ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΜΗΜΑΤΟΣ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΘΕΜΑΤΑ ΑΣΦΑΛΕΙΑΣ BIG DATA ΣΤΟ ΥΠΟΛΟΓΙΣΤΙΚΟ ΝΕΦΟΣ (CLOUD COMPUTING)

Χατζηκουντέλη Ζαχαρένια

Επιβλέπων καθηγητής: Ψάννης Κωνσταντίνος

Θεσσαλονίκη, 2018

Ευχαριστίες

Με την ολοκλήρωση του Μεταπτυχιακού Προγράμματος Σπουδών στο τμήμα της Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας, είναι απαραίτητο να εκφράσω τις ευχαριστίες μου σε όλους τους ανθρώπους, οι οποίοι με βοήθησαν και με στήριξαν στην πορεία αυτή.

Κυρίως, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα της διπλωματικής μου εργασίας, κ. Ψάννη Κωσταντίνο, για τις παρατηρήσεις, τις κατάλληλες υποδείξεις, καθώς επίσης και για την έγκαιρη βοήθεια που μου προσέφερε καθ' όλο το διάστημα εκπόνησης της παρούσας εργασίας. Η καθοδήγησή του και η κριτική του συνέβαλαν καθοριστικά κατά τη συγγραφή της εργασίας αυτής.

Τέλος, ένα μεγάλο ευχαριστώ σε όσους συνέβαλλαν, ακόμη και στον μικρότερο βαθμό, στην επιτυχημένη ολοκλήρωση της διπλωματικής μου εργασίας.

Πίνακας περιεχομένων

Ευχαριστίες……...2

Περίληψη……...5

Abstract……6

Κεφάλαιο 1: Εισαγωγή
1.1 Αιτιολόγηση του ερευνητικού θέματος …..……7
1.2 Χαρακτηριστικά των Μεγάλων Δεδομένων ……...8
1.3 Προκλήσεις των Μεγάλων Δεδομένων ……..9
1.4 Στόχος της έρευνας……10
1.5 Επισκόπηση της έρευνας……..11

Κεφάλαιο 2: Βιβλιογραφική ανασκόπηση
2.1 Εισαγωγή ……...12
2.2 Ανάπτυξη και καινοτομίες των Μεγάλων Δεδομένων……...12
2.3 Επεξεργασία των Μεγάλων Δεδομένων……15
2.4 Ασφάλεια των Μεγάλων Δεδομένων……20
2.5 Κριτήρια για την Σύγκριση της Βιβλιογραφικής Ανασκόπησης……...26

Κεφάλαιο 3: Big Data Analytics Software……..30

Κεφάλαιο 4: Παραδείγματα – Εφαρμογές των Μεγάλων Δεδομένων……....36

Κεφάλαιο 5: Κρυπτογράφηση – Encryption
5.1. Εισαγωγή στην κρυπτογράφηση……..39
5.2 Βασικά στοιχεία και λειτουργίες της κρυπτογράφησης……39

Κεφάλαιο 6: Σκοπός της έρευνας……43

Κεφάλαιο 7 : Σύγκριση και Ανάλυση
7.1 Κοινή χρήση ευαίσθητων δεδομένων σε πλατφόρμα μεγάλων δεδομένων….44
7.1.1 Πλαίσιο ασφαλής κοινής χρήσης ευαίσθητων δεδομένων……44
7.1.2 Heterogeneous Proxy Re-Encryption (H-PRE)……..44
7.1.3 Λειτουργίες υποβολής, αποθήκευσης και εξαγωγής δεδομένων……...45
7.1.4 Ασφαλής χρήση ευαίσθητων δεδομένων στο VMM…….45
7.2 Πολύ-κρυπτογράφηση δεδομένων……46
7.2.1 Εισαγωγή στην Πολύ-κρυπτογράφηση……..46
7.2.2 Παρουσίαση των δυο προσεγγίσεων……..47
7.2.3 Ανάλυση των αποτελεσμάτων ……..48
7.3 Σύγκριση των μεθόδων που παρουσιάστηκαν……..48

Κεφάλαιο 8: Cloud Computing – Νέφος
8.1 Εισαγωγή στο Cloud Computing……...50
8.2 Χαρακτηριστικά του Cloud Computing……50
8.3 Μοντέλα Υπηρεσιών (Service models)……51
8.3.1 Υποδομή ως υπηρεσία (Infrastructure as a Service -IaaS)……..52
8.3.2 Η πλατφόρμα ως υπηρεσία (Platform as a Service - PaaS)…….52
8.3.3 Το λογισμικό ως υπηρεσία (Software as a Service - SaaS)…….53
8.4 Μοντέλα ανάπτυξης (Deployment Models)……..54
8.5 Αρχιτεκτονική……...56
8.6 Ασφάλεια και προστασία της ιδιωτικότητας…….56

Κεφάλαιο 9 : CloudAnalyst
9.1 Εισαγωγή……..58
9.1.1 Ορολογία και συντομογραφίες……....58
9.2 Χαρακτηριστικά του προγράμματος προσομοίωσης……...59
9.3 Χρησιμοποιώντας το CloudAnalyst…….....59
9.3.1 Δημιουργία προσομοίωσης……...59
9.3.2 Οθόνες προσομοιωτή……60
9.3.2.1 Κύρια οθόνη με Simulation Panel……...60
9.3.2.2 Οθόνη Configure Simulation …….61
9.3.3 Εκτέλεση προσομοίωσης……...64
9.3.4 Οθόνη αποτελεσμάτων……64
9.4 Προσομοίωση εφαρμογής μεγάλης κλίμακας Internet που εκτελείται στο Cloud……..65
9.4.1 Διαμόρφωση προσομοίωσης…….66
9.5 Senario 1 - Απλή εφαρμογή στο Web που φιλοξενείται σε ένα ενιαίο κέντρο δεδομένων……...67
9.6 Senario 2 - Εφαρμογή στο Web που φιλοξενείται σε πολλαπλά κέντρα δεδομένων στον κόσμο……...……69
9.6.1 Περίπτωση 1: Δύο κέντρα δεδομένων με 25 VM στο κάθε ένα……70
9.6.2 Περίπτωση 2: Δύο κέντρα δεδομένων με 50 VM στο καθένα……...72
9.6.3 Περίπτωση 3: Δύο κέντρα δεδομένων με 50 VM το καθένα και φορτίο κατανομής κατά τη διάρκεια των ωρών αιχμής…….73
9.6.4 Περίπτωση 4: Εφαρμογή στο Web που φιλοξενείται σε 3 κέντρα δεδομένων με 50VM το καθένα…….74
9.6.5 Περίπτωση 5: Η εφαρμογή Web φιλοξενείται σε 3 κέντρα δεδομένων με 75, 50, 25 VM στο καθένα ……74
9.7 Σύνοψη αποτελεσμάτων προσομοίωσης……...76
9.8 Κύριες Παρατηρήσεις από τα Αποτελέσματα……..77

Κεφάλαιο 10: Συμπεράσματα - Μελλοντική Έρευνα…….78

Βιβλιογραφία ……..79

Περίληψη

Ο σκοπός της εργασίας αυτής είναι να ερευνήσει τα μεγάλα δεδομένα καθώς έχουν κατακλύσει τον κόσμο τα τελευταία χρόνια και αποτελούν σημαντικό κομμάτι της τεχνολογίας. Τα μεγάλα δεδομένα περιλαμβάνουν ευαίσθητα προσωπικά δεδομένα πράγμα που σημαίνει ότι ελλοχεύουν κίνδυνοι ως προς την προστασία τους. Για να ερευνηθεί αυτή η πλευρά έγινε εκτενής βιβλιογραφική έρευνα σε τρεις κατηγορίες που είναι σχετικές με την ανάπτυξη και τις καινοτομίες, την επεξεργασία και την ασφάλεια των μεγάλων δεδομένων. Στην συνέχεια έγινε σύγκριση μεταξύ δυο πηγών ως προς την ασφάλεια των μεγάλων δεδομένων με την χρήση της κρυπτογράφησης. Μετά από ανάλυση που έγινε παρουσιάστηκαν οι δύο μέθοδοι και συγκρίθηκαν ώστε να δούμε την κάθε σκοπιά πάνω στην ασφάλεια των δεδομένων. Επιπλέον, προχωρήσαμε στην σύνδεση του νέφους με τα μεγάλα δεδομένα καθώς αυτός ο τεράστιος όγκος δεδομένων αποθηκεύεται στο Cloud. Έγινε προσομοίωση με την χρήση προγράμματος προσημείωσης ώστε να μπορέσουμε να κατανοήσουμε τον νέφος και τον διαμοιρασμό των πόρων σε όλο τον κόσμο από έναν αριθμό χρηστών. Τα αποτελέσματα ήταν επιθυμητά και μπορούν να δεχτούν και περαιτέρω βελτίωση σε μελλοντική έρευνα.

Λέξεις κλειδιά: Μεγάλα Δεδομένα, νέφος, ασφάλεια, ευαίσθητα δεδομένα.

Abstract

The purpose of this work is to explore Big Data as they have overwhelmed the world and they form a significant part of technology in recent years. Big data include sensitive personal data, which means that risks loom over their protection. To investigate this aspect, extensive literature review has been carried out in three categories, which is Development and Innovation, Processing and Security of Big Data. Furthermore a comparison was made between two sources of Big Data security using encryption method. After analyzing them, the two sources were presented to look into every aspect of data security. Continuing, we look into the connection of Cloud and Big Data as this huge amount of data is stored in Cloud. A simulation was made using a footer program so we can understand the Cloud and resource sharing around the world by a number of users. The results were desirable and they can be further improved in a future study.

Keywords: Big Data, Cloud Computing, security, sensitive data

Κεφάλαιο 1: Εισαγωγή

1.1 Αιτιολόγηση του ερευνητικού θέματος

Η δημιουργία δεδομένων πλεόν πραγματοποιείται με γρήγορους ρυθμούς. Το 2010, ο κόσμος δημιούργησε πάνω από 1 ZettaByte δεδομένων και μέχρι το 2014, δημιουργήθηκαν 7 ZettaByte ετησίως. Μεγάλο μέρος αυτής της έκρηξης δεδομένων είναι το αποτέλεσμα μιας δραματικής αύξησης των συσκευών που βρίσκονται συνδεδεμένα με το διαδίκτυο, συμπεριλαμβανομένων των ενσωματωμένων αισθητήρων, των smartphones και των υπολογιστών και tablet. Όλα αυτά τα δεδομένα δημιουργούν νέες ευκαιρίες για "εξαγωγή περισσότερης αξίας" στην ανθρώπινη ζωή, την υγειονομική περίθαλψη, το πετρέλαιο και το φυσικό αέριο, την έρευνα, την επιτήρηση, τη χρηματοδότηση και πολλούς άλλους τομείς. Έτσι γίνεται η είσοδος στην εποχή των "μεγάλων δεδομένων".

Η International Data Corporation (IDC) πιστεύει ότι οι οργανισμοί που είναι σε καλύτερη θέση να λαμβάνουν επιχειρηματικές αποφάσεις σε πραγματικό χρόνο χρησιμοποιώντας τα Big Data θα αναπτυχθούν, ενώ όσοι δεν μπορούν να αγκαλιάσουν και να χρησιμοποιήσουν αυτή τη μετατόπιση, θα βρεθούν όλο και περισσότερο σε ανταγωνιστικό μειονέκτημα στην αγορά και θα αντιμετωπίσουν ενδεχόμενη αποτυχία.

Τα μεγάλα δεδομένα είναι ένας όρος που περιγράφει τον μεγάλο όγκο δεδομένων - τόσο δομημένων όσο και μη δομημένων - που κατακλύζουν μια επιχείρηση καθημερινά. Αλλά δεν είναι το ποσό των δεδομένων που είναι σημαντικό αλλά αυτό που κάνουν οι οργανισμοί με τα δεδομένα έχει σημασία. Μεγάλα δεδομένα μπορούν να αναλυθούν για ιδέες που οδηγούν σε καλύτερες αποφάσεις και στρατηγικές επιχειρηματικές κινήσεις.

Οι τεχνολογίες των Big Data περιγράφουν μια νέα γενιά τεχνολογιών και αρχιτεκτονικών, σχεδιασμένων έτσι ώστε οι οργανισμοί να μπορούν να εξάγουν σημαντικά στοιχεία από πολύ μεγάλους όγκους μιας ευρείας ποικιλίας δεδομένων, επιτρέποντας τη λήψη, ανακάλυψη και ανάλυση υψηλής ταχύτητας. Αυτός ο κόσμος των μεγάλων δεδομένων απαιτεί μια μετατόπιση της αρχιτεκτονικής υπολογιστών, έτσι ώστε οι χρηστες να μπορούν να χειριστούν τόσο τις απαιτήσεις αποθήκευσης δεδομένων όσο και την επεξεργασία μεγάλου διακομιστή που απαιτείται για την οικονομική ανάλυση μεγάλων όγκων δεδομένων.

Οι οργανισμοί βασίζονται σε ένα αυξανόμενο σύνολο εφαρμογών για την επικοινωνία και την παροχή υπηρεσιών και προϊόντων στις σημερινές απαιτητικές καταναλωτικές και επιχειρηματικές κοινότητες:

- Συλλέγουν, αποθηκεύουν και αναλύουν πιο λεπτομερείς πληροφορίες για περισσότερα προϊόντα, άτομα και συναλλαγές από ποτέ άλλοτε.
- Στηρίζονται σε εργαλεία ηλεκτρονικού ταχυδρομείου, εργαλεία συνεργασίας και κινητές συσκευές για την επικοινωνία και τη διεξαγωγή επιχειρήσεων με πελάτες και επιχειρηματικούς συνεργάτες.
- Δημιουργούν, λαμβάνουν και συλλέγουν μηνύματα που παράγονται από μηχάνημα και αισθητήρες, μερικές φορές σε πολύ μεγάλους όγκους, και οδηγούν τις λειτουργίες και τις επιχειρηματικές διαδικασίες από αυτά τα δεδομένα μηνυμάτων.

Αυτός ο πολλαπλασιασμός των δεδομένων οδηγεί επίσης στη ζήτηση για κεντρική χωρητικότητα αποθήκευσης για τη μεγιστοποίηση του ελέγχου και της χρησιμότητας των πληροφοριών που συλλέγονται. Στο πλαίσιο του νέου χαρακτήρα αυτής της έκρηξης δεδομένων, οι οργανισμοί σε όλο τον κόσμο εγκατέστησαν 6.1 ExaByte χωρητικότητας αποθήκευσης δίσκων το 2007. Έως το 2010, οι ετήσιες νέες εγκαταστάσεις ήταν 16.4 Exabyte , και μέχρι το 2014 φτάνει στα 79.8 ExaByte.

Τα μεγάλα δεδομένα αφορούν την αυξανόμενη πρόκληση που αντιμετωπίζουν οι οργανώσεις καθώς αντιμετωπίζουν μεγάλες και ταχέως αναπτυσσόμενες πηγές δεδομένων ή πληροφοριών που παρουσιάζουν επίσης ένα πολύπλοκο φάσμα προβλημάτων ανάλυσης και χρήσης. Αυτά μπορεί να περιλαμβάνουν μια υπολογιστική υποδομή που μπορεί να επικυρώνει και να αναλύει μεγάλους όγκους δεδομένων, την αξιολόγηση μικτών δεδομένων (δομημένων και αδόμητων) από πολλαπλές πηγές, την αντιμετώπιση απρόβλεπτου περιεχομένου χωρίς εμφανές σχήμα ή δομή και τέλος την ενεργοποίηση συλλογής, ανάλυσης και απαντήσεων σε πραγματικό χρόνο ή σχεδόν σε πραγματικό χρόνο. Οι τεχνολογίες των

Big Data περιγράφουν μια νέα γενιά τεχνολογιών και αρχιτεκτονικών, που αποσκοπούν στην οικονομική εξάπλωση της αξίας από πολύ μεγάλους όγκους μιας ευρείας ποικιλίας δεδομένων.

Νέες πηγές δεδομένων για τα Μεγάλα Δεδομένα αποτελούν οι βιομηχανίες που μόλις πρόσφατα άρχισαν να ψηφιοποιούν το περιεχόμενό τους. Σε όλες σχεδόν τις περιπτώσεις, οι ρυθμοί αύξησης των δεδομένων τα τελευταία χρόνια ήταν σχεδόν άπειροι, καθώς στις περισσότερες περιπτώσεις ξεκίνησε από το μηδέν. Οι βιομηχανίες περιλαμβάνουν:

- Μέσα ψυχαγωγίας: Η βιομηχανία των μέσων ενημέρωσης / ψυχαγωγίας μετακόμισε στην ψηφιακή καταγραφή, παραγωγή και παράδοση τα τελευταία χρόνια και τώρα συλλέγει μεγάλες ποσότητες εμπλουτισμένου περιεχομένου.
- Υγειονομική περίθαλψη: Ο κλάδος της υγειονομικής περίθαλψης κινείται γρήγορα σε ηλεκτρονικά ιατρικά αρχεία και εικόνες, τα οποία επιθυμεί να χρησιμοποιήσει για βραχυπρόθεσμη παρακολούθηση της δημόσιας υγείας και μακροπρόθεσμα προγράμματα επιδημιολογικής έρευνας.
- Επιστήμες της ζωής: Η αλληλουχία γονιδίων χαμηλού κόστους (<$ 1.000) μπορεί να δημιουργήσει δεκάδες terabytes πληροφοριών που πρέπει να αναλυθούν για να αναζητήσουν γενετικές παραλλαγές και πιθανή αποτελεσματικότητα θεραπείας.
- Παρακολούθηση βίντεο: Η παρακολούθηση βίντεο εξακολουθεί να μεταβαίνει από CCTV σε κάμερες IPTV και συστήματα εγγραφής που οι οργανισμοί θέλουν να αναλύσουν για πρότυπα συμπεριφοράς (ασφάλεια και βελτίωση υπηρεσιών).
- Μεταφορές, λιανική, επιχειρήσεις κοινής ωφέλειας και τηλεπικοινωνίες: Τα δεδομένα αισθητήρων παράγονται με γρήγορο ρυθμό από τους πομποδέκτες GPS, τους αναγνώστες RFID tag, τους έξυπνους μετρητές και τα κινητά τηλέφωνα. Τα δεδομένα χρησιμοποιούνται για τη βελτιστοποίηση των λειτουργιών και τη διευκόλυνση της επιχειρησιακής ευφυΐας (BI) για την πραγματοποίηση άμεσων επιχειρηματικών ευκαιριών.

Οι καταναλωτές συμμετέχουν ολοένα και περισσότερο σε μια αγορά υπηρεσιών αυτοεξυπηρέτησης, η οποία όχι μόνο καταγράφει τη χρήση καρτών αλλά μπορεί να συνδυαστεί όλο και περισσότερο με τα κοινωνικά δίκτυα και τα με τα δεδομένα με βάση τη γεωγραφική περιοχή, γεγονός που δημιουργεί ένα χρυσό ορυχείο δεδομένων ενεργών καταναλωτών.

1.2 Χαρακτηριστικά των Μεγάλων Δεδομένων

Οι τρεις ιδιότητες που παρουσιάζονται παρακάτω καθορίζουν την επέκταση ενός συνόλου δεδομένων κατά μήκος διαφόρων μέτωπων όπου αξίζει να ονομάζονται Μεγάλα Δεδομένα. Μια επέκταση που επιταχύνεται για τη δημιουργία ακόμα περισσότερων δεδομένων διαφόρων τύπων. Τα παρακάτω 3Vs ορίζουν εντελώς τα Μεγάλα Δεδομένα με παρόμοιο τρόπο.

- Όγκος δεδομένων (Data Volume):

Το μέγεθος των διαθέσιμων δεδομένων αυξάνεται με επικίνδυνα γρήγορο ρυθμό πράγμα που ισχύει για τις εταιρείες και για τα άτομα. Ένα αρχείο κειμένου είναι μερικά bytes, ένα αρχείο ήχου είναι μερικά mega byte ενώ μια ταινία πλήρους μήκους είναι μερικά giga bytes.

Περισσότερες πηγές δεδομένων προστίθενται σε συνεχή βάση. Για τις εταιρείες, όλα τα δεδομένα δημιουργήθηκαν εσωτερικά από τους υπαλλήλους, επί του παρόντος, τα δεδομένα δημιουργούνται από υπαλλήλους, συνεργάτες και πελάτες. Για μια ομάδα εταιρειών, τα δεδομένα παράγονται επίσης από μηχανές. Για παράδειγμα, εκατοντάδες εκατομμύρια έξυπνα τηλέφωνα στέλνουν μια ποικιλία πληροφοριών στην υποδομή του δικτύου. Αυτά τα δεδομένα δεν υπήρχαν πριν από κάποια χρόνια. Περισσότερες πηγές δεδομένων με μεγαλύτερο μέγεθος δεδομένων συνδυάζονται για την αύξηση του όγκου των

δεδομένων που πρέπει να αναλυθούν. Αυτό είναι ένα σημαντικό ζήτημα για όσους θέλουν να χρησιμοποιήσουν τα δεδομένα αυτά αντί να τα αφήσουν να εξαφανιστούν.

- Ταχύτητα δεδομένων (Data Velocity):

Αρχικά, οι εταιρείες ανέλυαν δεδομένα χρησιμοποιώντας μια διαδικασία ομαδοποίησης. Πιο αναλυτικά κάποιος παίρνει ένα κομμάτι των δεδομένων, υποβάλλει μια εργασία στο διακομιστή και περιμένει την παράδοση του αποτελέσματος. Αυτό το σχήμα λειτουργεί όταν ο ρυθμός εισερχόμενων δεδομένων είναι βραδύτερος από τον ρυθμό επεξεργασίας της κάθε ομάδας δεδομένων και όταν το αποτέλεσμα είναι χρήσιμο παρά την καθυστέρηση. Με τις νέες πηγές δεδομένων, όπως οι κοινωνικές και κινητές εφαρμογές, η διαδικασία της ομαδοποίησης σπάει. Τα δεδομένα ρέουν προς τον server σε πραγματικό χρόνο, με συνεχή τρόπο και το αποτέλεσμα είναι χρήσιμο μόνο εάν η καθυστέρηση είναι πολύ μικρή.

- Ποικιλία δεδομένων (Data Variety):

Από πίνακες και βάσεις δεδομένων excel, η δομή δεδομένων άλλαξε για να προσθέσει εκατοντάδες διαφορετικές μορφές όπως κείμενο, φωτογραφία, ηχητικό μύνημα, βίντεο, δεδομένα GPS, δεδομένα αισθητήρων, έγγραφα, SMS, pdf και διάφορα άλλα. Η δομή δεν μπορεί πλέον να επιβληθεί όπως στο παρελθόν, προκειμένου να διατηρηθεί ο έλεγχος της ανάλυσης, καθώς εισάγονται νέες εφαρμογές και νέες μορφές δεδομένων.

1.3 Προκλήσεις των Μεγάλων Δεδομένων

Η χρήση δεδομένων για τη δημιουργία επιχειρηματικής αξίας είναι ήδη πραγματικότητα σε πολλές βιομηχανίες με αποτέλεσμα να υπάρχει βελτίωση στην ανάλυσης και αύξηση στην συνδεσιμότητα μέσω της νέας τεχνολογίας και του λογισμικού που προσφέρουν σημαντικές ευκαιρίες. Ωστόσο, βλέπουμε πως εκτός από τα πλεονέκτηματα που υπάρχουν στην ανάπτυξη των μεγάλων δεδομένων αντιμετωπίζουν προκλήσεις όσον αφορά την αξιοποίηση της αξίας που πρέπει να προσφέρουν τα δεδομένα. Οι προκλήσεις που παρούσιαζονται παράκατω είναι τρεις από τις βασικές που πρέπει να επιλυθούν ώστε να υπάρχει βελτιώση και μεγαλύτερη ανάπτυξη.

- Ασφάλεια Δεδομένων (Data Security):

Καθώς τα Μεγάλα Δεδομένα αυξάνονται σε μέγεθος και ο ιστός των συνδεδεμένων συσκευών τείνει να εκραγεί, εκθέτει περισσότερα από τα δεδομένα σε πιθανές παραβιάσεις ασφαλείας. Πολλοί οργανισμοί παλεύουν ήδη με την ασφάλεια των δεδομένων, ακόμη και πριν από την πολυπλοκότητα που προστίθεται από το Big Data, πολλοί από τους οποίους προσπαθούν να συμβαδίσουν. Πρώτον, γίνονται προσπάθειες ώστε να κλείσει το χάσμα των δεξιοτήτων των μεγάλων δεδομένων και δεύτερον υπάρχουν λίγοι επαγγελματίες ασφάλειας δεδομένων με πείρα για να αισθάνονται σίγουροι ότι όλες οι επιχειρήσεις έχουν μια λαβή για την ασφάλεια των δεδομένων τους. Η μεγαλύτερη λύση ασφάλειας μπορεί τελικά να διαμένει στην ανάλυση των Μεγάλων Δεδομένων όπου οι απειλές μπορούν να ανιχνευθούν και ενδεχομένως να αποφευχθούν.

- Ιδιωτικότητα Δεδομένων (Data Privacy):

Όταν επικυρώθηκε η 4η Τοπολογία το 1791 για να δώσει στους Αμερικανούς τη «προσδοκία της ιδιωτικότητας», δεν υπήρχε κανένας τρόπος για αυτούς που το ενδραίωσαν τότε να φανταστούν τις επιπλοκές της τεχνολογίας του 21ου αιώνα. Δεν υπάρχει αμφιβολία, μεγάλου ώφελους από πολλές ευκολίες και ανακαλύψεις εξαιτίας των εφαρμογών και υπηρεσιών που υποστηρίζονται από τα Μεγάλα Δεδομένα, αλλά με κίνδυνο για την ιδιωτικότητά. Δεν υπάρχει κάποιος έλεγχος σχετικά με το ποια από τα προσωπικά στοιχεία των ανθρώπων χρησιμοποιούνται ούτε υπάρχει πλήρη προστασία από την συνεχή έκθεση στην τεχνολογία. Ακόμα και αν υπήρχε επιτυχία στην πλοήγηση στον σύγχρονο κόσμο χωρίς τεχνολογία είναι αρκετά δύσκολο και δεν θα υπήρχε πλήρης προστασία της ιδιωτικότητα.

Αν και με την έκβαση της εκκρεμούσας νομοθεσίας, οι αμερικανοί νομοθέτες θα μπορούσαν να ακολουθήσουν το προβάδισμα της Ευρωπαϊκής Ένωσης και να

δημιουργήσουν ένα περιβάλλον που προστατεύει τους ανθρώπους. Το 2018, ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) της ΕΕ θα εφαρμόσει πλήρως με πρωταρχικό στόχο την ανάκτηση του ελέγχου των προσωπικών δεδομένων των πολιτώ. Ο κανονισμός αυτός ισχύει για κάθε εταιρεία που κατέχει δεδομένα για κάθε πολίτη της Ευρωπαϊκής Ένωσης, και άλλες διεθνείς εταιρείες που επεξεργάζονται και κατέχουν προσωπικά δεδομένα ανεξαρτήτως της τοποθεσίας της εταιρείας.Οι κυρώσεις για μη συμμόρφωση ανέρχονται στο 4% του ετήσιου παγκόσμιου κύκλου εργασιών και αποτελεί σημαντικό παράγοντα για τη διατήρηση των εταιρειών.

