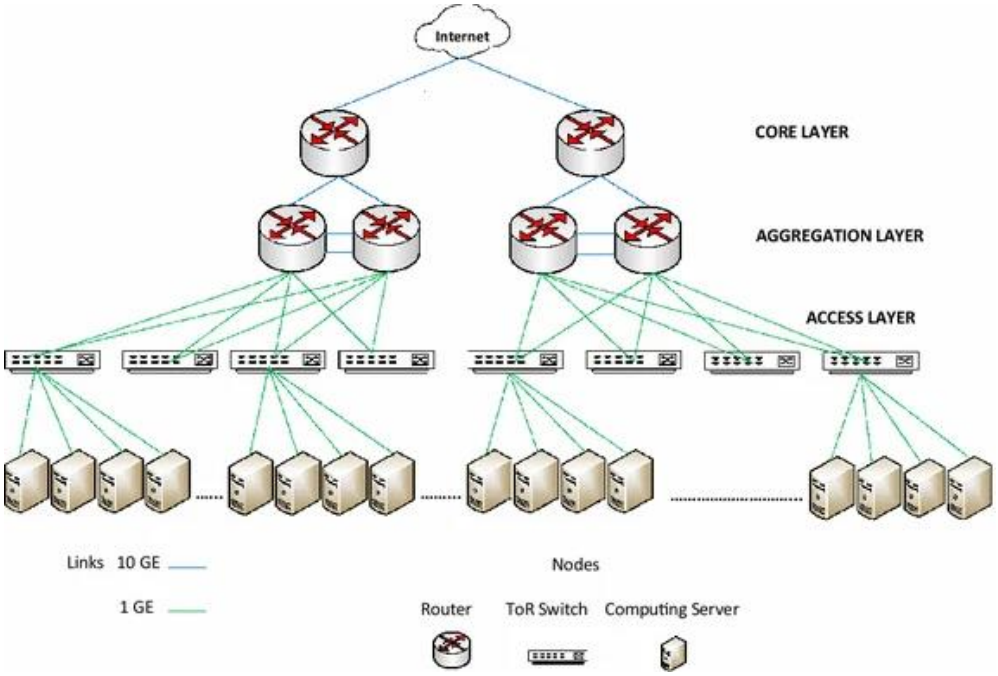
χωρητικότητα DCI. Επιπλέον, τα σύγχρονα συστήματα φωτονικών γραμμών προσφέρουν τη δυνατότητα επέκτασης του ωφέλιμου φάσματος από τη ζώνη C στη ζώνη L για διπλάσια χωρητικότητα ανά ίνα, επιτρέποντας περισσότερα μήκη κύματος μεταξύ των κέντρων δεδομένων σε περιοχές με περιορισμένες οπτικές ίνες ή με υψηλή ανάπτυξη. Είναι επίσης σημαντικό τα DCI πρέπει να είναι αξιόπιστα, ασφαλή, ακόμη και κρυπτογραφημένα, ώστε να αποφεύγονται δαπανηρές παραβιάσεις και απώλειες δεδομένων. Η κρυπτογράφηση και οι αυστηροί κανόνες για την πρόσβαση στα αποθηκευμένα δεδομένα χρησιμοποιούνται ευρέως για την προστασία από εισβολές, ενώ οι εξελίξεις στον εξοπλισμό οπτικών δικτύων πακέτων μπορούν να παρέχουν in flight κρυπτογράφηση, για την προστασία των δεδομένων καθώς αυτά ταξιδεύουν μέσω μιας οπτικής σύνδεσης DCI. Επίσης η σύνδεση μεταξύ δύο κέντρων δεδομένων πρέπει να είναι απλή και γρήγορη και η διαχείριση κάθε σύνδεσης να είναι αυτοματοποιημένη. Η χρήση των ανοιχτών API που βασίζονται σε πρότυπα και τα σύγχρονα μοντέλα δεδομένων είναι κρίσιμα για τη στροφή προς την αυτοματοποίηση, επειδή επιτρέπουν την ενσωμάτωση πλατφόρμας, τη δημιουργία σεναρίων και νέους τρόπους παρακολούθησης και διαχείρισης του δικτύου. Η μείωση του κόστους ανά bit για τη διασύνδεση αποτελεί συνεχές μέλημα, καθώς οι μεγάλες ροές δεδομένων που εισέρχονται και εξέρχονται από τα κέντρα δεδομένων πρέπει να μεταφέρονται όσο το δυνατόν πιο αποδοτικά. Για να μειωθεί το κόστος διασύνδεσης ανά bit, προσφέροντας παράλληλα μεγαλύτερη δυνατότητα κλιμάκωσης της χωρητικότητας,σημειώνεται πρόοδος στις συνεκτικές διεπαφές υψηλής ταχύτητας για την αύξηση της χωρητικότητας ανά μήκος κύματος και τη βελτίωση των επιδόσεων για την οδήγηση αυτών των bit σε μεγαλύτερες αποστάσεις.

2.6 Τοπολογίες DCN

Για την εδραίωση των DC διασυνδέσεων χρειάζεται να οριστεί η κατάλληλη μεθοδολογία-αρχιτεκτονική. Η ανάγκη για την δημιουργία τέτοιων αρχιτεκτονικών προήλθε απο τις δυσκολίες που υπήρχαν στα DCN’s όπως ο ανεπαρκής χώρος απομόνωσης στους μεταγωγείς και οι περιορισμοί του εύρους ζώνης, ώθησαν τους ερευνητές να προτείνουν τεχνικές για τη βελτίωση της απόδοσης του TCP ή να σχεδιάσουν νέα πρωτόκολλα μεταφοράς για το DCN. Μερικές από τις πιο αξιοσημείωτες αρχιτεκτονικές DCN είναι οι παραδοσιακές αρχιτεκτονικές τριών επιπέδων (three-tier), fat-tree και DCell.

2.6.1 Three-Tier

Η παραδοσιακή αρχιτεκτονική DCN τριών επιπέδων ακολουθεί μια τοπολογία δικτύου βασισμένη σε ένα δέντρο με πολλαπλές ρίζες, στο οποίο η ρίζα του δέντρου αποτελεί το επίπεδο πυρήνα, η μεσαία βαθμίδα αποτελεί το επίπεδο συνάθροισης και τα φύλλα του δέντρου αποτελούν το επίπεδο πρόσβασης. Οι servers στα χαμηλότερα επίπεδα συνδέονται απευθείας σε έναν switch του επιπέδου πρόσβασης οι οποίοι διαθέτουν μερικές θύρες 1-10 GigE για συνδεσιμότητα uplink. Οι switches επιπέδου συνάθροισης συνδέουν μεταξύ τους πολλαπλούς switches επιπέδου πρόσβασης. Όλοι οι switches επιπέδου συνάθροισης συνδέονται μεταξύ τους με switches επιπέδου πυρήνα οι οποίοι διαθέτουν θύρες 10 GigE. Οι switches επιπέδου πυρήνα είναι επίσης υπεύθυνοι για τη σύνδεση του κέντρου δεδομένων με το Διαδίκτυο. Η αρχιτεκτονική τριών επιπέδων είναι η πιο συνηθισμένη αρχιτεκτονική δικτύου που χρησιμοποιείται στα κέντρα δεδομένων. Ωστόσο, η αρχιτεκτονική τριών επιπέδων δεν είναι σε θέση να διαχειριστεί την αυξανόμενη ζήτηση της υπολογιστικής νέφους, καθώς αντιμετωπίζει προβλήματα που περιλαμβάνουν, την επεκτασιμότητα, την ανοχή σφαλμάτων, την ενεργειακή απόδοση και το εύρος ζώνης. Επίσης η αρχιτεκτονική τριών επιπέδων στα υψηλότερα στρώματα της τοπολογίας χρησιμοποιεί συσκευές δικτύου επιχειρηματικού επιπέδου οι οποίες είναι πολύ ακριβές και ενεργοβόρες. Στην εικόνα 2.4 φαίνεται η τοπολογία three-tier στην οποία τα core , aggregation και access layers αντιστοιχίζονται με τα επίπεδα πυρήνα, συνάθροισης και πρόσβασης αντίστοιχα. Οι μεταγωγείς στο επίπεδο πρόσβασης είναι χαμηλού κόστους μεταγωγείς Top Of Rack (TOR), οι οποίοι είναι μεταγωγείς Ethernet που συνδέουν διακομιστές στο ίδιο rack μέσω συνδέσεων 1 GigE.Το επίπεδο πυρήνα διαθέτει μεταγωγείς επιπέδου πυρήνα και έναν ή περισσότερους δρομολογητές συνόρων που παρέχουν συνδεσιμότητα μεταξύ του δικτύου του κέντρου δεδομένων και του Διαδικτύου. Κανονικά το στρώμα συνάθροισης διαθέτει έναν εξισορροπιστή φορτίου.



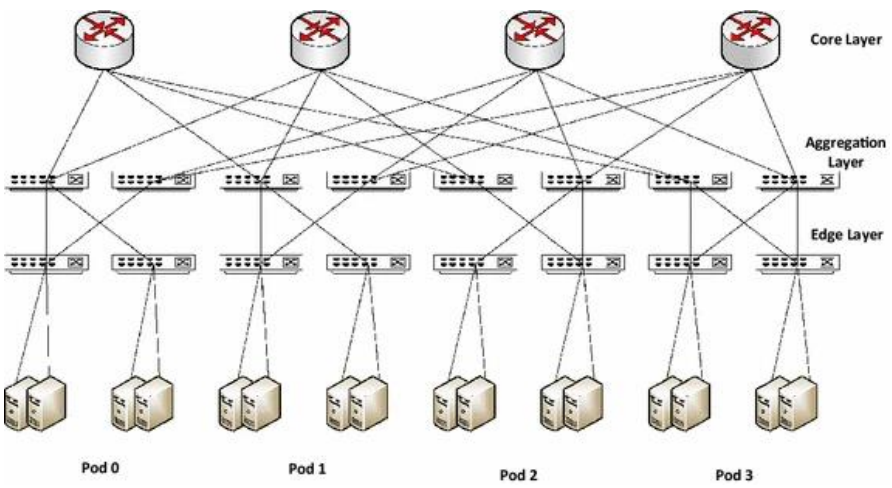
Εικόνα 2.4: Τοπολογία Three-Tier.

Πηγή: https://springerplus.springeropen.com/articles/10.1186/s40064-016-2454-4

2.6.2 Fat-Tree

Τα fat-tree switch fabrics γνωστά και ως αρχιτεκτονική leaf spine ή αρχιτεκτονική 2-tier leaf spine είναι ένα από τα πιο συνηθισμένα switch fabrics που υπάρχουν στα κέντρα δεδομένων. Η σχεδίαση spine και leaf αρχικά υλοποιήθηκε σαν ένας τρόπος βελτίωσης της απόδοσης κατά τη διαχείριση της κίνησης east-west, η οποία επιτυγχάνεται σε μεγάλο βαθμό με τη μείωση του αριθμού των "αλμάτων" μεταξύ δύο οποιονδήποτε συσκευών στο δίκτυο σε μόνο ένα άλμα καθώς κάθε switch φύλλου στο δίκτυο έχει άμεση σύνδεση με κάθε switch σπονδυλικής στήλης, όπως φαίνεται στην εικόνα 2.5. Η διαφορά της με την three-tier αρχιτεκτονική είναι πως η αρχιτεκτονική leaf-spine ουσιαστικά συγκεντρώνει τα επίπεδα πυρήνα και συνάθροισης σε ένα επίπεδο, το spine ενώ το επίπεδο leaf είναι ανάλογο με το επίπεδο πρόσβασης. Το επίπεδο Spine αποτελείται από switches που εκτελούν δρομολόγηση είναι η ραχοκοκαλιά του δικτύου, όπου κάθε μεταγωγέας Leaf είναι διασυνδεδεμένος με κάθε μεταγωγέα Spine. Το στρώμα φύλλου αποτελείται από switches πρόσβασης που συνδέονται με συσκευές όπως διακομιστές, τείχη προστασίας, εξισορροπητές φορτίου και δρομολογητές. Αυτός ο σχεδιασμός χρησιμοποιεί αναζήτηση διαδρομής δύο επιπέδων για να βοηθήσει τη δρομολόγηση πολλαπλών διαδρομών.

Προκειμένου να αποφευχθεί η συμφόρηση σε μία μόνο θύρα λόγω της συγκέντρωσης της κυκλοφορίας σε ένα υποδίκτυο και να διατηρηθεί ο αριθμός των προθεμάτων σε περιορισμένο αριθμό, χρησιμοποιούνται πίνακες δρομολόγησης δύο επιπέδων που κατανέμουν την εξερχόμενη κυκλοφορία από ένα pod ομοιόμορφα μεταξύ των switches πυρήνα χρησιμοποιώντας τα χαμηλής τάξης bits της διεύθυνσης IP προορισμού.

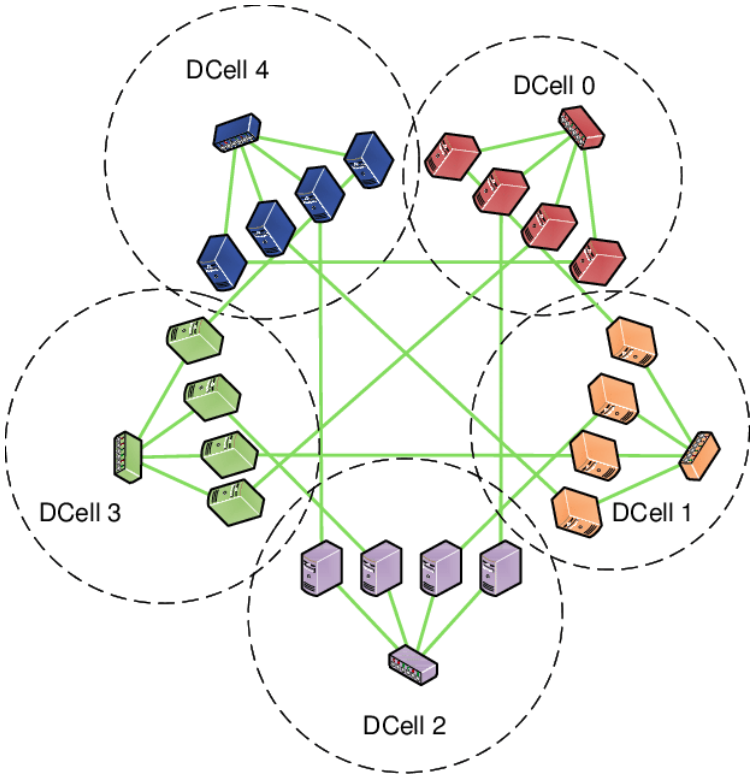


Εικόνα 2.5: Τοπολογία Fat-Tree
Πηγή: https://springerplus.springeropen.com/articles/10.1186/s40064-016-2454-4

2.6.3 DCell

Το DCell είναι μια υβριδική server-centric αρχιτεκτονική DCN, όπου ένας διακομιστής συνδέεται απευθείας με έναν άλλο διακομιστή ο οποίος είναι εξοπλισμένος με πολλαπλές κάρτες διασύνδεσης δικτύου (NIC). Το DCell έχει τέσσερα ουσιαστικά στοιχεία που συνεργάζονται για την αντιμετώπιση προκλήσεων. Αυτά είναι η κλιμακούμενη δομή του δικτύου DCell, η απόδοση και ο κατανεμημένος αλγόριθμος δρομολόγησης που εκμεταλλεύεται τη δομή DCell, είναι επίσης ανεκτικό σε σφάλματα αφού δεν έχει σημείο αποτυχίας και η κατανεμημένη ανοχή σε σφάλματα εκτελεί δρομολόγηση συντομότερης διαδρομής ακόμη και όταν υπάρουν σοβαρές αποτυχίες συνδέσεων ή κόμβων. Επιπλέον παρέχει υψηλότερη χωρητικότητα δικτύου από το παραδοσιακό δέντρο για διάφορους τύπους υπηρεσιών. Αποτελέσματα από θεωριτική ανάλυση, προσομοιώσεις και πειράματα δείχνουν ότι το DCell είναι μια βιώσιμη δομή διασύνδεσης για κέντρα δεδομένων. Το DCell ουσιαστικά ακολουθεί μια δομημένη ιεραρχία απο cells. Ένα cell είναι η βασική μονάδα και

το δομικό στοιχείο της τοπολογίας DCell που διαρθρώνεται σε πολλαπλά επίπεδα, στα οποία ένα cell υψηλότερου επιπέδου περιέχει πολλαπλά cells χαμηλότερου επιπέδου. Το cell0 είναι το δομικό στοιχείο της τοπολογίας DCell, το οποίο περιέχει n διακομιστές και έναν switch δικτύου. Το switch δικτύου χρησιμοποιείται μόνο για τη σύνδεση του server εντός του cell0. Ένα cell1 περιέχει k=n+1 cells0, και ομοίως ένα cell2 περιέχει k * n + 1 dcell1, όπως φαίνεται στην εικόνα 2.6. Το DCell είναι μια εξαιρετικά επεκτάσιμη αρχιτεκτονική, όπου ένα DCell τεσσάρων επιπέδων με μόνο έξι διακομιστές στο cell0 μπορεί να φιλοξενήσει περίπου 3,26 εκατομμύρια διακομιστές. Εκτός από την πολύ υψηλή επεκτασιμότητα, η αρχιτεκτονική DCell απεικονίζει πολύ υψηλή δομική ευρωστία. Ωστόσο, το εύρος ζώνης διατομής και η καθυστέρηση δικτύου είναι ένα σημαντικό ζήτημα στην αρχιτεκτονική DCell DCN.



Εικόνα 2.6: Τοπολογία DCell

Πηγή: https://www.researchgate.net/figure/DCell-data-center-architecture_fig3_220018693

ΚΕΦΑΛΑΙΟ 3

ΑΡΧΕΣ ΤΟΥ DISAGGREGATION

ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό παρουσιάζεται ο όρος του disaggregation σύμφωνα με την ευρεία έννοια αλλά και ειδικότερα στον τομέα των δικτύων. Πιο συγκεκριμένα δίνεται έμφαση στις απαιτήσεις όσον αφορά την CPU, storage και μνήμη αλλά και τον ρόλο του scheduler.

3.1 Disaggregation

Πρόκειται για τον διαχωρισμό των επιμέρους συστατικών εντός κάποιου συστήματος με σκοπό την καλύτερη λειτουργία και απόδοση. Στον τομέα των δικτύων και πιο συγκεκριμένα στα Data Centers, σημαίνει την αποδόμηση της επικοινωνίας των κύριων συστατικών μέσα στους servers. Η βασική ιδέα πίσω από το disaggregation των πόρων είναι η διάσπαση των μονολιθικών servers που παραδοσιακά διατηρούν όλους τους πόρους τους σε φυσικά κυκλώματα σε ξεχωριστές "διαχωρισμένες" δεξαμενές πόρων που συνδέονται με διαύλους υψηλής ταχύτητας (π.χ. GPUs, μνήμη ,αποθήκευση) και είναι φυσικά διακριτοί. Η δομή αυτή προϋποθέτει διασυνδέσεις μικρών αποστάσεων και υψηλών ταχυτήτων μέσω καλωδίων υψηλής ποιότητας, όπως οι οπτικές ίνες ώστε να προσφέρει πολυ υψηλότερες ταχύτητες. Ουσιαστικά αυτό που πρέπει να επιτευχθεί στα σύγχρονα Data Centers είναι η αξιοποιήση των ανεκμετάλλευτων πόρων η οποία δεν μπορεί να πραγματοποιηθεί με την παραδοσιακή δομή. Υπάρχουν δυο είδη διαχωρισμού στα οποία γίνονται μελέτες και έρευνες. Πρόκειται για τον μερικό και ολικό διαχωρισμό. Ενώ οι λύσεις μερικού διαχωρισμού χρησιμοποιούνται ήδη εδώ και αρκετά χρόνια, θα πρέπει να σημειωθεί ότι στην προκειμένη αρχιτεκτονική, οι πόροι της CPU και της μνήμης εξακολουθούν να είναι συνδεδεμένοι ως υπολογιστικός κόμβος, προκαλώντας το πρόβλημα της περιορισμένης χρήσης των πόρων της CPU/της μνήμης. Η έννοια της πλήρως διαχωρισμένης αρχιτεκτονικής, προβλέπει πως δεν υπάρχουν πλέον φυσικά "κουτιά" που ενσωματώνουν διαφορετικούς τύπους πόρων. Αντ' αυτού, ο ίδιος τύπος πόρων σχηματίζει μια μονάδα, δηλαδή μια λεπίδα πόρων, ένα rack ή ακόμη και μια συστάδα. Η πλήρως διαχωρισμένη αρχιτεκτονική επιτρέπει στους χειριστές DC να αντικαθιστούν/αναβαθμίζουν οποιονδήποτε τύπο πόρου όταν είναι απαραίτητο και έχουν

μεγάλες δυνατότητες βελτίωσης της χρήσης των πόρων. Ωστόσο, μέχρι σήμερα η επικοινωνία μεταξύ διαφορετικών πόρων αντιμετωπίζει σοβαρά προβλήματα όσον αφορά την καθυστέρηση και το απαιτούμενο εύρος ζώνης μετάδοσης. Ειδικότερα, οι διασυνδέσεις CPU-μνήμης στα πλήρως διαχωρισμένα κέντρα δεδομένων απαιτούν εξαιρετικά χαμηλή καθυστέρηση και εξαιρετικά υψηλό εύρος ζώνης μετάδοσης, προκειμένου να αποφευχθεί η υποβάθμιση της απόδοσης των εφαρμογών που εκτελούνται.

3.2 Απαιτήσεις CPU, Memory, Storage

3.2.1 Χαρακτηριστικά CPU

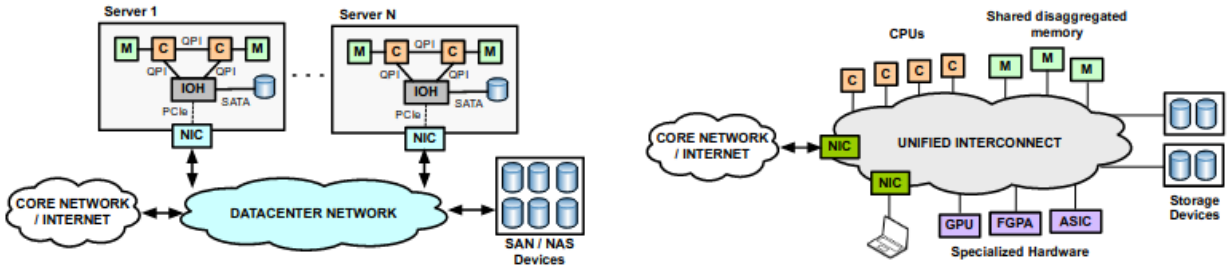
Ένας βασικός παράγοντας που επιτρέπει (ή προς το παρόν εμποδίζει) το disagreggation είναι το δίκτυο, αφού η διάσπαση της CPU από τη μνήμη και τον σκληρό δίσκο απαιτεί την επικοινωνία μεταξύ των πόρων μέσω δικτύου ενώ παλαιότερα ήταν εντός ενός κλασικού διακομιστή. Μια ενδεικτική λύση προς την σωστή κατεύθυνση θα ήταν μερική διάσπαση CPU-μνήμης, όπου κάθε CPU έχει κάποια τοπική μνήμη. Υποστηρίζεται ότι αυτό είναι ένα λογικό ενδιάμεσο βήμα προς το πλήρες disaggregation CPU-μνήμης. Ένας αβίαστος τρόπος για τη διάσπαση του ζεύγους CPU-μνήμης, είναι αντί να επανασχεδιάζεται κάθε πίνακας πόρων, ο διαχωρισμός γίνεται στο επίπεδο του τομέα τροφοδοσίας(power supply domain level). Με άλλα λόγια, η CPU και η μνήμη εξακολουθούν να μοιράζονται την ίδια πλακέτα, αλλά η τροφοδοσία τους διαχωρίζεται σε τομείς. Σε γενικές γραμμές, το disagregation υποδηλώνει ότι κάθε server blade περιέχει ένα συγκεκριμένο πόρο με άμεση σύνδεση με το δίκτυο. Μια εξαίρεση σε αυτή την αυστηρή αποσύνδεση είναι οι λεπίδες CPU: Κάθε CPU blade διατηρεί κάποια ποσότητα τοπικής μνήμης(local memory)[2] που δρα ως κρυφή μνήμη για την απομακρυσμένη μνήμη(remote memory)[3] που προορίζεται για τους πυρήνες της συγκεκριμένης λεπίδας. Έτσι, η διάκριση CPU-μνήμης μπορεί να θεωρηθεί ως επέκταση της ιεραρχίας μνήμης ώστε να περιλαμβάνει ένα απομακρυσμένο επίπεδο, το οποίο μοιράζονται όλες οι λεπίδες CPU. Ενώ υποθέτουμε ότι ο μερικός διαχωρισμός CPU-μνήμης θα είναι ο κανόνας, προχωράμε ένα βήμα περαιτέρω και αξιολογούμε τον τρόπο με τον οποίο η ποσότητα της τοπικής μνήμης επηρεάζει τις απαιτήσεις του δικτύου όσον αφορά το εύρος ζώνης του δικτύου αλλά και την καθυστέρηση, καθώς και τους χρόνους ολοκλήρωσης ροής στο επίπεδο μεταφοράς.

Communication	Latency(ns)	Bandwidth(Gbps)
CPU – CPU	10	500
CPU – Memory	20	500
CPU – Disk (SSD)	10^4	5
CPU – Disk (HDD)	10^6	1

Πίνακας 3.1: Σύνηθες latency και μέγιστες απαιτήσεις εύρους ζώνης σε έναν παραδοσιακό server. Οι αριθμοί διαφέρουν ανάλογα με το hardware.
Πηγή: http://www.cs.cornell.edu/~ragarwal/pubs/disaggregation.pdf

3.2.2 Storage Disaggregation

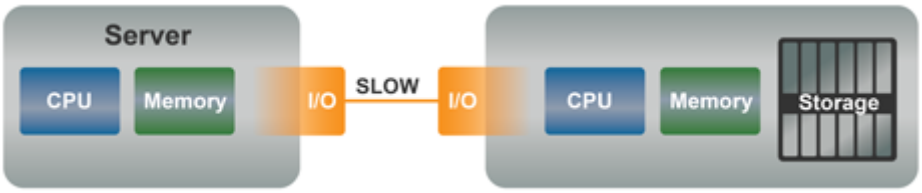
Με τη διαθεσιμότητα νέων και γρήγορων τεχνολογιών διασύνδεσης, το disaggregation του storage από τους διακομιστές μειώνει σημαντικά το συνολικό κόστος της επένδυσης στα κέντρα δεδομένων με διάφορους τρόπους, βελτιώνει την αποτελεσματικότητα της χρήσης του ίδιου του storage, ενώ προσθέτει στην ανθεκτικότητα των στοίβων αποθήκευσης και επιτρέπει τον σχεδιασμό του μέλλοντος των κέντρων δεδομένων. Το 2018, ο συνολικός όγκος δεδομένων που δημιουργήθηκε, καταγράφηκε, αντιγράφηκε και καταναλώθηκε στον κόσμο ήταν 33 zettabytes (ZB) - το ισοδύναμο 33 τρισεκατομμυρίων gigabytes. Το ποσό αυτό αυξήθηκε σε 59ZB το 2020 και προβλέπεται να φθάσει το απίστευτο ποσό των 175ZB μέχρι το 2025. Βλέποντας αυτούς τους ασύλληπτους αριθμούς γίνεται αντιληπτό πως η ανάγκη για αναβάθμιση των Data Centers είναι πλεόν επιτακτική, ενώ επίσης μπορεί να επιτευχθεί χωρίς τεράστιες επενδύσεις προς την κατεύθυνση της αναβάθμισης σε ευέλικτες, κλιμακούμενες λύσεις. Για να κατανοηθεί αυτή η εξέλιξη, είναι χρήσιμο να μελετηθεί η ιστορία των Data Centers και να σημειωθεί η τεχνολογική λογική των αλλαγών στην πορεία.



(a) Current datacenter (b) Disaggregated datacenter

Πηγή: https://www.semanticscholar.org/paper/Network-Requirements-for-Resource-Disaggregation-Gao-Narayan/e403d8d55e3cdc85fbe64b38b8374b3392e8a1df/figure/1

Original Data Center Architecture

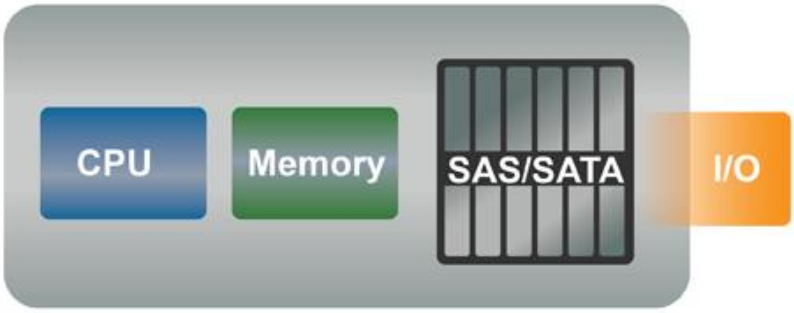


Εικόνα 3.1: Αρχιτεκτονική παραδοσιακού Data Center

Πηγή: https://www.datacenterknowledge.com/archives/2013/10/18/storage-disaggregation-in-the-data-center

Η πρώτη έκδοση ενός κέντρου δεδομένων δεν ήταν τίποτα περισσότερο από έναν κεντρικό υπολογιστή που περιείχε μια CPU, μια κρυφή μνήμη και αποθηκευτικό χώρο(Εικόνα 3.1). Πριν την ανάπτυξη του δικτύου όλες οι λειτουργίες του κέντρου δεδομένων περιέχονταν σε μια κεντρική δομή. Αφού εισήχθη το δίκτυο, ο διαχωρισμός των στοιχείων αποθήκευσης από τα υπολογιστικά στοιχεία του δικτύου έγινε μια αρκετά διαδεδομένη τακτική. Αυτό είχε το πλεονέκτημα ότι επέτρεπε αυξημένη, αποκλειστική αποθήκευση, η οποία μπορούσε να αξιοποιηθεί καλύτερα από ό,τι αν ήταν συνδυασμένη με την CPU. Ωστόσο, ο πολλαπλασιασμός των δεδομένων που αναπτύσσεται την τελευταία δεκαετία και η αντίστοιχη ζήτηση για ανάλυση δεδομένων άλλαξαν και πάλι τη σύνθεση του τυπικού κέντρου δεδομένων. Οι υπάρχουσες τεχνολογίες διασύνδεσης ήταν υπερβολικά αργές για να ανταποκριθούν στις απαιτήσεις για επεξεργασία σε πραγματικό χρόνο των μεγάλων ποσοτήτων δεδομένων που οδηγούσαν σε σχετικές απαντήσεις με αναλυτικές πληροφορίες. Τα περισσότερα αιτήματα για ανάλυση δεδομένων χρειάζονταν εβδομάδες για να εκπληρωθούν και μέχρι τότε ήταν πολύ αργά για να αξιοποιηθούν οι πληροφορίες , καθώς

είχαν προκύψει νέα δεδομένα. Για να αντιμετωπιστεί η κακή απόδοση της διασύνδεσης, οι λύσεις για την βελτιστοποίηση των κέντρων δεδομένων άρχισαν να προσφέρουν αποθηκευτικό χώρο ενσωματωμένο στους διακομιστές(Εικόνα 3.2). Μειώνοντας την απόσταση μεταξύ του υπολογιστικού συστήματος και της αποθήκευσης σχεδόν στο μηδέν, οι εταιρείες απέκτησαν τη δυνατότητα άμεσης πρόσβασης στα δεδομένα, επιτρέποντας πολύ ταχύτερη ανάλυση και ενισχύοντας τις ικανότητες λήψης επιχειρηματικών αποφάσεων.