Οι εταιρείες πρέπει να κάνουν ό, τι μπορούν, ώστε να είναι διαφανείς και να βοηθήσουν τους καταναλωτές να καταλάβουν ποια δεδομένα συλλέγουν και για ποιο σκοπό. Το οικοσύστημα Big Data καθίσταται όλο και πιο σύνθετο με το Διαδίκτυο των πραγμάτων και τις συνδεδεμένες συσκευές. Οι εταιρείες που είναι ανοικτές και οικοδομούν εμπιστοσύνη θα είναι όλο και πιο σημαντικές για τους πελάτες τους.

- Διάκριση Δεδομένων (Data Discrimination):

Η διάκριση των ανθρώπων με βάση τα δεδομένα που έχουν στη ζωή τους είναι κάτι πολύ συνηθισμένο. Με τον τρόπο αυτό και τα μεγάλα δεδομένα βοηθούν τις επιχειρήσεις να πετύχουν εμπορικά και να παρέχουν υπηρεσίες, επίσης κάνοντας διακρίσεις. Υπάρχει η γενική αποδοχή από τους καταναλωτές ότι αναλύονται και αξιολογούνται λεπτομερέστερα τα δεδομένα τους με αποτέλεσμα να έχουν καλύτερες συνθήκες διαβίωσης, όμως αυτή η γνώση καθιστά πιο δύσκολο για κάποιους ανθρώπους να αποκτήσουν τις πληροφορίες ή τους πόρους που χρειάζονται.

Υπάρχουν ισχύοντες νόμοι για την προστασία των καταναλωτών, όπως ο νόμος Fair Credit Reporting Act και ο νόμος της Federal Trade Commission, οι οποίοι ισχύουν για την ανάλυση Μεγάλων Δεδομένων. Οι εταιρείες πρέπει να τηρούν αυτές τις πράξεις και τους νόμους περί ίσων ευκαιριών για να βεβαιωθούν ότι συμμορφώνονται. Επιπλέον, οι εταιρείες πρέπει να ελέγξουν τα δεδομένα τους για να εξασφαλίσουν εάν είναι αντιπροσωπευτικό το δείγμα των καταναλωτών, εάν οι αλγόριθμοι δίνουν προτεραιότητα στη δικαιοσύνη, και εάν υπάρχει έλεγχος των αποτελέσματων των μεγάλων δεδομένων σε σχέση με τις παραδοσιακά εφαρμοζόμενες στατιστικές πρακτικές.

Καθώς η εξέλιξη των μεγάλων δεδομένων συνεχίζεται, αυτές οι τρεις προκλήσεις που αφορούν την Προστασία των Δεδομένων, την Ασφάλεια των Δεδομένων και τη Διάκριση των Δεδομένων θα είναι στοιχεία προτεραιότητας που θα συμφιλιωθούν για τις ομοσπονδιακές και τις κρατικές κυβερνήσεις, τους ιδιοκτήτες επιχειρήσεων, τους ειδικούς των Μεγάλων Δεδομένων και τους καταναλωτές. Ο κατακλυσμός των πληροφοριών που συλλέγονται και η ταχεία αλλαγή της τεχνολογίας καθιστά ακόμα πιο δύσκολη την επίλυση αυτών των ζητημάτων, οπότε για αρκετό καιρό θα αποτελούν μεγάλες ανησυχίες για τα δεδομένα.

1.4 Στόχος της έρευνας

Αφού παρουσιάστηκαν τα βασικά στοιχεία των Μεγάλων Δεδομένων και αναλύσαμε την έννοια από κάθε πλευρά, επόμενο βήμα είναι να εξηγηθεί ο στόχος της έρευνα και σε πιο κομμάτι των Μεγάλων Δεδομένων θα στηριχτεί. Οι προκλήσεις που έχουν να ανατιμετωπίσουν είναι αρκετές και σε μεγάλο βαθμό μιας και μιλάμε για δεδόμενα με μεγάλο όγκο και πληροφορίες. Τα δεδομένα αυτά μπορεί να περιλαμβάνουν προσωπικές πληροφορίες ανθρώπων, πληροφορίες ασθενών εαν αφορούν θέματα υγείας, πληροφορίες σχετικά με ολόκληρους οργανισμούς και εταιρείες, κρατικά δεδομένα και δεδομένα ακόμη και από το πιο μικρό στοιχείο που θα μπορούσε να υπάρχει. Από την ίδια τη φύση των επιχείρησεων και οργανισμών, ορισμένα από τα μεγάλα δεδομένα που αποθηκεύονται και επεξεργάζονται είναι ευαίσθητα. Για τον λόγο αυτό γεννιούνται κάποια ερωτήματα σχετικά με την πρόσβαση στα δεδομένα μέσα στο περιβάλλον που βρίσκονται και εάν το περιβάλλον και τα δεδομένα είναι ευάλωτα στις απειλές του κυβερνοχώρου.

Στην εποχή της ανάπτυξης της τεχνολογίας και των επιστημών η ασφάλεια εκτελεί καθοριστικό ρόλο καθώς υπάρχουν πολλές απειλές και προβλήματα που πρέπει να αντιμετωπιστούν. Στόχος της έρευνα είναι να παρουσιαστούν όλα τα θέματα ασφάλειας των Μεγάλων Δεδομένων και να δωθούν λύσεις και απάντησεις στα ερωτήματα που έχουν δημιουργηθεί παραθέτωντας τεχνικές και μεθόδους που μπορούν να βοηθήσουν στην επίλυση του προβλήματος.
Το ευρύ ερευνητικό ερώτημα που ταιριάζει με αυτό το στόχο είναι:

" Τι σημαίνει η ασφάλεια για τα Μεγάλα Δεδομένα και ποιοι τρόποι υπάρχουν ώστε να δωθούν λύσεις στην πρόκληση αυτή ; "

Δεν υπάρχει χρόνος για σπατάλη όταν πρόκειται για επανεξέταση της ασφάλειας για το περιβάλλον των Μεγάλων Δεδομένων. Ο αριθμός και το εύρος των παραβιάσεων των δεδομένων συνεχίζει να αυξάνεται αμείωτα, με αύξηση κατά 40% των παραβιάσεων το 2016 που αναφέρθηκε από το Identity Theft Resource Center. Πρέπει να κατανοήσουμε και να δώσουν προτεραιότητα στην εφαρμογή καλύτερης ασφάλειας για μεγάλα δεδομένα διότι, το τελευταίο πράγμα που θα ήθελε κάποιος να ακούσει είναι ότι υπήρξε μεγάλη παραβίαση των δεδομένων του.

Παράλληλα τα Μεγάλα Δεδομένα συνδέονται άμεσα με την ύπαρξη του Νέφους καθώς αποθηκεύονται εκεί για να μπορούν να είναι διαθέσιμα όταν ζητηθούν. Μεγάλοι όγκοι δεδομένων αποθηκεύονται στο Cloud καθημερινά όπου περιλαμβάνονται και ευαίσθητα προσωπικά δεδομένα, τα οποία κινδυνεύουν άμεσα. Για αυτό τον λόγο είναι σημαντικό να εξετάσουμε και αυτή την τεχνολογία ώστε να δούμε πως λειτουργεί και πως εξυπηρετεί τους χρήστες σε όλο τον κόσμο σε σχέση με τα δεδομένα τους.

1.5 Επισκόπηση της έρευνας

Στην εργασία αυτή παρουσιάζουμε μια σύντομη εισαγωγή των μεγάλων δεδομένων στην πρώτη ενότητα. Η υπόλοιπη εργασία είναι ως εξής: Στο δεύτερο κεφάλαιο γίνεται η ανασκόπηση της βιβλιογραφίας. Στο τρίτο γίνεται παρουσίαση των διάφορων software analytics που υπάρχουν σχετικά με τα Big Data. Στο τέταρτο παραθέτονται διάφορα παραδείγματα και εφαρμογές σχετικά με τα μεγάλα δεδομένα. Στο πέμπτο γίνεται μια εισαγωγή στην κρυπτογράφηση και στο επόμενο η σύγκριση και ανάλυση των δυο βιβλιογραφικών αναφορών σχετικά με την ασφάλεια των μεγάλων δεδομένων. Στην συνέχεια υπάρχει μια εκτενής εισαγωγή και επεξήγηση του Cloud Computing και των στοιχείων του. Στο όγδοο κεφάλαιο βλέπουμε την χρήση του CloudAnalyst με διάφορες εφαρμογές καθώς και τα αποτελέσματα τους. Ολοκληρώνοντας, παρουσιάζονται τα συμπεράσματα αυτής της εργασίας, και προτάσεις για μελλοντική έρευνα.

Κεφάλαιο 2: Βιβλιογραφική ανασκόπηση

2.1 Εισαγωγή

Η ανάλυση των Μεγάλων Δεδομένων έχει απασχολήσει τα τελευταία χρόνια τους επιστήμονες. Μία από τις πρώτες προσπάθειες αξιολόγησης των Μεγάλων Δεδομένων ήταν μια σημαντική ανακάλυψη στα συστήματα σχεσιακής βάσης δεδομένων**.** Ο Codd (1970) ήταν ο άνθρωπος που αρχικά κατάλαβε το σχεσιακό μοντέλο διαχείρισης βάσεων δεδομένων το οποιό αποτελεί μια πρόημη αναφορά στα Μεγάλα Δεδομένα. Στην πρώτη ενότητα της ερευνάς του, παρουσίαζει τις ανεπάρκειες των μοντέλων που ήδη υπάρχουν, καθώς είναι μη εισαγόμενα, μορφοποιημένα συστήματα δεδομένων που παρέχουν στους χρήστες δομημένα αρχεία ή ελαφρώς γενικότερα μοντέλα δεδομένων. Στην δεύτερη ενότητα, αναφέρεται σε ορισμένες πράξεις σχετικά με τις σχέσεις στα προβλήματα πλεονασμού και συνέπειας στο μοντέλο των χρηστών.

2.2 Ανάπτυξη και καινοτομίες των Μεγάλων Δεδομένων

Οι Demchenko et al. (2012) στην παρούσα έρευνα ασχολούνται με τις προκλήσεις που επιβάλλει η Big Data Science στη σύγχρονη και μελλοντική Scientific Data Infrastructure (SDI). Αναφέρονται σε διαφορετικές επιστημονικές κοινότητες για τον καθορισμό απαιτήσεων σχετικά με τη διαχείριση δεδομένων, τον έλεγχο πρόσβασης και την ασφάλεια και εισάγει το μοντέλο Scientific Data Lifecycle Management (SDLM) που περιλαμβάνει όλα τα σημαντικά στάδια και αντικατοπτρίζει τις ιδιαιτερότητες της διαχείρισης δεδομένων στην σύγχρονη ηλεκτρονική επιστήμη. Παρουσιάζουν την προταση για το μοντέλο γενικής αρχιτεκτονικής SDI που παρέχει τη βάση για την οικοδόμηση διαλειτουργικών δεδομένων ή SDI με επίκεντρο το έργο, χρησιμοποιώντας σύγχρονες τεχνολογίες και βέλτιστες πρακτικές.

Η έρευνα των Boyd και Crawford (2012) στηρίζεται στην αρχή της εποχής των Μεγάλων Δεδομένων. Οι επιστήμονες υπολογιστών, οι φυσικοί, οι οικονομολόγοι, οι μαθηματικοί, οι πολιτικοί επιστήμονες, οι βιοπληροφορικοί, οι κοινωνιολόγοι και άλλοι μελετητές φωνάζουν για την πρόσβαση στις τεράστιες ποσότητες πληροφοριών που παράγονται από και για τους ανθρώπους και τις αλληλεπιδράσεις τους. Γεννιούνται ερωτήματα, όπως εαν θα βοηθήσουν τα μεγάλης κλίμακας δεδομένα στο να δημιουργηθούν καλύτερα εργαλεία, υπηρεσίες και δημόσια αγαθά, εαν θα βοηθήσουν οι αναλύσεις των δεδομένων στην κατανόηση των διαδικτυακών κοινοτήτων και των πολιτικών κινήσεων , που πρέπει να βρεθεί μια απάντηση. Δεδομένης της ανόδου των Big Data ως κοινωνικοτεχνικό φαινόμενο, είναι απαραίτητο να διερευνήθει με κριτικό πνεύμα για τις υποθέσεις και τις προκαταλήψεις του. Παρουσιάζονται έξι προκλήσεις που προκαλούν συζητήσεις για τα θέματα του Big Data: ένα πολιτισμικό, τεχνολογικό και επιστημονικό φαινόμενο που στηρίζεται στην αλληλεπίδραση της τεχνολογίας και της ανάλυσης που προκαλεί εκτεταμένη ρητορική ουτοπία.

Το Big Data είναι μια νέα ετικέτα που δίνεται σε ένα διαφοροποιημένο πεδίο πληροφορικής, στο οποίο τα σύνολα δεδομένων είναι τόσο μεγάλα ώστε να είναι δύσκολο να εργαστούν αποτελεσματικά. Ο όρος έχει χρησιμοποιηθεί κυρίως σε δύο πλαίσια, πρώτον ως τεχνολογική πρόκληση όταν ασχολούμαστε με τομείς υψηλής έντασης δεδομένων όπως η φυσική, η αστρονομία ή η αναζήτηση στο διαδίκτυο και, δεύτερον, ως κοινωνιολογικό πρόβλημα όταν συλλέγονται και εξορύσσονται δεδομένα από εταιρείες όπως το Facebook, το Google, οι εταιρείες κινητής τηλεφωνίας, οι αλυσίδες καταστημάτων λιανικής και οι κυβερνήσεις. Οι Smith et al. (2012) εξετάζουν το δεύτερο θέμα που αναφέρθηκε από μια νέα οπτική γωνία, δηλαδή πώς μπορεί ο χρήστης να αποκτήσει επίγνωση των προσωπικών δεδομένων σε σχετικό τμήμα των Big Data που είναι δημοσίως διαθέσιμο στον κοινωνικό ιστό. Η ποσότητα των μέσων που δημιουργούνται από τους χρήστες που ανεβαίνουν στον ιστό επεκτείνεται γρήγορα και είναι πέρα από τις δυνατότητες οποιουδήποτε ανθρώπου να κοσκινίσει όλα αυτά για να δει ποια μέσα επηρεάζουν την ιδιωτικότητά μας. Με βάση την ανάλυση των κοινωνικών μέσων στο Flickr, το Locr, το Facebook και το Google+,

ελέγχθηκαν οι συνέπειες της ιδιωτικότητας και των δυνατοτήτων της αναδυόμενης τάσης των κοινωνικών μέσων.

Οι Sagiroglu et al. (2013) παρουσιάζουν μια γενική εικόνα του περιεχομένου, του πεδίου εφαρμογής, των δειγμάτων, των μεθόδων, των πλεονεκτημάτων και των προκλήσεων των Μεγάλων Δεδομένων καθώς εξετάζουν και το θέμα της προστασίας της ιδιωτικότητας σε αυτό. Τα Μεγάλα Δεδομένα είναι ένας όρος για μαζικά σύνολα δεδομένων που έχουν μεγάλη και περίπλοκη δομή με τις δυσκολίες αποθήκευσης, ανάλυσης και οπτικοποίησης για περαιτέρω διαδικασίες ή αποτελέσματα. Η διαδικασία της έρευνας σε τεράστιες ποσότητες δεδομένων για την αποκάλυψη κρυφών προτύπων και μυστικών συσχετισμών ονομάζονται ανάλυση Μεγάλών Δεδομένων. Για το λόγο αυτό, οι μεγάλες υλοποιήσεις δεδομένων πρέπει να αναλύονται και να εκτελούνται όσο το δυνατόν ακριβέστερα.

Οι βάσεις δεδομένων που κρύβονται πίσω από τις υπηρεσίες και τις εφαρμογές ιστού τροφοδοτούνται διαρκώς με πληροφορίες σχετικά με τις κινήσεις και τα πρότυπα επικοινωνίας μας και μια σημαντική διάσταση της ζωής μας, ποσοτικοποιημένη σε πρωτοφανή επίπεδα, αποθηκεύεται σε αυτές τις τεράστιες αποθήκες στο διαδίκτυο. Η Sandra Gonzalez-Bailon (2013) εξετάζει μερικές από τις συνέπειες αυτού του χείμαρρου δεδομένων για την έρευνα των κοινωνικών επιστημών και για τους τύπους των ερωτήσεων που μπορούμε να ζητήσουμε από τον κόσμο που κατοικούμε. Ο σκοπός του άρθρου είναι διττός: να εξηγήσει γιατί, παρά τα δεδομένα, η θεωρία εξακολουθεί να έχει σημασία για την οικοδόμηση αξιόπιστων ιστοριών για το τι αποκαλύπτουν τα δεδομένα, και να δείξει πώς αυτό επιτρέπει στους κοινωνικούς επιστήμονες να επανεξετάσουν τα παλιά ερωτήματα στη διασταύρωση των νέων τεχνολογιών και προσεγγίσεων. Επίσης εξετάζει τον τρόπο με τον οποίο η έρευνα των μεγάλων δεδομένων μπορεί να μετασχηματίσει τη χάραξη πολιτικής, εστιάζοντας στο πώς μπορεί να μας βοηθήσει να βελτιώσουμε την επικοινωνία και τη διακυβέρνηση σε τομείς που σχετίζονται με την πολιτική.

Οι Chen και Zhang (2014) επικεντρώνονται στο ότι τα Big Data έχουν προσελκύσει τεράστια προσοχή από τους ερευνητές στις επιστήμες της πληροφορίας, στους πολιτικούς και στους υπεύθυνους για τη λήψη αποφάσεων σε κυβερνήσεις και επιχειρήσεις. Δεδομένου ότι η ταχύτητα της ανάπτυξης πληροφοριών αυξάνεται στις αρχές αυτού του νέου αιώνα, τα υπερβολικά δεδομένα προκαλούν μεγάλα προβλήματα στους ανθρώπους. Ωστόσο, υπάρχουν πολλές δυναμικές και άκρως χρήσιμες πληροφορίες κρυμμένες στον τεράστιο όγκο των δεδομένων. Αφενός, τα Big Data είναι εξαιρετικά πολύτιμα για την αύξηση της παραγωγικότητας στις επιχειρήσεις και οδηγούν σε εξελικτικές ανακαλύψεις σε επιστημονικούς κλάδους, από την άλλη πλευρά, τα Big Data κρύβουν πολλές προκλήσεις, όπως είναι οι δυσκολίες στη συλλογή δεδομένων, η αποθήκευση, η ανάλυση και η απεικόνιση των δεδομένων. Στόχος της παρούσας έρευνας είναι να παρουσιάσει μια πιο εμπεριστατωμένη άποψη για τα μεγάλα δεδομένα, συμπεριλαμβανομένων εφαρμογών, μεγάλων ευκαιριών και προκλήσεων, καθώς και τεχνολογίες που υιοθετούμε σήμερα για την αντιμετώπιση των σχετικών προβλημάτων που προκύπτουν.

Ο Nir Kshetri (2014) στην έρευνα του έχει σκοπός να αναδείξει το κόστος, τα οφέλη και τις εξωτερικές συνέπειες που σχετίζονται με τη χρήση μεγάλων δεδομένων από τους οργανισμούς. Συγκεκριμένα, οι έρευνες δείχνουν ότι διάφορα εγγενή χαρακτηριστικά των μεγάλων δεδομένων σχετίζονται με την προστασία της ιδιωτικότητας, την ασφάλεια και την ευημερία των καταναλωτών. Η έρευνα ασχολείται επίσης με το πώς τα προσωπικά δεδομένα, η ασφάλεια και τα ευεργετικά αποτελέσματα των μεγάλων δεδομένων είναι πιθανό να διαφέρουν μεταξύ των καταναλωτών σε διαφορετικά επίπεδα πολυπλοκότητας, ευπάθειας και τεχνολογικής καταλληλότητας.

Η ανάπτυξη της ακαδημαϊκής γνώσης έχει προχωρήσει τους τελευταίους αιώνες με τη χρήση μικρών μελετών δεδομένων που χαρακτηρίζονται από δείγματα δεδομένων που δημιουργήθηκαν για να απαντήσουν σε συγκεκριμένες ερωτήσεις. Είναι σημαντικό, ωστόσο, ότι τα μικρά δεδομένα θα γίνουν ολοένα και περισσότερο πολύ μεγαλύτερα από την ανάπτυξη νέων υποδομών δεδομένων που συγκεντρώνουν, κλιμακώνουν και συνδέουν μικρά δεδομένα προκειμένου να δημιουργήσουν μεγαλύτερα σύνολα δεδομένων.Οι Kitchin και Lauriault (2014) εξετάζουν τη λογική και την αξία των μικρών μελετών δεδομένων, τη σχέση τους με τα αναδυόμενα μεγάλα δεδομένα και την επιστήμη των δεδομένων και τις συνέπειες

της κλιμάκωσης μικρών δεδομένων σε υποδομές, με έμφαση στα παραδείγματα χωρικών δεδομένων.

Οι Chen και Lin (2014) παρέχουν μια σύντομη επισκόπηση της βαθιάς μάθησης (Deep Learning) και επισημαίνουν τις τρέχουσες ερευνητικές προσπάθειες και τις προκλήσεις για μεγάλα δεδομένα ως τις μελλοντικές τάσεις. Η βαθιά εκμάθηση είναι σήμερα ένας εξαιρετικά ενεργός ερευνητικός χώρος στην κοινωνική μάθηση και την αναγνώριση προτύπων. Με το τεράστιο μέγεθος των διαθέσιμων σήμερα δεδομένων, τα μεγάλα δεδομένα προσφέρουν μεγάλες ευκαιρίες και δυναμικό μετασχηματισμό για διάφορους τομείς. Από την άλλη πλευρά, παρουσιάζει επίσης πρωτοφανείς προκλήσεις για την αξιοποίηση δεδομένων και πληροφοριών. Καθώς τα δεδομένα γίνονται όλο και μεγαλύτερα, η βαθιά εκμάθηση έρχεται να διαδραματίσει βασικό ρόλο στην παροχή λύσεων προγνωστικών αναλυτικών δεδομένων για μεγάλα δεδομένα.

Οι Kim et al. (2014) αναφέρονται στην έρευνά τους στα Μεγάλα Δεδομένα, καθώς είναι ένας γενικός όρος για το τεράστιο ποσό των ψηφιακών δεδομένων που συλλέγονται από όλες τις πηγές. Σχεδόν το 90% των δεδομένων παγκοσμίως δημιουργήθηκαν τα τελευταία δύο χρόνια, με 2,5 bytes quintillion δεδομένων να προστίθενται κάθε μέρα. Επιπλέον, περίπου το 90% του δεν είναι δομημένο. Ακόμα, το συντριπτικό ποσό των μεγάλων δεδομένων από το Web και την νέφωση προσφέρουν νέες ευκαιρίες για ανακάλυψη, δημιουργία αξίας και πλούσια business intelligence για την υποστήριξη αποφάσεων σε οποιαδήποτε οργάνωση. Μεγάλα δεδομένα σημαίνουν επίσης νέες προκλήσεις με την πολυπλοκότητα, την ασφάλεια και τους κινδύνους για την ιδιωτικότητα, όπως καθώς και την ανάγκη για νέες τεχνολογίες και ανθρώπινες δεξιότητες.

Οι Fan et al. (2014) παρέχουν ανασκοπήσεις σχετικά με τα χαρακτηριστικά γνωρίσματα του Big Data και τον τρόπο με τον οποίο αυτά τα χαρακτηριστικά επηρεάζουν την αλλαγή παραδειγμάτων σε στατιστικές και υπολογιστικές μεθόδους καθώς και αρχιτεκτονικές υπολογιστών. Παρέχουν επίσης διάφορες νέες προοπτικές στην ανάλυση και τον υπολογισμό των μεγάλων δεδομένων. Συγκεκριμένα, υπογραμμίζουν τη βιωσιμότητα της πιο αραιής λύσης σε σύνολο υψηλής εμπιστοσύνης και επισημαίνουν ότι οι εξωγενείς υποθέσεις στις περισσότερες στατιστικές μεθόδους για το Big Data δεν μπορούν να επικυρωθούν λόγω τυχαίας ενδογενείας. Μπορούν να οδηγήσουν σε λανθασμένα στατιστικά συμπεράσματα και κατά συνέπεια σε λανθασμένα επιστημονικά συμπεράσματα.

Το 90% των σημερινών δεδομένων στον κόσμο έχει δημιουργηθεί τα τελευταία δύο χρόνια. Εκτός από τις προκλήσεις που θέτει τόσο τεράστιο όγκο δεδομένων, όπως η αποθήκευση, η αναζήτηση, η κοινή χρήση, η ανάλυση και η απεικόνιση, υπάρχουν και πολλές ευκαιρίες για τον κόσμο καθώς γίνεται ολοένα και πιο ψηφιοποιημένο. Οι Bagheri και Shaltooki (2015) παρουσιάζουν τα μεγάλα δεδομένα και υπογραμμίζουν τις βασικές έννοιές και τις σύγχρονες εφαρμογές καθώς και τις ερευνητικές προκλήσεις και προτείνει κατευθύνσεις έρευνας για το μέλλον.

Οι Andreu-Perez et al.(2015) παρέχουν μια επισκόπηση των πρόσφατων εξελίξεων των μεγάλων δεδομένων στο πλαίσιο της βιοϊατρικής και της πληροφορικής για την υγεία. Περιγράφουν τα βασικά χαρακτηριστικά των μεγάλων δεδομένων καιπώς η ιατρική και η υγειονομική πληροφορική, η βιοπληροφορική, η πληροφορική αισθητήρων και η πληροφορική απεικόνισης θα επωφεληθούν από μια ολοκληρωμένη προσέγγιση που θα συνδυάζει διάφορες πτυχές των εξατομικευμένων πληροφοριών από ένα ευρύ φάσμα πηγών δεδομένων. Αναμένεται ότι οι πρόσφατες εξελίξεις στα μεγάλα δεδομένα θα επεκτείνουν τις γνώσεις για τη δοκιμή νέων υποθέσεων σχετικά με τη διαχείριση κάποιας νόσου από τη διάγνωση μέχρι την πρόληψη της εξατομικευμένης θεραπείας. Ωστόσο, η άνοδος των μεγάλων δεδομένων εγείρει προκλήσεις όσον αφορά την προστασία της ιδιωτικότητας, την ασφάλεια, την κατοχή δεδομένων και την διαχείριση δεδομένων. Εξετάζουν, μερικές από τις υπάρχουσες δραστηριότητες και τις μελλοντικές ευκαιρίες που σχετίζονται με τα μεγάλα δεδομένα για την υγεία, περιγράφοντας μερικά από τα βασικά ζητήματα που πρέπει να αντιμετωπιστούν.

Τα τελευταία χρόνια, η ταχεία ανάπτυξη του Διαδικτύου, του Ίντερνετ των πραγμάτων και του Cloud Computing οδήγησε στην εκρηκτική ανάπτυξη δεδομένων σε όλες σχεδόν τις βιομηχανίες και τις επιχειρήσεις. Οι Jin et al. (2015) παρουσιάζουν πρώτα εν

συντομία την έννοια των μεγάλων δεδομένων, συμπεριλαμβανομένου του ορισμού, των χαρακτηριστικών και της αξίας τους. Στη συνέχεια, προσδιορίζεται από διαφορετικές οπτικές γωνίες η σημασία και οι ευκαιρίες που προσφέρονται καθώς και αντιπροσωπευτικές πρωτοβουλίες μεγάλων δεδομένων σε όλο τον κόσμο. Παρουσιάζουν μια περιγραφή των προκλήσεων, καθώς δίνονται και πιθανές λύσεις για την αντιμετώπιση αυτών των προκλήσεων.