Εικόνα 3.2: Αποθήκευση συγκεντρωμένη με υπολογισμό (Aggregated Server)
Πηγή: https://www.datacenterknowledge.com/archives/2013/10/18/storage-disaggregation-in-the-data-center

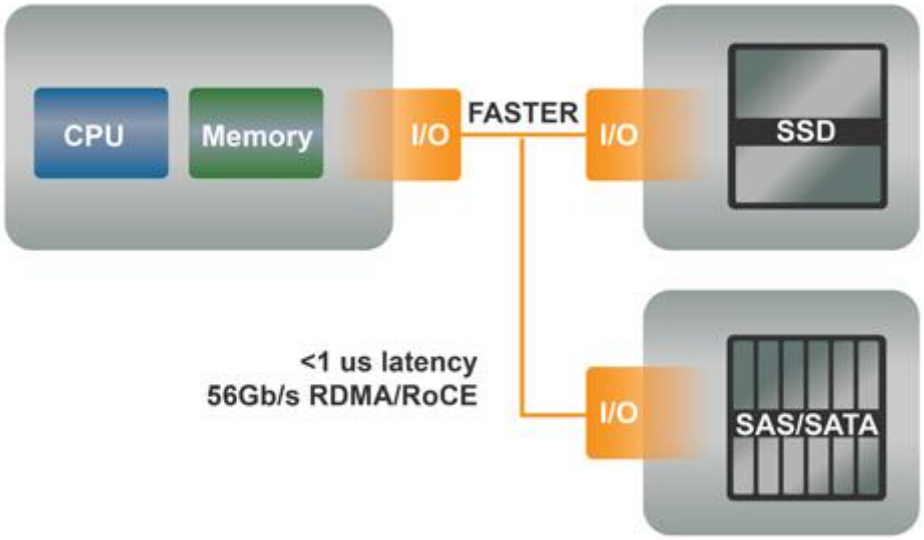
Παρ'όλα αυτά η αλλαγή σε συγκεντρωτικά κέντρα δεδομένων δεν ήταν εύκολη. Οι νέοι διακομιστές προσέφεραν λιγότερη ευελιξία, υψηλότερο κόστος και περισσότερη σπατάλη αποθηκευτικού χώρου σε σχέση με τους διαχωρισμένους προκατόχους τους. Όταν ο σκληρός δίσκος στερεάς κατάστασης (SSD) έγινε η τεχνολογία αποθήκευσης της επιλογής (Εικόνα 3.3), προσφέροντας ακόμη ταχύτερες επιδόσεις μεταξύ υπολογισμού και αποθήκευσης, το κόστος διατήρησης του συγκεντρωτικού κέντρου δεδομένων έγινε ακόμη πιο ακριβό. Πρέπει να σημειωθεί όμως πως η αύξηση την απαιτήσεων συνεπάγεται με αύξηση του κόστους.

Εικόνα 3.3: Αποθήκευση Solid-State Drive

Πηγή: https://www.datacenterknowledge.com/archives/2013/10/18/storage-disaggregation-in-the-data-center

Disaggregation

Με την ανάπτυξη των τεχνολογιών διασύνδεσης όπως το InfiniBand[4], το RDMA[5] και το RoCE[6] (RDMA over Converged Ethernet), έχει καταστεί δυνατή η αποστολή και λήψη δεδομένων με ταχύτητες έως και 56Gb/s, ενώ σύντομα θα ακολουθήσουν και τα 100Gb/s. Επιπλέον, με καθυστέρηση μικρότερη από 1 μικροδευτερόλεπτο ακόμη και σε τόσο υψηλές ταχύτητες διασύνδεσης , ουσιαστικά δεν υπάρχει σχεδόν καθόλου καθυστέρηση. Έτσι ερχόμαστε πιο κοντά στην δημιουργία disaggregated Data Center (Εικόνα 3.4). Καθώς είναι πλέον δυνατή η μετακίνηση του αποθηκευτικού χώρου μακριά από τον υπολογισμό χωρίς καμία επιβάρυνση στην απόδοση. Η ανάλυση δεδομένων εξακολουθεί να είναι δυνατή σε σχεδόν πραγματικό χρόνο, επειδή η διασύνδεση μεταξύ της αποθήκευσης και του υπολογιστικού συστήματος είναι αρκετά γρήγορη για να υποστηρίζει τέτοιες απαιτήσεις. Επιπλέον, ο τύπος του αποθηκευτικού χώρου που χρησιμοποιείται μπορεί πλέον να προσαρμόζεται στα δεδομένα που περιέχονται σε αυτόν. Ενώ είναι λογικό να χρησιμοποιείται αποθήκευση SSD για δεδομένα που πρέπει να είναι άμεσα διαθέσιμα, υπάρχει επίσης πληθώρα δεδομένων σε αποθηκευτικό χώρο με σπάνια πρόσβαση, για τα οποία η αποθήκευση SSD είναι περιττά δαπανηρή και ανεπαρκώς χρησιμοποιούμενη. Για τέτοια δεδομένα backend, τα οποία απαιτούν μεγάλη χωρητικότητα αλλά λιγότερη ταχύτητα, είναι λογικό να χρησιμοποιείται η πιο αργή αλλά πολύ φθηνότερη αποθήκευση SAS[7] ή SATA[8].



Εικόνα 3.4: Η διασύνδεση υψηλής ταχύτητας επιτρέπει το Disaggregation του αποθηκευτικού χώρου (storage) με μηδενική απώλεια στην απόδοση.
Πηγή: https://www.datacenterknowledge.com/archives/2013/10/18/storage-disaggregation-in-the-data-center

3.2.3 Memory Disaggregation

Τα τελευταία χρόνια παρατηρείται η ανισορροπία στη χρήση μνήμης σε πολλά εικονικά νέφη και κέντρα δεδομένων παραγωγής. Αυτή η διαχρονική διακύμανση της χρήσης της μνήμης αποτελεί μια σημαντική αιτία για υπερβολική σελιδοποίηση και thrashing σε εικονικούς διακομιστές, παρόλο που υπάρχει επαρκής αδρανής μνήμη στον ίδιο κόμβο ή στο σύμπλεγμα Cloud. Ο διαχωρισμός της μνήμης είναι μια αναδυόμενη έρευνα και προσπάθεια για την αντιμετώπιση αυτών των ανισορροπιών χρήσης μνήμης. Η διάσπαση μνήμης αποσυνδέει τη φυσική μνήμη που κατανέμεται σε εικονικούς διακομιστές (π.χ. Vms/containers/executors) κατά την αρχικοποίησή τους από τη διαχείριση κατά το χρόνο εκτέλεσης της μνήμης. Η διάσπαση αυτή, αποσκοπεί στο να επιτρέψει στον διακομιστή που βρίσκεται υπό υψηλή πίεση μνήμης να χρησιμοποιεί την αδρανή μνήμη είτε από άλλους διακομιστές που φιλοξενούνται στον ίδιο φυσικό κόμβο (διαχωρισμός μνήμης σε επίπεδο κόμβου) είτε από απομακρυσμένους κόμβους στον ίδια συστάδα (διαχωρισμός μνήμης σε επίπεδο συστάδας).

Διαχωρισμός μνήμης σε επίπεδο κόμβου

Ένας τυπικός κόμβος στο ένα Cloud cluster μπορεί να φιλοξενεί πολλαπλούς εικονικούς διακομιστές (VMs, εμπορευματοκιβώτια ή εκτελεστές JVM). Μια ευρέως υιοθετημένη προσέγγιση για τη διαχείριση των πόρων μνήμης είναι να κατανέμεται ίσο ποσό χωρητικότητας μνήμης σε όλους τους εικονικούς διακομιστές σε έναν κόμβο, με βάση την εκτιμώμενη ζήτηση σε χρόνο αιχμής στον χρόνο αρχικοποίησης της συστάδας. Δεδομένου ότι οι περισσότερες εφαρμογές τρέχουν σε ένα σύμπλεγμα εικονικών διακομιστών, υποθέτουμε ότι οι εφαρμογές μπορούν να "νοικιάσουν" μια διαχωρισμένη μνήμη η οποία είναι ενεργοποιημένη σε συστάδα. Έτσι σε αυτόν την συστάδα οι εικονικοί διακομιστές μαζί με τους αντίστοιχους φυσικούς κόμβους θα συμμετέχουν σε ένα σύστημα διαχωρισμού μνήμης. Ουσιαστικά κάθε φυσικός κόμβος στη συστάδα συμφωνεί να δωρίσει κάποιο μέρος της της φυσικής του DRAM[9], καταχωρώντας κάποια μνήμη που προορίζεται για δίκτυο RDMA[10]. Αυτή η καταχωρισμένη περιοχή μνήμης θα χρησιμοποιείται για τη διατήρηση δύο τύπων διαχωρισμένων δεξαμενών μνήμης σε κάθε κόμβο: η δεξαμενή ρυθμιστικού διαχωρισμού αποστολής και η δεξαμενή ρυθμιστικού διαχωρισμού λήψης. Το εκάστοτε ποσοστό μνήμης το οποίο δωρίζουν οι εικονικοί διακομιστές σαν πόρος κοινής μνήμης, μπορεί να χρησιμοποιηθεί κατόπιν αιτήματος από οποιονδήποτε από τους εικονικούς διακομιστές. Ωστόσο θα πρέπει να σημειωθεί πως υπάρχει η πιθανότητα κάποιος απο τους εικονικούς διακομιστές που φιλοξενούνται στον ίδιο φυσικό κόμβο μπορεί να μην ανήκουν απαραίτητα σε αυτήν την συστάδα και συνεπώς δεν θα είναι οι συμμετέχοντες στη συστάδα διαχωρισμένου συστήματος μνήμης. Όταν ένας εικονικός διακομιστής χρειάζεται να χρησιμοποιήσει πρόσθετη μνήμη από τη διαχωρισμένη δεξαμενή μνήμης, ο τοπικός πελάτης (client) διαχωρισμένης μνήμης (LDMC[11]) που εκτελείται στον εικονικό διακομιστή θα στείλει ένα αίτημα put στην αντίστοιχη διαχωριστή διαχωρισμένης μνήμης σε επίπεδο κόμβου (LDMS[12]), ο οποίος θα ελέγχει αν διαθέτει επαρκή χώρο στη συντονισμένη κοινόχρηστη μνήμη του κόμβου μνήμης για να εξυπηρετήσει αυτό το αίτημα put. Εάν όχι, ο εν λόγω LDMS θα αλληλεπιδρά με τον διαχειριστή κόμβου για να αποκτήσει πρόσθετο ελεύθερο χώρο μνήμης στην κοινόχρηστη μνήμη. Όταν η λειτουργία put ολοκληρώνεται, ενημερώνονται ο διαχωρισμένος πίνακας σελίδων μνήμης που διατηρείται από τον διαχειριστή κόμβων μαζί με το LDMS .Εάν όμως, δεν υπάρχει επαρκής ελεύθερη μνήμη στην κοινόχρηστη δεξαμενή μνήμης, ο διαχειριστής κόμβων θα επικοινωνήσει με τον απομακρυσμένο πελάτη διαχωρισμένης μνήμης (RDMC), ο οποίος επιλέγει τον/τους

απομακρυσμένο/ους κόμβο/ους της ίδιας ομάδας για διαθέσιμη μνήμη απο τον διαχωρισμό της μνήμης που έχει προκύψει.Το αίτημα put του εικονικού διακομιστή θα εξυπηρετείται τώρα από τον επιλεγμένο απομακρυσμένο κόμβο που εκτελείται στον τοπικό κόμβο μέσω του RDMC.

Διαχωρισμός μνήμης σε επίπεδο συστάδας

Ο διαχωρισμός μνήμης σε επίπεδο συστάδας ουσιαστικά πρόκειται για την δυνατότητα που έχει η μνήμη όλων των κόμβων σε μια συστάδα υπολογιστών να εκτίθεται διαφανώς ως μια ενιαία κοινόχρηστη δεξαμενή μνήμης σε επίπεδο συστάδας για τη διακριτική ευχέρεια όλων των εφαρμογών που εκτελούνται στην ίδια υπολογιστική συστάδα. Ακολουθεί ένα παράδειγμα για την καλύτερη κατανόηση της λειτουργίας της συγκεκριμένης τακτικής. Έστω ότι οι κόμβοι Α και Β είναι δύο οποιοιδήποτε κόμβοι στη συστάδα. Όταν ένας εικονικός διακομιστής στον κόμβο Α χρειάζεται να επεκτείνει την τοπική του μνήμη στην συντεταγμένη από κόμβο διαχωρισμένη μνήμη, εάν ο κόμβος Α αποφασίσει να επιλέξει τον κόμβο Β ως την απομακρυσμένη διαχωρισμένη μνήμη για να καλύψει τη μνήμη ζήτηση αυτού του εικονικού διακομιστή, τότε κατά τη λήψη των δεδομένων εισόδου από τον εικονικό διακομιστή, ο κόμβος Α θα τοποθετήσει τα δεδομένα στη δική του δεξαμενή απομονωτών αποστολής διαχωριζόμενης μνήμης (DM) σε επίπεδο συστάδας. Στην περίπτωση αυτή, ο κόμβος Α λειτουργεί ως ο διαχωρισμένος κόμβος πελάτη μνήμης (RDMC) σε όλη τη συστάδα. Ενω, ο κόμβος Β θα χρησιμεύσει ως απομακρυσμένος διακομιστής διαχωρισμένης μνήμης (RDMS) επιπέδου κόμβου-συστάδας. Ο κόμβος Α αποστέλλει δεδομένα μέσω λειτουργίας εγγραφής RDMA στον κόμβο Β στην απομακρυσμένη απομονωμένη μνήμη. Όταν ο εικονικός διακομιστής χρειάζεται να διαβάσει τα δεδομένα που υπάρχουν στον κόμβο Β, θα στείλει ένα αίτημα ανάγνωσης με το αναγνωριστικό καταχώρησης δεδομένων στον τοπικό διαχωρισμένο διακομιστή μνήμης (LDMS) στον κόμβο Α, χρησιμοποιώντας το διαχωρισμένο χάρτη μνήμης, ο κόμβος Α γνωρίζει ότι η καταχώρηση δεδομένων βρίσκεται στον κόμβο Β, και συνεπώς εκδίδει ένα αίτημα ανάγνωσης RDMA στον κόμβο Β. Υπάρχουν ορισμένες σημαντικές σχεδιαστικές αποφάσεις που πρέπει να γίνουν προσεκτικά για να διατηρηθούν οι επιθυμητές ιδιότητες του συστήματος, συμπεριλαμβανομένων των επιδόσεων, της επεκτασιμότητας, της αξιοπιστίας, εξισορρόπησης μνήμης, καταχώρησης μνήμης όπως επίσης, σύνδεση, ορθότητα και συνέπεια του συστήματος διαχωρισμένης μνήμης.