Λόγω της ταχείας εξέλιξης των τεχνολογιών της πληροφορίας, τα μεγάλα δεδομένα, εξελίσσονται με χαρακτηριστικά τα 4V (όγκος, ποικιλία, ακρίβεια και ταχύτητα), προσφέρουν σημαντικά οφέλη καθώς και πολλές προκλήσεις. Ένα σημαντικό πλεονέκτημα των Big Data είναι να παρέχουν έγκαιρη ενημέρωση και ενεργητικές υπηρεσίες για τον άνθρωπο. Ο πρωταρχικός στόχος Wang et al. (2017) είναι να αναθεωρήσει την τρέχουσα κατάσταση των μεγάλων δεδομένων από τις πτυχές της οργάνωσης, της ολοκλήρωσης, της επεξεργασίας, της ασφάλειας, των αναλύσεων και των εφαρμογών, και στη συνέχεια να παρουσιάσει ένα νέο πλαίσιο για την παροχή υψηλής ποιότητας αποκαλούμενων υπηρεσιών ως Big Data-as-a-service. Το πλαίσιο αποτελείται από τρία επίπεδα, συγκεκριμένα της ανίχνευσης, της νέφωσης και το επίπεδο εφαρμογής. Τέλος, συζητούνται ορισμένες προκλήσεις σχετικά με το προτεινόμενο πλαίσιο.

Οι Bi και Cochran (2017) παρουσιάζουν τις πρόσφατες εξελίξεις στο Διαδίκτυο των Πραγμάτων (IoT) και τις εφαρμογές του, καθώς και ερευνώνται οι επιπτώσεις των πρόσφατα αναπτυγμένων Big Data (BD). Η ανάλυση των Μεγάλων Δεδομένων (BDA) έχει χαρακτηριστεί ως μια κρίσιμη τεχνολογία για την υποστήριξη της απόκτησης δεδομένων, την αποθήκευση και την ανάλυση σε συστήματα διαχείρισης δεδομένων στη σύγχρονη κατασκευή. Ο σκοπός της εργασίας που παρουσιάζεται είναι η αποσαφήνιση των απαιτήσεων των συστημάτων πρόβλεψης, και να εντοπίσουμε τις προκλήσεις και τις ευκαιρίες της έρευνας στο BDA για την υποστήριξη του cloudbased πληροφοριακά συστήματα.

2.3 Επεξεργασία των Μεγάλων Δεδομένων

Οι Ji et al. (2012) εισάγουν πολλές τεχνικές επεξεργασίας Μεγάλων Δεδομένων από πλευράς συστήματος και εφαρμογών. Αρχικά, από την πλευρά της διαχείρισης δεδομένων στο Cloud και των μηχανισμών επεξεργασίας δεδομένων, παρουσιάζουμε τα βασικά ζητήματα της επεξεργασίας, συμπεριλαμβανομένης της πλατφόρμας Cloud Computing, της αρχιτεκτονικής Cloud, της βάσης δεδομένων Cloud και του συστήματος αποθήκευσης δεδομένων. Έπειτα, πραγματοποιείται εισαγωγή σε στρατηγικές βελτιστοποίησης MapReduce και σχετικές εφαρμογές. Τέλος, διερευνούνται ανοικτά ζητήματα και προκλήσεις καθώς και ερευνητικές κατευθύνσεις σχετικά με τη επεξεργασία των Μεγάλων Δεδομένων σε περιβάλλον υπολογιστικού νέφους.

Οι Wu et al. (2013) παρουσιάζουν ένα θεώρημα HACE που χαρακτηρίζει τους τομείς της επανάστασης του Big Data, και προτείνει ένα μοντέλο επεξεργασίας Μεγάλων Δεδομένων, από την άποψη της εξόρυξης δεδομένων. Αυτό το μοντέλο που βασίζεται σε δεδομένα περιλαμβάνει ζήτηση από την συνάθροιση πηγών πληροφόρησης, εξόρυξης και ανάλυσης, προσομοίωσης χρηστών και θέματα ασφάλειας και προστασίας της ιδιωτικότητας. Αναλύονται τα δύσκολα ζητήματα στο πλαίσιο του μοντέλου των δεδομένων και επίσης στην επανάσταση των Big Data.

Λόγω του μεγάλου μεγέθους των δεδομένων καθίσταται πολύ δύσκολο να πραγματοποιηθεί αποτελεσματική ανάλυση χρησιμοποιώντας τις υπάρχουσες παραδοσιακές τεχνικές. Τα μεγάλα δεδομένα λόγω των διαφόρων ιδιοτήτων που έχουν, όπως ο όγκος, η ταχύτητα, η ποικιλία, η μεταβλητότητα, η αξία και η πολυπλοκότητα, οδήγησαν σε πολλές προκλήσεις. Δεδομένου ότι τα μεγάλα δεδομένα είναι μια πρόσφατη επερχόμενη τεχνολογία στην αγορά η οποία μπορεί να αποφέρει τεράστια οφέλη στις επιχειρηματικές οργανώσεις, είναι απαραίτητο να τεθούν στο φως διάφορες προκλήσεις και ζητήματα που συνδέονται με την προσαρμογή στην τεχνολογία αυτή. Οι Katal et al.(2013) εισάγουν τη μεγάλη τεχνολογία δεδομένων μαζί με τη σημασία της στον σύγχρονο κόσμο. Οι διάφορες προκλήσεις και ζητήματα προσαρμογής και αποδοχής της μεγάλης τεχνολογίας δεδομένων, των εργαλείων

της (Hadoop) συζητούνται επίσης λεπτομερώς μαζί με τα προβλήματα που αντιμετωπίζει το Hadoop. Τέλος, ολοκληρώνεται η έρευνα τους με τις καλές πρακτικές δεδομένων που πρέπει να ακολουθηθούν.

Οι Cuzzocrea et al.(2013) στην ερευνά τους δείχουν μεγάλο ενδιαφέρον για το Big Data και την αυξηση τους, κυρίως λόγω ενός εκτεταμένου αριθμού ερευνητικών προβλημάτων που συνδέονται στενά με εφαρμογές και συστήματα πραγματικής ζωής, όπως η μοντελοποίηση, η επεξεργασία, η αναζήτηση και η εξόρυξη γνώσης μεγάλης κλίμακας. Αναλύονται τρεις σημαντικές πτυχές της έρευνας μεγάλων δεδομένων, δηλαδή OLAP πάνω από τα μεγάλα δεδομένα, την εγγραφή μεγάλων δεδομένων και την προστασία προσωπικών δεδομένων. Παρουσιάζονται επίσης οι μελλοντικές κατευθύνσεις έρευνας, οπότε καθορίζεται εμμέσως μια ερευνητική ατζέντα που στοχεύει στην αντιμετώπιση μελλοντικών προκλήσεων σε αυτόν τον τομέα της έρευνας.

Οι Kim et al. (2013) περιγράφουν το υπόβαθρο των μεγάλων δεδομένων, την εξόρυξη δεδομένων και τα χαρακτηριστικά γνωρίσματα των μεγάλων δεδομένων καθώς προτείνεται μεθοδολογία επιλογής χαρακτηριστικών για την προστασία των μεγάλων δεδομένων. Η εξαγωγή πολύτιμων πληροφοριών είναι ο κύριος στόχος της ανάλυσης μεγάλων δεδομένων, επομένως, η συνάφεια μεταξύ χαρακτηριστικών ενός συνόλου δεδομένων είναι ένα πολύ σημαντικό στοιχείο για την ανάλυση των δεδομένων. Εστιάζουν σε δύο πράγματα, πρώτον, η καταλληλότητα των χαρακτηριστικών στα μεγάλα δεδομένα αποτελεί βασικό στοιχείο για την εξαγωγή πληροφοριών. Δεύτερον, είναι αδύνατο να προστατευθούν όλα τα μεγάλα δεδομένα και τα χαρακτηριστικά τους. Θεωρούμε μεγάλα δεδομένα ως ένα μόνο αντικείμενο που έχει τα δικά του χαρακτηριστικά. Υποθέτουμε ότι ένα χαρακτηριστικό που έχει μεγαλύτερη συνάφεια είναι πιο σημαντικό από άλλα χαρακτηριστικά.

Οι Chen et al. (2014) σε αυτή την έρευνα εξετάζεται το ιστορικό και η τεχνολογία των Μεγάλων Δεδομένων. Καταρχήν παρουσιάζεται το γενικό υπόβαθρο και αναθεωρούνται οι σχετικές τεχνολογίες, όπως είναι οι υπολογισμοί, το Διαδίκτυο των πραγμάτων, τα κέντρα δεδομένων και το Hadoop. Στη συνέχεια, επικεντρώνεται στις τέσσερις φάσεις της αλυσίδας των Μεγάλων Δεδομένων, δηλαδή στην παραγωγή δεδομένων, την απόκτηση δεδομένων, την αποθήκευση και την ανάλυση δεδομένων. Τέλος, εξετάζονται οι διάφορες αντιπροσωπευτικές εφαρμογές, όπως η διαχείριση επιχειρήσεων, το Διαδίκτυο των Πραγμάτων, τα διαδικτυακά κοινωνικά δίκτυα, οι εφαρμογές στο διαδίκτυο, η συλλογική νοημοσύνη και το έξυπνο δίκτυο. Αυτή η έρευνα ολοκληρώνεται με μια συζήτηση για ανοιχτά προβλήματα και μελλοντικές κατευθύνσεις.

Η αυξανόμενη διαθεσιμότητα και ο ρυθμός ανάπτυξης των βιοϊατρικών πληροφοριών, γνωστών και ως «μεγάλα δεδομένα», παρέχουν μια ευκαιρία για μελλοντικά εξατομικευμένα προγράμματα ιατρικής που θα βελτιώσουν σημαντικά την περίθαλψη των ασθενών. Ο Fabricio F.Costa (2014) στηρίζεται στις πρόσφατες εξελίξεις της τεχνολογίας της πληροφορίας (IT) που εφαρμόζονται στη βιοϊατρική καθώς μεταβάλλουν το τοπίο της ιδιωτηκότητας και των προσωπικών πληροφοριών, με τους ασθενείς να αποκτούν μεγαλύτερο έλεγχο των πληροφοριών τους για την υγεία. Ωστόσο, πρέπει να αντιμετωπιστούν συγκεκριμένες προκλήσεις για να ενσωματωθούν οι τρέχουσες ανακαλύψεις στην ιατρική πρακτική. Παρουσιάζονται οι σημαντικές καινοτομίες που επιτεύχθηκαν στο συνδυασμό της ωματικής και των κλινικών δεδομένων υγείας όσον αφορά την εφαρμογή τους στην εξατομικευμένη ιατρική. Επίσης εξετάζονται οι προκλήσεις που σχετίζονται με τη χρήση μεγάλων δεδομένων στη βιοϊατρική και τη μεταγραφική επιστήμη.

Οι K.U. και M. David (2014) παρουσιάζουν τη βιβλιογραφική ανασκόπηση για το Big Mining Data και τα θέματα και τις προκλήσεις με έμφαση στα διακεκριμένα χαρακτηριστικά του Big Data. Επίσης παραθέτουν μερικές μεθόδους για την αντιμετώπιση των μεγάλων δεδομένων. Τα δεδομένα έχουν καταστεί αναπόσπαστο μέρος κάθε οικονομίας, της βιομηχανίας, της οργάνωσης, της επιχειρηματικής λειτουργίας και του ατόμου. Τα μεγάλα δεδομένα είναι ένας όρος που χρησιμοποιείται για την αναγνώριση των συνόλων δεδομένων, των οποίων το μέγεθος είναι πέρα από την ικανότητα των τυπικών εργαλείων λογισμικού βάσης δεδομένων να αποθηκεύουν, να διαχειρίζονται και να αναλύουν.

Γίνεται παρουσιάση μιας βιβλιογραφικής έρευνας και ενός εκπαιδευτικού σσυστήματος για τις πλατφόρμες ανάλυσης Μεγάλων Δεδομένων. Ο Hu et al. (2014) αναφέρεται σε ένα συστηματικό πλαίσιο για την αποσύνθεση μεγάλων συστημάτων δεδομένων σε τέσσερις διαδοχικές ενότητες, δηλαδή την παραγωγή δεδομένων, την απόκτηση δεδομένων, την αποθήκευση δεδομένων και την ανάλυση δεδομένων. Αυτά τα τέσσερα στοιχεία αποτελούν μια μεγάλη αλυσίδα αξίας δεδομένων. Μετά από αυτό, υπάρχει μια λεπτομερή έρευνα πολυάριθμων προσεγγίσεων και μηχανισμών από τις ερευνητικές και βιομηχανικές κοινότητες με επικρατέστερο πλαίσιο το Hadoop. Τέλος, δίνονται διάφορα σημεία αναφοράς αξιολόγησης και πιθανές κατευθύνσεις έρευνας για συστήματα Μεγάλων Δεδομένων.

Η διαδεδομένη και εκθετικά αυξανόμενη κυκλοφορία δεδομένων παρουσιάζει επικείμενες προκλήσεις σε όλες τις πτυχές του σχεδιασμού του ασύρματου συστήματος. Ο Bi et al. (2015) ερευνούν τις προκλήσεις και τις ευκαιρίες στο σχεδιασμό κλιμακούμενων ασύρματων συστημάτων για την αποδοχή της εποχής των μεγάλων δεδομένων. Από τη μία πλευρά, εξετάζονται οι προηγμένες αρχιτεκτονικές δικτύωσης και οι τεχνικές επεξεργασίας σήματος, προσαρμόσιμες για τη διαχείριση μεγάλης κυκλοφορίας δεδομένων σε ασύρματα δίκτυα. Από την άλλη πλευρά, αντί να παρακολουθούνται τα μεγάλα δεδομένα ως ανεπιθύμητη επιβάρυνση, εισάγονται μέθοδοι για την αξιοποίηση της τεράστιας κυκλοφορίας δεδομένων, για την οικοδόμηση ενός ασύρματου δικτύου μεγάλης ευαισθητοποίησης δεδομένων με καλύτερη ποιότητα ασύρματων υπηρεσιών και νέες κινητές εφαρμογές.

Το Cloud computing είναι μια ισχυρή τεχνολογία για την εκτέλεση μαζικής κλίμακας και πολύπλοκων εργασιών με την χρήση υπολογιστή, που εξαλείφει την ανάγκη διατήρησης δαπανηρού υλικού, αποκλειστικού χώρου και λογισμικού. Ο Hashem et al. (2015) επικεντρώθηκαν στην μαζική αύξηση της κλίμακας δεδομένων ή τα μεγάλα δεδομένα που παράγονται μέσω του cloud computing. Η διαχείριση των μεγάλων δεδομένων είναι μια δύσκολη και χρονοβόρα εργασία που απαιτεί μεγάλη υπολογιστική υποδομή για την επιτυχή επεξεργασία και ανάλυση των δεδομένων. Εισάγονται ο ορισμός, τα χαρακτηριστικά και η ταξινόμηση των μεγάλων δεδομένων για το cloud computing, επίσης αναφέρεται η σχέση μεταξύ των μεγάλων δεδομένων και του cloud computing, των μεγάλων συστημάτων αποθήκευσης δεδομένων και της τεχνολογίας Hadoop. Επιπλέον, διερευνώνται οι ερευνητικές προκλήσεις με έμφαση στην επεκτασιμότητα, τη διαθεσιμότητα, την ακεραιότητα, τον μετασχηματισμό, την ποιότητα των δεδομένων, την ετερογένεια, το ιδιωτικό απόρρητο, τα νομικά και ρυθμιστικά θέματα τους.

Οι Σαπουντζή και Ψάννης (2016) στην έρευνα που πραγματοποίησαν επικεντρώθηκαν στο Online Social Network's (OSN) το οποίο εκτινάχθηκε με την εξάπλωση της τάσης των Μεγάλων Δεδομένων. Η εξαγωγή πληροφοριών από τέτοια δεδομένα έχει γίνει μια πολύ διευρυμένη πολυεπιστημονική περιοχή που απαιτεί επιστημονικά εργαλεία και εμπειρογνωμοσύνη. Σε απάντηση αυτής της χαοτικής αναδυόμενης επιστήμης των κοινωνικών δεδομένων, η έρευνα παρέχει μια εξελιγμένη ταξινόμηση των state-of-the-art πλαισίων λαμβάνοντας υπόψη την ποικιλομορφία των πρακτικών μεθόδων και των τεχνικών. Αυτή είναι η πρώτη απόπειρα που απεικόνισε όλο το φάσμα της ανάλυσης δικτύων κοινωνικών δεδομένων και των σχετικών πλαισίων τους. Η έρευνα επιδεικνύει προκλήσεις και μελλοντικές κατευθύνσεις με έμφαση στην εξόρυξη κειμένων και την ελπιδοφόρα λεωφόρο υπολογιστικής νοημοσύνης.

Οι Στεργίου και Ψάννης (2016) αναφέρονται σε δυο νέες τεχνολογίες, το Mobile Cloud Computing όπου τόσο η αποθήκευση δεδομένων όσο και η επεξεργασία δεδομένων λειτουργούν εκτός της κινητής συσκευής και το Internet of Things που στηρίζεται στην αλληλεπίδραση και στην συνεργασία μεταξύ των αντικειμένων που στέλνονται μέσω των ασύρματων δικτύων. Στην παρούσα εργασία παρουσιάζεται μια έρευνα για το IoT και το Cloud Computing με έμφαση στα ζητήματα ασφάλειας και των δύο τεχνολογιών.Συγκεκριμένα, συνδυάζονται οι δύο προαναφερθείσες τεχνολογίες (δηλ. Cloud Computing και IoT) για να εξετάσουμε τα κοινά χαρακτηριστικά και για να ανακαλύψουμε τα οφέλη της ενσωμάτωσής τους. Τέλος, εξετάζονται οι προκλήσεις ασφάλειας της ενσωμάτωσης του IoT και του Cloud Computing καθώς βλέπουμε πώς η τεχνολογία Cloud Computing βελτιώνει τη λειτουργία του IoT.

Οι Su, Xu και Qi (2016) προτείνουν ένα νέο πλαίσιο για την παροχή μεγάλων δεδομένων κινητής τηλεφωνίας σε κινητά κοινωνικά δίκτυα που βασίζονται στο περιεχόμενο. Καταρχάς, μελετώνται τα χαρακτηριστικά και οι προκλήσεις των κινητών δεδομένων. Παρουσιάζεται η αρχιτεκτονική δικτύου που βασίζεται στο περιεχόμενο για την παροχή μεγάλων δεδομένων κινητής τηλεφωνίας, όπου κάθε στοιχείο αποτελείται από πακέτα ενδιαφέροντος και πακέτα δεδομένων, αντίστοιχα. Στη συνέχεια, πώς να γίνεται η επιλογή των πακέτων από τον κόμβο πράκτορα για προώθηση και πως ο κόμβος αναμετάδοσης μετάδίδει τα πακέτα δεδομένων καθορίζοντας προτεραιότητες των πακέτων ενδιαφέροντος και των πακέτων δεδομένων. Τέλος, τα αποτελέσματα προσομοίωσης δείχνουν την απόδοση του πλαισίου με ποικίλες παραμέτρους.

Τα κινητά κυψελοειδή δίκτυα έχουν γίνει τόσο οι γεννήτριες όσο και οι μεταφορείς μαζικών δεδομένων. Η ανάλυση Μεγάλων Δεδομένων μπορεί να βελτιώσει την απόδοση των κινητών κυψελοειδών δικτύων και να μεγιστοποιήσει τα έσοδα των φορέων εκμετάλλευσης. Οι He et al. (2016) εισάγουν ένα ενιαίο μοντέλο δεδομένων βασισμένο σε τυχαία θεωρία πινάκων και μηχανικής μάθησης. Παρουσιάζουν ένα αρχιτεκτονικό πλαίσιο για την εφαρμογή ανάλυσης σε Μεγάλα Δεδομένων σε κινητά δίκτυα κινητής τηλεφωνίας. Επιπλέον, περιγράφουν παραδείγματα, όπως μεγάλα δεδομένα σηματοδοσίας, μεγάλα δεδομένα κυκλοφορίας, μεγάλα δεδομένα θέσης και μεγάλα ετερογενή δεδομένα σε κινητά κυψελοειδή δίκτυα. Τέλος, παραθέτουν μια σειρά από ανοικτές ερευνητικές προκλήσεις ανάλυσης σε δεδομένα στοιχεία σε δίκτυα κινητής τηλεφωνίας.

Η έρευνα του Ira S. Rubinstein (2017) περιλαμβάνει τρια μέρη. Το πρώτο μέρος εξετάζει λεπτομερέστερα τα μεγάλα δεδομένα και διερωτάται εάν καταστρατηγεί τον παραδοσιακό νόμο περί ιδιωτικότητας, υπονομεύοντας τις βασικές αρχές και τις ρυθμιστικές υποθέσεις. Εάν ισχυεί η παραπάνω υπόθεση τότε , οι ρυθμιστικές αρχές πρέπει να εξετάσουν νέες προσεγγίσεις πέραν εκείνων που αντικατοπτρίζονται στην τρέχουσα σκέψη τους. Το δεύτερο μέρος αναλύει κατά πόσο ο προτεινόμενος κανονισμός αντιμετωπίζει επιτυχώς τις προκλήσεις του Big Data. Αυτό περιλαμβάνει μια προσεκτική εξέταση της νέας και αναθεωρημένης διάταξης για την αυτοματοποιημένη λήψη αποφάσεων ή τη δημιουργία προφίλ. Το τρίτο μέρος ξεκινά με την περιγραφή μιας «μετατόπισης ελέγχου» που βρίσκεται ήδη σε εξέλιξη λόγω της εμφάνισης ενός νέου επιχειρηματικού μοντέλου που βασίζεται σε «Υπηρεσίες Προσωπικών Δεδομένων» ή PDSes. Στη συνέχεια, εξετάζει σε ποιο βαθμό το PDSs είναι τεχνολογικά εφικτό, διανοητικά συνεκτικό και ρεαλιστικό από επιχειρηματική άποψη. Τέλος, προσφέρει μερικές προκαταρκτικές συστάσεις σχετικά με τον τρόπο με τον οποίο τα θεσμικά όργανα της ΕΕ θα μπορούσαν να προωθήσουν τα PDS, παρέχοντας περισσότερες ρυθμιστικές αρχές έτσι ώστε να πειραματίζονται με κώδικες δεοντολογίας.

Τα δεδομένα αξιοπιστίας, χρησιμοποιούνται παραδοσιακά για σκοπούς όπως η δημιουργία προβλέψεων για το κόστος εγγύησης και η βελτιστοποίηση του κόστους λειτουργίας και συντήρησης των συστημάτων. Στην τρέχουσα (και μελλοντική) γενιά πολλών προϊόντων, η φύση των δεδομένων αξιοπιστίας πεδίου αλλάζει δραματικά. Συγκεκριμένα, τα προϊόντα μπορούν να εξοπλίζονται με αισθητήρες που μπορούν να χρησιμοποιηθούν για την καταγραφή πληροφοριών σχετικά με τον τρόπο και τον χρόνο και υπό ποιες περιβαλλοντικές και λειτουργικές συνθήκες χρησιμοποιούνται τα προϊόντα. Οι Meeker και Hong (2017) εξετάζουν ορισμένες εφαρμογές όπου χρησιμοποιούνται δεδομένα αξιοπιστίας και διερευνώνται μερικές από τις δυνατότητες χρήσης σύγχρονων δεδομένων για την παροχή ισχυρότερων στατιστικών μεθόδων για τη λειτουργία και την πρόβλεψη της απόδοσης των συστημάτων στο πεδίο. Παρουσιάζουν ορισμένα παραδείγματα πρόσφατων τεχνικών εξελίξεων που έχουν σχεδιαστεί για να χρησιμοποιηθούν σε τέτοιες εφαρμογές και περιγράφουν τις εναπομένουσες προκλήσεις.

Η έρευνα του Ben Williamson (2017) αποτελείται από μια ανάλυση του εργαστηρίου Lytics, του Πανεπιστημίου Stanford και του κέντρου Digital Data, Analytics και Adaptive Learning, ενός μεγάλου ερευνητικού κέντρου της εταιρείας εμπορικής εκπαίδευσης Pearson. Αυτά τα ιδρύματα καθίστανται μεθοδολογικοί φύλακες με την ικανότητα να διεξάγουν νέες μορφές εκπαιδευτικής έρευνας χρησιμοποιώντας μεθόδους μεγάλων δεδομένων και αλγοριθμικών δεδομένων. Το κεντρικό επιχείρημα είναι ότι καθώς η επιστήμη έχει μεταναστεύσει από το ακαδημαϊκό εργαστήριο στον εμπορικό τομέα, η κατοχή των μέσων

παραγωγής αναλύσεων εκπαιδευτικών δεδομένων έχει συγκεντρωθεί στις δραστηριότητες των κερδοσκοπικών εταιρειών. Κατά συνέπεια, νέες θεωρίες μάθησης ενσωματώνονται στα εργαλεία που παρέχουν, με τη μορφή τεχνολογιών εξατομίκευσης που βασίζονται σε αλγόριθμους, οι οποίες μπορούν να πωληθούν σε σχολεία και πανεπιστήμια. Η έρευνα εξετάζει δύο θέματα: (1) πώς πρέπει να αναθεωρηθεί η εκπαίδευση σε σχέση με τις αλγοριθμικές μεθόδους και τα δεδομένα επιστημονικών επιστημολογιών και (2) πώς μετατοπίζεται η πολιτική οικονομία της εκπαίδευσης καθώς η παραγωγή γνώσης συγκεντρώνεται σε δεδομένα.

Οι Mahrt και Scharkow (2017) ασχολήθηκαν με τις μεθοδολογικές πτυχές της ανάλυσης των Big Data αναφορικά με την εφαρμοσιμότητα και τη χρησιμότητά τους στην έρευνα ψηφιακών μέσων. Πραγματοποιήθηκε ανασκόπηση μιας ποικίλης βιβλιογραφίας σχετικά με τη μεθοδολογία, εξετάζονται οι συνέπειες της χρήσης των Big Data σε διάφορα στάδια της ερευνητικής διαδικασίας. Υποστηρίζεται ότι οι ερευνητές πρέπει να εξετάσουν εάν η ανάλυση τεράστιων ποσοτήτων δεδομένων είναι θεωρητικά δικαιολογημένη, δεδομένου ότι μπορεί να περιοριστεί σε ισχύ και πεδίο εφαρμογής και ότι οι αναλύσεις μικρού μεγέθους μπορούν να παράγουν σωστά αποτελέσματα όταν χρησιμοποιούν σωστά δειγματοληψία, μέτρηση και αναλυτικές διαδικασιές.

Στην παρούσα έρευνα τους, οι Sivinski et al (2017) προτείνουν ένα πλαίσιο για τον προσδιορισμό της ανταγωνιστικής σημασίας των δεδομένων. Το πλαίσιο αυτό εξετάζει πρώτα αν τα μέρη κατέχουν ή ελέγχουν τα σχετικά δεδομένα. Η δεύτερη πτυχή είναι κατά πόσον τα σχετικά δεδομένα διατίθενται στο εμπόριο ως προϊόν ή ως εισροή για προϊόντα. Η τρίτη θεώρηση είναι κατά πόσο τα σχετικά δεδομένα είναι αποκλειστικά του ιδιοκτήτη ή του διαχειριστή για προϊόντα ή υπηρεσίες και για ανταγωνιστική εισαγωγή Η τελευταία σκέψη είναι αν υφίστανται λογικά διαθέσιμα υποκατάστατα των σχετικών δεδομένων ή εάν τα δεδομένα είναι μοναδικά.

Τα Hadoop και Spark χρησιμοποιούνται ευρέως για επεξεργασία δεδομένων μεγάλης κλίμακας με αποτελεσματικό και ανθεκτικό σε σφάλματα τρόπο σε ιδιωτικά ή δημόσια νέφη. Αυτά τα μεγάλα συστήματα επεξεργασίας δεδομένων χρησιμοποιούνται ευρέως από πολλές βιομηχανίες, π.χ.,Google, Facebook και Amazon, για την επίλυση μιας μεγάλης κατηγορίας προβλημάτων. Ωστόσο, όλα αυτά τα δημοφιλή συστήματα έχουν ένα σημαντικό μειονέκτημα όσον αφορά τους τοπικά κατανεμημένους υπολογισμούς, οι οποίοι τους εμποδίζουν στην εφαρμογή της γεωγραφικά κατανεμημένης επεξεργασίας δεδομένων. Για τον λόγο αυτό γίνεται έρευνα από τον Dolev et al. (2017) για τις προκλήσεις και τις απαιτήσεις στο σχεδιασμό της γεωγραφικά κατανεμημένης επεξεργασίας δεδομένων καθώς και για την επεξεργασία ροής (συστήματα βασισμένα σε Spark) και SQL-επεξεργασμένα γεωδιανεμημένα πλαίσια, μοντέλα και αλγόριθμους με τα γενικά θέματα.

Η ποσότητα των ψηφιακών δεδομένων που παράγονται σε όλο τον κόσμο αυξάνεται εκθετικά. Υπάρχουν τεράστια κίνητρα, από τις εκστρατείες μάρκετινγκ μέχρι την εγκληματολογία και την έρευνα στις κοινωνικές επιστήμες, που παρακινούν την επεξεργασία όλο και μεγαλύτερων δεδομένων, ώστε να αντλούν πληροφορίες και γνώσεις ώστε να βελτιωθούν οι διαδικασίες και τα οφέλη. Συνεπώς, αυξάνεται όλο και περισσότερο η ανάγκη για πιο αποτελεσματικά συστήματα υπολογιστών προσαρμοσμένα σε τέτοιες μεγάλες εφαρμογές δεδομένων. Τέτοιες προσαρμοσμένες αρχιτεκτονικές αναμένεται να περιλαμβάνουν την ετερογένεια για να ταιριάζουν καλύτερα σε κάθε φάση του υπολογισμού. Για τον λόγο αυτό ο Goudarzi (2017) μελετά την κατάσταση της τεχνολογίας καθώς και των μελλοντικών μεγάλης κλίμακας υπολογιστικών αρχιτεκτονικών που προσαρμόζονται για την επεξεργασία μεγάλων εφαρμογών δεδομένων στο μοντέλο MapReduce.

Η πρόβλεψη των τιμών ηλεκτρικής ενέργειας αποτελεί σημαντικό μέρος του έξυπνου δικτύου, επειδή καθιστά το έξυπνο δίκτυο οικονομικά αποδοτικό. Ωστόσο, οι υπάρχουσες μέθοδοι για την πρόβλεψη των τιμών μπορεί να είναι δύσκολο να αντιμετωπιστούν με τεράστια δεδομένα τιμών στο δίκτυο. Ο Wang et al.(2017) για την επίλυση ενός τέτοιου προβλήματος, αναπτύσσει ένα νέο πρότυπο πρόβλεψης τιμών ηλεκτρικής ενέργειας. Συγκεκριμένα, στο προτεινόμενο μοντέλο ενσωματώνονται τρεις ενότητες. Κατ 'αρχάς, με τη συγχώνευση του Random Forest (RF) και του Algorithm Relief-F και δεύτερον με την ενσωμάτωση της λειτουργίας πυρήνα και της Principle Component Analysis (KPCA) Τέλος,

για την πρόβλεψη της ταξινόμησης των τιμών, παρουσιάζεται ένας ταξινομητή SVM (Differential Evolution Machine) βάσει διαφορικής εξέλιξης (DE).

2.4 Ασφάλεια των Μεγάλων Δεδομένων

Ο Tankard (2012) στην έρευνα που πραγματοποίησε αναφέρεται στα Μεγάλα Δεδομένα και στα προβλήματα ασφάλειας που δημιουργούνται. Ο όρος Μεγάλα Δεδομένα έχει έρθει σε χρήση πρόσφατα για να αναφέρεται στην συνεχώς αυξανόμενη ποσότητα πληροφοριών που αποθηκεύουν, επεξεργάζονται και αναλύουν οι οργανισμοί, λόγω του αυξανόμενου αριθμού των πηγών πληροφοριών που χρησιμοποιούνται. Τα προβλήματα που υπάρχουν αναλύονται και αναπτύσονται ώστε στην συνέχεια να δωθούν λύσεις. Αναπτύσει μια ολιστική προσέγγιση για την καλύτερη διαχείριση και τέλος δίνονται προτάσεις έτσι ώστε να περιοριστούν τα προβλήματα της ασφάλειας.

Οι Cárdenas et al. (2013) μας παρουσιάζουν τα Μεγάλα Δεδομένα και την ραγδαία ανάπτυξη τους καθώς αυξάνονται και οι ρυθμοί που αναπτύσεται η τεχνολογία και η πληροφορίες στο διαδίκτυο. Αυτοί οι αριθμοί αυξάνονται καθώς οι επιχειρήσεις επιτρέπουν την καταγραφή συμβάντων σε περισσότερες πηγές, προσλαμβάνοντας περισσότερους υπαλλήλους, με την ανάπτυξη περισσότερων συσκευών και με την χρήση περισσότερου λογισμικού. Δυστυχώς, αυτός ο όγκος και η ποικιλία των δεδομένων γρήγορα γίνονται συντριπτικές με αποτέλεσμα οι υπάρχουσες αναλυτικές τεχνικές να μην λειτουργούν καλά και τυπικά να παράγουν πολλά ψευδή στοιχεία. Πραγματοποιούν αναφορά στις εξελίξεις της αναλυτική των Μεγάλων Δεδομένων καθώς και οι προκλήσεις που εμφανίζονται καταλήγοντας στην ασφάλεια και τα μέτρα αντιμετώπισης που πρέπει να ληφθούν για τα προβλήματα που υπάρχουν.

Οι Mahmood και Afzal (2013) στην ερευνά τους αναλύουν την ταχεία ανάπτυξη του Διαδικτύου καθώς οδηγεί σε μια εκθετική αύξηση του τύπου και της συχνότητας των επιθέσεων στον κυβερνοχώρο. Για να αντιμετωπιστεί αυτό το πρόβλημα, η έρευνα εστιάζεται στο Security Analytics, δηλαδή στην εφαρμογή τεχνικών των Big Data Analytics στην ασφάλεια του κυβερνοχώρου. Τα Analytics μπορεί να βοηθήσουν τους διαχειριστές δικτύων ιδιαίτερα στην παρακολούθηση και επίβλεψη ροών δικτύου σε πραγματικό χρόνο και ανίχνευσης σε πραγματικό χρόνο τόσο των κακόβουλων όσο και των ύποπτων μοτίβων. Παρουσιάζουν μια ολοκληρωμένη έρευνα σχετικά με την εξέλιξη της τέχνης του Security Analytics, δηλαδή την περιγραφή, την τεχνολογία, τις τάσεις και τα εργαλεία του.

Ο Alexandru Adrian Tole (2013) στην ερευνά του αντιμετωπίζει τις προκλήσεις που δημιουργούν τα Μεγάλα Δεδομένα. Το ποσό των δεδομένων που ταξιδεύουν στο διαδίκτυο σήμερα, όχι μόνο είναι μεγάλο, αλλά είναι πολύπλοκο επίσης. Οι εταιρείες, τα ιδρύματα, το σύστημα υγειονομικής περίθαλψης, χρησιμοποιούν στοιχεία για τη δημιουργία αναφορών, προκειμένου να διασφαλιστεί η συνέχεια όσον αφορά τις υπηρεσίες που πρέπει να προσφέρουν. Η διαδικασία πίσω από τα αποτελέσματα που ζητούν οι εν λόγω φορείς αποτελούν πρόκληση για τους προγραμματιστές λογισμικού και τις εταιρείες που παρέχουν υποδομή πληροφορικής. Η πρόκληση είναι να χειριστούμε έναν εντυπωσιακό όγκο δεδομένων που πρέπει να παραδοθεί με ασφάλεια μέσω του διαδικτύου και να φθάσει στον προορισμό του άθικτο.

Οι Sedayao et al. (2014) αναλύουν τις εμπειρίες και τα προβλήματα που αντιμετωπίστηκαν όταν έγινε συνδυασμός σε τεχνικές ανωνυμίας, προστασίας της ιδιωτικότητας και ανάλυσης των δεδομένων των χρηστών, προστατεύοντας παράλληλα την ταυτότητές τους. Η ομάδα Human Factors Engineering χρησιμοποιήσε αρχεία καταγραφής πρόσβασης στο Web και εργαλεία Big Data για τη βελτίωση της χρηστικότητας της ισχυρής εσωτερικής δικτυακής πύλης της Intel. Προκειμένου να προστατευθεί το ιδιωτικό απόρρητο των εργαζομένων της Intel, καταργήθηκαν οι προσωπικές πληροφορίες αναγνώρισης (PII) από το αρχείο καταγραφής χρήσης της πύλης, αλλά με τρόπο που δεν επηρέασε τη χρήση των εργαλείων μεγάλων δεδομένων για να γίνει η ανάλυση. Δημιουργήθηκε μια ανοικτή αρχιτεκτονική για ανωνυμία, η οποία επιτρέπει τη χρήση ποικίλων εργαλείων τόσο για τον εντοπισμό όσο και για την επαναπροσδιορισμό αρχείων καταγραφής του ιστού. Αποδείχθηκε ότι οι τεχνικές Big Data θα μπορούσαν να αποφέρουν οφέλη στο επιχειρηματικό περιβάλλον

ακόμα και όταν εργάζονταν σε ανώνυμα δεδομένα. Διαπιστώθηκε επίσης ότι, παρά την αποκάλυψη προφανών PII όπως ονόματα χρηστών και διευθύνσεις IP, τα ανώνυμα δεδομένα ήταν ευάλωτα σε επιθέσεις συσχέτισης. Πραγματοποιήθηκαν διορθώσεις των τρωτών σημείων και διαπίστώθηκε ότι οι πληροφορίες για το User Agent (Browser / OS) συσχετίζονται έντονα με μεμονωμένους χρήστες. Ενώ το δακτυλικό αποτύπωμα του προγράμματος περιήγησης ήταν γνωστό πριν, έχει επιπτώσεις στα εργαλεία και τα προϊόντα που χρησιμοποιούνται για την εξακρίβωση της ταυτότητας των δεδομένων επιχείρησης. Τέλος, συμπεραίνουμε ότι τα Μεγάλα Δεδομένα, η ανωνυμία και το απόρρητο μπορούν να συνδυαστούν με επιτυχία, αλλά απαιτεί ανάλυση συνόλων δεδομένων για να διασφαλιστεί ότι η ανωνυμία δεν είναι ευάλωτη σε επιθέσεις συσχέτισης.

Οι Kupwade et al. (2014) στην παρούσα εργασία, παρουσιάζουν τα πλέον σύγχρονα θέματα ασφάλειας και ιδιωτικότητας σε Μεγάλα Δεδομένα, σε εφαρμογές στη βιομηχανία της υγειονομικής περίθαλψης. Το νέο κύμα ψηφιοποίησης των ιατρικών αρχείων έχει οδηγήσει σε μια μεταβολή στον κλάδο της υγειονομικής περίθαλψης με αποτέλεσμα, να σημειώθει αύξηση στον όγκο των δεδομένων από την άποψη της πολυπλοκότητασκαι της ποικιλομορφίας. Τα Μεγάλα Δεδομένα αναδεικνύονται ως μια εύλογη λύση για τον μετασχηματισμό της βιομηχανίας της υγειονομικής περίθαλψης ώστε να μειωθεί το κόστος και τελικά να οδηγήσει σε οικονομική ανάπτυξη. Παράλληλα η αξιοποίηση των Μεγάλων Δεδομένων στην βιομηχανία της υγειονομικής περίθαλψης, τα ζητήματα ασφάλειας και ιδιωτικότητας βρίσκονται στο επίκεντρο, καθώς οι αναδυόμενες απειλές και τα τρωτά σημεία συνεχίζουν να αυξάνονται.

Οι Matturdi et al. (2014) εξετάζουν στην έρευνα τους πρώτα τα τεράστια οφέλη και τις προκλήσεις της ασφάλειας και της ιδιωτικότητας των Big Data, και στη συνέχεια, παρουσιάζουν κάποιες πιθανές μεθόδους και τεχνικές για να διασφαλιστεί η ασφάλεια και η ιδιωτικότητα των Μεγάλων Δεδομένων. Επικεντρώθηκαν σε αυτό το ερευνητικό πλαίσιο διότι τα Big Data γίνονται σταδιακά ένα κορυφαίο θέμα έρευνας των επιχειρήσεων και έχει εφαρμογές παντού σε πολλές βιομηχανίες.

Οι Gupta et al. (2014) πρότεινουν ένα μοντέλο ασφάλειας που δεν παρουσιάζει χαρακτηριστικά ελέγχου ασφαλείας και πρόσβασης τα οποία χρησιμοποιούνται κατά τη στιγμή της προέλευσης του Big Data. Ο ρυθμός αύξησης των συσκευών που αφορούν δεδομένα και η παραγωγή δεδομένων αυξήθηκε εκθετικά, γεγονός που οδήγησε στη μαζική έκρηξη της διαθεσιμότητας δεδομένων τόσο σε δομημένη ή αδόμητη μορφή που είναι επίσης γνωστή ως μεγάλα δεδομένα. Η πρόκληση είναι να αναλυθεί αυτή η αδόμητη μορφή δεδομένων και κατηγοριών σε μια πολύ συγκεκριμένη μορφή. Το εργαλείο σχεδιάστηκε για τη διεκπεραίωση μεγάλων δεδομένων χωρίς να λαμβάνεται υπόψη το χαρακτηριστικό ασφάλειας και ελέγχου πρόσβασης. Επιπλέον, λόγω του τεράστιου μεγέθους των δεδομένων γίνεται ένα πολύ κουραστικό καθήκον να ενσωματωθεί το χαρακτηριστικό ασφαλείας σε μεταγενέστερο στάδιο. Επομένως, η ασφάλεια πρέπει να ενσωματωθεί στο αρχικό στάδιο ή στο στάδιο του σχεδιασμού.

Η ανάπτυξη των τεχνολογιών εξόρυξης δεδομένων θέτουν σοβαρούς κινδύνους για την ασφάλεια των ευαίσθητων πληροφοριών του ατόμου. Ένα αναδυόμενο ερευνητικό θέμα στην εξόρυξη δεδομένων, γνωστό ως privacy preserving (PPDM), έχει μελετηθεί εκτενώς τα τελευταία χρόνια. Η βασική ιδέα του PPDM είναι να τροποποιήσει τα δεδομένα με τέτοιο τρόπο ώστε να διεξάγει αποτελεσματικά τους αλγορίθμους εξόρυξης δεδομένων χωρίς να υπάρχει συμβιβασμός στην ασφάλεια των ευαίσθητων πληροφοριών που περιέχονται στα δεδομένα. Τα θέματα ιδιωτικότητας που σχετίζονται με την εξόρυξη δεδομένων ερευνούνται από τον Lei et al. (2014) με διάφορες προσεγγίσεις που μπορούν να βοηθήσουν στην προστασία ευαίσθητων πληροφοριών. Εντοπίζονται τέσσερις διαφορετικοί τύποι χρηστών που εμπλέκονται σε εφαρμογές εξόρυξης δεδομένων, ο παροχέας δεδομένων, ο συλλέκτης, αυτός που κάνει την εξόρυξη και ο υπεύθυνος λήψης αποφάσεων όπου για τον κάθε ένα παρουσιάζονται μεθόδοι που μπορούν να υιοθετηθούν για την προστασία ευαίσθητων πληροφοριών. Εξετάζονται επίσης θεωρητικές προσεγγίσεις οι οποίες προτείνονται για την ανάλυση των αλληλεπιδράσεων μεταξύ διαφορετικών χρηστών σε ένα σενάριο εξόρυξης δεδομένων, το καθένα από τα οποία έχει την δική του εκτίμηση για τις ευαίσθητες πληροφορίες.

Η Elisa Bertino (2015) στην έρευνα που πραγματοποιήθηκε εισάγει ένα ερευνητικό πρόγραμμα για την ασφάλεια και την προστασία της ιδιωτικότητας στα μεγάλα δεδομένα. Εξετάζει τις ερευνητικές προκλήσεις και τις κατευθύνσεις σχετικά με την εμπιστευτικότητα των δεδομένων, την ιδιωτικότητα και την αξιοπιστία στο πλαίσιο μεγάλων δεδομένων. Τα βασικά ερευνητικά ζητήματα που αναλύονται περιγράφουν τον τρόπο που συνδιάζεται η ασφάλεια με την προστασία της ιδιωτικότητας, την έννοια της ιδιοκτησίας δεδομένων και τον τρόπο επιβολής του ελέγχου πρόσβασης στα μεγάλα δεδομένα. Η αντιμετώπιση αυτών των προκλήσεων απαιτεί διεπιστημονική έρευνα που προέρχεται από πολλούς διαφορετικούς τομείς, όπως η επιστήμη της πληροφορικής και η μηχανική, τα συστήματα πληροφοριών, η στατιστική, τα οικονομικά, οι κοινωνικές επιστήμες και ο ανθρώπινος παράγοντας. Όλες αυτές οι προοπτικές είναι απαραίτητες για την επίτευξη αποτελεσματικών λύσεων στο πρόβλημα της ιδιωτικότητας και της ασφάλειας στην εποχή των μεγάλων δεδομένων και ιδιαίτερα στο πρόβλημα του συνδυασμού αυτών των δύο.

Οι χρήστες αποθηκεύουν τεράστια ποσά ευαίσθητων δεδομένων σε μια μεγάλη πλατφόρμα δεδομένων. Η ανταλλαγή ευαίσθητων δεδομένων θα βοηθήσει τις επιχειρήσεις να μειώσουν το κόστος της παροχής εξατομικευμένων υπηρεσιών στους χρήστες και να παρέχουν υπηρεσίες δεδομένων προστιθέμενης αξίας. Ωστόσο, η ασφαλής ανταλλαγή δεδομένων είναι προβληματική. Οι Dong et al. (2015) !προτείνουν ένα πλαίσιο για την ασφαλή ανταλλαγή δεδομένων σε μια μεγάλη πλατφόρμα δεδομένων, συμπεριλαμβανομένης της ασφαλούς παράδοσης, της αποθήκευσης, της χρήσης και της καταστροφής σε μια πλατφόρμα ανταλλαγής δεδομένων ημι-εμπιστοσύνης. Παρουσιάζεται ένας αλγόριθμος ανα-κρυπτογράφησης και μια μέθοδο προστασίας, η οποία παρέχει υποστήριξη για την υλοποίηση των λειτουργιών του συστήματος.

Οι Li et al. (2016) αναφέρονται στα Financial Service Institutions (FSIs) όπου αναλύεαται μια καινοτόμα προσέγγιση του συνδυασμού των συστημάτων νέφωσης με αυτα των Μεγάλων Δεδομένων ώστε να διαφοροποιήσουν την προσφορά υπηρεσιών με αποτελεσματικό τρόπο. Ωστόσο, η ασφάλεια εξακολουθεί να είναι ένα μεγάλο ζήτημα στο οποίο η διαθεσιμότητα υπηρεσιών συχνά έρχεται σε σύγκρουση με τους περιορισμούς ασφαλείας. Η πρόταση που παρουσιάζεται είναι η χρήση των τεχνικών SBAC (Semantic Based Control Access Control) για την απόκτηση ασφαλών χρηματοπιστωτικών υπηρεσιών σε μεγάλα δεδομένα πολυμέσων στον τομέα του Cloud Computing. Η προτεινόμενη προσέγγιση ονομάζεται IntercroSsed Secure Big Multimedia Model (2SBM), η οποία έχει σχεδιαστεί για να εξασφαλίζει πρόσβαση μεταξύ διαφόρων μέσων μέσα από τις πολλαπλές πλατφόρμες cloud. Οι κύριοι αλγόριθμοι που υποστηρίζουν το προτεινόμενο μοντέλο περιλαμβάνουν τον αλγόριθμο Ontology-Based Access Recognition (OBAR) και τον αλγόριθμο Semantic Information Matching (SIM).

Οι Manogaran et al. (2016) αναφέρεται στην προστασία των Μεγάλων Δεδομένων στο περιβάλλον του Cloud. Το νέφος χρησιμοποιείται όλο και περισσότερο για την αποθήκευση και επεξεργασία των μεγάλων δεδομένων όμως υπάρχουν αρκετά θέματα προστασίας και ασφάλειας που πρέπει να αντιμετωπισθούν. Οι παραδοσιακοί μηχανισμοί ασφαλείας που χρησιμοποιούν την κρυπτογράφηση δεν είναι ούτε αποτελεσματικοί ούτε κατάλληλοι για την προστασία μεγάλων δεδομένων στο Cloud. Στην έρευνα αυτή, παρουσιάζονται οι προκλήσεις και οι πιθανές λύσεις για την προστασία μεγάλων δεδομένων στο υπολογιστικό σύννεφο. Καθώς και προτείνεται η Meta Cloud Data Storage Architecture για την προστασία μεγάλων δεδομένων στο περιβάλλον των σύννεφων. Αυτό το πλαίσιο διασφαλίζει την αποτελεσματική επεξεργασία των μεγάλων δεδομένων στο περιβάλλον του cloud computing και κερδίζει περισσότερα επιχειρηματικά δεδομένα.

Οι Παπαδόπουλος και Ψάννης (2016) προτείνουν μια γενική προσέγγιση ενός μοντέλου πραγματικού χρόνου κρυπτογράφησης δεδομένων και watermarking για εικόνες, βίντεο και ηχητικό σήμα. Στόχος είναι η ανάπτυξη ισχυρών παραλλαγών ασφάλειας και επέκτασης χωρητικότητας της Στεγανογραφίας σε πραγματικό χρόνο RT(real time) ή κοντά σε πραγματικό χρόνο NRT(near real time) εικόνας, βίντεο και ηχητικού σήματος. Επιπλέον, αυτή η έρευνα περιέχει την πρόταση συγκεκριμένων υποθέσεων υποδειγμάτων όπως την Στεγανογραφία του συνόλου των Οπτικά Κρυπτογράφημένων σχεδίων σε ασπρόμαυρες εικόνες. Εισάγεται μια νέα μέθοδος που ονόμαζεται εν συντομία SMLSB (Sequential

Multiple LSB) η οποία αποτελεί μια διαφορετική προσέγγιση όσον αφορά την κρυπτογράφιση δεδομένων σε πραγματικό και μη πραγματικό χρόνο, γεγονός που δίνει ισχυρό πλεονέκτημα, ενώ παρέχει διάφορους τρόπους μεθόδων ανάλογα τις απαιτήσεις.

Οι Strang και Sun (2016) στην ερευνά τους συγκέντρωσαν 79.012 άρθρα από το 1916 έως το 2016 σχετικά με τα Μεγάλα Δεδομένα για να καθορίσουν ποια θέματα μελετήθηκαν και πόσα βιβλία επικεντρώθηκαν στις λέξεις-κλειδιά που σχετίζονται με την προστασία της ιδιωτικής ζωής ή την ασφάλεια. Η ανάλυση έδειξε ότι το παράδειγμα των Μεγάλων Δεδομένων ξεκίνησε στα τέλη του 2011 και η ερευνητική παραγωγή αυξήθηκε εκθετικά από το 2012, η οποία προσέγγισε μια κατανομή Weibull που συγκέντρωσε το 82% της διακύμανσης ($p < .01$). Διαπιστώθηκε ότι υπήρχαν 13 κυρίαρχα θέματα που κάλυπταν το 49% της παραγωγής των Μεγάλων Δεδομένων σε περιοδικά κατά την περίοδο 2011-2016, αλλά τα θέματα της ιδιωτικότητας και της ασφάλειας αντιπροσώπευαν μόνο το 2% και η τάση αυτή μειώθηκε πρόσφατα σε λιγότερο από 1%. Ως εκ τούτου, το αποτέλεσμα της έρευνας είναι ότι πρέπει να προωθήθει μια μεγαλύτερη έρευνα για την προστασία της ιδιωτικής ζωής και των δεδομένων.

Οι Yin et al. (2016) !επικεντρώνουν την ερευνά τους στα Μεγάλα Δεδομένα και τις συνεχείς προκλήσεις ασφάλειας και προστασίας των δεδομένων που εμφανίζονται. Προτείνεται το βελτιωμένο μοντέλο που ενσωματώνουν την K-ανωνυμία (K-anonymity) με την L-ποικιλομορφία (L- diversity) και μπορούν να λύσουν το πρόβλημα της μη ισορροπημένης κατανομής ευαίσθητων χαρακτηριστικών. Ο αλγόριθμος ομαδοποίησης K-μέλους (K-member) μπορεί να μεταφράσει το πρόβλημα της ανωνυμίας στο πρόβλημα της ομαδοποίησης και να βρει ένα σύνολο κλάσεων ισοδυναμίας στις οποίες τα αρχεία θα γενικευτούν στην ίδια τιμή. Χρησιμοποιήθηκε ο αλγόριθμος ομαδοποίησης του K- μέλους για να υλοποιήθει το βελτιωμένο μοντέλο ανωνυμίας που μπορεί να μειώσει τον χρόνο εκτέλεσης αλγορίθμου και την απώλεια πληροφοριών. Η ενσωμάτωση του μοντέλου ανωνυμίας και του αλγόριθμου ομαδοποίησης καθιστά τη διαδικασία γενίκευσης πιο αποτελεσματική, πράγμα που είναι ιδιαίτερα σημαντικό για Μεγάλα Δεδομένα.