Πλήρης έναντι μερικού διαχωρισμού μνήμης. (Full v.s. Partial Memory Disaggregation)

Μπορούμε να κατηγοριοποιήσουμε περαιτέρω τις δυνατότητες διαχωρισμού μνήμης σε πλήρη διαχωρισμό μνήμης και μερικό διαχωρισμό μνήμης. Από τεχνολογική άποψη, η πλήρης διαίρεση μνήμης δεν είναι εφικτή μέχρι σήμερα για διάφορους λόγους. Πρώτον, η DRAM μνήμη είχε δύο κύρια χαρακτηριστικά που επηρεάζουν την απόδοση του διακομιστή: χωρητικότητα μνήμης και ταχύτητα μνήμης. Δεύτερον, οι επεξεργαστές υπολογιστών απαιτούν εξαιρετικά γρήγορη πρόσβαση στην μνήμη. Με βάση τις πιο σύγχρονες τεχνολογίες δικτύωσης, η ταχύτητα της τοπικής μνήμης παραμένει προς το παρόν πολύ ταχύτερη από το δίκτυο. Τρίτον, εάν ένα σύστημα λειτουργεί αργά λόγω της έλλειψης τοπικής μνήμης DRAM, αλλά ο επεξεργαστής μπορεί να διαβάσει δεδομένα από την τοπική μνήμη ή την απομακρυσμένη μνήμη πολύ πιο γρήγορα από ό,τι ένα εξωτερικό σκληρό δίσκο, τότε η προσθήκη περισσότερης μνήμης ή η χρήση μνήμης αποτελεί απλά μια ευκαιριακή λύση. Αυτό οφείλεται στο γεγονός ότι όταν ένα σύστημα δεν έχει αρκετή μνήμη DRAM πρέπει να μεταφέρει τα δεδομένα που έχουν ξεχειλίσει στο σκληρό δίσκο, γεγονός που μπορεί να επιβραδύνει σημαντικά την απόδοση του συστήματος. Έτσι, ο πλήρης διαχωρισμός της μνήμης σε επίπεδο συστάδας θα είναι εφικτός όταν η ταχύτητα πρόσβασης στην απομακρυσμένη μνήμη είναι συγκρίσιμη με ταχύτητα της τοπικής μνήμης. Ο μερικός διαχωρισμός μνήμης αναφέρεται στη δυνατότητα διάσπασης της μνήμης μόνο για κάποια βοηθητικά προγράμματα. Αυτό ισχύει τόσο για την τοπική μνήμη στο επίπεδο κόμβου όσο και στην απομακρυσμένη μνήμη σε επίπεδο συστάδας. Η μερική διάσπαση μνήμης σε επίπεδο συστάδας αναφέρεται στην δυνατότητα χρήσης της αδρανούς μνήμης στους απομακρυσμένους κόμβους που βρίσκονται στην ίδια συστάδα. Με τον όρο μερική, εννοούμε ότι αν ο επεξεργαστής υπερβαίνει τη χωρητικότητα της μνήμης του, και αν μπορεί να γράψει δεδομένα σε (και να διαβάσει δεδομένα από) απομακρυσμένη μνήμη πολύ πιο γρήγορα από ό,τι σε έναν εξωτερικό σκληρό δίσκο, τότε η χρήση μνήμης σε επίπεδο συστάδας διαχωρισμού παρουσιάζει μια βιώσιμη ευκαιρία για την ενίσχυση της απόδοσης της εφαρμογής. Η μερική διάσπαση μνήμης σε επίπεδο κόμβου αναφέρεται στην δυνατότητα να επιτρέπεται σε έναν διακομιστή (VM, container ή JVM executor) να ανταποκριθεί στην παροδική υψηλή πίεση μνήμης του αξιοποιώντας την αδρανή μνήμη από άλλα VM, εμπορευματοκιβώτια ή JVM executors στον ίδιο φυσικό κόμβο (host).

3.3 Scheduler

Σε ένα σύγχρονο disaggregated κέντρο δεδομένων ο scheduler είναι το στοιχείο που είναι υπεύθυνο για την εύρεση διαθέσιμων πόρων, την εφαρμογή των αντίστοιχων πολιτικών και την απόφαση σε σε ποιον server θα διατεθεί το VM αλλά και να παρέχει πόρους δικτύου. Έτσι είναι σημαντικό μια συνολική εικόνα του περιβάλλοντος και να γνωρίζει τα ειδικά χαρακτηριστικά και τις απαιτήσεις του disaggregation. Σύμφωνα με την έρευνα "The Benefits of a Disaggregated Data Centre: A Resource Allocation Approach" η δημιουργία ενός κατάλληλου για disaggregation scheduler έγινε με τις εξής υποθέσεις. Εφόσον με τις τρέχουσες τεχνολογίες η επικοινωνία CPU-CPU δεν είναι δυνατή και η ισχύς των επεξεργαστών συμβαδίζει με την εξέλιξη του Moore Law[13], είναι ασφαλές να υποθέσουμε ότι δεν θα υπάρξει καμία εφαρμογή που να απαιτεί περισσότερους πυρήνες από όσους μπορεί να παρέχει ένας server. Επιπλεόν γίνεται η παραδοχή πως σε κάθε CPU blade είναι διαθέσιμη κάποια μνήμη cache. Ο scheduler λαμβάνοντας υπόψην τις παραδοχές αυτές διασφαλίζει την ομαλή λειτουργία και την τήρηση των αυστηρών απαιτήσεων δικτύου ενός disaggregated DC, μέσω συνεχών επαναλήψεων φιλτραρίσματος, της ιεράρχησης και της ταξινόμησης του δικτύου και υπολογιστικών πόρων και λαμβάνει τη βέλτιστη απόφαση σύμφωνα με την εκάστοτε πολιτική. Τα VM's στέλνουν αιτήματα τα οποία ποικίλουν σε απαιτήσεις σε πυρήνες CPU, μέγεθος RAM και εύρος ζώνης δικτύου για κάθε στόχο στον οποίο έχουν σκοπό να συνδεθούν. Ο scheduler καθώς δεν γνωρίζει για αυτά τα αιτήματα εκτελεί σε πραγματικό χρόνο μια διαδικασία λήψης μοναδικής απόφασης, βάσει της τρέχουσας κατάστασης του disaggregated DC. Τα βήματα που πραγματοποιεί ο scheduler για τον καθορισμό των απαραίτητων πόρων είναι τα εξής:

1) Αρχικά φιλτράρει όλες τις CPU blades και υποδικνύει εκείνες που έχουν διαθέσιμους πυρήνες και εύρος ζώνης δικτύου για επικοινωνία μεταξύ VM's.

2) Έπειτα τις ιεραρχεί ανάλογα με την δυνατότητα τους να φιλοξενήσουν το VM με την καταχωρημένη μνήμη.

3) Για κάθε CPU που επιλέγεται, ο scheduler βρίσκει τη διαδρομή που ικανοποιεί την ακόλουθη εξίσωση για κάθε μονοπάτι το οποίο συνδέει την εκάστοτε CPU blade με τα αντίστοιχα VM's:

$$min(wpd * pd + wpb * pb + wph * ph)$$

όπου $wppd + wpb + wph = 1$ και pd, pb, ph είναι μονοπάτια καθυστέρησης, το εύρος ζώνης και τα άλματα αντίστοιχα.

4) Ταξινομεί τις CPU blades σύμφωνα με την απόδοση, όπως έχει οριστεί απο το σύστημα στα βάρη των CPU:

$$wcp * cp + wcu * cu + wspb * spb + wspd * spd$$

όπου $wcp + wcu + wspb + wspd = 1$ και cp, cu, spb, spd τα οποία είναι η προτεραιότητα και η χρησιμοποίηση της CPU blade, το εύρος ζώνης και η καθυστέρηση διαδρομής, αντίστοιχα.

Τέλος, ο scheduler λαμβάνει τον ταξινομημένο κατάλογο των διαθέσιμων blades και προσπαθεί να διαθέσει μια RAM blade. Εάν βρει μια RAM blade διαθέσιμη σταματά και διανέμει τους πόρους εκεί που χρειάζεται.

Στα σύγχρονα κέντρα δεδομένων, η υπερδέσμευση της CPU και της RAM χρησιμοποιείται για την αύξηση των εικονικών υπολογιστικών πόρων. Στα disaggregation ισχύει ο ίδιος ακριβώς κανόνας ο οποίος μπορεί να αυξήσει ακόμη περισσότερο τη χρήση των πόρων. Εκεί ως εκ τούτου, ο αλγόριθμος του scheduler λαμβάνει υπόψην του τα ποσοστά της υπερδέσμευσης κατά τον υπολογισμό των διαθέσιμων πόρων. Σύμφωνα με τα αποτελέσματα της προσομοίωσης παρατηρείται ότι ο συγκεκριμένος αλγόριθμος του scheduler μπορεί να βελτιώσει σημαντικά τη χρήση των πόρων σε σύγκριση με τους ήδη υπάρχοντες αλγορίθμους τελευταίας, με αποτέλεσμα την μειώση της ενεργειακής κατανάλωση. Επιπλέον, αποδεικνύεται ότι η εφαρμογή των προσεγγίσεων αυτών για το scheduling σε ένα disaggregated δεδομένο κέντρο θα έχει καταστροφικές επιπτώσεις στη χρήση του υλικού αλλά και ως προς το κόστος.

ΚΕΦΑΛΑΙΟ 4

DISAGGREGATION ΚΑΙ ΟΠΤΙΚΑ ΔΥΚΤΙΑ

ΕΙΣΑΓΩΓΗ

Μέσα απο αυτό το κεφάλαιο αποτυπώνεται πως τα οπτικά δίκτυα είναι το κατάλληλο μέσο προς την σωστή κατεύθυνση για την υλοποίηση των πλήρως διαχωρισμένων κέντρων δεδομένων. Αρχικά παρατίθενται πληροφορίες σχετικά με τα οπτικά δίκτυα και τις λειτουργίες τους. Έπειτα γίνεται ανάλυση για τις διάφορες τεχνολογίες και αρχιτεκτονικές που χρησιμοποιούνται, ενώ επιπλέον παρουσιάζονται κάποιες προτάσεις για την επίτευξη του πλήρους διαχωρισμού μέσα από κάποιες μελέτες και έρευνες.

4.1 Οπτικά δίκτυα

Ένα οπτικό δίκτυο είναι ένα σύστημα επικοινωνίας που χρησιμοποιεί σήματα φωτός, αντί για ηλεκτρονικά, για την αποστολή πληροφοριών μεταξύ δύο ή περισσότερων σημείων. Τα σημεία μπορεί να είναι υπολογιστές, μεγάλα αστικά κέντρα ή ακόμη και κέντρα δεδομένων στο παγκόσμιο σύστημα τηλεπικοινωνιών. Τα οπτικά δίκτυα περιλαμβάνουν οπτικούς πομπούς και δέκτες, καλώδια οπτικών ινών, οπτικούς διακόπτες και άλλα οπτικά εξαρτήματα. Επίσης μπορούν να λάβουν διάφορες μορφές. Τα δίκτυα από σημείο σε σημείο δημιουργούν μόνιμες συνδέσεις μεταξύ δύο ή περισσότερων σημείων, ώστε οποιοδήποτε ζεύγος κόμβων να μπορεί να επικοινωνεί μεταξύ του. Τα δίκτυα από σημείο σε πολλαπλά σημεία μεταδίδουν τα ίδια σήματα ταυτόχρονα σε πολλούς διαφορετικούς κόμβους. Τα δίκτυα μεταγωγής, όπως το τηλεφωνικό σύστημα, περιλαμβάνουν μεταγωγείς που δημιουργούν προσωρινές συνδέσεις μεταξύ ζευγών κόμβων. Τα βασικά δομικά στοιχεία αυτών των δικτύων είναι καλώδια οπτικών ινών που μεταφέρουν σήματα από κόμβο σε κόμβο, με μεταγωγείς που τα κατευθύνουν στον προορισμό τους. Ένα οπτικό σήμα αποτελείται από μια σειρά παλμών που παράγονται από την απενεργοποίηση και ενεργοποίηση μιας δέσμης λέιζερ. Η ταχύτητά του εξαρτάται από το πόσο γρήγορα μπορεί να ενεργοποιηθεί και να απενεργοποιηθεί η δέσμη και από το πόσο εξαπλώνονται οι παλμοί σε μήκος κατά τη μετάδοση, ένα φαινόμενο που ονομάζεται διασπορά. Η ποσότητα της διασποράς εξαρτάται από τον τύπο της ίνας, το μήκος της ίνας και τη φύση του οπτικού

σήματος. Όσο μεγαλύτερη είναι η διασπορά, τόσο πιο δύσκολη είναι η διάκριση μεταξύ γειτονικών παλμών. Με την τρέχουσα τεχνολογία, διαφορετικοί τύποι οπτικών ινών μπορούν να συνδυαστούν για να μειώσουν τα φαινόμενα διασποράς, επιτρέποντας τη μετάδοση με ταχύτητα 10 gigabits ανά δευτερόλεπτο για μερικές χιλιάδες χιλιόμετρα. Για την επίτευξη ταχύτερων ταχυτήτων μετάδοσης, διερευνούνται τρόποι ενεργής αντιστάθμισης της διασποράς. Μια ενιαία ίνα μπορεί να μεταδίδει ταυτόχρονα πολλά ξεχωριστά σήματα σε διαφορετικά μήκη κύματος του φωτός, μια τεχνική που ονομάζεται πολυπλεξία με διαίρεση μήκους κύματος. Αυτό είναι ανάλογο με τη μετάδοση πολλών ραδιοφωνικών και τηλεοπτικών σημάτων μέσω του αέρα σε διαφορετικές συχνότητες. Ο μέγιστος αριθμός των οπτικών καναλιών περιορίζεται από το τμήμα του φάσματος που χρησιμοποιείται για κάθε κανάλι και από το συνολικό διαθέσιμο φάσμα. Οι συσκευές που ονομάζονται "αποπολυπλέκτες" διαχωρίζουν τα οπτικά κανάλια και τα διανέμουν σε ξεχωριστούς οπτικούς δέκτες. Οι αποπολυπλέκτες τεμαχίζουν το φάσμα σε πολύ στενά κομμάτια, απομονώνοντας κάθε οπτικό κανάλι από τα γειτονικά. Ο πολλαπλασιασμός του αριθμού των οπτικών καναλιών επί τον ρυθμό δεδομένων σε κάθε οπτικό κανάλι δίνει τη συνολική ικανότητα μετάδοσης μιας ίνας. Σε εργαστηριακά πειράματα έχουν μεταδοθεί περισσότερα από 10 τρισεκατομμύρια bits (10 terabits) ανά δευτερόλεπτο μέσω περισσότερων από 100 χιλιομέτρων οπτικών ινών. Ωστόσο, οι εμπορικοί ρυθμοί μετάδοσης συνήθως δεν υπερβαίνουν μερικές εκατοντάδες gigabits ανά δευτερόλεπτο.