Τα Μεγάλα Δεδομένα έχουν ως αποτέλεσμα το αυξανόμενο ενδιαφέρον τόσο για τον επιστημονικό όσο και για τον βιομηχανικό τομέα. Ωστόσο, προτού να χρησιμοποιήσουμε την τεχνολογία τωνμεγάλων δεδομένων σε τεράστιες εφαρμογές, θα πρέπει να διερευνηθεί ένα βασικό θέμα: η ασφάλεια και η ιδιωτικότητα. Οι Ye et al. (2016) πραγματοποιείται έρευνα σχετικά με την ασφάλεια και την προστασία της ιδιωτικότητας στα μεγάλα δεδομένα. Πρώτον, περιγράφονται οι επιδράσεις των χαρακτηριστικών των μεγάλων δεδομένων στην ασφάλεια των πληροφοριών και στην ιδιωτικότητα. Στη συνέχεια συζητούνται και εξετάζονται θέματα και ζητήματα σχετικά με την ασφάλεια. Επιπλέον, μελετάται η δημοσίευση δεδομένων για τη διαφύλαξη της ιδιωτικότητας, λόγω της μελλοντικής χρήσης, ιδίως στη λειτουργία των τηλεπικοινωνιών.

Οι προηγμένες τεχνολογίες επικοινωνιών και επεξεργασίας δεδομένων προσφέρουν μεγάλα οφέλη στο έξυπνο δίκτυο. Ωστόσο, οι απειλές στον κυβερνοχώρο επίσης εκτείνονται από το σύστημα πληροφοριών στο έξυπνο δίκτυο. Τα υφιστάμενα έργα ασφάλειας για το έξυπνο δίκτυο εστιάζονται στις παραδοσιακές μεθόδους προστασίας και ανίχνευσης. Ωστόσο, πολλές απειλές εμφανίζονται σε πολύ σύντομο χρονικό διάστημα και παραβλέπονται εξερχόμενοι από τα εξαρτήματα ασφάλειας. Επιπλέον, είναι πολύ αργά να αναλάβουμε δράση για να υπερασπιστούμε τις απειλές μόλις ανιχνευθούν και οι ζημίες θα είναι δύσκολο να αποκατασταθούν. Προκειμένου να αντιμετωπιστεί αυτό το ζήτημα ο Wu et al. (2016) προτείνει έναν μηχανισμό επίγνωσης της κατάστασης ασφάλειας, ο οποίος βασίζεται στην ανάλυση μεγάλων δεδομένων στο έξυπνο δίκτυο. Η αναλυτική μέθοδος που βασίζεται σε ασαφές σύμπλεγμα και την θεωρία παιγνίων ενσωματώνονται απρόσκοπτα στην εκτέλεση της ανάλυσης της κατάστασης ασφαλείας για το έξυπνο δίκτυο. Τα αποτελέσματα προσομοίωσης και πειραματισμού δείχνουν τα πλεονεκτήματα του σχεδίου όσον αφορά την υψηλή απόδοση και το χαμηλό ποσοστό σφάλματος για την επίγνωση της κατάστασης ασφάλειας.

Η βιοϊατρική έρευνα συχνά περιλαμβάνει τη μελέτη δεδομένων ασθενών που περιέχουν προσωπικές πληροφορίες. Η ακατάλληλη χρήση αυτών των δεδομένων μπορεί να οδηγήσει σε διαρροή ευαίσθητων πληροφοριών, οι οποίες μπορούν να θέσουν σε κίνδυνο την

ιδιωτικότητας των ασθενών. Το πρόβλημα της διαφύλαξης της ιδιωτικότητας των ασθενών έχει λάβει αυξημένες επιφυλάξεις στην εποχή των μεγάλων δεδομένων. Έχουν αναπτυχθεί πολλές μέθοδοι προστασίας για την προστασία από διάφορα μοντέλα επίθεσης. Ο Wang et al. (2016) εξετάζει σχετικά θέματα στο πλαίσιο της βιοϊατρικής έρευνας. Γίνεται ανάλυση τεχνολογίων διατήρησης της ιδιωτηκότητας που σχετίζονται με τη σύνδεση καταγραφής, την παραγωγή συνθετικών δεδομένων και την προστασία της ιδιωτηκότητας των γονιδιωματικών δεδομένων. Επίσης γίνεται αναφορά για τις δεοντολογικές συνέπειες της ιδιωτικότητας των Μεγάλών Δεδομένων στη βιοϊατρική και τις παρούσες προκλήσεις στις μελλοντικές κατευθύνσεις έρευνας για τη βελτίωση της ασφάλειας των δεδομένων στη βιοϊατρική έρευνα.

Έχουν αναπτυχθεί διάφοροι μηχανισμοί διατήρησης της ιδιωτικής ζωής για την προστασία της σε διαφορετικά στάδια (π.χ. παραγωγή δεδομένων, αποθήκευση δεδομένων και επεξεργασία δεδομένων) ενός μεγάλου κύκλου ζωής δεδομένων. Οι Mehmood et.al (2017) παρέχουν μια συνολική εικόνα των μηχανισμών διαφύλαξης της ιδιωτικής ζωής στα Μεγάλα Δεδομένα και παρουσιάζουν τις προκλήσεις για τους μηχανισμούς αυτούς. Συγκεκριμένα, απεικονίζεται η υποδομή των Μεγάλων Δεδομένων και των σύγχρονων μηχανισμών διαφύλαξης της ιδιωτικής ζωής σε κάθε στάδιο του μεγάλου κύκλου ζωής των δεδομένων.

Οι Dubey και Srivastava (2017) στην έρευνα τους εξέτασαν τη μεγάλη πρόκληση για τα μεγάλα δεδομένα και αυτή είναι η ασφάλεια των δεδομένων. Ακόμη και με τα τεράστια πλεονεκτήματα, ο κλάδος λαμβάνει πίσω θέση για τη μετάβαση από την κανονική βάση δεδομένων σε μεγάλα δεδομένα λόγω της ανησυχίας για την προστασία της ιδιωτικότητας. Ακόμα και πολλοί μεγάλοι οργανισμοί δεν θεωρούν τα μεγάλα δεδομένα ως μια ασφαλή επιλογή δεδομένου ότι τα δεδομένα μπορούν να προσεγγιστούν από οποιονδήποτε. Ορισμένες διαφορετικές μέθοδοι όπως η τεχνική κρυπτογράφησης-αποκρυπτογράφησης, η ανωνυμία έχουν προταθεί από τους ερευνητές που εργάζονται για να ξεπεράσουν τη σοβαρή απειλή του Big Data, δηλαδή της ασφάλειας των δεδομένων. Αλλά, δυστυχώς, λόγω των τριών στοιχείων των Μεγάλων Δεδομένων, όπως δηλώνεται από την Gartner, δηλαδή Velocity, Volume και Variety, αυτές οι μέθοδοι δεν αποδείχθηκαν πλεονεκτικές. Για τους λόγους που αναφέρθηκαν τα διάφορα ζητήματα ανησυχίας των μεγάλων δεδομένων, που θα ερευνηθούν θα έχουν επίκεντρο κυρίως την ασφάλεια.

Οι Gupta, Arachchilage και Ψάννης (2017) σε ερευνά τους επικεντρώθηκαν στις απειλές για την ασφάλεια των συστημάτων και των δικτύων οι οποιές αυξάνονται ραγδαία με την ανάπτυξη της τεχνολογίας. Μία τέτοια σοβαρή απειλή είναι το "phishing", στο οποίο οι επιτιθέμενοι προσπαθούν να κλέψουν τα διαπιστευτήρια του χρήστη χρησιμοποιώντας ψεύτικα μηνύματα ηλεκτρονικού ταχυδρομείου ή ιστοσελίδες ή και τα δύο. Αρχικά αναλυέται λεπτομερώς η ιστορία των επιθέσεων phishing και τα κίνητρα των εισβολέων καθώς ταξινομούνται και οι διάφοροι τύποι phishing επιθέσεων. Στην συνέχεια παραθέτονται οι λύσεις που προτείνονται στη βιβλιογραφία για την προστασία των χρηστών από το ηλεκτρονικό ψάρεμα με βάση τις επιθέσεις που εντοπίστηκαν στη ταξινόμησή. Ολοκληρώνοντας, αναπτύχθηκαν οι επιπτώσεις των επιθέσεων ηλεκτρονικού "ψαρέματος" στο Διαδίκτυο των πραγμάτων (IoT) καθώς και οι προκλήσεις που εξακολουθούν να υπάρχουν, οι οποίες είναι σημαντικές για την καταπολέμηση των απειλών ηλεκτρονικού "ψαρέματος".

Η ανάλυση των Μεγάλων Δεδομένων έχει καταστεί βασικός παράγοντας καινοτομίας και ανταγωνιστικότητας. Μαζί με την παγκόσμια αύξηση του πληθυσμού και την τάση γήρανσης του πληθυσμού στις ανεπτυγμένες χώρες, το ποσοστό της χρήσης της εθνικής ιατρικής περίθαλψης αυξάνεται. Λόγω του γεγονότος ότι τα μεμονωμένα ιατρικά δεδομένα είναι συνήθως διασκορπισμένα σε διαφορετικά ιδρύματα και οι μορφές τους είναι ποικίλες, η ενσωμάτωση αυτών των δεδομένων που συνεχίζουν να αυξάνονται είναι πρόκληση. Πρέπει να εξεταστούν ορισμένα ζητήματα προκειμένου να χρησιμοποιηθεί το cloud computing για την ταχεία ενσωμάτωση μεγάλων ιατρικών δεδομένων σε βάσεις δεδομένων για εύκολη ανάλυση, αναζήτηση και φιλτράρισμα μεγάλων δεδομένων για την απόκτηση πολύτιμων πληροφοριών. Οι Yang et al. (2017) στην έρευνα τους δημιουργούν ένα σύστημα αποθήκευσης cloud με το HBase του Hadoop για την αποθήκευση και ανάλυση μεγάλων δεδομένων των ιατρικών αρχείων. Τα δεδομένα των ιατρικών αρχείων αποθηκεύονται στην

πλατφόρμα βάσης δεδομένων HBase για μεγαλύτερη ανάλυση. Αυτό το σύστημα εκτελεί επεξεργασία δεδομένων ιατρικών αρχείων μέσω του Hadoop MapReduce και παρέχει λειτουργίες, όπως αναζήτηση λέξεων-κλειδιών, φιλτράρισμα δεδομένων και βασικά στατιστικά στοιχεία για τη βάση δεδομένων HBase. Πραγματοποιείται χρήση του Put και του μηχανισμού CompleteBulkload για την εισαγωγή ιατρικών δεδομένων. Από τα πειραματικά αποτελέσματα διαπιστώθηκε ότι όταν το μέγεθος του αρχείου είναι μικρότερο από 300MB, χρησιμοποιείται η μέθοδος Put και όταν το μέγεθος του αρχείου είναι μεγαλύτερο από 300MB, χρησιμοποιείται ο μηχανισμός Complete-Bulkload για να βελτιώσει την απόδοση της εισαγωγής δεδομένων σε βάση δεδομένων. Παρέχεται μια διεπαφή ιστού που επιτρέπει στους χρήστες να αναζητούν δεδομένα, να φιλτράρουν σημαντικές πληροφορίες μέσω του ιστού, να αναλύουν και να μετατρέπουν τα δεδομένα σε κατάλληλες μορφές που θα είναι χρήσιμες για το ιατρικό προσωπικό και τα ιδρύματα.

Οι Στεργίου και Ψάννης (2017) σε μια πρόσφατη μελέτη τους ερευνούν τα Big Data και το Cloud Computing ως προς την βελτίωση και βελτιστοποίηση των ζητήματων της ιδιωτικότητας και ασφάλειας. Μελέτησαν αυτές οι δύο νέες τεχνολογίες καθώς αναπτύσσονται με μεγάλο ρυθμό στον τομέα των ασύρματων τηλεπικοινωνιών και αντιμετωπίζουν ζητήματα ασφάλειας και προστασίας της ιδιωτικότητα για τα οποία πρέπει να βρεθεί μια λύση. Συγκεκριμένα, συνδυάζεται η λειτουργικότητα των δύο τεχνολογιών με σκοπό να εξεταστούν τα πιο συχνά χαρακτηριστικά και να βρεθούν τα οφέλη που σχετίζονται με θέματα ασφάλειας της ενσωμάτωσής τους. Παρουσιάζουν μια νέα μέθοδο ενός αλγορίθμου που μπορεί να χρησιμοποιηθεί για τη βελτίωση της ασφάλειας του Cloud Computing μέσω της χρήσης αλγορίθμων που μπορούν να παρέχουν περισσότερη προστασία της ιδιωτικότητας στα δεδομένα που σχετίζονται με την τεχνολογία Big Data. Καταλήγουν, παρουσιάζοντας μια έρευνα σχετικά με τις προκλήσεις της ενσωμάτωσης των Big Data και του Cloud Computing σχετικά με το επίπεδο ασφαλειάς τους.

Ο Wu et al. (2017) παρουσιάζει πρώτα τον ορισμό της συσχετιζόμενης διακριτικής εμπιστευτικότητας για την αξιολόγηση του πραγματικού επιπέδου προστασίας προσωπικών δεδομένων ενός μεμονωμένου συνόλου δεδομένων που επηρεάζεται από τα άλλα σύνολα δεδομένων. Στη συνέχεια, κατασκευάστηκε ένα μοντέλο παιχνιδιών πολλαπλών παικτών, στο οποίο εκδίδει κάθε σύνολο δεδομένων από το διαφορικό απορρήτου. Γίνεται ανάλυση της ύπαρξης και τη μοναδικότητας της καθαρής ισορροπίας Nash. Τέλος, αναφέρεται μια έννοια, δηλαδή την τιμή της αναρχίας, για την αξιολόγηση της αποτελεσματικότητας της καθαρής ισορροπίας Nash.

Ο Bou-Harb et al. (2017) ασχολείται με τα προβλήματα των δεδομένων και την επίγνωση της κατάστασης του κυβερνοχώρου, αναλύοντας 910 GB πραγματικής κλίμακας κυκλοφορία , η οποία συλλέχθηκε παθητικά παρακολουθώντας σχεδόν 16,5 εκατομμύρια διευθύνσεις IP σκοτεινού δικτύου. Ένα νέο μοντέλο έχει σχεδιαστεί χρησιμοποιώντας μια κατανεμημένη πολυνηματική προσέγγιση, καθιστώντας την λειτουργική και εξαιρετικά αποτελεσματική στα μεγάλα δεδομένα. Η προσέγγιση μειώνει μοναδικά τη διαστατικότητα αυτών των μεγάλων δεδομένων χρησιμοποιώντας τα τεχνουργήματα της, αντί να επεξεργάζεται τα πραγματικά ανεπεξέργαστα δεδομένα. Αυτό επιτυγχάνεται με την εξαγωγή και την ανάλυση των χρονικών σειρών ανίχνευσης χρησιμοποιώντας τις επίσημες μεθόδους που έχουν τις ρίζες τους στον μετασχηματισμό Fourier και στο φιλτράρισμα Kalman. Το μοντέλο εξυγίανσης του σκοτεινού δικτύου και η προσέγγιση της συσχέτισης είναι σημαντικής αξίας, δεδομένης της θέσης τους που είναι ιδιαίτερα εφαρμόσιμη στον τομέα των μετρήσεων του Διαδικτύου για την ασφάλεια στον κυβερνοχώρο.

Τα σύνθετα συστήματα Big Data σε σύγχρονους οργανισμούς καθίστανται προοδευτικά στόχοι επίθεσης από υπάρχοντες και αναδυόμενες απειλές. Με την συνεχώς αυξανόμενη τάση των εγκλημάτων στον κυβερνοχώρο και των περιστατικών που οφείλονται σε αυτά τα τρωτά σημεία, είναι απαραίτητη η αποτελεσματική διαχείριση για τις σύγχρονες οργανώσεις ανεξάρτητα από το μέγεθός τους. Αξιοποιώντας την πλούσια αλλά και πολύπλοκη ευπάθεια στα δεδομένα οι Tang, Alazab και Luo (2017) πραγματοποιήσαν σημαντική μελέτη όχι μόνο για το χειρισμό των επίμονων μεταβλητών στα δεδομένα αλλά και για την περαιτέρω αποκάλυψη της δομής πολλών μεταβλητών μεταξύ των διαφόρων κινδύνων τρωτότητας. Μέσα από τις εκτενείς εμπειρικές μελέτες αποδείχθηκε ότι ένα

σύνθετο μοντέλο μπορεί να καταγράψει αποτελεσματικά και να διατηρήσει μακροπρόθεσμη εξάρτηση μεταξύ διαφορετικής ευπάθειας.

Λόγω της πολυπλοκότητας και του όγκου, η εξωτερική ανάθεση κρυπτογραφικών στοιχείων σε ένα νέφος θεωρείται μία από τις πιο αποτελεσματικές προσεγγίσεις για αποθήκευση Μεγάλων Δεδομένων. Παρ 'όλα αυτά, ο κάτοχος δεδομένων έχει κρίσιμες προκλήσεις ώστε η μεγάλη αποθήκευση δεδομένων βασισμένη σε νέφος να είναι και αποτελεσματική. Ο Hu et al. (2017) ! προτείνει ένα ασφαλές και επαληθεύσιμο σύστημα ελέγχου πρόσβασης που βασίζεται στο κρυπτοσύστημα NTRU για μεγάλη αποθήκευση δεδομένων σε νέφοι. Χρησιμοποιείται ο αλγόριθμος αποκρυπτογράφησης NTRU για να ξεπεραστούν οι αποτυχίες αποκρυπτογράφησης του αρχικού NTRU και στη συνέχεια αναλύεται η ορθότητα, η ασφαλεία και η υπολογιστική αποδοτικότητα. Επιτρέπει, στον κάτοχο δεδομένων και τους κατάλληλους χρήστες να επαληθεύουν αποτελεσματικά τη νομιμότητα ενός χρήστη για πρόσβαση στα δεδομένα και έναν χρήστη να επικυρώνει τις πληροφορίες που παρέχονται από άλλους χρήστες για σωστή ανάκτηση ιστότοπου. Η αυστηρή ανάλυση δείχνει ότι το πρόγραμμά μπορεί να εμποδίσει τους επιλέξιμους χρήστες να εξαπατήσουν και να αντισταθούν σε διάφορες επιθέσεις.

Η ανάπτυξη της τεχνολογίας και η εισαγωγή σε νέες έννοιες και τεχνικές οδήγησαν στην έρευνα τους και την εξέλιξη τους. Ο Κ.Ψάννης et.al (2017-2018) έχουν ερευνήσει έννοιες όπως το Cloud και τα Μεγάλα Δεδομένα με την χρήση αλγόριθμών και τεχνικών ώστε να εξελίξουν και να παρουσιάσουν νέες ιδέες. Επίσης, παρουσιάζουν και μεθόδους ώστε να συνδέσουν τα παραπάνω με το Διαδίκτυο των Πραγμάτων. Μια επίσης νέα τεχνολογία άμεσα συνδεδεμένη με τους μεγάλους όγκους δεδομένων που πρέπει να διαχειριστούν. Καταλήγουν στην συνεχή εξέλιξη και πρόοδο με προτάσεις καινοτομιών και εφαρμογών.

2.5 Κριτήρια για την Σύγκριση της Βιβλιογραφικής Ανασκόπησης

Αυτή η ενότητα επισημαίνει διάφορες παραμέτρους που παίζουν ζωτικό ρόλο στην αποδοχή οποιουδήποτε στοιχείου που χρησιμοποιήθηκε στην βιβλιογραφική ανασκόπηση. Επομένως, τα στοιχεία αυτά σχετικά με την ασφάλεια των Μεγάλων Δεδομένων που αντιμετωπίζουν τις περισσότερες από τις ακόλουθες παραμέτρους θεωρούνται ως προεξέχοντα.

- Context Awareness **(Επίγνωση Πλαισίου)**

Η επίγνωση του πλαισίου αναφέρεται στην ευαισθητοποίησή σχετικά με τις οντότητες και τις παραμέτρους που μπορούν να επηρεάσουν την απόφαση υπολογισμού της πορείας. Ετυμολογικά, με τον όρο πλαίσιο (context), αναφερόμαστε σε οτιδήποτε περιβάλλει και ορίζει κάτι άλλο. Εξειδικεύοντας τον όρο στον τομέα της πληροφορικής, αυτό το «κάτι άλλο» μπορεί να είναι μία υπόθεση, ένα άτομο, ένα υπολογιστικό σύστημα. Κοιτάζοντας τα προσεκτικά, θα βρούμε αρκετά σημεία κοινά, τα οποία θα μας βοηθήσουν στο να αναγνωρίσουμε κάποια συγκεκριμένα πρότυπα τα οποία ακολουθούνται. Σχεδόν πάντα θα αναγνωρίζουμε κάποιο στοιχείο και κάποια κατάσταση που ορίζεται. Πλαίσιο είναι κάθε πληροφορία που μπορεί να χρησιμοποιηθεί για να χαρακτηρίσει την κατάσταση μίας οντότητας. Την προαναφερθείσα πληροφορία μπορούμε να την χωρίσουμε σε δύο κατηγορίες με βάση την πολυπλοκότητά της: Σε στοιχειώδη πλαίσια και σε μοντέλα πλαισίου. Κατά κύριο λόγο, είναι πολύ σημαντικό για ένα μοντέλο εφαρμογής να βρίσκεται σε ένα συνειδητό περιβάλλον, καθώς μπορεί να προκύψουν περιπτώσεις όπου η απόδοση των εφαρμογών υποβαθμίζεται.

- **Generality** (Γενικότητα)

Η γενικότητα ενός μοντέλου εφαρμογής αναφέρεται στην υποστήριξή του για ένα φάσμα εφαρμογών. Στην πράξη, υπάρχουν πολλαπλές μορφές εφαρμογών με διαφορετικές απαιτήσεις και συμπεριφορά πόρων. Για παράδειγμα, οι εργασίες για την εκμάθηση των ιών, τα αρχεία για την αναζήτηση και την ανάκτηση νέων δεδομένων για τα νέα μηνύματα είναι καθήκοντα καθυστέρησης που δεν απαιτούν αλληλεπίδραση χρήστη. Αφού ολοκληρωθούν οι εργασίες, τα αποτελέσματα μπορούν να συγχρονιστούν. Κατά συνέπεια, είναι αρκετά δύσκολο για ένα μοντέλο εφαρμογής να υποστηρίζει πολλαπλούς τύπους εφαρμογών. Ωστόσο, πρέπει να καταβληθούν προσπάθειες για το σχεδιασμό μοντέλων εφαρμογών για να υποστηρίζουν όλους τους τύπους εφαρμογών.

- **Complexity (**Πολυπλοκότητα)

Η πολυπλοκότητα χαρακτηρίζει τη συμπεριφορά ενός συστήματος ή ενός μοντέλου του οποίου τα στοιχεία αλληλεπιδρούν με πολλούς τρόπους και ακολουθούν τους τοπικούς κανόνες, πράγμα που σημαίνει ότι δεν υπάρχει λογική υψηλότερη οδηγία για τον ορισμό των διαφόρων πιθανών αλληλεπιδράσεων. Η πολυπλοκότητα γενικά χρησιμοποιείται για να χαρακτηρίσει κάτι με πολλά μέρη, όπου τα μέρη αλληλεπιδρούν μεταξύ τους με πολλαπλούς τρόπους, καταλήγοντας σε μια υψηλότερη τάξη εμφάνισης μεγαλύτερη από το άθροισμα των τμημάτων της. Ακριβώς όπως δεν υπάρχει απόλυτος ορισμός της "νοημοσύνης", δεν υπάρχει απόλυτος ορισμός της "πολυπλοκότητας", η μόνη συναίνεση μεταξύ των ερευνητών είναι ότι δεν υπάρχει συμφωνία σχετικά με τον συγκεκριμένο ορισμό της πολυπλοκότητας. Η μελέτη αυτών των σύνθετων συνδέσεων σε διάφορες κλίμακες είναι ο κύριος στόχος της σύνθετης θεωρίας των συστημάτων.

- **Scalability** (Επεκτασιμότητα)

Η επεκτασιμότητα είναι ένα σημαντικό χαρακτηριστικό των Μεγάλων Δεδομένων. Επομένως, τα μοντέλα εφαρμογών γιατα Μεγάλα Δεδομένα πρέπει να υποστηρίζουν την ανάπτυξη εφαρμογών που μπορούν να κλιμακωθούν για να καλύψουν τις απρόβλεπτες απαιτήσεις των χρηστών. Επιπλέον, τα μοντέλα εφαρμογών πρέπει να βελτιώσουν τις υποστηριζόμενες λειτουργίες ώστε να ενσωματώσουν έγκαιρα νέους τύπους εφαρμογών. Παρ 'όλα αυτά, τα μοντέλα εφαρμογών για νέους κινητούς υπολογιστές πρέπει επίσης να είναι κλιμακωτά όσον αφορά την υιοθεσία. Ωστόσο, η κλιμάκωση δεν εξαρτάται μόνο από το μοντέλο εφαρμογής και εξαρτάται σε κάποιο βαθμό από την πλατφόρμα ανάπτυξης. Επομένως, τα προαναφερθέντα ζητήματα επεκτασιμότητας πρέπει να λαμβάνονται υπόψη κατά την ανάπτυξη ή την υιοθέτηση μοντέλων εφαρμογών.

- **Enabling Technologies** (Ενεργοποίηση τεχνολογιών)

Η αξιοποίηση των άφθονων δεδομένων για την επίτευξη πιο έξυπνων αποφάσεων απαιτεί νέα σκέψη. Η έννοια των Μεγάλων Δεδομένων δεν σημαίνει ότι οι επιχειρήσεις είχαν μόνο «μικρά δεδομένα» πριν. Σημαίνει ριζική μετατόπιση των περιβαλλόντων δεδομένων από την άποψη του όγκου, της ταχύτητας και της ποικιλίας. Τα Μεγάλα Δεδομένα δημιουργούν μια αρχιτεκτονική αναταραχή όπου τα συστήματα, η αποθήκευση και το λογισμικό συνδέονται και αξιοποιούνται. Η κινητήρια δύναμη για τα Μεγάλα Δεδομένα είναι το λογισμικό και οι πλατφόρμες για την υποδομή και την ανάλυση. Για παράδειγμα δύο αρχιτεκτονικές είναι οι εκτεταμένες RDBMS και MapReduce / Hadoop. Το Hadoop είναι η κύρια υποδομή που χρησιμοποιείται για τη διανομή, την καταλογογράφηση, διαχείριση και διερεύνηση δεδομένων σε πολλαπλούς κόμβους υπηρεσίων με οριζόντια κλίμακα.

- **Methodology** (Μεθοδολογία)

Η μεθοδολογία έρευνας αναφέρεται στις παραμέτρους της ερευνητικής προσπάθειας του ερευνητή, οι οποίες αφορούν στις γενικές μεθοδολογικές προσεγγίσεις, στις μεθόδους, στις τεχνικές, στα μέσα, στα υλικά και στις διαδικασίες που θα επιλέξει για τη διεξαγωγή της έρευνας του. Δηλαδή, μεθοδολογία έρευνας είναι η κατανόηση της επιστημονικής ερευνητικής διαδικασίας: πώς θα σχεδιαστεί μια έρευνα ή και πώς θα πραγματοποιηθεί μια έρευνα. Η μεθοδολογία έχει την έννοια του όρου ως συστηματική μελέτη αρχών που διέπουν επιστημονική και φιλοσοφική διερεύνηση. Υπό την δεύτερη αυτή άποψη η μεθοδολογία προσδιορίζει την όλη θεματική ύλη του όρου με την αρχική σημασία του.

- **Tested Application** (Δοκιμή Εφαρμογής)

Καθώς αναπτύσσονται τα Μεγάλα Δεδομένα δημιουργούνται ολοένα και περισσότερο νέες εφαρμογές και καινοτομίες. Οι εφαρμογές αυτές αποτελούν την βάση της ανάπτυξης των Μεγάλων Δεδομένων καθώς όσες περισσότερες υλοποιούνται τόσο μεγαλύτερη η εξέλιξη που επέρχεται. Ο έλεγχος και οι δοκιμές των εφαρμογών βοηθούν πολύ στην βελτίωση και την επέκταση τους, καθώς μπορούν να διαγνωστούν τυχόν προβλήματα και λάθη. Επιπλέον οι δοκιμές μπορούν να γίνονται σε συγκεκριμένο περιβάλλον για το οποίο έχει γίνει η εφαρμογή καθώς και για διάφορους σκοπούς. Το πιο σημαντικό στοιχείο είναι οτι μπορεί να υπάρξει κάποια μελλοντική έρυνα και βελτίωση της κάθε εφαρμογής και έρευνας πάντα συμβαδίζοντας με την ανάπτυξη των Μεγάλων Δεδομένων.

- **Future Proposed Work** (Μελλοντική Προτεινόμενη Έρευνα)

Ένα στοιχείο άμμεσα συνδεδεμένο με την ανάπτυξη της τεχνολογίας, των Μεγάλων Δεδομένων και ότι συνεπάγεται με αυτά είναι η έρευνα. Σχεδόν σε όλες τις ερευνητικές εργασίες υπάρχει μια πρόταση για μελλοντική έρευνα και εξέλιξη της παρούσας. Κάτι τέτοιο αποτελεί έναν θετικό παράγοντα καθώς υπάρχουν βλέψεις για περεταίρω προέκταση μιας έρευνας ή μιας εφαρμογής.

Πίνακας 1.Σύγκριση της Βιβλιογραφικής Ανασκόπησης

Model	Ca	Ge	Co	Sc	Et	Methodology			Ta	Fw
Jin et al. (2017)	H	M	M	H	L	Theoretical Review			Opportunities & Significance of Big Data	How to Make a Big Data Project Successful
Bi and Cochran (2017)	M	L	H	H	M	Theoretical Review			Clarification the requirements of predictive systems & identification research challenges to support information systems	Parallel and Distributed Systems
Su, Xu and Qi (2017)	H	M	H	H	M	Theoretical Review & Use of CCNs			A framework to deliver mobile Big Data over content-centric mobile social networks	Dynamic mobile big data, Resource allocation, Privacy and security
Meeker and Hong (2017)	H	M	M	H	H	Applications & Examples of Recent Technical Developments			Extending existing models for reliability to take advantage of System operating Data	Software Applications, Methods for products With multiple degradation Performance Characteristics
Dubey et al. (2017)	L	M	M	M	M	Theoretical Review & RSA and AES encryption techniques			Issues of Big Data on Data Security	An encryption & Decryption algorithm where t data stored only encrypted.
Yin et al (2016)	H	L	H	H	H	Anonymization & Clustering			K-member clustering algorithm to improve anonymity model	
	Ca: Context Awareness Ge: Generality Co: Complexity Sc: Scalability Et: Enabling Technologies Ta: Tested Application Fw: Future Work					H: High M: Medium L: Low				

Κεφάλαιο 3: Big Data Analytics Software

Η ανάλυση δεδομένων είναι η διαδικασία εξέτασης μεγάλων και ποικίλων συνόλων δεδομένων όπως τα Μεγάλα Δεδομένα, για την αποκάλυψη κρυφών μοτίβων, άγνωστων συσχετισμών, τάσεων της αγοράς, προτιμήσεων πελατών και άλλων χρήσιμων πληροφοριών που μπορούν να βοηθήσουν τους οργανισμούς να λαμβάνουν πιο ενημερωμένες επιχειρηματικές αποφάσεις**.** Ο χρόνος ανάλυσης των δεδομένων στον οποίο ζούμε χαρακτηρίζεται επαναστατικός. Οι επιχειρήσεις αντιμετωπίζουν τεράστιες ποσότητες και ποικιλίες δεδομένων από τη μια πλευρά και όλο και πιο γρήγορες προσδοκίες για ανάλυση από την άλλη. Η κοινότητα των πωλητών αποκρίνεται παρέχοντας εξαιρετικά κατανεμημένες αρχιτεκτονικές και νέα επίπεδα μνήμης και επεξεργαστικής ισχύος. Επίσης, γίνεται εκμετάλλευση του μοντέλο αδειοδότησης ανοιχτού κώδικα, το οποίο δεν είναι καινούργιο, αλλά γίνεται ολοένα και περισσότερο αποδεκτό και αναζητείται από επαγγελματίες διαχείρισης δεδομένων. Για την ανάλυση των δεδομένων σημαντικό ρόλο εκτελούν τα λογισμικά που χρησιμοποιούνται για τον σκοπό αυτό. Παρακάτω θα αναλύσουμε τα πιο γνωστά από αυτά και θα δούμε τα χαρακτηριστικά τους και τις λειτουργίες τους.

- **MapR Converged Data Platform**

Η MapR Converged Data Platform ενσωματώνει το Hadoop, το Spark και το Apache Drill με δυνατότητες βάσεων δεδομένων σε πραγματικό χρόνο, με παγκόσμια ροή συμβάντων και κλιμακωτή αποθήκευση για την εξουσία μιας νέας γενιάς μεγάλων εφαρμογών δεδομένων. Προσφέρει ασφάλεια σε επίπεδο επιχείρησης, αξιοπιστία και απόδοση σε πραγματικό χρόνο, μειώνοντας δραστικά τόσο το υλικό όσο και το λειτουργικό κόστος των πιο σημαντικών εφαρμογών και δεδομένων των χρηστών του. Το MapR υποστηρίζει δεκάδες έργα ανοιχτού κώδικα και έχει δεσμευτεί να χρησιμοποιεί βιομηχανικά πρότυπα APIs για την παροχή μιας μεθόδου φρικτότητας, αναπτύσσοντας και αξιοποιώντας συναρπαστικές νέες εφαρμογές που μπορούν να ικανοποιήσουν τις πιο μεγάλες απαιτήσεις παραγωγής. Αποτελείται από το Enterprise-Grade Platform Services, το Open Source Engines & Tools και το Commercial Engines &Applications.

Το MapR Platform Services αποτελει τις βασικές δυνατότητες χειρισμού δεδομένων του MapR Converged Data Platform. Ο φιλικός προς τις επιχειρήσεις σχεδιασμός του παρέχει ένα γνωστό σύνολο υπηρεσιών διαχείρισης αρχείων και δεδομένων, συμπεριλαμβανομένου μιας παγκόσμιας ονοματοδοσίας, υψηλής διαθεσιμότητας, προστασίας δεδομένων, ελέγχου πρόσβασης, απόδοσης σε πραγματικό χρόνο, παρακολούθησης και διαχείρισης. Επιπλέον, το MapR πακετάρει ένα ευρύ σύνολο έργων πηγαίου κώδικα Apache που επιτρέπουν μεγάλες εφαρμογές δεδομένων. Ο στόχος είναι να παρέχει μια ανοικτή πλατφόρμα η οποία επιτρέπει στους χρήστες να επιλέξουν το σωστό εργαλείο για την εργασία οτ υλοποιούν. Οι δοκιμές του MapR ενσωματώνει έργα ανοιχτού κώδικα όπως Hive, Pig, Apache HBase και Mahout.

- **IBM Big Data Analytics**

Η IBM Big Data προσφέρει στους χρήστες της μια αρχιτεκτονική επόμενης γενιάς για μεγάλα δεδομένα και στοιχείαανάλυσης, τα οποία προσφέρουν νέες επιχειρηματικές ιδέες μειώνοντας σημαντικά το κόστος αποθήκευσης και συντήρησης. Οι επιστήμονες δεδομένων της IBM διαχωρίχουν τα Μεγάλα Δεδομένα σε τέσσερις διαστάσεις, όπως ο όγκος, η ποικιλία, η ταχύτητα και η ειλικρίνεια. Μια εταιρεία θα είναι σε θέση να παραμείνει στην κορυφή όλων των αλλαγών, συμπεριλαμβανομένων των αναλύσεων Hadoop, της ροής της αποθήκευσης αναλυτικών στοιχείων, της ενσωμάτωσης καθώς και της διακυβέρνησης. Οι εταιρείες θα είναι σε θέση να κερδίσουν κορυφαία απόδοση βάσεων δεδομένων σε πολλαπλές φόρμες εργασίας, μειώνοντας παράλληλα το κόστος διαχείρισης, αποθήκευσης, ανάπτυξης και διακομιστή, χρησιμοποιώντας το IBM Big Data.

Επίσης πραγματοποιώντας εξαιρετική ταχύτητα με βελτιστοποιημένες δυνατότητες για φόρμες εργασίας αναλυτικών στοιχείων υπάρχει ωφέλεια από συστήματα που έχουν

επιδιορθωθεί με φόρτο εργασίας και μπορούν να λειτουργούν για πολλές ώρες, είναι ένα άλλο χαρακτηριστικό από τη διαχείριση δεδομένων της IBM. Οι επιχειρήσεις θα είναι επίσης σε θέση να μεγιστοποιήσουν πλήρως την ισχύ του Apache Hadoop στην επιχείρηση με επιταχυντές εφαρμογών, αναλύσεις, οπτικοποίηση, αναπτυξιακά εργαλεία, βελτιωτικά χαρακτηριστικά και χαρακτηριστικά ασφαλείας. Με τη δυνατότητα διαχείρισης περιεχομένου IBM Big Data, επιτρέπει στην επιχείρηση να διαθέτει ολοκληρωμένο κύκλο ζωής περιεχομένου και διαχείριση εγγράφων με οικονομικά αποδοτικό έλεγχο των υφιστάμενων και νέων τύπων περιεχομένου με κλίμακα, ασφάλεια και σταθερότητα. Οι χρήστες θα μπορούν να καταγράφουν και να αναλύουν όλα τα δεδομένα όλη την ώρα και ακριβώς καθώς και, να αποθηκεύουν λιγότερα, να αναλύουν περισσότερα και να λαμβάνουν καλύτερες αποφάσεις πιο γρήγορα.

- **Cloudera Enterprise Big Data**

Το Cloudera Enterprise Big Data τροφοδοτείται από Apache Hadoop. Το Cloudera Enterprise είναι η ταχύτερη, ευκολότερη και πιο ασφαλής σύγχρονη πλατφόρμα δεδομένων. Από τις αναλύσεις έως την επιστήμη των δεδομένων, ο καθένας μπορεί τώρα να αποκτήσει αποτελέσματα από οποιαδήποτε δεδομένα και σε οποιοδήποτε περιβάλλον, όλα μέσα σε μια ενιαία, κλιμακούμενη πλατφόρμα. Το Cloudera παρέχει τη σωστή πλατφόρμα για να ικανοποιήσει τις ανάγκες των χρηστών, οι χρήστες μπορούν να φέρουν τους μηχανικούς δεδομένων και τους επιστήμονες δεδομένων τους για να χτίσουν αγωγούς σε πραγματικό χρόνο, να επεξεργαστούν ταχύτητα δεδομένα και να αναπτύξουν και να εκπαιδεύσουν μοντέλα δεδομένων. Εκσυγχρονίζει την αρχιτεκτονική πληροφορικής της εταιρείας, ώστε να παρέχει τη δυνατότητα ELT και υψηλής απόδοσης Analytics SQL για αναφορές, εξερεύνηση και αυτοεξυπηρέτηση επιχειρηματικής ευφυΐας.

Με την αποτελεσματική λύση data-in-motion του Cloudera, θα παρέχει τα κατάλληλα εργαλεία για να λαμβάνει, να επεξεργάζεται και να εξυπηρετεί δεδομένα, διατηρώντας παράλληλα την ασφάλεια. Το Cloudera Manager είναι το εργαλείο διαχείρισης του Hadoop, το οποίο εμπιστεύονται οι επαγγελματίες και τις μεγαλύτερες αξίες του Hadoop. Με έξυπνες προεπιλογές και μοναδικές προσαρμογές παρακολούθησης, απλοποιεί δραστικά τις λειτουργίες του συμπλέγματος. Σχεδιασμένο με ένα επεκτάσιμο θεμέλιο, ενσωματώνεται γρήγορα και απρόσκοπτα τόσο με τα εργαλεία τρίτων όσο και με τα νεότερα στοιχεία Hadoop για ενοποιημένες, αξιόπιστες λειτουργίες.

- **1010data**

Το 1010data έχει σχεδιαστεί για να παρέχει ισχυρές πληροφορίες για όλα τα δεδομένα των χρηστών του χρησιμοποιώντας ένα απλοποιημένο σύστημα το οποίο είναι ευέλικτο και απίστευτα γρήγορο. Το καλύτερο από όλα, βάζει τη δύναμη των δεδομένων και αναλύσεων στα χέρια των χρηστών του παρέχοντας περισσότερα για αυτούς και τις ομάδες τους. Οι χρήστες μπορούν να φορτώνουν, να μετασχηματίζουν και να ενσωματώνουν δεδομένα από οποιαδήποτε και όλες τις πηγές, συμπεριλαμβανομένων ισχυρών τρισδιάστατων συνόλων δεδομένων διαθέσιμων απευθείας στην πλατφόρμα 1010data. Μπορούν να εκμεταλλευτούν τις διεπαφές προσαρμοσμένες για κάθε τύπο χρήστη και απόδοση του συστήματος που προσφέρουν χρόνους απόκρισης ερωτημάτων αστραπής πράγμα που επιτρέπει στους χρήστες να ζητούν και να απαντούν σε οποιαδήποτε αναλυτική ερώτηση.

Με το 1010data οι χρήστες μπορούν να δίνουν πληροφορίες σε όσους τις χρειάζονται μέσω ισχυρών αναφορών επιχειρήσεων, με τυποποιημένες KPIs και καθοδηγούμενες ad hoc δυνατότητες. Ως εκ τούτου, μπορούν να πραγματοποιήσουν οπτικοποίηση δεδομένων σε όλα τα μεγάλα δεδομένα τους, απευθείας στην πλατφόρμα. Τέλος, το 1010data επιτρέπει στους χρήστες της να συνεργάζονται με αξιόπιστους συνεργάτες χρησιμοποιώντας κοινά δεδομένα και αναλυτικά στοιχεία για να κερδίσουν ένα ανταγωνιστικό πλεονέκτημα Το 1010data δίνει τη δυνατότητα στους πελάτες να συνδιάζουν δεδομένα και αναλύσεις σε κάθε διαδικασία, με αποτέλεσμα πιο έξυπνες και πιο πολύτιμες

αποφάσεις. Επίσης, εξαλείφει την ανάγκη μετακίνησης δεδομένων μεταξύ διαφορετικών εργαλείων, εξαλείφοντας την αναποτελεσματικότητα και τη σύγχυση.

- **SAP Big Data Analytics**

Η SAP Big Data Analytics εισάγει το SAP HANA 2 που είναι μια πλατφόρμα πληροφορικής που επιτρέπει στους χρήστες της να επιταχύνουν την επιχειρησιακή τους ευφυΐα και να απλοποιήσουν την εμφάνιση της πληροφορικής. Παρέχοντας τη συγχώνευση για όλες τις ανάγκες των εταιρικών δεδομένων, το SAP HANA απομακρύνει το βάρος της διατήρησης ξεχωριστών συστημάτων δεδομένων, έτσι ώστε η επιχείρηση να μπορεί να τρέχει ζωντανά και να λαμβάνει καλύτερες επιχειρηματικές αποφάσεις στη νέα ψηφιακή οικονομία. Από την άλλη πλευρά, οι προγραμματιστές εφαρμογών μπορούν επίσης να χρησιμοποιήσουν το SAP HANA 2 για την παροχή έξυπνων εφαρμογών που εκμεταλλεύονται την προηγμένη αναλυτική επεξεργασία και την ενδυνάμωση όλων των χρηστών, συμπεριλαμβανομένων των εργαζομένων, των πελατών και των ειδικών με βαθύτερη γνώση οποιονδήποτε δεδομένων από οπουδήποτε.

Το SAP Big Data Analytics παρέχει προηγμένη επεξεργασία δεδομένων για κείμενο, χωροταξικά, γραφήματα και σειρές δεδομένων σε ένα σύστημα. Παρέχει επίσης βαθύτερες γνώσεις με ισχυρές δυνατότητες πρόβλεψης και μηχανικής μάθησης. Η πλατφόρμα ενσωματώνει την εικονικοποίηση δεδομένων, συμπεριλαμβανομένης της ενοποίησης και της αναπαραγωγής, καθώς και τις δυνατότητες ποιότητας για την ταχεία πρόσβαση και ανάλυση δεδομένων από οποιαδήποτε πηγή.

- **Oracle Big Data Analytics**

Το Oracle Big Data Analytics βοηθά τις επιχειρήσεις να διαχειρίζονται τις λειτουργίες τους όταν πρόκειται για την αυξανόμενη ποσότητα δεδομένων που ο κόσμος έχει να προσφέρει. Το πρόγραμμα περιλαμβάνει την ικανότητα αξιοποίησης των πλεονεκτημάτων των μεγάλων δεδομένων στο νέφος, την επεκτασιμότητα, την αξιοπιστία και την ευελιξία σε ολόκληρο το περιβάλλον και την προστασία των επενδύσεων και δεξιοτήτων στην εποχή των μεγάλων δεδομένων και του νέφους. Οι μεγάλες λύσεις της Oracle για τη βιομηχανία δεδομένων απευθύνονται στην αυξανόμενη ζήτηση για ενοποιημένη αρχιτεκτονική ολοκλήρωσης, διαχείρισης, αναλύσεων και εφαρμογών για την κάλυψη των μοναδικών επιχειρηματικών απαιτήσεων κάθε κλάδου.

Η Oracle μπορεί να βοηθήσει τις επιχειρήσεις να προχωρήσουν στην τεχνολογική τους υποδομή με λύσεις που ενσωματώνονται από εφαρμογές. Αυτό επιτρέπει επίσης στην επιχείρηση να δημιουργεί και να προσφέρει απλόχερα εμπειρίες στον χρήστη σε μία ασφαλή πλατφόρμα, για οποιαδήποτε συσκευή, σε οποιαδήποτε εφαρμογή. Δημιουργήται αποτελεσματικότητα για την επεξεργασία συμβάντων και εντολών μεγάλης ταχύτητας και μεγάλου όγκου.

- **Hortonworks Data Platform**

Η Hortonworks Data Platform είναι η μοναδική αληθινή, ασφαλής, επιχειρησιακά έτοιμη ανοικτή πηγή Apache Hadoop που βασίζεται σε μια κεντρική αρχιτεκτονική (YARN). Η πλατφόρμα αυτή καλύπτει τις πλήρεις ανάγκες των δεδομένων σε κατάσταση αναπαύσεως, εξουσιοδοτεί τις εφαρμογές πελατών σε πραγματικό χρόνο και παρέχει ισχυρές αναλύσεις που επιταχύνουν τη λήψη αποφάσεων και την καινοτομία. Περιλαμβάνει ένα ευέλικτο φάσμα επεξεργασίας που ενδυναμώνουν τους χρήστες να αλληλεπιδρούν με τα ίδια δεδομένα με πολλούς τρόπους, ταυτόχρονα. Αυτό σημαίνει ότι οι εφαρμογές μπορούν να αλληλεπιδρούν με τα δεδομένα με τον καλύτερο τρόπο: από διαδραστική SQL ή πρόσβαση χαμηλού λανθάνοντος χρόνου με NoSQL.

Αναδυόμενες περιπτώσεις χρήσης για την επιστήμη των δεδομένων, την αναζήτηση και την ροή υποστηρίζονται επίσης με Apache Spark, Storm και Kafka. Όταν πρόκειται για την ασφάλεια, υπάρχουν κρίσιμα χαρακτηριστικά για την πιστοποίηση ταυτότητας, την εξουσιοδότηση, την υπευθυνότητα και την προστασία των δεδομένων ώστε να διασφαλιστεί η ασφάλεια της πλατφόρμας δεδομένων σε αυτές τις βασικές απαιτήσεις. Η συνεπής αυτή προσέγγιση σε όλες τις δυνατότητες της επιχείρησης Hadoop, η πλατφόρμα δεδομένων Hortonworks διασφαλίζει επίσης ότι ο χρήστης μπορεί να ενσωματώσει και να επεκτείνει τις τρέχουσες λύσεις ασφάλειας για να παρέχει μια ενιαία, συνεπή και ασφαλή ομπρέλα πάνω από τη σύγχρονη αρχιτεκτονική δεδομένων της εταιρεία.

- **DataStax BigData**

Το DataStax BigData Enterprise επιταχύνει την ικανότητα της εταιρείας να παρέχει αξία σε πραγματικό χρόνο σε επική κλίμακα παρέχοντας ένα ολοκληρωμένο και λειτουργικά απλό επίπεδο διαχείρισης δεδομένων με μια μοναδική αρχιτεκτονική που βασίζεται στην Apache Cassandra**.** Οι χρήστες μπορούν να γράψουν δεδομένα μία φορά και να έχουν πρόσβαση σε αυτά χρησιμοποιώντας μια ποικιλία φόρτων εργασίας ή μοτίβα πρόσβασης, όλα από μια ενιαία συνεκτική λύση. Η μηχανή ευρετηρίασης επιτρέπει στους χρήστες της να βρίσκουν γρήγορα και εύκολα δεδομένα χρησιμοποιώντας σύνθετα ερωτήματα και περιλαμβάνει υποστήριξη για αναζήτηση υποσυνάρτησης, ασαφή και πλήρους κειμένου.

Το DataStax BigData Enterprise βασίζεται στη λειτουργικότητα που βρίσκεται στο Apache Spark, εξαλείφει πάρα πολύ τα σημεία αποτυχίας, αυξάνει την απόδοση και παρέχει πλήρη ενσωμάτωση στις δυνατότητες αναζήτησης και γραφικών που υπάρχουν στο DSE. Τέλος, παρέχει επίσης μια ισχυρή πλατφόρμα πολλαπλών μοντέλων με υποστήριξη για key-value, πίνακες, JSON / Document και graph. Αυτή η δυνατότητα επιτρέπει στους χρήστες να γράφουν δεδομένα σε μια ενιαία λύση και να έχουν πρόσβαση σε αυτήν χρησιμοποιώντας ποικίλες μεθόδους που βασίζονται στις ανάγκες της εφαρμογής τους.

- **Informatica PowerCenter Big Data Edition**

Το Informatica PowerCenter Big Data Edition είναι ένα εξαιρετικά κλιμακωτό, υψηλής απόδοσης λογισμικό ολοκλήρωσης δεδομένων για επιχειρήσεις, το οποίο χρησιμοποιεί την οπτική ανάπτυξη για να δημιουργήσει ροές δεδομένων ETL που τρέχουν natively στο Hadoop. Οι ροές δεδομένων μπορούν να επαναχρησιμοποιηθούν και να συνεργαστούν με άλλους προγραμματιστές και αναλυτές με ένα κοινό ολοκληρωμένο αναπτυξιακό περιβάλλον (IDE). Το PowerCenter Big Data Edition επιτρέπει την πρόσβαση σε όλους τους τύπους δεδομένων μεγάλων συναλλαγών, συμπεριλαμβανομένων των δεδομένων RDBMS, OLTP, OLAP, ERP, CRM,νέφους και άλλων δεδομένων σχετικά με τα κοινωνικά μέσα, αρχεία καταγραφής, τοποθεσίες Web δεδομένων, αισθητήρων μηχανών, ιστολόγια, έγγραφα, ηλεκτρονικά ταχυδρομεία και άλλα μη δομημένα ή πολυ-δομημένα δεδομένα.

Παρέχει μια εκτεταμένη βιβλιοθήκη προ-διαμορφωμένων δυνατοτήτων μετασχηματισμού στο Hadoop, συμπεριλαμβανομένων μετατροπών τύπου δεδομένων και χειρισμών σειράς, φίλτρων, δρομολογητών, συνόλων και πολλών άλλων.

- **Kognitio Analytical Platform**

Η Kognitio Analytical Platform είναι μια μαζική παράλληλη επεξεργασία (MPP), όχι μόνο SQL, τεχνολογί λογισμικού που είναι βελτιστοποιημένη για φορτία μεγάλου όγκου δεδομένων και μεγάλου όγκου αναλύσεων. Είναι πρωτοπόρος σε αναλυτικές αναλύσεις μεγάλων δεδομένων, υψηλής απόδοσης, για Data Science & Business Intelligence. Χτίζοντας και αναπτύσσοντας τεχνολογία και υπηρεσίες για πάνω από δύο δεκαετίες, η Kognitio συνεχίζει να υπερβαίνει τις προσδοκίες των πελατών της, να πρωτοστατεί στην καινοτομία και να παραδίδει εγκαταστάσεις παραγωγής για την ανάλυση Big Data. Μια διαισθητική,

μαζικά παράλληλη πλατφόρμα in-memomry, που συνδυάζει SQL και NoSQL, διαβάζει τα Big Data απευθείας από υπάρχουσες πλατφόρμες εμμονής, όπως Hadoop, DW και νέφους, ενσωματώνοντας ταυτόχρονα τις επιχειρησιακές εφαρμογές μέσω βιομηχανικών APIs.