Η επίτευξη αυτών των υψηλών ρυθμών δεδομένων και των πολλαπλών καναλιών απαιτεί εξελιγμένα εξαρτήματα. Τα λέιζερ ημιαγωγών - τα οποία παράγουν τους παλμούς φωτός που χρησιμοποιούνται σχεδόν σε όλα τα συστήματα επικοινωνιών οπτικών ινών - πρέπει να εκπέμπουν μόνο ένα πολύ στενό εύρος μηκών κύματος για να περιορίσουν τη διασπορά. Οι ίνες σχεδιάζονται επίσης για να περιορίζουν τη διασπορά.

Ενισχυτές

Οι πιο καθαρές οπτικές ίνες μπορούν να μεταδώσουν σήματα σε απόσταση μεγαλύτερη των 100 χιλιομέτρων χωρίς ενίσχυση - πολύ μακρύτερα από τα χάλκινα καλώδια. Όταν το σήμα πρέπει να καλύψει μεγαλύτερη απόσταση, περνάει από έναν οπτικό ενισχυτή, ο οποίος πολλαπλασιάζει την ισχύ του οπτικού σήματος. Οι πιο ευρέως χρησιμοποιούμενοι οπτικοί ενισχυτές είναι οπτικές ίνες με ντοπάρισμα ατόμων έρβιου, ενός στοιχείου σπάνιων γαιών που απορροφά φωτεινή ενέργεια από ένα εξωτερικό λέιζερ άντλησης. Στη συνέχεια,

τα άτομα ερβίου απελευθερώνουν αυτή την ενέργεια για να ενισχύσουν τα ασθενή οπτικά σήματα σε ολόκληρη τη ζώνη μηκών κύματος που εκπέμπει το λέιζερ. Με προσεκτικό έλεγχο, μια σειρά από δεκάδες ενισχυτές οπτικών ινών μπορεί να μεταδώσει σήματα χιλιάδες χιλιόμετρα στον ωκεανό.

Οπτικοί διακόπτες

Οι οπτικοί διακόπτες μπορούν να λειτουργούν σε ένα μόνο μήκος κύματος ή σε όλα τα μήκη κύματος που μεταδίδονται μέσω μιας ίνας. Ένα σταθερό φίλτρο, θα μπορούσε να αντικατασταθεί από έναν διακόπτη που επιλέγει ένα από διάφορα φίλτρα για να εκτρέψει το επιθυμητό μήκος κύματος στο ενδιάμεσο σημείο. Ένα τρίτο είδος διακόπτη διαχωρίζει τα μήκη κύματος σε ξεχωριστές δέσμες και ένα κινούμενο κάτοπτρο κατευθύνει ένα ή περισσότερα από τα μήκη κύματος σε διαφορετική κατεύθυνση. Άλλοι οπτικοί διακόπτες αλλάζουν ταυτόχρονα όλα τα μήκη κύματος που διέρχονται από μια ίνα. Για παράδειγμα είναι ένα κάτοπτρο στην έξοδο της ίνας που μπορεί να γέρνει μεταξύ δύο διαφορετικών θέσεων για να ανακατευθύνει όλα τα οπτικά κανάλια σε περίπτωση θραύσης της ίνας. Τα προηγούμενα παραδείγματα ονομάζονται "αμιγώς οπτικοί" διακόπτες επειδή λειτουργούν με φωτεινά σήματα. Μια διαφορετική κατηγορία διακοπτών μετατρέπει τα οπτικά σήματα σε ηλεκτρονική μορφή η οποία μπορεί να αλλάξει ηλεκτρονικά- το ηλεκτρονικό σήμα τροφοδοτεί στη συνέχεια έναν οπτικό πομπό για να παράγει ένα νέο οπτικό σήμα. Αυτοί ονομάζονται οπτο-ηλεκτρο-οπτικοί διακόπτες. Καθώς η τεχνολογία συνεχίζει να εξελίσσεται, τα οπτικά δίκτυα θα πρέπει να μετατρέπουν τα σήματα από το ένα μήκος κύματος στο άλλο. Αυτό μπορεί να γίνει τώρα με οπτικο-ηλεκτρο-οπτικούς μετατροπείς μήκους κύματος που μετατρέπουν το οπτικό σήμα εισόδου σε ηλεκτρονική μορφή για να οδηγήσουν έναν πομπό στο δεύτερο μήκος κύματος. Οι αμιγώς οπτικοί μετατροπείς μήκους κύματος έχουν επιδειχθεί στο εργαστήριο, αλλά δεν χρησιμοποιούνται ακόμη σε πρακτικά συστήματα. Θα χρειαστούν επίσης πηγές λέιζερ που μπορούν να συντονίζονται σε πολλά διαφορετικά μήκη κύματος- έχουν καταδειχθεί διάφοροι τύποι και ορισμένοι βρίσκονται σε εμπορική παραγωγή.

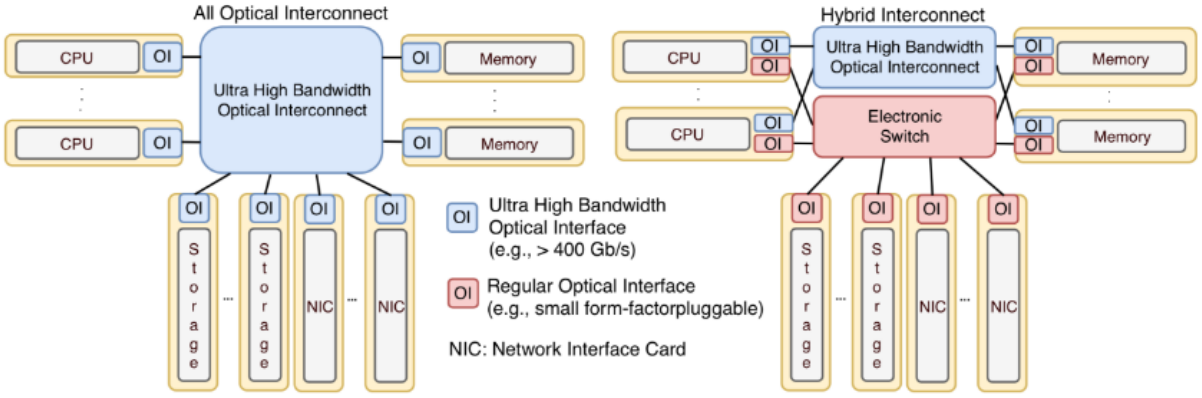
Παθητικό οπτικό δίκτυο (Passive optical network)

Το παθητικό οπτικό δίκτυο (PON) είναι ένα δίκτυο οπτικών ινών που χρησιμοποιεί τοπολογία σημείου προς πολλαπλά σημεία και οπτικούς διαχωριστές για την παροχή δεδομένων από ένα μόνο σημείο μετάδοσης σε πολλαπλά τελικά σημεία χρηστών. Το παθητικό, σε αυτό το πλαίσιο, αναφέρεται στην κατάσταση μη τροφοδοσίας της ίνας και των στοιχείων διαχωρισμού. Σε αντίθεση με ένα ενεργό οπτικό δίκτυο, ηλεκτρική ισχύς απαιτείται μόνο στα σημεία αποστολής και λήψης, καθιστώντας ένα PON εγγενώς αποδοτικό από άποψη κόστους λειτουργίας. Τα παθητικά οπτικά δίκτυα χρησιμοποιούνται για την ταυτόχρονη μετάδοση σημάτων τόσο στην ανάντη όσο και στην κατάντη κατεύθυνση προς και από τα τελικά σημεία των χρηστών.

Ενεργό οπτικό δύκτιο (Active Optical Network)

Το ενεργό οπτικό δίκτυο, χρησιμοποιεί κυρίως μια αρχιτεκτονική δικτύου σημείο-προς-σημείο (PTP) και κάθε χρήστης μπορεί να έχει μια αποκλειστική γραμμή οπτικής ίνας. Ενεργό οπτικό δίκτυο σημαίνει ότι κατά τη μετάδοση των σημάτων αναπτύσσεται εξοπλισμός μεταγωγής, όπως δρομολογητές και συγκεντρωτές μεταγωγής, ενεργές οπτικές συσκευές κ.λπ. από το κεντρικό γραφείο στη μονάδα διανομής χρήστη. Αυτός ο εξοπλισμός μεταγωγής οδηγείται από την ηλεκτρική ενέργεια για τη διαχείριση της διανομής σημάτων και της κατεύθυνσης σημάτων για συγκεκριμένους πελάτες. Οι ενεργές οπτικές συσκευές περιλαμβάνουν πηγές φωτός (λέιζερ), οπτικούς δέκτες, μονάδες οπτικών πομποδεκτών, οπτικούς ενισχυτές.

4.2 Πλήρης διαχωρισμός σε επίπεδο rack (rack-scale disaggregated DCs)



Σχήμα 4.2: Πλήρως διαχωρισμένο rack με αμιγώς οπτική διασύνδεση (αριστερά) και υβριδική διασύνδεση (δεξιά)

Πηγή:https://www.diva-portal.org/smash/get/diva2:1292615/FULLTEXT01.pdf

Η συγκεκριμένη ενότητα επικεντρώνεται στα πλήρως διαχωρισμένα κέντρα δεδομένων σε κλίμακα rack, καθώς και παρουσιάζει και αναλυεί δύο υποψήφιες αρχιτεκτονικές για πλήρη disaggregation μέσω οπτικών δικτύων. Οι αρχιτεκτονικές βασίζονται στο Σχήμα 4.1. Σε κάθε αρχιτεκτονική, κάθε τύπος πόρων, όπως CPU, μνήμη, αποθηκευτικός χώρος, κάρτα διασύνδεσης δικτύου (NIC) και μονάδα επεξεργασίας γραφικών GPU είναι πλήρως αποσυνδεδεμένοι μεταξύ τους. Επίσης και στις δυο αρχιτεκτονικές αντί οι πόροι να βρίσκονται εξ ολοκλήρου στα server baldes υπάρχουν στα resource blades τα οποία διασυνδέονται μέσω των οπτικών διεπαφών σε ένα rack και περιλαμβάνουν μόνο έναν συγκεκριμένο τύπο πόρων. Στην περίπτωση της αμιγώς οπτικής διασύνδεσης, κάθε resource blade απαιτεί μια οπτική διεπαφή και ταυτόχρονα όλες οι λεπίδες συνδέονται με μια οπτική διασύνδεση μέσω μιας οπτικής σύνδεσης. Όλοι οι τύποι επικοινωνιών πόρων (όπως CPU-μνήμη, μνήμη-αποθήκευση, μνήμη-NIC κ.λπ.), οι οποίοι παλαιότερα βρίσκονταν στους διαύλους της μητρικής πλακέτας του ολοκληρωμένου server, πλέον υπάρχουν στις εξωτερικές οπτικές διαδρομές που δημιουργούνται μεταξύ των resource blades. Αυτό σημαίνει ότι οι οπτικές διεπαφές στα resource blades είναι σημαντικό να ικανοποιούν τις κρίσιμες απαιτήσεις όσον αφορά την καθυστέρηση και το εύρος ζώνης των επικοινωνιών μεταξύ των πόρων, ώστε να αποφεύγεται η υποβάθμιση των επιδόσεων στις

εφαρμογές που εκτελούνται. Οι οπτικές διεπαφές για την αμιγώς οπτική διασύνδεση καθώς και η ίδια η διασύνδεση εμφανίζονται με μπλε χρώμα στο Σχήμα 4.1, αντιπροσωπεύοντας έτσι τις δυνατότητές τους να υποστηρίζουν όλους τους τύπους επικοινωνιών πόρων, ιδίως τις πιο απαιτητικές σε εύρος ζώνης, δηλαδή CPU-μνήμη, όπου η μνήμη νέας γενιάς με υψηλές επιδόσεις απαιτεί συνήθως μέγιστο εύρος ζώνης μεγαλύτερο από 400 Gb/s. Σε αντίθεση με την πρώτη αρχιτεκτονική ωστόσο, η δεύτερη περιλαμβάνει δύο τύπους διασυνδέσεων σε ένα rack. Ο ένας είναι οπτική διασύνδεση εξαιρετικά υψηλού εύρους ζώνης μόνο για την επικοινωνία CPU-μνήμης και ο άλλος είναι ένας διακόπτης(switch) για τους πόρους επικοινωνίας που τυπικά δεν έχουν τόσο αυστηρές απαιτήσεις επιδόσεων. Για τις CPU blades και μνήμης απαιτούνται δύο τύποι οπτικών διεπαφών, δηλαδή οπτικές συνδέσεις εξαιρετικά υψηλού εύρους ζώνης που εξυπηρετούν την οπτική διασύνδεση, και η τυπική οπτική διασύνδεση που συνδέεται με το διακόπτη(switch). Τα blades αποθήκευσης και κάρτας δικτύου NIC μπορούν να εξοπλιστούν μόνο με την κανονική οπτική διασυνδέση, ενώ οι πόροι επικοινωνίας που σχετίζονται με αυτούς τους δύο τύπους χειρίζονται αποκλειστικά από τον ηλεκτρονικό switch. Η κύρια διαφορά μεταξύ αυτών των δύο αρχιτεκτονικών είναι η πρόσθετη κανονική οπτική διεπαφή και το switch που εντοπίζονται στη δεύτερη αρχιτεκτονική. Επίσης η καλωδίωση στην πρώτη αρχιτεκτονική είναι λιγότερο πολύπλοκη από ό,τι στη δεύτερη, δεδομένου ότι μπορεί να υπάρχει μόνο μία ίνα για κάθε λεπίδα πόρου. Ωστόσο, λόγω του γεγονότος ότι κάθε επικοινωνία από ή προς ενός resource blade χειρίζεται απο την ενιαία οπτική διασύνδεση, έτσι ο συντονισμός της επικοινωνίας είναι πιο πολύπλοκος. Επιπλέον, υπάρχουν ήδη εμπορικά προϊόντα (π.χ, InfiniBand που έχουν απομακρυσμένη άμεση πρόσβαση στη μνήμη (RDMA) από τη Mellanox) που μπορούν να εφαρμοστούν σε αυτήν την αρχιτεκτονική, δεδομένου ότι είναι σε θέση να ικανοποιήσουν τις απαιτήσεις της καθυστέρησης και του εύρους ζώνης της αποθήκευσης και των επικοινωνιών που σχετίζονται με τη NIC.

4.3 Οπτική μετάδοση για επικοινωνία μεταξύ πόρων

Η τεχνολογία οπτικής μετάδοσης θεωρείται η μόνη δυνατή λύση λόγω της δυνατότητάς της να προσφέρει εξαιρετικά υψηλό εύρος ζώνης και χαμηλή καθυστέρηση ,για την ικανοποίηση των κρίσιμων απαιτήσεων επικοινωνίας μεταξύ των πόρων, ιδίως της επικοινωνίας μεταξύ CPU και μνήμης. Υπάρχουν δύο κατηγορίες οπτικής μετάδοσης. Η πρώτη είναι η διαμόρφωση της έντασης και άμεση ανίχνευση(IM/DD) και η δεύτερη είναι το συνεκτικό

σύστημα το οποίο ενώ έχει εφαρμοστεί ευρέως στην μετάδοση μεγάλων αποστάσεων, το υψηλό κόστος και η πολυπλοκότητα το καθιστά δύσκολο να είναι προσιτό για εφαρμογές μικρής εμβέλειας. Από την άλλη πλευρά, το IM/DD έχει το πλεονεκτήματα της απλής εγκατάστασης του συστήματος και έχει θεωρείται πολλά υποσχόμενο για την παροχή υψηλού εύρους ζώνης για τα DC.

Modulation	Wavelength band (nm)	Data rate per fiber	Multiplexing	Reach	Optical link	Transceiver	Pre-FEC BER	Reference
DMT	1550	4 x 87 Gb/s	WDM	20km	SMF	SiP	3.8e-2	[7]
NRZ	850	6x40 Gb/s	SDM	7m	MMF	VCSEL	1e-12	[9]
NRZ/EDB	1550	7x100 Gb/s	SDM	10km	MCF	EAM	5e-5	[10]
NRZ	1310	8x4x25 Gb/s	SDM/WDM	1.1km	MCF	VCSEL	1e-12	[11]
PAM4	1550	7X149 Gb/s	SDM	1 km	MCF	VCSEL	3.8e-3	[12]

DMT: Discrete multitone modulation; WDM: wavelength division multiplexing; SMF: single mode fiber; SiP: silicon photonics
NRZ: Non-return-to-zero; SDM: spatial division multiplexing; MMF: multi mode fiber; VCSEL: vertical-cavity surface emitting laser
EDB: electrical duo-binary; MCF: multi core fiber; EAM: electro-absorption modulator; PAM4: 4-level pulse amplitude modulation

Πίνακας 4.1: Οπτική μετάδοση μικρής εμβέλειας.
Πηγή: https://www.diva-portal.org/smash/get/diva2:1292615/FULLTEXT01.pdf

Στον πίνακα 4.1 παρουσιάζονται ενημερωμένα έργα για οπτικές επικοινωνίες μικρής εμβέλειας πέραν των 200Gb/s, όπου χρησιμοποιούνται διορθώσεις σφάλματος (FEC), διαφορετικές μορφές διαμόρφωσης, πολυπλεξίας τύποι πομποδεκτών, σήματα τεχνικών επεξεργασίας, υποδεικνύοντας πιθανές τεχνικές για την επικοινωνία μεταξύ πόρων στα πλήρως διαχωρισμένα Dcs. Όπως φαίνεται απ τον πίνακα για να επιτευχθεί χαμηλό κόστος και χαμηλή κατανάλωση ενέργειας ανά bit, είναι προτιμότερη η χρήση υψηλών τιμών ανά λωρίδα ρυθμού δεδομένων. Η μετάδοση πέραν των 100 Gb/s έχει επιτευχθεί με τη χρήση απλούστερων μορφών διαμόρφωσης, π.χ. non-return to zero σε-off-keying (NRZ-OOK)[1] , επιπροσθέτως ο παλμός πλάτους 4 επιπέδων(PAM4)[2] και το discretemultitone (DMT)[3] είναι άλλες δύο επιλογές διαμόρφωσης, οι οποίες μπορούν να επιτύγχουν υψηλή αποδοτικότητα εύρους ζώνης. Η χωρητικότητα ανά ίνα της διασύνδεσης μπορεί να ενισχυθεί περαιτέρω εφόσον αξιοποιηθούν τεχνικές πολυπλεξίας όπως η πολυπλεξία με διαίρεση μήκους κύματος (WDM)[4], ή η πολυπλεξία χωρικής διαίρεσης (SDM)[5] που βασίζεται σε πολυπύρηνη/πολυτροπική ίνα και ο συνδυασμός τους. Το WDM σύστημα συνεπάγεται υψηλό κόστος του πομποδέκτη ενώ η προσέγγιση SDM μπορεί να είναι ακριβή λόγω της χρήσης προηγμένων τεχνολογιών ινών. Η ίνα μονής λειτουργίας (SMF) επιτρέπει τη σχετικά μακρά απόσταση επικοινωνίας με λεπτούς πομποδέκτες ενώ οι πομποδέκτες χαμηλού

κόστους μπορούν να χρησιμοποιηθούν μαζί με πολύτροπες ίνες (MMF), ωστόσο σε αυτήν την περίπτωση το σήμα, το εύρος ζώνης και η απόσταση μετάδοσης είναι περιορισμένα. Στο πλήρως διαχωρισμένο DC, οι πομποδέκτες θα πρέπει να είναι μικροί και απλοί στην υλοποίηση ή να ενσωματώνουν τη μονάδα πόρου σε έναν οικονομικά αποδοτικό τρόπο. Οι δύο κύριοι υποψήφιοι για την αντιμετώπιση των προκλήσεων όσον αφορά το κόστος απο την πλευρά του πομποδέκτη είναι η κάθετη κοιλότητα επιφανειακής εκπομπής λέιζερ (VCSEL)[6] και τα φωτονικά κυκλώματα πυριτίου (SiP)[7]. Τα σωστά σχεδιασμένα VCSEL είναι σε θέση να λειτουργούν χωρίς πρόσθετη παρακολούθηση πάνω από ένα ευρύ φάσμα θερμοκρασιών με ελάχιστη μεταβολή στην απόδοση , γεγονός που είναι κατάλληλο για τα Κέντρα Δεδομένων δεδομένου ότι οι θερμοκρασίες μπορεί να ποικίλλουν πολύ ανάλογα με το διαφορετικό φορτίο εργασίας. Ο χαρακτήρας της επιφανειακής εκπομπής επιτρέπει επίσης την πυκνή δισδιάστατη κατασκευή των VCSEL και την κάθετη ολοκλήρωση με άλλα στοιχεία, επομένως η συσκευασία απλοποιείται, έτσι το σύνολο της μετάδοσης μονάδας καθίσταται μικρό και αρκετά εύκολο για να ενσωματωθεί και να υλοποιηθεί στη μονάδα πόρων. Επιτρέπει επίσης τη δοκιμή και τον έλεγχο σε επίπεδο πλακιδίων, η οποία μειώνει το κόστος κατασκευής, μειώνοντας έτσι συνολικά το κόστος της υποδομής των Κέντρων Δεδομένων. Το SiP μαζί με το WDM επιτρέπουν υψηλό ρυθμό δεδομένων με τη χρήση της κατασκευής πυριτίου μεγάλου όγκου αλλά και καλή αξιοπιστία καθώς έχουν ήδη καταδειχθεί 100 Gb/s ανά μεμονωμένο κανάλι IM/DD. Η εξέλιξη από το 100G Ethernet στο 400G Ethernet στα δίκτυα των Κέντρων Δεδομένων καθιστά το πλεονέκτημα του SiP περισσότερο προφανές. Ενδεικτικά υπάρχουν ήδη λύσεις 400G που βασίζονται σε SiP από βιομηχανίες, όπως Intel, Luxtera και Acacia, υποδεικνύοντας την δυνατότητα υποστήριξης των μεταδόσεων σε διαχωρισμένα Κέντρα Δεδομένων.Επιπλέον, η ελαχιστοποίηση της καθυστέρησης είναι απαραίτητη για την πρακτική ανάπτυξη της οπτικής διασύνδεσης στα διαχωρισμένα Κ.Δ.. Συγκρίνοντας με συνδέσεις μεγάλων αποστάσεων, η καθυστέρηση διάδοσης είναι προφανώς χαμηλότερη στα Κ.Δ. επομένως η ελαχιστοποίηση αυτή μπορεί να επιτευχθεί πιο εύκολα.Εκτός από τον επιπλέον χρόνο επεξεργασίας που εισάγεται από τις μονάδες DSP, η τυπική καθυστέρηση FEC συμβάλλει σημαντικά στη συνολική επιβάρυνση του συστήματος. Όσον αφορά τις αυστηρές απαιτήσεις των pre-FEC bit και την ευαισθησία του δέκτη η τυπική σκληρή απόφαση FEC (HD-FEC) (σε επίπεδο δεκάδων νανοδευτερολέπτων, π.χ. 51 ns της 802.3bj KR FEC) ή καινοτόμοι κώδικες χαμηλής καθυστέρησης αποτελούν την βέλτιστη λύση. Στο άμμεσο μέλλον τα 400 Gb/s και 800 Gb/s

θα εδραιωθούν ως οι στάνταρ τυποποιημένοι ρυθμοί δεδομένων, οι οποίοι θα επιτρέψουν υψηλότερη ταχύτητα ανά λωρίδα. Ωστόσο αν και η τελευταία λέξη της τεχνολογίας οπτικών μεταδόσεων που παρατίθενται στον πίνακα 4.1 είναι σε θέση να επιτύχουν ρυθμό δεδομένων έως και 800 Gb/s ανά ίνα,το αν αυτοί οι ρυθμοί δεδομένων είναι επαρκείς για για τα fully disaggregated D.C.'s παραμένει ένα ανοικτό ερώτημα.

4.4 Αρχιτεκτονική dRedBox

Η συγκεκριμένη αρχιτεκτονική αποτελείται από dRacks (disaggregated Racks) που στεγάζουν πολλαπλά διασυνδεδεμένα dBoxes. Κάθε dBox φιλοξενεί τα εξής:

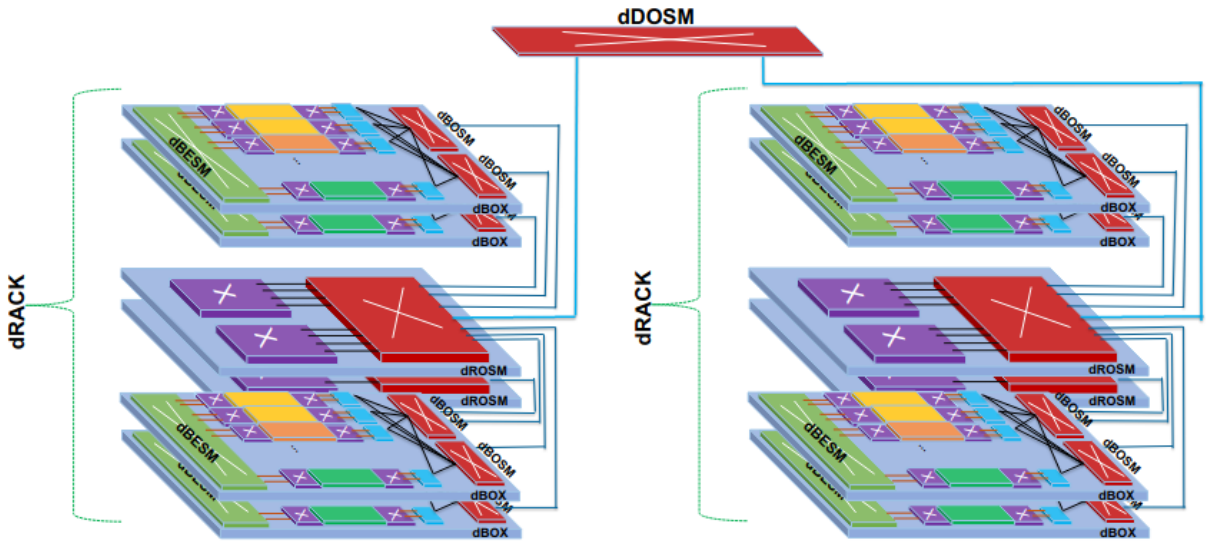
α) ένα αυθαίρετο συνδυασμό dBricks συνδέσιμων components υπολογισμού-μνήμης - επιταχυντή,

β) έναν switch διασταυρούμενων σημείων για την τη συνδεσιμότητα των dBox και

γ) ένα σύνολο μικροσκοπικών οπτικών διακοπτών για την εσωτερική δικτύωση των dBox.

Κάθε dBox ουσιαστικά είναι μια μονάδα 2U η οποία υποστηρίζει έως και 16 dBricks και βρίσκεται πάνω στο rack. Κάθε dBrick υποστηρίζει είτε επεξεργασία γενικού σκοπού (dCompubrick) ή μνήμη τυχαίας προσπέλασης (dMembrick) ή επιταχυντή εφαρμογών (dAccelbrick). Όλα τα dBricks είναι διασυνδέονται μεταξύ τους στο ίδιο dBox μέσω του switch L1 και του οπτικού διακόπτη κυκλώματος. Ο οπτικός διακόπτης λειτουργεί με τεχνολογία διεύθυνσης δέσμης. Η επικοινωνία μεταξύ των εκάστοτε dBricks σε διαφορετικά dBox πραγματοποιείται αυστηρά μέσω μεταγωγής οπτικών κυκλωμάτων. Επιπλέον είναι σημαντικό το γεγονός ότι κάθε dBrick εκτός από την κύρια λειτουργία που επιτελεί (υπολογισμός/μνήμη/επιτάχυνση) χρησιμοποιεί και μια μονάδα chip για την εκτέλεση λειτουργιών δικτύωσης πέρα από την απλή διασύνδεση, όπως οι παραδοσιακές κάρτες διασύνδεσης δικτύου, έτσι κάθε μεμονωμένο dBrick μπορεί να ενσωματώσει και να υποστηρίξει προώθηση, μεταγωγή και συγκέντρωση σε επίπεδο πακέτου είτε σε επίπεδο κυκλώματος. Ωστόσο πρέπει να εξεταστεί κατά πόσο είναι εφικτή η επέκταση του disaggregated συστήματος από ένα μόνο dRack σε ένα κέντρο δεδομένων, με μια αρχιτεκτονική δικτύου επέκτασης, καθώς το εν λόγω σύστημα πρέπει επίσης να αντιμετωπίζει τις ποικίλες απαιτήσεις δικτύου για τους διάφορους τύπους επικοινωνιών δηλαδή CPU προς μνήμη, CPU προς χρήστη, CPU προς επιταχυντή, CPU προς αποθήκευση. Η αρχιτεκτονική του κέντρου δεδομένων dRedBox βασίζεται στην τοπολογία

τριών επιπέδων, όπως φαίνεται στην εικόνα Σχήμα 4.3.