Πίνακας 2. Λογισμικά ανάλυσης των Μεγάλων Δεδομένων

Oracle Big Data Analytics	***SAP HANA***	***MapR Platform***	***DataStax Enterprise (DSE)***
Is a comprehensive data management layer enables you to work with all data types and technologies. Seamlessly integrate big data with your existing data, applications and reports.	Is an in-memory computing platform that lets you accelerate business processes, deliver more business intelligent and simplify IT environment.	Delivers enterprise grade security, reliability and real-time performance while dramatically lowering both hardware and operational costs of applications and data.	Accelerate the ability to deliver real-time value at epic scale by providing a comprehensive and operational simple management layer with a unique architecture on Apache Cassandra.
Big Data Platform	BigData Platform	BigData Platform	BigData Platform
Application Development and System Integration SQL Cloud Service Oracle Big Data Discovery Oracle R Advanced Analytics for Hadoop Business Intelligent Cloud Service Data Visualization Cloud Service	Multi-tier Storage Database Services Analytics Processing App Development Data Access	Real-Time High Availability Unified Security Multi-Tenancy Disaster Recover Global Namespace Self-Healing Management & Monitoring	Multi-Model & Graph DataStax OpsCentre DataStax Studio
Proprietary Software	Proprietary Software	Proprietary Software	Proprietary Software
US $75.00/Month	Contact for pricing	Converged Community Edition: Free	Contact for pricing
Subscription	Subscription	Subscription	Subscription
Free Trail Available	Free Trail Available	Free Trail Available	Free Trail Available
Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)	Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)	Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)	Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)
Oracle Big Data Analytics	SAP BigData Analytics	MapR Converged Data Platform	DataStax BigData

IBM Big Data Analytics	***Hortonworks***	***1010data***	***Cloudera Enterprise Big Data***
Solve the challenge with a zone architecture optimized for Big Data. The next generation architecture for big data and analytics delivers new business insights while reducing storage and costs.	Is an in-memory computing platform that lets you accelerate business processes, deliver more business intelligent and simplify IT environment.	Is design to deliver powerful insights on all of data using a single, simplified system. It's democratic, flexible and incredibly fast.	Is the fastest, easiest and most secure modern data platform. From analytics to data science, anyone can get results from any data and across any environment.
Big Data Platform	BigData Platform	BigData Platform	BigData Platform
Data Management & Warehouse Hadoop System Stream Computing Content Management Information Integration & Governance	Data Management Data Access Security Cloudbreak Data Governance & Integration	Application Development Analysis & Modeling Reporting & Visualization Data Sharing & Monetization	Easy to Manage Data Engineering Operational Database Analytic Database Secure without Compromise
Proprietary Software	Proprietary Software	Proprietary Software	Proprietary Software
Contact for pricing	Contact for pricing	Converged Community Edition: Free	Contact for pricing
Subscription	Subscription	Subscription	Subscription
Free Trail Available	Free Trail Available	Free Trail Available	Free Trail Available
Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)	Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)	Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)	Small (<50 employees), Medium (50 to 1000 employees), Enterprise (>1001 employees)
IBM Big Data	Hortonworks Data Platform	MapR Converged Data Platform	Cloudera Enterprise Big Data

Informatica PowerCenter Big Data Edition	***Kognitio Analytical Platform***
Is highly scalable, high performance enterprise data integration software which uses visual development environment to build ELT data flows that run natively on Hadoop.	Is a scale-out-in-memory, massively parallel processing (MPP), software technology that is optimized for low-latency large volume data load and high complex analytical workloads.
Big Data	Big Data Analytics, Big Data Platform
Informatica	Kognitio

Κεφάλαιο 4: Παραδείγματα – Εφαρμογές των Μεγάλων Δεδομένων

Τα Big Data έχουν εισέλθει σαν θύελλα στον κόσμο. Με τα τεράστια ποσά δεδομένων που προέρχονται από διάφορες ψηφιακές πηγές, η σημασία των αναλυτικών στοιχείων έχει αυξηθεί σημαντικά, κάνοντας τις εταιρείες να αξιοποιήσουν τα σκοτεινά δεδομένα που θεωρήθηκαν άχρηστα όλα αυτά τα χρόνια. Δεδομένου ότι οι εταιρείες είναι υποχρεωμένες να παρέχουν αποτελέσματα εν κινήσει, η σημασία των μεγάλων δεδομένων έχει πολλαπλασιαστεί σε όλους τους κλάδους με ταχείς ρυθμούς. Δημιουργείται το ερώτημα σχετικά με την διαφημιστική εκστρατεία για τα μεγάλα δεδομένα. Οι λόγοι για τους οποίους κάθε εταιρεία τάσσεται υπέρ της υιοθέτησης μεγάλων δεδομένων είναι αρκετοί, τους οποίους θα δούμε αναλυτικά στον παρακάτω πίνακα.

Reasons	**Big Data benefits**
Timely	Gain instant insights from diverse data sources
Better analytics	Improvement of business performance through real-time analytics
Vast amount of data	Big data technologies manage huge amounts of data
Insights	Can provide better insights with the help of unstructured and semi-structured data
Decision-making	Helps mitigate risk and make smart decision by proper risk analysis

Η εισαγωγή των Μεγάλων Δεδομέων έχουν ωθήσει πολλές βιομηχανίες στην υιοθέτηση και προώθηση τους. Τέτοιες βιομηχανίες είναι : Υπηρεσίες του δημόσιου τομέα, Υπηρεσίες στην υγειονομική περίθαλψη, Υπηρεσίες για την εκπαίδευση, Ασφαλιστικές υπηρεσίες, Βιομηχανικοί και φυσικοί πόροι, Υπηρεσίες μεταφοράς, Τραπεζικοί τομείς και ανίχνευση απάτης. Παρακάτω θα πραγματοποιήθει εκτενής ανάλυση σχετικά με τους τομείς αυτούς και την σύνδεσή τους με τα Μεγάλα Δεδομένα.

- **Συνεισφορά των Μεγάλων Δεδομένων στον Δημόσιο Τομέα**

Στους δημόσιους τομείς, οι μεγάλες αντιπαραθέσεις είναι η συγχώνευση και η ικανότητα των μεγάλων δεδομένων διαφόρων μονάδων του δημόσιου τομέα και των συνδικάτων. Τα μεγάλα δεδομένα παρέχουν ένα ευρύ φάσμα διευκολύνσεων στους κυβερνητικούς τομείς, συμπεριλαμβανομένης της διερεύνησης της εξουσίας, της αναγνώρισης λαθών, της διασυνδεδεμένης διερεύνησης της φυσικής κατάστασης, της διερεύνησης της οικονομικής προαγωγής και της οικολογικής οχύρωσης.

Τα Μεγάλα δεδομένα χρησιμοποιούνται ακόμη και για να εξεταστούν οι λοιμώξεις από το FDA. Τα μεγάλα αποτελέσματα δεδομένων είναι γρήγορα, τα οποία οδηγούν σε ταχύτερη ευεξία. Επίσης, κατά τη διερεύνηση ενός τεράστιου όγκου κοινοτικών καταγγελιών, χρησιμοποιούνται τα μεγάλα στατιστικά στοιχεία δεδομένων. Αυτά τα ίδια αναλυτικά στοιχεία χρησιμοποιούνται κατά τη διάρκεια των στατιστικών για τον έλεγχο της υγείας επειγόντως και επιμελώς για την ταχύτερη διεκπεραίωση της παραγωγής και για να συνειδητοποιήσουν δυσπιστία ή ψευδείς δηλώσεις.

- **Συνεισφορά των Μεγάλων Δεδομένων στην Υγειονομική Περίθαλψη**

Τα μεγάλα δεδομένα είναι σε εκτεταμένη χρήση στον τομέα της ιατρικής και της υγειονομικής περίθαλψης. Καθώς η τεχνολογία εξελίσεται το κόστος της υγειονομικής περίθαλψης αυξάνεται όλο και περισσότερο. Τα Μεγάλα Δεδομένα είναι ένα μεγάλο χέρι βοήθειας σε αυτό το θέμα. Είναι μεγάλη βοήθεια ακόμη και για τους ιατρούς να παρακολουθούν την ιστορία όλων των ασθενών. Η σύνδεση με το ιστορικό του ασθενούς μπορεί να προσεγγιστεί μόνο από τον ασθενή και τον ειδικό του γιατρό.

Μόλις ένας ασθενής αντιμετωπίσει κάποιο πρόβλημα, το όνομά του και τα δεδομένα του θα αποθηκευτούν στην βάση δεδομένων με ασφάλεια για πάντα και όποτε απαιτείται, ο γιατρός μπορεί να έχει μια άποψη γι 'αυτό. Ένας μεγάλος αριθμός ιατρικών συσκευών είναι εκεί που είναι μεγάλα δεδομένα προσανατολισμένα. Σήμερα τα δεδομένα χρησιμοποιούνται σε τέτοιο βαθμό ώστε ο γιατρός να συνταγογραφεί τα φάρμακα χωρίς να επισκέπτεται τον ασθενή, γνωρίζοντας τον καρδιακό παλμό και τη θερμοκρασία μέσω του ρολογιού παρακολούθησης της καρδιάς και της θερμοκρασίας που τοποθετείται στο χέρι του ασθενούς που παραμένει σε απομακρυσμένο μέρος.

Ένα παράδειγμα είναι τα nanobots είναι μικροσκοπικά ρομπότ που αναπτύσσονται, τα οποία θα αυξήσουν την ανοσία στο ανθρώπινο σώμα καταπολεμώντας τα βακτήρια και άλλα επιβλαβή μικρόβια. Έχουν τους δικούς τους αισθητήρες και θα είναι σπουδαίοι στην παροχή χημειοθεραπείας. Τα nanobots είναι μεγάλα ρομπότ βιοτεχνολογίας που θα χρησιμοποιηθούν για τη μεταφορά οξυγόνου, την καταστροφή μικροβίων και την ανανέωση των ιστών.

- **Συνεισφορά των Μεγάλων Δεδομένων στις Ασφαλιστικές Υπηρεσίες**

Η ανεπάρκεια των τροποποιημένων υπηρεσιών, η απουσία προσαρμοσμένης χρέωσης και η ανάγκη παροχής υπηρεσιών με τεράστιες ποσότητες σε νέα θραύσματα και σε συγκεκριμένα τμήματα της αγοράς είναι μερικές από τις κύριες προκλήσεις. Τα Μεγάλα Δεδομένα είναι το τεχνολογικό εργαλείο που χρησιμοποιείται στην παραγωγή για να προσφέρει γνώση αγοραστών για προϊόντα προβολής και απλούστερων προϊόντων, εντοπίζοντας και προβλέποντας συμπεριφορά αγοραστών από πλευράς πλευράς πληροφορίες που προέρχονται από ιστότοπους διαδικτύου, συμπεριλαμβανομένων των κοινωνικών μέσων ενημέρωσης καθώς και βίντεο εγγραφής.

Τα Μεγάλα Δεδομένα καθιστούν δυνατή την καλύτερη διατήρηση των αγοραστών από τους ασφαλιστικούς οργανισμούς. Στη διεκδίκηση αξιώσεων, χρησιμοποιήθηκαν παρεκκλίνουσες επιχειρηματικές αναλύσεις μεγάλων δεδομένων για την παροχή ταχύτερης εξυπηρέτησης, δεδομένου ότι τεράστιες ποσότητες πληροφοριών μπορούν να επεξεργαστούν ιδιαίτερα κατά την περίοδο υπογραφής. Η ανακάλυψη της απάτης έχει επίσης βελτιωθεί. Κατά τη διάρκεια των γιγαντιαίων δεδομένων από τους ψηφιακούς αγωγούς και τα κοινωνικά μέσα, ο έλεγχος σε πραγματικό χρόνο των χορδών σε όλη τη σειρά των παραμέτρων χρησιμοποιείται για να παρέχει πληροφορίες.

- **Συνεισφορά των Μεγάλων Δεδομένων σε Βιομηχανικούς και Φυσικούς Πόρους**

Η υψηλή ζήτηση των φυσικών πηγών σε στην γη προκαλεί τον υψηλό όγκο καθώς και την ταχύτητα των Μεγάλων Δεδομένων. Παρομοίως, μια μεγάλη ποσότητα δεδομένων που ξεκινάει από την κατασκευασμένη βιομηχανία είναι ανεκμετάλλευτη. Τα αχρησιμοποίητα δεδομένα αποφεύγουν την προηγμένη υπεροχή των εμπορευμάτων, την ικανότητα εξουσίας, την αξιοπιστία και τα βελτιωμένα όρια εισοδήματος. Στη βιομηχανία φυσικού πλούτου, τα μεγάλα δεδομένα καθιστούν δυνατή την αναλυτική μοντελοποίηση για τη διατήρηση της δημιουργίας κρίσεων που χρησιμοποιείται για την κατανάλωση και την ενσωμάτωση τεράστιων ποσοτήτων πληροφοριών από γεωγραφικές πληροφορίες, γραφικές πληροφορίες, χειρόγραφα και χρονολογικά στατιστικά στοιχεία. Τα Μεγάλα Δεδομένα έχουν επίσης φορεθεί στην εξεύρεση λύσης στην ανάπτυξη αντιπαραθέσεων και στην ανάπτυξη επιθετικών βελτιώσεων στη μέση των πρώην οικισμών.

- **Συνεισφορά των Μεγάλων Δεδομένων στις Μεταφορές**

Σήμερα, οι τεράστιες ποσότητες στατιστικών στοιχείων από τα τοπικά δίκτυα που προσανατολίζονται προς την περιοχή και οι στατιστικές ταχύτητας από τις τηλεπικοινωνίες έχουν επηρεάσει πολύ τις πολιτικές ταξιδιού. Δυστυχώς, η διερεύνηση της εκτίμησης της

πολιτικής ταξιδιού δεν έχει αναπτυχθεί ακόμα. Συνήθως, η εκπροσώπηση των μεταφορικών απαιτήσεων προσανατολίζεται και πάλι σε αρχιτεκτονικές φρέσκων κοινωνικών μέσων που δεν έχουν προσδιοριστεί καλά. Ορισμένοι ισχυρισμοί Μεγάλων Δεδομένων από τον δημόσιο τομέα, τις ιδιωτικές ενώσεις και την προσωπική χρήση περιλαμβάνουν:

- Ο ιδιωτικός τομέας χρησιμοποιεί τα Μεγάλα Δεδομένα για τη διαχείριση της κυκλοφορίας, την προετοιμασία της κατεύθυνσης, τις ρυθμίσεις διανοητικής μεταφοράς και τη διοίκηση υπερπληθυσμού
- Ο ιδιωτικός τομέας χρησιμοποιεί τα Μεγάλα Δεδομένα για τη διαχείριση εισοδήματος, τις βιομηχανικές βελτιώσεις, την εφοδιαστική και για εύλογο όφελος
- Η προσωπική χρήση των Μεγάλων Δεδομένων περιλαμβάνει την πρόβλεψη κατεύθυνσης για τη συσσώρευση στο πετρέλαιο και την περίοδο, για δραστηριότητες περιηγήσεων βλέποντας τα αξιοθέατα κλπ.

- **Συνεισφορά των Μεγάλων Δεδομένων σε Τραπεζικές ζώνες και Ανίχνευση απάτης**

Τα Μεγάλα Δεδομένα χρησιμοποιούνται εξαιρετικά στην ανίχνευση απάτης στους τραπεζικούς τομείς. Στους τραπεζικούς τομείς, καθώς υλοποιούνται τα Μεγάλα Δεδομένα, διαπιστώνει όλα τα καθήκοντα που έχουν κάνει κακό. Ανιχνεύει την κακή χρήση των πιστωτικών καρτών, την κατάχρηση των χρεωστικών καρτών, την αρχειοθέτηση των επιθεωρήσεων, την αντιμετώπιση κινδύνων επιχειρηματικών κινδύνων, τη σαφήνεια των επιχειρήσεων, τη μεταβολή των στατιστικών πελατών, τις δημόσιες αναλύσεις για τις επιχειρήσεις, τις αναλύσεις δράσης IT και τις αναλύσεις ολοκλήρωσης της στρατηγικής πληροφορικής. Η SEC χρησιμοποιεί αυτά τα Μεγάλα Δεδομένα για να παρακολουθήσει όλες τις εμπορικές κινήσεις της αγοράς.

Αυτή τη στιγμή χρησιμοποιούν αναλυτές δικτύων και φυσικούς επεξεργαστές ομιλίας για να αντιληφθούν την παράνομη επιχειρηματική δραστηριότητα στις οικονομικές αγορές. Οι έμποροι λιανικών πωλήσεων, οι τράπεζες ιδιωτικών και δημόσιων φορέων, τα αμοιβαία κεφάλαια και άλλοι στη νομισματική αγορά κάνουν χρήση μεγάλων δεδομένων για επιχειρησιακές αναλύσεις που χρησιμοποιούνται στις μεγάλες επιχειρήσεις, διάσταση αντίδρασης, προγνωστικά Analytics κλπ. Στις επιχειρήσεις τα μεγάλα δεδομένα βοηθούν πολύ στη γνώση των αγορών τους πελάτες και τις τακτικές CRM των ανταγωνιστών ώστε να μπορούν να τις εφαρμόσουν στις επιχειρήσεις τους για να βελτιώσουν τις πωλήσεις.

Κεφάλαιο 5: Κρυπτογράφηση - Encryption

5.1. Εισαγωγή στην κρυπτογράφηση

Η κρυπτογράφηση και η διαχείριση κλειδιών θα πρέπει να θεωρηθεί ως ο ακρογωνιαίος λίθος οποιαδήποτε στρατηγική ασφάλειας δεδομένων και ιδαίτερα των δεδομένων μεγάλης κλίμακας. Η κρυπτογράφηση μπορεί να μειώσει δραματικά τους σχετικούς κινδύνους με συμβιβασμούς δεδομένων. Σύμφωνα με τον Ευρωπαϊκό Οργανισμό για την Ασφάλεια Δικτύων και Πληροφοριών (ENISA), υπάρχουν κάποια στοιχεί τα οποία είναι απαραίτητα για την προστασία δεδομένα μεγάλης κλίμακας, κάποια από αυτά είναι τα ακόλουθα: [tankard 2017]

- Η κρυπτογράφηση δεδομένων σε διαμετακόμιση και σε ηρεμία, εξασφαλίζουν την εμπιστευτικότητα και την ακεραιότητα των δεδομένων.
- Πρέπει να αναπτύσσεται μια σωστή λύση διαχείρισης κλειδιού κρυπτογράφησης, λαμβάνοντας υπόψη τις τεράστιες ποσότητες συσκευών που πρέπει να καλύπτονται.
- Εξέταση του χρονικού πλαισίου για το οποίο πρέπει να διατηρούνται τα δεδομένα και η προστασία.
- Σχεδίαση βάσεων δεδομένων με εμπιστευτικότητα.

Όλα τα ευαίσθητα δεδομένα θα πρέπει να είναι κρυπτογραφημένα, συμπεριλαμβανομένων της βάσης δεδομένων, των υπολογιστικών φύλλων, τα έγγραφα λέξεων, τις παρουσιάσεις και τα αρχεία. Πολύ συχνά, τα δεδομένα ενδέχεται να απομακρυνθούν από τον οργανισμό στον οποίο ανήκουν, ίσως να επικοινωνούν μεταξύ υπαλλήλων και επιχειρηματικών συνεργατών ή να τοποθετούνται στο cloud για αποθήκευση. Όταν τα δεδομένα μετακινούνται έξω από έναν οργανισμό, είναι ζωτικής σημασίας τα κλειδιά κρυπτογράφησης να παραμείνουν εντός αυτού για να αποτρέψουν οποιονδήποτε να αποκτήσει ακατάλληλα πρόσβαση στα κλειδιά, κάτι που θα τους επιτρέψει να αποκρυπτογραφήσουν και να διαβάσουν τα δεδομένα. Αν τα κλειδιά δεν προστατεύονται, οι υπάλληλοι του παρόχου υπηρεσιών θα μπορούσαν να έχουν πρόσβαση στα δεδομένα ή θα μπορούσαν να υπόκεινται σε απαιτήσεις κυβερνητικών οργανισμών για την παράδοση των δεδομένων, συχνά χωρίς τη γνώση του οργανισμού που κατέχει τα δεδομένα. Η διασφάλιση ότι τα κλειδιά κρυπτογράφησης δεν αποθηκεύονται με κρυπτογραφημένα δεδομένα θα βοηθήσει επίσης στην αποτροπή της υπονόμευσης των δεδομένων από τους χάκερ.

Τα προγράμματα μεγάλων δεδομένων ωφελούν τους οργανισμούς με πολλούς τρόπους, οδηγώντας την ανταγωνιστικότητα και την καινοτομία. Αλλά μπορούν επίσης να αυξήσουν τους κινδύνους ασφαλείας εξαιτίας της τεράστιας ευαισθησίας πληροφοριών που συχνά περιλαμβάνεται στα τεράστια σύνολα δεδομένων που αναλύονται. Η ασφάλεια δεδομένων είναι απαραίτητη για κάθε οργανισμό για την προστασία της επιχείρησης. Η κρυπτογράφηση πρέπει να αποτελεί βασικό μέρος κάθε μεγάλου περιβάλλοντος δεδομένων, ώστε να διασφαλίζεται ότι οι ευαίσθητες πληροφορίες προστατεύονται επαρκώς.

5.2 Βασικά στοιχεία και λειτουργίες της κρυπτογράφησης

Η κρυπτογράφηση είναι η μέθοδος με την οποία το απλό κείμενο ή οποιοσδήποτε άλλος τύπος δεδομένων μετατρέπεται από μια αναγνώσιμη μορφή σε μια κωδικοποιημένη έκδοση που μπορεί να αποκωδικοποιηθεί από μια άλλη οντότητα μόνον εάν έχει πρόσβαση σε ένα κλειδί αποκρυπτογράφησης. Η κρυπτογράφηση είναι μια από τις πιο σημαντικές μεθόδους για την παροχή ασφάλειας δεδομένων, ειδικά για την προστασία από άκρο σε άκρο των δεδομένων που μεταδίδονται μέσω δικτύων. [rekha]

Τα μη κρυπτογραφημένα δεδομένα, συχνά είναι αναφερόμενα ως απλό κείμενο, κρυπτογραφούνται χρησιμοποιώντας έναν αλγόριθμο κρυπτογράφησης και ένα κλειδί κρυπτογράφησης. Αυτή η διαδικασία δημιουργεί κρυπτογράφημα το οποίο μπορεί να προβληθεί μόνο στην αρχική του μορφή αν αποκρυπτογραφηθεί με το σωστό κλειδί. Η αποκρυπτογράφηση είναι απλά το αντίστροφο της κρυπτογράφησης, ακολουθώντας τα ίδια

βήματα, αλλά αντιστρέφοντας τη σειρά με την οποία εφαρμόζονται τα κλειδιά. Οι σημερινοί πιο ευρέως χρησιμοποιούμενοι αλγόριθμοι κρυπτογράφησης εμπίπτουν σε δύο κατηγορίες: συμμετρικές και ασύμμετρες.

- **Συμμετρική κρυπτογραφία**

Τα συμπιεσμένα κλειδιά κρυπτογράφησης, που αναφέρονται επίσης ως "μυστικό κλειδί", χρησιμοποιούν ένα μόνο κλειδί, το οποίο μερικές φορές αναφέρεται ως κοινόχρηστο μυστικό επειδή το σύστημα που κάνει την κρυπτογράφηση πρέπει να το μοιράζεται με οποιαδήποτε οντότητα που προτίθεται να αποκρυπτογραφήσει τα κρυπτογραφημένα δεδομένα. Ο πιο ευρέως χρησιμοποιούμενος κρυπτογράφος συμμετρικού κλειδιού είναι το Advanced Encryption Standard (AES), το οποίο σχεδιάστηκε για την προστασία των διαβαθμισμένων πληροφοριών της κυβέρνησης.

Η κρυπτογράφηση με συμμετρικό κλειδί είναι συνήθως πολύ πιο γρήγορη από την ασύμμετρη κρυπτογράφηση, αλλά ο αποστολέας πρέπει να ανταλλάξει το κλειδί που χρησιμοποιείται για την κρυπτογράφηση των δεδομένων με τον παραλήπτη πριν ο παραλήπτης μπορεί να εκτελέσει αποκρυπτογράφηση στο κρυπτογράφημα. Η ανάγκη για ασφαλή διανομή και διαχείριση μεγάλου αριθμού κλειδιών σημαίνει ότι οι περισσότερες κρυπτογραφικές διαδικασίες χρησιμοποιούν έναν συμμετρικό αλγόριθμο για την αποτελεσματική κρυπτογράφηση δεδομένων, αλλά χρησιμοποιούν έναν ασύμμετρο αλγόριθμο για την ασφαλή ανταλλαγή του μυστικού κλειδιού.

- **Ασύμμετρη κρυπτογραφία**

Η ασύμμετρη κρυπτογραφία, γνωστή και ως κρυπτογραφία δημόσιου κλειδιού, χρησιμοποιεί δύο διαφορετικά αλλά μαθηματικά συνδεδεμένα κλειδιά, ένα δημόσιο και ένα ιδιωτικό. Το δημόσιο κλειδί μπορεί να μοιραστεί με όλους, ενώ το ιδιωτικό κλειδί πρέπει να παραμείνει μυστικό. Ο αλγόριθμος κρυπτογράφησης RSA είναι ο πιο ευρέως χρησιμοποιούμενος αλγόριθμος δημόσιου κλειδιού, εν μέρει επειδή τόσο το δημόσιο όσο και το ιδιωτικό κλειδί μπορούν να κρυπτογραφήσουν ένα μήνυμα. το αντίθετο κλειδί από εκείνο που χρησιμοποιείται για την κρυπτογράφηση ενός μηνύματος χρησιμοποιείται για την αποκρυπτογράφηση του. Αυτό το χαρακτηριστικό παρέχει μια μέθοδο που εξασφαλίζει όχι μόνο την εμπιστευτικότητα, αλλά και την ακεραιότητα, την αυθεντικότητα και τη μη δυνατότητα χρήσης ηλεκτρονικών επικοινωνιών και δεδομένων σε κατάσταση ηρεμίας μέσω της χρήσης ψηφιακών υπογραφών.

- **Οφέλη κρυπτογράφησης**

Ο πρωταρχικός σκοπός της κρυπτογράφησης είναι να προστατεύσει την εμπιστευτικότητα των ψηφιακών δεδομένων που είναι αποθηκευμένα σε συστήματα υπολογιστών ή να μεταδοθεί μέσω του διαδικτύου ή οποιουδήποτε άλλου δικτύου υπολογιστών. Ορισμένοι οργανισμοί είτε συνιστούν είτε απαιτούν κρυπτογραφημένα ευαίσθητα δεδομένα, προκειμένου να αποτρέπεται η πρόσβαση τρίτων μερών ή φορέων που απειλούν την πρόσβαση στα δεδομένα. Για παράδειγμα, το πρότυπο ασφαλείας δεδομένων βιομηχανικής κάρτας πληρωμών απαιτεί από τους εμπόρους να κρυπτογραφούν τα δεδομένα της κάρτας πληρωμών των πελατών όταν είναι ταυτόχρονα αποθηκευμένα σε κατάσταση ηρεμίας και μεταδίδονται μέσω δημόσιων δικτύων.

Οι σύγχρονοι αλγόριθμοι κρυπτογράφησης διαδραματίζουν επίσης ζωτικό ρόλο στη διασφάλιση της ασφάλειας των συστημάτων πληροφορικής και των επικοινωνιών, καθώς μπορούν να παρέχουν όχι μόνο την εμπιστευτικότητα αλλά και τα ακόλουθα βασικά στοιχεία ασφάλειας:

- Έλεγχος ταυτότητας: η προέλευση ενός μηνύματος μπορεί να επαληθευτεί.
- Ακεραιότητα: Απόδειξη ότι τα περιεχόμενα ενός μηνύματος δεν έχουν αλλάξει από την αποστολή του.

- Μη αναδημοσίευση: ο αποστολέας ενός μηνύματος δεν μπορεί να αρνηθεί την αποστολή του μηνύματος.

- **Τύποι κρυπτογράφησης**

Η παραδοσιακή κρυπτογραφία δημόσιου κλειδιού εξαρτάται από τις ιδιότητες των μεγάλων πρωτευόντων αριθμών και από την υπολογιστική δυσκολία του factoring αυτών των prime. Η κρυπτογράφηση ελλειπτικής καμπύλης (ECC) επιτρέπει ένα άλλο είδος κρυπτογραφίας δημόσιου κλειδιού που εξαρτάται από τις ιδιότητες της εξίσωσης ελλειπτικής καμπύλης. Οι προκύπτοντες αλγόριθμοι κρυπτογράφησης μπορούν να γίνουν ταχύτεροι και αποτελεσματικότεροι και μπορούν να παράγουν συγκρίσιμα επίπεδα ασφαλείας με μικρότερα κρυπτογραφικά κλειδιά. Ως αποτέλεσμα, οι αλγόριθμοι ECC εφαρμόζονται συχνά σε διαδικτυακές συσκευές συσκευών και σε άλλα προϊόντα με περιορισμένους υπολογιστικούς πόρους.