Εικόνα 4.3: Αρχιτεκτονική Δυκτίου Disaggregated Data Center σε cluster-scale
Πηγή:
https://www.researchgate.net/publication/322867312_Optically_Disaggregated_Data_Centers_With_Minimal_Remote_Memory_Latency_Technologies_Architectures_and_Resource_Allocation_Invited

Η συγκεκριμένη τοπολογία θεωρεί τις παρακάτω μονάδες οπτικής μεταγωγής που ονομάζονται dBOSM (disaggregated Box Optical Switch Module), dROSM (disaggregated Rack Optical Swtich Module) και dDOSM (disaggregated Data Centre Optical Switch Module). Οι οπτικοί switches dBOSM οι οποίοι έχουν μικρό αριθμό θυρών 48 ή 96 μπορούν να φιλοξενήσουν έως και 16 dBricks στην βαθμίδα 1. Από την άλλη πλευρά, οι dROSM και οι dDOSM switches στη βαθμίδα 2 και στη βαθμίδα 3 απαιτούν μεγάλο port 384x384. Το switch της βαθμίδας 2 εκτός από την διάφανη συνδεσημότητα των βαθμίδων 1 και 3 παρέχει επίσης πρόσβαση σε συνδεόμενα προγραμματιζόμενα πακέτα μεταγωγών. Οι υπηρεσίες μεταγωγής πακέτων μπορούν να υποστηριχθούν με τη χρήση προγραμματιζόμενης μεταγωγής πακέτων ή κυκλωμάτων που υποστηρίζονται στο dBrick ή είναι προσαρτημένες ως συναρμολογούμενες μονάδες στο dBOSM. Ο κυριότερος λόγος για τη χρήση ενός

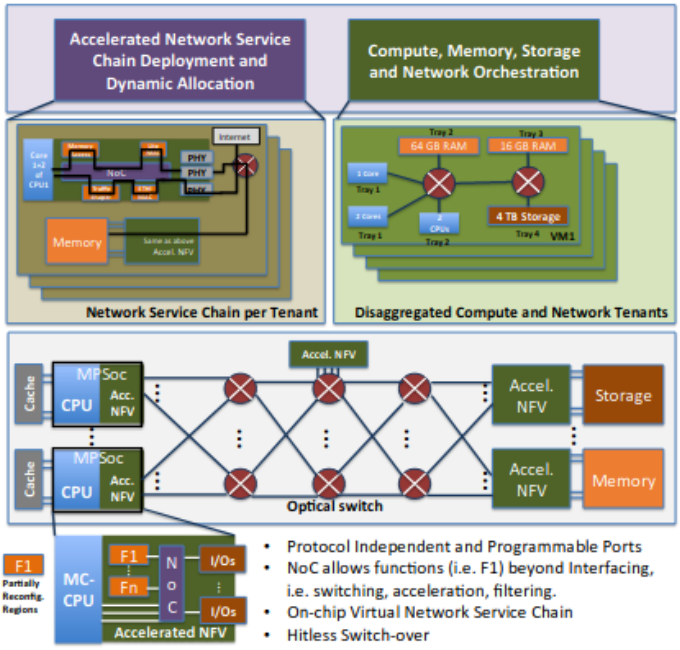
αμιγούς κυκλώματος switching δικτύου (είτε ηλεκτρικού για επικοινωνία εντός του dBox είτε οπτικού για επικοινωνία εντός/εκτός dBox) είναι πως παρέχει τη χαμηλότερη δυνατή καθυστέριση επικοινωνίας CPU με μνήμη. Η προτεινόμενη αρχιτεκτονική διερευνά την υποστήριξη και των δύο DDR4 (μνήμη παράλληλης βάσης) και οπτικά συνδεδεμένης HMC (σειριακή μνήμη) dBricks. Ωστόσο, προτείνει τη χρήση HMC με οπτική σύνδεση ως ιδανική επιλογή, λόγω των ακόλουθων χαρακτηριστικών της σειριακής επικοινωνίας και της εγγενούς συμβατότητας με υψηλής ταχύτητας διασύνδεση και λειτουργίας δικτύου, τη δυνατότητα παροχής πολύ χαμηλών καθυστερήσεων, και την εξάλειψη των πρόσθετου chip (π.χ. ASIC, FPGA, MPSoC) στο dBrick μνήμης στον ελεγκτή μνήμης του κεντρικού υπολογιστή όπως στην περίπτωση DDR4 που με τη σειρά του μειώνει το κόστος, την κατανάλωση ενέργειας και το αποτύπωμα. Ο επόμενος λόγος για τη χρήση του switch οπτικού κυκλώματος είναι η ικανότητά του να επιτρέπει scalability εντός των αρχιτεκτονικών. Επιπλέον, η χρήση ενιαίου οπτικού διακόπτη υπερευρείας ζώνης που προσφέρει εξαιρετικά χαμηλή εισαγωγή απώλειας, δηλαδή 1dB/διασταυρούμενης σύνδεσς, σε συνδυασμό με μονόπλευρες ίνες επιτρέπει τον μέγιστο αριθμό διαφανών μεταπηδήσεων μέσω ενός συστήματος που επιτρέπει στην αρχιτεκτονική να κλιμακώνεται scale-up και scale-out. Ένα άλλο εξίσου σημαντικό πλεονέκτημα της χρήσης switch οπτικών κυκλωμάτων είναι ότι η συγκεκριμένη αρχιτεκτονική λαμβάνει υπόψη τα dBricks όχι ως καθαρές μονάδες υπολογισμού, μνήμης ή επιταχυντή με απλές διεπαφές εισόδου-εξόδου, αλλά ως στοιχεία που επίσης παρέχουν μια σειρά λειτουργιών δικτύωσης, δηλαδή μεταγωγή κυκλωμάτων-πακέτων, προώθησης, παρακολούθησης, ανάλυσης, queuing, επιτάχυνση και υποστήριξη ποικίλων πρωτοκόλλων. Ένας συνδυασμός λειτουργιών, όπως το service chain, μπορεί να συνταχθεί και να συνδεθεί με ένα σύνολο θυρών που είναι διαθέσιμες σε έναν ενιαίο πολυεπεξεργαστή System on Chip (MPSoC). Έτσι, ο συνδυασμός των switch οπτικών κυκλωμάτων στον πυρήνα του δικτύου και του υλικού προγραμματιζόμενης μεταγωγής στο τσιπ για κάθε ένα από τα τελικά σημεία επιτρέπει την αποτελεσματική αναδιαμόρφωση λειτουργίας και τοπολογίας. Ετσι η προτεινόμενη αρχιτεκτονική υπερτερεί στην πιθανότητα φραγής, το κόστος και την κατανάλωση ενέργειας συμβατικής υβριδικής αρχιτεκτονικής πακέτων-κυκλωμάτων. Ειδικότερα, ο συνδυασμός αυτός δημιουργεί μια εξαιρετικά ευέλικτη και εύκολα προγραμματιζόμενη αρχιτεκτονική που μπορεί να εξυπηρετήσει ποικίλες ανάγκες δικτύωσης CPU-to-CPU, CPU-to-Memory, CPU-προς-Storage για την εξυπηρέτηση ενός

αιτήματος VM με disaggregated resources. Για παράδειγμα στην εικόνα 4.4 φαίνεται ότι μια τέτοια αρχιτεκτονική μπορεί να υποστηρίζει

α) Μεταγωγή κυκλώματος on-chip με εξαιρετικά χαμηλή καθυστέρηση μέσω πρωτοκόλλου lite για τη μεταφορά δεδομένων στη μνήμη και οπτική μεταγωγή κυκλώματος για N CPU προς M μονάδες μνήμης που συνδέονται με θύρες για την ικανοποίηση των απαιτήσεων εύρους ζώνης,

β) Μεταγωγή πακέτων με βάση το Ethernet στο τσιπ για τη μεταφορά δεδομένων απο VM σε άλλες VM ή τελικούς χρήστες σε συγκεκριμένες θύρες,

γ) Δίκτυο αποθήκευσης ή πρωτόκολλα αποθήκευσης, π.χ. FibreChannel, για την εξυπηρέτηση μεταφοράς δεδομένων από τη CPU στο Storage.



Εικόνα 4.4: Αρχιτεκτονική λειτουργιών δικτύου και υπολογισμού

Πηγή:

https://www.researchgate.net/publication/322867312_Optically_Disaggregated_Data_Centers_With_Minimal_Remote_Memory_Latency_Technologies_Architectures_and_Resource_Allocation_Invite

Για την υλοποιήση της συγκεκριμένης αρχιτεκτονικής αναπτύχθηκαν οι εξής αλγόριθμοι. First Fit (FF) είναι ο απλούστερος αλγόριθμος που αναπτύχθηκε. Με αυτόν τον αλγόριθμο, οι πόροι κατανέμονται με βάση το ποίος “χωράει” πρώτος, γεγονός που σημαίνει ότι κάθε

αίτηση θα κατανεμηθεί στον πρώτο διαθέσιμο κόμβο (κόμβους) (dBricks) κατά τη διάρκεια της διαδικασία προσδιορισμού των πόρων. Δεδομένου ότι η διαθεσιμότητα του δικτύου δεν λαμβάνεται υπόψη κατά τη διαδικασία κατανομής, μπορεί κάποιοι κόμβοι να διαθέτουν επαρκείς πόρους ΤΠ αλλά να αδυνατούν να ικανοποιήσουν τις απαιτήσεις εύρους ζώνης. Best Fit (BF) μπορεί να θεωρηθεί ως βέλτιστος αλγόριθμος σε σύγκριση με τον προηγούμενο. Στο Best Fit για κάθε αίτημα επιλέγεται ο καλύτερος συνδιασμός πόρων, δεδομένου ότι διαφορετικοί τύποι πόρων (CPUs,μνήμη ή storage) αναζητούνται και κατανέμονται ανεξάρτητα. Το κύριο πλεονέκτημα του είναι ότι μπορεί να μεταπηδά σε διαφορετικά racks για κάθε πόρο το οποίο δεν είναι δυνατό με τον αλγόριθμο FF0.

4.5 Εμπορικές Αρχιτεκτονικές

Οι δημοφιλέστερες αρχιτεκτονικές οπτικών δικτύων διασύνδεσης για Κ.Δ. προτείνονται από ακαδημαϊκά και ερευνητικά κέντρα. Ωστόσο, ορισμένες εταιρείες έχουν προβεί πρόσφατα σε εμπορικά προϊόντα που στοχεύουν στα DCN που βασίζονται σε αμιγώς οπτικές διασυνδέσεις. Για παράδειγμα, η Calient Technologies είναι μεταξύ των πρώτων εταιριών που έχει εμπορευματοποιήσει την οπτική διασύνδεση ρητά για κέντρα δεδομένων. Η Calient προσφέρει μια υβριδική λύση πακέτου-κυκλώματος στην οποία το δίκτυο αποτελείται τόσο από πακέτα μεταγωγής όσο και από OCS.Η υβριδική προσέγγιση απαιτεί την υιοθέτηση ενός δικτύου καθορισμένου από λογισμικό (SDN), το οποίο μπορεί να διαχωρίσει το επίπεδο ελέγχου από το επίπεδο δεδομένων. Η Calient χρησιμοποιεί επί του παρόντος το πρότυπο OpenFlow για την υποδομή SDN. ‘Αλλη μια εταιρία που προσφέρει αρχιτεκτονική οπτικών δικτύων είναι η Plexxi. Η Plexxi είναι μια νεοσύστατη εταιρεία που πρόσφατα εισήγαγε έναν οπτικό μεταγωγέα που στοχεύει στα DCN. Το switch της Plexxi ενσωματώνει τη μεταγωγή Ethernet με ένα κεντρικό SDN που βασίζεται στο επίπεδο ελέγχου. Οι διακόπτες Plexxi είναι βασικά διασυνδεδεμένοι σε μια τοπολογία δακτυλίου χρησιμοποιώντας οπτικό πολυπλέκτη τεχνολογίας LightRail. Το κύριο πλεονέκτημα αυτής της προσέγγισης είναι ότι αντικαθιστά τις παραδοσιακές μεταγωγές ιεραρχίες με ένα κλιμακούμενο σύστημα υψηλού εύρους ζώνης και χαμηλής καθυστέρησης. Η αρχιτεκτονική επίπεδου δακτυλίου επιτρέπει τη γραμμική κλιμάκωση, έτσι με κάθε πρόσθετο switch προσθέτει χωρητικότητα δικτύου. Στην πραγματικότητα, αν και οι Calient και Plexxi παρέχουν διαφορετικές αρχιτεκτονικές, αυτές οι δύο αρχιτεκτονικές μπορούν να συνδυαστούν προκειμένου να παρέχουν πιο ευέλικτες τοπολογίες και κλιμακούμενες λύσεις

με ακόμη μικρότερη καθυστέρηση.Σε μια τέτοια υβριδική αρχιτεκτονική, μεταγωγείς κέντρων δεδομένων θα συνδέονται μέσω του μεταγωγέα Plexxi, ο οποίος με τη σειρά του θα συνδεόταν με το Calient μέσω θυρών 10GbE ή 40GbE. Οι μεταγωγείς Plexxi συνδέονται μεταξύ τους μέσω μιας οπτικής διασύνδεσης, και όταν υπάρχει μεγάλη ροή κίνησης, οι μεταγωγείς Plexxi την παρακάμπτουν από τον μεταγωγέα εισόδου στον οπτικό ιστό της Calient προκειμένου να μειωθεί η καθυστέρηση. Το κύριο πλεονέκτημα αυτής της προσέγγισης είναι ότι προστατεύει το δίκτυο από τη συμφόρηση και εγγυάται επίσης υψηλό εύρος ζώνης με χαμηλή καθυστέρηση για μεγάλου όγκου ροές δεδομένων. Ορισμένες εταιρείες παρέχουν αποκλειστικά οπτικές αρχιτεκτονικές που βασίζονται σε προηγμένες οπτικές τεχνολογίες μεταγωγής. Για παράδειγμα, η Polatis προσφέρει ένα οπτικό δίκτυο διασύνδεσης για ενδοεπικοινωνία κέντρων δεδομένων. Ο οπτικός διακόπτης Polatis βασίζεται σε πιεζοηλεκτρικό OCS και σε σύστημα διεύθυνσης δέσμης τεχνολογίας. Ως εκ τούτου, το παρεχόμενο σύστημα βασίζεται σε έναν συγκεντρωτικό οπτικό διακόπτη που μπορεί να αναδιαμορφωθεί με βάση την κυκλοφορία του δικτύου. Τα σημαντικότερα χαρακτηριστικά του είναι η σχετικά χαμηλή ισχύς του διακόπτη, η κατανάλωση και η ικανότητα του ρυθμού δεδομένων που σημαίνει ότι μπορεί να υποστηρίξει 10, 40 και 100 Gb/s. Το μόνο μειονέκτημα αυτής της εμπορικής αρχιτεκτονικής είναι ότι βασίζεται σε οπτικά MEMS διακόπτες, και συνεπώς έχει αυξημένη αναδιαμόρφωση χρόνου (σύμφωνα με τα δελτία δεδομένων ο μέγιστος χρόνος μεταγωγής είναι μικρότερος από 20 ms).

ΚΕΦΑΛΑΙΟ 5

ΜΕΛΛΟΝΤΙΚΕΣ ΕΞΕΛΙΞΕΙΣ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

ΔΗΜΙΟΥΡΓΙΑ ΛΥΣΗΣ DISAGGREGATION

Οι αναδυόμενες τεχνολογίες, όπως το 5G, δημιουργούν την απαίτηση για τις εταιρείες να δημιουργήσουν μια σειρά από micro Data Centers στην κορυφή του δικτύου κινητής τηλεφωνίας, προκειμένου να επεξεργάζονται ταχύτερα και αποτελεσματικότερα τα δεδομένα πιο κοντά στην πηγή. Σύμφωνα με το www.networkworld.com τα βασικά βήματα που πρέπει να ακολουθήσουν οι οργανισμοί όταν προσεγγίζουν το disaggregation είναι τα εξής:

Βήμα 1: Σχεδιασμός και επικύρωση

Είναι ουσιώδους σημασίας, να γίνει ο σχεδιασμός και η επικύρωση των σχεδίων υποδομής πριν από οποιαδήποτε αγορά καθώς έτσι επιλέγονται τα κατάλληλα υλικά για disaggregation. Οι εκάστοτε οργανισμοί θα πρέπει να βασίζονται σε συνεργάτες που διαθέτουν αποδεδειγμένα σχέδια και αρχιτεκτονικές που περιλαμβάνουν βασική τεχνολογία αυτοματισμού και εικονικοποίησης μαζί με λογισμικό παρακολούθησης για τη μείωση του κινδύνου και τη σημαντική επιτάχυνση του χρόνου διάθεσης στην αγορά.

Βήμα 2: Διαμόρφωση και δοκιμή

Οι εικονικοποιημένες και διαχωρισμένες λύσεις απαιτούν πολλά κινούμενα μέρη, ενώ επίσης η κλιμάκωση γίνεται το μεγαλύτερο εμπόδιο. Οπότε οποιοσδήποτε συνεργάτης επιλεγεί πρέπει να είναι απόλυτα εξοικειωμένος με τις μοναδικές ανάγκες της αλυσίδας εφοδιασμού σας, προκειμένου να βελτιωθεί η διαθεσιμότητα των εξαρτημάτων. Ωστόσο η ύπαρξη των κατάλληλων εγκαταστάσεων είναι σημαντική για την προμήθεια, την ενσωμάτωση και τη δοκιμή της λύσης, ώστε να διασφαλιστεί ότι λειτουργεί όπως προβλέπεται σε πραγματικές συνθήκες.

Βήμα 3: Ανάπτυξη

Αφού έχουν γίνει οι απαραίτητες δοκιμές, ο εκάστοτε συνεργάτης πληροφορικής θα πρέπει να είναι σε θέση να διαμορφώνει ολόκληρα racks, να τα τοποθετεί σε κιβώτια και να τα αποστέλλει ως ολοκληρωμένη, προελεγμένη λύση. Έτσι ο πελάτης λαμβάνει τα racks και απλά τα συνδέει.

Βήμα 4: Λειτουργία και διαχείριση

Μετά την εγκατάσταση θα πρέπει να υπάρχει συνεχής παρακολούθηση και διαχείριση του disaggregation, όπως και αποκατάσταση προβλημάτων σε περίπτωση που προκύψουν αλλά και υποστήριξη για τη βελτιστοποίηση της αρχιτεκτονικής στο μέλλον.

Disaggregated Operating System

Για την σωστή υλοποίηση του disaggregation χρειάζεται να υπάρχει και το κατάλληλο λειτουργικό σύστημα. Η εταιρία WukLab προτείνει ένα νέο μοντέλο λειτουργικού συστήματος που ονομάζεται splitkernel για τη διαχείριση των disaggregated συστημάτων. Το splitkernel παρέχει τις παραδοσιακές λειτουργικότητες του λειτουργικού συστήματος σε χαλαρά συνδεδεμένες οθόνες, όπου σε καθεμία από τις οποίες εκτελείται και διαχειρίζεται ένα στοιχείο του hardware. Ένας splitkernel εκτελεί επίσης την κατανομή πόρων και τον χειρισμό αποτυχιών ενός κατανεμημένου συνόλου στοιχείων υλικού. Χρησιμοποιώντας το μοντέλο splitkernel, κατασκευάστηκε το LegoOS, ένα νέο λειτουργικό σύστημα σχεδιασμένο για τη διάσπαση πόρων υλικού. Το LegoOS εμφανίζεται στους χρήστες ως ένα σύνολο κατανεμημένων servers. Εσωτερικά, μια εφαρμογή χρήστη μπορεί να καλύπτει πολλαπλά στοιχεία υλικού επεξεργαστή, μνήμης και αποθήκευσης. Τα αποτελέσματα της αξιολόγησής δείχνουν ότι η απόδοση του LegoOS σε x86-64 είναι συγκρίσιμη με μονολιθικούς servers Linux, ενώ βελτιώνει σε μεγάλο βαθμό τη συσκευασία πόρων και μειώνει το ποσοστό αποτυχίας σε σχέση με τα μονολιθικά clusters.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Μέσα σε μόλις δύο δεκαετίες, τα Κέντρα Δεδομένων έχουν κλιμακωθεί από το μέγεθος ενός δωματίου στο μέγεθος ενός ολόκληρου κτηρίου δίνοντας τη δυνατότητα να εξυπηρετήσουν την εκθετικά αυξανόμενη ανάγκη αποθήκευσης. Εκτός όμως από την αποθήκευση, τα σύγχρονα DC’s αναβαθμίζονται επίσης για να μπορούν να εξυπηρετούν περισσότερες υπηρεσίες. Είναι πιο συνδεδεμένα από ποτέ και μπορούν να ανταποκριθούν στις ανάγκες του σύγχρονου επιχειρηματικού κόσμου καθώς έχουν προκύψει νέες λύσεις γύρω από την αρχιτεκτονική τους που μπορούν να προσφέρουν ανταγωνιστικά πλεονεκτήματα στους χρήστες μέσω πιο βελτιστοποιημένων επιδόσεων. Έτσι τα Κέντρα Δεδομένων έχουν γίνει πλέον κρίσιμα στοιχεία μιας σύγχρονης υποδομής IT. Ωστόσο μετά τα πρόσφατα γεγονότα της πανδημίας αλλά και της εισβολής στην Ουκρανία το μέλλον είναι αβέβαιο ακόμα και για τα Data Centers. Οι παράγοντες που επηρεάζουν το μελλόν αυτό ποικίλουν και κάποιοι απο

αυτούς είναι η κλιματική αλλαγή, η ανθεκτικότητα και διαθεσιμότητα,οι συγχωνεύσεις και εξαγορές οι οποίες μετατοπίζονται στην αλυσίδα εφοδιασμού των κέντρων δεδομένων και η υγρή ψύξη. Τα Κέντρα Δεδομένων προσφέρουν πλέον συγκλίνουσες προς το cloud υποδομές και η τάση κινείται περαιτέρω προς το νεφός αποκλειστικά. Αυτό επιφέρει πολλά πλεονεκτήματα για τις λειτουργίες του Data Center ενώ επίσης συμφέρει και τους πελάτες της διανομής όπως αναλύθηκε στο κεφάλαιο 1.Για παράδειγμα, ο κίνδυνος βλάβης υλικού, ταλαιπωρούσε τις εταιρείες με τον κίνδυνο απώλειας δεδομένων και προσπαθούσαν να επαναφέρουν την υποδομή τους. Οι απομονωμένες προσεγγίσεις για τη διαχείριση των servers ήταν μια άλλη πρόκληση που καθιστούσε τις λειτουργίες του Data Center ακριβές και περίπλοκες. Με τις cloud υποδομές, η διαδικασία διαχείρισης ενός DC οργανώνεται σε ένα ενιαίο περιβάλλον εργασίας που χρησιμοποιείται για τη διαχείριση της υποδομής, και στη διατήρηση της ασφάλειας των δεδομένων στο cloud. Ενώ η διαχείριση καθίσταται ευκολότερη μέσω της ενοποίηση των λειτουργιών στο cloud. Όσον αφορά τα DCN ένα σημαντικό κριτήριο για την υιοθέτηση των οπτικών διασυνδέσεων από τους φορείς εκμετάλλευσης κέντρων δεδομένων τα επόμενα χρόνια είναι η αποτελεσματική ανάπτυξη κλιμακούμενων λύσεων που μπορούν να φιλοξενήσουν τις εκατοντάδες χιλιάδες servers φυσικά ή εικονικά σε ένα κέντρο δεδομένων. Επίσης ο αναδυόμενος τομέας της οπτικής διασύνδεσης δικτύων έχει ανοίξει νέους ορίζοντες για τις υπερυψηλής χωρητικότητας δίκτυα κέντρων δεδομένων που μπορούν να προσφέρουν χαμηλή καθυστέρηση και μειωμένη ισχύ κατανάλωση ενέργειας. Υπάρχουν αρκετές αρχιτεκτονικές που υπόσχονται να προσφέρουν σημαντικά πλεονεκτήματα σε σχέση με τα τρέχοντα δίκτυα DCN που βασίζονται σε switches. Ωστόσο, προκειμένου να υιοθετηθούν ευρέως από τους φορείς εκμετάλλευσης κέντρων δεδομένων, οι τεχνολογίες οπτικών διασυνδέσεων πρέπει να ξεπεράσουν αρκετές σημαντικές προκλήσεις, όπως η ανάγκη για ενισχυμένη επεκτασιμότητα και ανθεκτικότητα, καθώς και μειωμένο κόστος.

Σκοπός της παρούσας εργασίας ήταν να αποσαφηνιστεί και να αναλυθεί η έννοια των Disaggregated Data Centers, τα οποία παρότι είναι ακόμα υπό ανάπτυξη αναμένεται να επιτύχουν πολύ καλύτερη αξιοποίηση των πόρων σε σύγκριση με τα παραδοσιακά κέντρα δεδομένων. Στο κεφάλαιο αυτό αναλύονται τα συμπεράσματα τα οποία προέκυψαν απο την έρευνα για την εργασία, καθώς και πως προβλέπεται σύμφωνα με τα σημερινά δεδομένα το μέλλον των Data Centers. Είναι σημαντικό να σημειωθεί πως ακόμη και με εξαιρετικά υψηλές

ταχύτητες οπτικής μετάδοσης, η χωρητικότητα της επικοινωνίας μεταξύ των πόρων δεν μπορεί να θεωρηθεί απεριόριστη. Συμπεραίνεται επίσης πως τα οφέλη των Disaggregated DC's μπορεί να μην αποτελούν ουτοπία ή ακόμη και να μην υπάρχουν, λόγω του γεγονότος ότι το εύρος ζώνης που παρέχεται από τις τεχνολογίες επικοινωνίας οπτικών ινών δεν επαρκεί. Επομένως χρειάζεται παραπάνω έρευνα ώστε να επιτευχθεί οικονομικά αποδοτική οπτική μετάδοση μικρής εμβέλειας με μεγαλύτερο εύρος ζώνης. Ωστόσο υπάρχει επιτακτική ανάγκη για την επιτυχία του fully disaggregation καθώς έχει πρόσφατα αναδειχθεί ως μια ισχυρή λύση για την αντιμετώπιση της πρόκλησης της αυξανόμενης κατανάλωσης πόρων των κέντρων δεδομένων. Και ιδιαίτερα απ' την στιγμή που αυτές οι εγκαταστάσεις καταναλώνουν ολοένα και μεγαλύτερη ποσότητα ενέργειας, με ορισμένα από τα μεγαλύτερα κέντρα δεδομένων στον κόσμο να απαιτούν περισσότερα από 100 MW. Όπως αναφέρθηκε και στο κεφάλαιο 4 το όραμα του διαχωρισμένου κέντρου δεδομένων, προβλέπει πως οι υπολογιστικές μονάδες (CPUs, GPUs) αποσυνδέονται από την ιεραρχία μνήμης, με όλα τα στοιχεία να συνδέονται μέσω του datacenter fabric. Έτσι αποτρέπεται η αύξηση του κόστους εφόσον δίνεται έμφαση στην κατανομή των πόρων just-in-time και στην επαναχρησιμοποίηση με την *"pay for what you need, use only what you need"* φιλοσοφία να κυριαρχεί.

ΒΙΒΛΙΟΓΡΑΦΙΑ

(1) https://www.diva-portal.org/smash/get/diva2:1292615/FULLTEXT01.pdf
(2) https://www.zdnet.com/article/how-system-disaggregation-would-reorganize-it-and-how-arm-may-benefit/
(3) https://www.infinera.com/blog/is-network-router-disaggregation-inevitable/tag/access-and-aggregation/
(4) http://www.cs.cornell.edu/~ragarwal/pubs/disaggregation.pdf
(5) https://www.datacenterknowledge.com/archives/2013/10/18/storage-disaggregation-in-the-data-center
(6) **https://www.researchgate.net/profile/Qi-Zhang-126/publication/336945502_Memory_Disaggregation_Research_Problems_and_Opportunities/links/5f6370e4458515b7cf39c9f7/Memory-Disaggregation-Research-Problems-and-Opportunities.pdf**
(7) **https://people.csail.mit.edu/alizadeh/courses/6.888/papers/disagg.pdf**
(8) **https://packetpushers.net/demystifying-dcn-topologies-clos-fat-trees-part1/**
(9) **https://community.fs.com/blog/data-center-switch-wiki-usage-buy-tips.html**
(10) **https://www.networkworld.com/article/3643388/data-center-disaggregation-101-how-to-build-a-new-edge-computing-infrastructure.html**
(11) **https://ieeexplore.ieee.org/abstract/document/7842314**
(12) **https://www.datacenterdynamics.com/en/analysis/disaggregated-data-centers-great-idea-but-not-just-yet/**
(13) **https://datacenterfrontier.com/the-eight-trends-that-will-shape-the-data-center-industry-in-2022/**
(14) **https://www.sigarch.org/the-time-is-ripe-for-disaggregated-systems/**
(15) **https://www.researchgate.net/profile/Georgios-Zervas-2/publication/322867312_Optically_Disaggregated_Data_Centers_With_Minimal_Remote_Memory_Latency_Technologies_Architectures_and_Resource_Allocation_Invited/links/5b0295dea6fdccf9e4f6e7d8/Optically-Disaggregated-Data-Centers-With-Minimal-Remote-Memory-Latency-Technologies-Architectures-and-Resource-Allocation-Invited.pdf?origin=publication_detail**
(16) **https://arxiv.org/pdf/2104.04060.pdf**
(17) **https://www.cse.iitb.ac.in/~mythili/teaching/cs641_autumn2015/references/guest_lecture_Oct23/dc-survey-commg.pdf**
(18) https://springerplus.springeropen.com/articles/10.1186/s40064-016-2454-4
(19) https://dc.mynetworkinsights.com/what-is-three-tier-architecture-and-fat-tree-architecture/
(20) https://www.ciena.com/insights/what-is/What-is-DCI.html
(21) Cloud Computing. Αρχές, Τεχνολογία & Αρχιτεκτονική Thomas Erl. Εκδώσεις: Μ. Γκιούρδας.
(22) **Διαφάνειες Υπολογιστικής Νέφους ΠΑΔΑ**
(23) Διαφάνειες Δικτύωσης ορισμένης απο Λογισμικό ΠΑΔΑ