Καθώς η ανάπτυξη των κβαντικών υπολογιστών συνεχίζει να προσεγγίζει την πρακτική εφαρμογή, η κβαντική κρυπτογραφία θα γίνει πιο σημαντική. Η κβαντική κρυπτογραφία εξαρτάται από τις κβαντικές μηχανικές ιδιότητες των σωματιδίων για την προστασία των δεδομένων. Συγκεκριμένα, η αρχή της αβεβαιότητας του Heisenberg υποδηλώνει ότι οι δύο ιδιότητες ταυτοποίησης ενός σωματιδίου - η θέση του και η ορμή του - δεν μπορούν να μετρηθούν χωρίς να αλλάξουν οι τιμές αυτών των ιδιοτήτων. Ως αποτέλεσμα, τα κβαντικά κωδικοποιημένα δεδομένα δεν μπορούν να αντιγραφούν επειδή κάθε προσπάθεια πρόσβασης στα κωδικοποιημένα δεδομένα θα αλλάξει τα δεδομένα. Ομοίως, οποιαδήποτε απόπειρα αντιγραφής ή πρόσβασης στα δεδομένα θα προκαλέσει αλλαγή στα δεδομένα, ειδοποιώντας έτσι τα εξουσιοδοτημένα μέρη της κρυπτογράφησης ότι έχει σημειωθεί επίθεση.

Η κρυπτογράφηση χρησιμοποιείται για την προστασία δεδομένων που είναι αποθηκευμένα σε ένα σύστημα (κρυπτογράφηση στη θέση ή κρυπτογράφηση σε ηρεμία). Πολλά πρωτόκολλα διαδικτύου ορίζουν μηχανισμούς κρυπτογράφησης δεδομένων που μετακινούνται από ένα σύστημα σε άλλο (δεδομένα υπό διαμετακόμιση).

Ορισμένες εφαρμογές που αφορούν τη χρήση κρυπτογράφησης από άκρο σε άκρο (E2EE) για να διασφαλιστεί ότι τα δεδομένα αποστέλλονται μεταξύ δύο μερών δεν μπορούν να προβληθούν από έναν εισβολέα που παρεμποδίζει το κανάλι επικοινωνίας. Η χρήση ενός κρυπτογραφημένου κυκλώματος επικοινωνίας, όπως παρέχεται από το Transport Layer Security (TLS) μεταξύ λογισμικού web client και web server, δεν είναι πάντα αρκετό για να ασφαλιστεί το E2EE. Τυπικά, το πραγματικό περιεχόμενο που μεταδίδεται κρυπτογραφείται από το λογισμικό πελάτη πριν μεταφερθεί σε ένα web client και αποκρυπτογραφείται μόνο από τον παραλήπτη.

- **Κρυπτογραφικές λειτουργίες κατακερματισμού**

Η κρυπτογράφηση είναι συνήθως μια αμφίδρομη λειτουργία, που σημαίνει ότι ο ίδιος αλγόριθμος μπορεί να χρησιμοποιηθεί για την κρυπτογράφηση του απλού κειμένου και για την αποκρυπτογράφηση του κρυπτογραφικού κειμένου. Μια κρυπτογραφική λειτουργία κατακερματισμού μπορεί να θεωρηθεί ως ένας τύπος μονόδρομης λειτουργίας κρυπτογράφησης, που σημαίνει ότι η έξοδος της λειτουργίας δεν μπορεί εύκολα να αντιστραφεί για να ανακτήσει την αρχική είσοδο. Οι λειτουργίες Hash χρησιμοποιούνται συνήθως σε πολλές πτυχές της ασφάλειας για τη δημιουργία ψηφιακών υπογραφών και ελέγχων ακεραιότητας δεδομένων. Παίρνουν ένα ηλεκτρονικό αρχείο, ένα μήνυμα ή ένα μπλοκ δεδομένων και παράγουν ένα σύντομο ψηφιακό αποτύπωμα του περιεχομένου που ονομάζεται digest ή hash. Οι βασικές ιδιότητες μιας ασφαλούς λειτουργίας κρυπτογραφικής κατακερματισμού είναι:

- Το μήκος εξόδου είναι μικρό σε σύγκριση με την είσοδο
- Ο υπολογισμός είναι γρήγορος και αποτελεσματικός για κάθε είσοδο
- Οποιαδήποτε αλλαγή στην είσοδο επηρεάζει πολλά bits εξόδου
- Τιμή μονής κατεύθυνσης - η είσοδος δεν μπορεί να προσδιοριστεί από την έξοδο

- Ισχυρή αντίσταση σύγκρουσης - δύο διαφορετικές εισόδους δεν μπορούν να δημιουργήσουν την ίδια έξοδο

Οι ψηφιακές κρυφές λειτουργίες έχουν βελτιστοποιηθεί για το hashing: Χρησιμοποιούν μεγάλα πλήκτρα και μπλοκ, μπορούν να αλλάξουν αποτελεσματικά τα πλήκτρα σε κάθε μπλοκ και έχουν σχεδιαστεί και ελεγχθεί για αντοχή σε επιθέσεις σχετιζόμενου κλειδιού. Τα κρυπτογραφικά στοιχεία γενικού σκοπού που χρησιμοποιούνται για κρυπτογράφηση τείνουν να έχουν διαφορετικούς σχεδιαστικούς στόχους. Για παράδειγμα, ο συμπιεσμένος κρυπτογραφικός κατάλογος συμμετρικού κλειδιού AES θα μπορούσε επίσης να χρησιμοποιηθεί για τη δημιουργία τιμών κατακερματισμού, αλλά τα μεγέθη του κλειδιού και του μπλοκ τον καθιστούν ως αναποτελεσματικό.

Κεφάλαιο 6 : Σκοπός της έρευνας

Η νέα εποχή των Δεδομένων Μεγάλης Κλίμακας έχει κάνει την είσοδο της πράγμα που αντιλαμβανόμαστε ολοένα και περισσότερο. Ο μεγάλος όγκος των δεδομένων αποτελεί κύριο στοιχείο που όμως εγγυμωνεί πολλούς κινδύνους και προκλήσεις όπως αναλύσαμε σε προηγούμενο κεφάλαιο. Για την ορθή αντιμετώπιση αυτών των κινδύνων επικεντρωθήκαμε στην ασφάλεια των μεγάλων δεδομένων και σε τρόπους όπου μπορούν να διασφαλίσουν τα δεδομένα από επιθέσεις. Ένας τέτοιος τρόπος ώστε να υπάρχει προστασία και ασφάλεια των δεδομένων είναι η κρυπτογράφησή τους ή μέρος αυτών ώστε να μην είναι εκτεθυμένα σε επιθέσεις και κακόβουλες ενέργειες.

Έχει γίνει εκτενής περιγραφή σχετικά με την έννοια της κρυπτογράφησης και τα χαρακτηριστηκά της ώστε να έχουμε τα επιθυμητά αποτελέσματα. Το πρόβλημα το οποίο θα προσπαθήσουμε να θεραπεύσουμε αφορά την ασφάλεια των μεγάλων δεδομένων με την χρήση της κρυπτογράφησης. Η κρυπτογράφηση δεδομένων μεταφράζει δεδομένα σε μια άλλη μορφή ή κώδικα, έτσι ώστε μόνο τα άτομα που έχουν πρόσβαση σε ένα μυστικό κλειδί ή ο κωδικό πρόσβασης μπορούν να τα διαβάσουν. Επί του παρόντος, η κρυπτογράφηση είναι μία από τις πιο δημοφιλείς και αποτελεσματικές μεθόδους ασφάλειας δεδομένων που χρησιμοποιούν οι οργανισμοί.

Οι λύσεις προστασίας δεδομένων για την κρυπτογράφηση μπορούν να παρέχουν κρυπτογράφηση συσκευών, email και δεδομένων. Οι εταιρείες και οι οργανισμοί αντιμετωπίζουν την πρόκληση της προστασίας δεδομένων και την πρόληψη της απώλειας τους καθώς οι εργαζόμενοι χρησιμοποιούν πιο συχνά εξωτερικές συσκευές, αφαιρούμενα μέσα και εφαρμογές ιστού ως μέρος των καθημερινών επιχειρησιακών τους διαδικασιών. Τα ευαίσθητα δεδομένα ιδαίτερα όταν το μέγεθος τους είναι μεγάλο ενδέχεται να μην είναι πλέον υπό τον έλεγχο και προστασία της εταιρείας, καθώς οι εργαζόμενοι αντιγράφουν δεδομένα σε αφαιρούμενες συσκευές ή φορτώνονται στο νέφος. Ως αποτέλεσμα, οι καλύτερες λύσεις πρόληψης απώλειας δεδομένων αποτρέπουν την κλοπή δεδομένων και την εισαγωγή κακόβουλου λογισμικού. Για να γίνει αυτό, πρέπει επίσης να διασφαλίσουν ότι οι συσκευές και οι εφαρμογές χρησιμοποιούνται σωστά και ότι τα δεδομένα είναι εξασφαλισμένα με αυτόματη κρυπτογράφηση ακόμη και μετά την αναχώρησή τους από την επιχείρηση.

Η κρυπτογράφηση δεδομένων μπορεί να φαίνεται σαν μια αποθαρρυντική, περίπλοκη διαδικασία, υπάρχουν όμως λογισμικά πρόληψης απώλειας δεδομένων και τεχνικές, ειδικά όταν μιλάμε για μεγάλα δεδομένα που χειρίζονται αξιόπιστα σε καθημερινή βάση τα δεδομένα εταιρείων ιδιαίτερα όταν είναι ευαίσθητα, μπορούν να δώσουν λύσεις και ολοκληρωμένα αποτελέσματα ώστε να επιτευχθεί ο στόχος της ασφάλειάς τους.

Κεφάλαιο 7 : Σύγκριση και Ανάλυση

7.1 Κοινή χρήση ευαίσθητων δεδομένων σε πλατφόρμα μεγάλων δεδομένων.

7. 1.1 Πλαίσιο ασφαλής κοινής χρήσης ευαίσθητων δεδομένων

Ο όρος των μεγάλων δεδομένων αναφέρεται σε πολύ μεγάλες ποσότητες δεδομένων που είναι αποθηκευμένες είτε από επιχειρήσεις είτε από διάφορους οργανισμούς. Ένα τέτοιο μέσο αποθήκευσης είναι και οι μεγάλες πλατφόρμες δεδομένων όπου αποθηκεύονται πληροφορίες, που αποτελούν δεδομένα ανθρώπων τα οποία μπορεί να είναι ευαίσθητα και θα πρέπει να εξασφαλιστεί η ιδιωτικότητα τους ώστε να μην έρθουν σε κοινή χρήση. Η οικοδόμηση ασφαλών καναλιών για έναν πλήρως ευαίσθητο κύκλο ζωής δεδομένων απαιτεί την εξέταση τεσσάρων πτυχών των προβλημάτων ασφάλειας: αξιόπιστη υποβολή, ασφαλής αποθήκευση, χρήση χωρίς κίνδυνο και ασφαλή καταστροφή. Ένας πάροχος υπηρεσιών μιας πλατφόρμας νέφους που χρησιμοποιεί δεδομένα πρέπει διασφαλίζει την ασφάλεια των δεδομένων με τη λήψη και τη χρήση της προσθήκης ασφαλείας.

Το νέφος δεν μπορεί να εμπιστευτεί ότι το αποκρυπτογραφημένο κείμενο θα διαρρεύσει τις προσωπικές πληροφορίες των χρηστών οπότε υιοθετείται η τεχνολογία προστασίας της διαδικασίας που βασίζεται σε ένα VMM (Virtual Machine Monitor), παρακάμπτοντας το λειτουργικό σύστημα και παρέχοντας προστασία δεδομένων απευθείας στη διαδικασία του χρήστη. Η μονάδα διαχείρισης κλειδιών του VMM χρησιμοποιείται για την αποθήκευση δημόσιων κλειδιών, έτσι όταν εκτελείται ένα πρόγραμμα, το συμμετρικό κλειδί στο κάτω μέρος του κύριου προγράμματος θα αποκρυπτογραφείται δυναμικά από τη μονάδα διαχείρισης κλειδιών.

Η βασική ροή του πλαισίου για την παροχή ασφάλειας έχει ως εξής. Πρώτον, επιχειρήσεις που έχουν ευαίσθητες πληροφορίες έχουν προκαθορισμένους παρόχους υπηρεσιών με τους οποίους πρέπει να μοιραστούν αυτές τις ευαίσθητες πληροφορίες και στη συνέχεια να υποβάλουν και να αποθηκεύσουν τα αντίστοιχα κρυπτογραφημένα δεδομένα σε μια πλατφόρμα μεγάλων δεδομένων. Δεύτερον, πρέπει στα δεδομένα να γίνει η χρήση του PRE (Proxy re-encryption) αλγόριθμου. Στη συνέχεια, οι πάροχοι υπηρεσιών της πλατφόρμας cloud που θέλουν να μοιραστούν τις ευαίσθητες πληροφορίες κατεβάζουν και αποκρυπτογραφούν τα αντίστοιχα δεδομένα χρησιμοποιώντας την ασφαλή προσθήκη με ευαίσθητα δεδομένα απορρήτου που εκτελούνται σε αυτό το διάστημα. Ολοκληρώνοντας, γίνετε χρήση ενός ασφαλούς μηχανισμού για την καταστροφή των δεδομένων που εξακολουθούν να αποθηκεύονται προσωρινά στο νέφος. Εν ολίγοις, το πλαίσιο προστατεύει αποτελεσματικά την ασφάλεια ολόκληρου του κύκλου ζωής των ευαίσθητων δεδομένων ενώ παράλληλα οι κάτοχοι τους έχουν πλήρη έλεγχο των δικών τους δεδομένων.

7.1.2 Heterogeneous Proxy Re-Encryption (H-PRE)

Μια κοινή και δημοφιλής μέθοδος για την εξασφάλιση της ασφάλειας της υποβολής δεδομένων σε μια ημι-αξιόπιστη πλατφόρμα δεδομένων είναι η κρυπτογράφηση των δεδομένων πριν από την υποβολή δεδομένων στην πλατφόρμα. Ορισμένες λειτουργίες παρέχονται χρησιμοποιώντας μια προσθήκη ασφαλείας. Για τη διασφάλιση της ασφαλούς αποθήκευσης, σχεδιάστηκε η Ετερογενή Επαναπρογράμμιση Proxy (H-PRE), η οποία υποστηρίζει ετερογενή μετασχηματισμό από Κρυπτογράφηση με βάση την ταυτότητα (IBE) σε κρυπτογράφηση δημόσιου κλειδιού (PKE). Το H-PRE είναι συμβατό με την παραδοσιακή κρυπτογραφία. Η προσθήκη είναι να μετασχηματιστούν τα δεδομένα κρυπτογράφησης που μεταφορτώνει ο ιδιοκτήτης σε κρυπτογραφημένο κείμενο, το οποίο ο χρήστης δεδομένων μπορεί να αποκρυπτογραφήσει χρησιμοποιώντας το δικό του ιδιωτικό κλειδί.

Το H-PRE περιλαμβάνει τρεις τύπους αλγορίθμων, την παραδοσιακή κρυπτογράφηση με βάση την ταυτότητα (συμπεριλαμβανομένων των SetupIBE, KeyGenIBE, EncIBE και DecIBE), την επανα-κρυπτογράφηση (συμπεριλαμβανομένων των λειτουργιών KeyGenRE, ReEnc και ReDec), ενώ το τελευταίο είναι τα παραδοσιακά cryptosystems

δημόσιου κλειδιού (συμπεριλαμβανομένων των KeyGenPKE, EncPKE και DecPKE). Η βασική διαδικασία H-PRE είναι απλή. Ο κάτοχος δεδομένων κρυπτογραφεί ευαίσθητα δεδομένα χρησιμοποιώντας μια τοπική προσθήκη ασφαλείας και στη συνέχεια μεταφορτώνει τα κρυπτογραφημένα δεδομένα σε μια μεγάλη πλατφόρμα δεδομένων. Τα δεδομένα μετατρέπονται στο κρυπτογραφημένο κείμενο που μπορεί να αποκρυπτογραφηθεί από έναν συγκεκριμένο χρήστη μετά τις υπηρεσίες PRE. Εάν ένας SESP είναι ο συγκεκριμένος χρήστης, τότε το SESP μπορεί να αποκρυπτογραφήσει τα δεδομένα χρησιμοποιώντας το δικό του ιδιωτικό κλειδί για να αποκτήσει το αντίστοιχο καθαρό κείμενο.

7.1.3 Λειτουργίες υποβολής, αποθήκευσης και εξαγωγής δεδομένων

Ο κάτοχος δεδομένων κρυπτογραφεί τα δεδομένα τοπικά, χρησιμοποιώντας πρώτα τον συμμετρικό αλγόριθμο κρυπτογράφησης Standard Advanced Encryption Standard (AES) για την κρυπτογράφηση των δεδομένων υποβολής και στη συνέχεια, χρησιμοποιώντας τον αλγόριθμο PRE για την κρυπτογράφηση του συμμετρικού κλειδιού των δεδομένων. Αυτά τα αποτελέσματα αποθηκεύονται όλα στα κατανεμημένα δεδομένα. Εν τω μεταξύ, αν ο κάτοχος δεδομένων μοιράζεται τα ευαίσθητα δεδομένα με άλλους χρήστες, ο κάτοχος δεδομένων πρέπει να εξουσιοδοτήσει τα ευαίσθητα δεδομένα τοπικά και να δημιουργήσει το κλειδί PRE, το οποίο αποθηκεύεται στον διακομιστή κλειδιών εξουσιοδότησης. Στη πλατφόρμα μεγάλων δεδομένων, ο διακομιστής PRE επανακρυπτογραφεί και μετατρέπει τον αρχικό κρυπτογραφητή χρησιμοποιώντας το πλήκτρο PRE. Στη συνέχεια, παράγεται το κρυπτογράφημα PRE, το οποίο μπορεί να κρυπτογραφηθεί από τους (εξουσιοδοτημένους) χρήστες δεδομένων. Εάν ο χρήστης δεδομένων θέλει να χρησιμοποιήσει τα δεδομένα στην πλατφόρμα μεγάλων δεδομένων, ο χρήστης δεδομένων θα στείλει αιτήματα δεδομένων στην πλατφόρμα και στη συνέχεια ερώτημα εάν υπάρχουν αντίστοιχα δεδομένα στον κοινόχρηστο χώρο. Εάν υπάρχουν τέτοια δεδομένα, ο χρήστης δεδομένων έχει πρόσβαση και κάνει λήψη του. Η λειτουργία στη μεγάλη πλατφόρμα δεδομένων είναι ανεξάρτητη και διαφανής για τους χρήστες. Επιπλέον, οι υπολογιστικοί πόροι της μεγάλης πλατφόρμας δεδομένων είναι πιο ισχυροί από αυτούς του πελάτη. Ως εκ τούτου, μπορούμε να βάλουμε PRE υπολογιστικά γενικά στη μεγάλη πλατφόρμα δεδομένων για να βελτιώσουμε την εμπειρία των χρηστών. Το σύστημα PRE περιλαμβάνει την υποβολή δεδομένων, την αποθήκευση (κοινή χρήση) και την εξαγωγή δεδομένων.

7.1.4 Ασφαλής χρήση ευαίσθητων δεδομένων στο VMM

Για την διασφαλίση της ασφαλής εκτέλεσης μιας εφαρμογής στο νέφος, γίνεται χρήση του ιδιωτικού χώρου μιας διαδικασίας χρήστη που βασίζεται σε ένα VMM, ώστε να γίνει εξαγωγή ευαίσθητων προσωπικών δεδομένών στην πλατφόρμα μεγάλων δεδομένων. Μια ευαίσθητη διαδικασία πρέπει να αποτρέπει τις απειλές από ένα VMM διαχείρισης και από ένα μη αξιόπιστο επίπεδο λειτουργικού συστήματος κάτω από αυτό. Το εξουσιοδοτημένο υλικό κάτω χρησιμοποιεί τη λειτουργία TPM, διασφαλίζοντας ότι το VMM είναι αξιόπιστο. Με το κρυπτογράφημα PRE που υπολογίζεται σε μια πλατφόρμα μεγάλων δεδομένων, ο ιδιωτικός χώρος μνήμης των διαδικασιών στην πλατφόρμα του νέφους μπορεί να εγγυηθεί την ασφάλεια των δεδομένων στη μνήμη και στο σκληρό δίσκο. Το VMM παρέχει ιδιωτικό χώρο μνήμης, δηλαδή η διαδικασία εκτελείται σε ιδιωτικό χώρο μνήμης του οποίου η μνήμη δεν είναι προσβάσιμη από το λειτουργικό σύστημα ή άλλες εφαρμογές. Η μέθοδος απομόνωσης μνήμης διασφαλίζει την ιδιωτικότητα και την ασφάλεια δεδομένων στη μνήμη. Επιπλέον, τα δεδομένα που χρησιμοποιούνται και αποθηκεύονται στο δίσκο είναι κρυπτοκείμενο. Το VMM αποκρυπτογραφεί ή κρυπτογραφεί κατά την ανάγνωση ή την εγγραφή δεδομένων, αντίστοιχα. Ως αποτέλεσμα, ένας συνδυασμός αυτών των δύο μέτρων μπορεί να προστατευθεί χρησιμοποιώντας το VMM, είτε το πρόγραμμα χρήστη εκτελείται στη μνήμη είτε αποθηκεύεται στο δίσκο.

Χρησιμοποιούμε την τεχνολογία προστασίας της διαδικασίας που βασίζεται σε ένα VMM, προσφέροντας απευθείας προστασία δεδομένων τη διαδικασία χρήστη. Για την

προστασία της ασφάλειας των δεδομένων κατά τη διαδικασία αλληλεπίδρασης στην πλατφόρμα του cloud, πρέπει να ολοκληρωθούν τα παρακάτω βήματα.

- Δημιουργία ενός αξιόπιστου περιβάλλοντος και διαύλων

Κατά τη διάρκεια της διαδικασίας εκκίνησης, η πλατφόρμα του νέφους πρέπει να μετρήσει το λογισμικό εκκίνησης μέσω αξιόπιστης τεχνολογίας υπολογιστών. Επομένως, οι χρήστες του cloud πρέπει να διασφαλίζουν την ακεραιότητα του VMM, δηλαδή οι χρήστες πρέπει να διασφαλίζουν ότι το VMM είναι έμπιστος. Μετά τη διαδικασία εκκίνησης, ο διακομιστής σύννεφων θα αποστέλλει απομακρυσμένη επαλήθευση στον χρήστη για να εξασφαλίσει τη σχέση εμπιστοσύνης μεταξύ τους. Στην πραγματικότητα, το VMM αποκρίνεται στο αίτημα στο τέλος του διακομιστή cloud. Κατ 'αρχάς, το SESP στέλνει ένα αίτημα ακεραιότητας στο διακομιστή σύννεφο, δεύτερον, το VMM παράγει ένα κλειδί συνεδρίας (Ksess). Μετά από μια σύνθετη διεργασία επικοινωνίας μεταξύ αυτών και υπολογισμό κάποιον στοιχείων, εάν οι τιμές είναι συνεπείς, οι επικοινωνίες είναι ασφαλείς. Ως αποτέλεσμα, και οι δύο πλευρές των επικοινωνιών καθορίζουν ένα κλειδί συνόδου. Στη συνέχεια, και οι δύο πλευρές της επικοινωνίας θα κρυπτογραφηθούν χρησιμοποιώντας το κλειδί συνόδου.

- Μεταφόρτωση και εξαγωγή δεδομένων

Οι χρήστες του νέφους (SESP) εξαγάγουν τα ευαίσθητα δεδομένα από τη μεγάλη πλατφόρμα δεδομένων μέσω της ανάκτησης, ενώ αρχικά υποθέτουμε ότι το νέφος δεν είναι αξιόπιστο. Η μεταφορτωμένη εκτελέσιμη εφαρμογή και τα δεδομένα πρέπει να κρυπτογραφηθούν πριν το SESP χρησιμοποιήσει το νέφος. Τα δεδομένα που λαμβάνονται από τη πλατφόρμα μεγάλων δεδομένων είναι το PRE ciphertext, το οποίο μπορεί να αποκρυπτογραφηθεί κατά τη διάρκεια του χρόνου εκτέλεσης. Τέλος, κρυπτογραφημένα εκτελέσιμα αρχεία και αρχεία δεδομένων μεταφορτώνονται στον εξυπηρετητή του νέφους.

- Εκτέλεση προγράμματος

Στη διαδικασία εκτέλεσης της εφαρμογής στην πλατφόρμα του νέφους, η δυναμική προστασία δεδομένων και η κρυπτογράφηση είναι παρόμοια με την διαδικασία προστασίας του χώρου μνήμης. Το VMM λειτουργεί ως η γέφυρα ανταλλαγής δεδομένων μεταξύ του λειτουργικού συστήματος και της διαδικασίας χρήστη. Όταν το λειτουργικό σύστημα αντιγράφει τα δεδομένα από το χώρο μνήμης, το VMM εκτελεί τη λειτουργία αντιγραφής, επειδή το λειτουργικό σύστημα δεν διαθέτει δικαιώματα ανάγνωσης και εγγραφής. Όταν τα δεδομένα αντιγράφονται στο ιδιωτικό χώρο μνήμης της διαδικασίας, το VMM αποκρυπτογραφεί τα δεδομένα χρησιμοποιώντας το αντίστοιχο συμμετρικό κλειδί AES. Αντίθετα, όταν τα δεδομένα στον ιδιωτικό χώρο μνήμης της διαδικασίας αντιγράφονται προς τα έξω, το VMM κρυπτογραφεί τα δεδομένα χρησιμοποιώντας το αντίστοιχο συμμετρικό κλειδί AES. Επομένως, τα δεδομένα χρήστη που είναι αποθηκευμένα στο δίσκο είναι σε μορφή κρυπτοκειμένου. Ως εκ τούτου, ο ιδιωτικός χώρος της διαδικασίας χρήστη ενεργεί ως σημείο ισορροπίας του μηχανισμού ασφαλείας μεταξύ του ιδιοκτήτη δεδομένων και του χρήστη, ωφελώντας ταυτόχρονα την αποφυγή της διαρροής ευαίσθητων πληροφοριών.

7.2 Πολύ-κρυπτογράφηση δεδομένων

7.2.1 Εισαγωγή στην Πολύ-κρυπτογράφηση

Η εξέλιξη της κρυπτογράφησης μπορεί να παρέχει καλύτερη ασφάλεια απ 'ότι μια ενιαία ρουτίνα κρυπτογράφησης. Σε αυτό το πλαίσιο, η πολυ-κρυπτογράφηση ήρθε στο φως με πολύ σημαντικά αποτελέσματα σε επίπεδο ασφάλειας και ταχύτητα επεξεργασίας. Η πολλαπλή κρυπτογράφηση είναι η διαδικασία κρυπτογράφησης ενός ήδη κρυπτογραφημένου μηνύματος μία ή περισσότερες φορές, είτε χρησιμοποιώντας τον ίδιο είτε διαφορετικό αλγόριθμο. Παρουσιάζεται η πολυ-κρυπτογράφηση με τις προσεγγίσεις των RC6 και XTEA.

Στην κρυπτογραφία, το RC6 (Rivest Cipher 6) είναι ένας συμμετρικός κρυπτογραφικός αποκλεισμός κλειδιού που προέρχεται από το RC5. Σχεδιάστηκε από τον Ron Rivest, τον Matt Robshaw, τον Ray Sidney και τον Yiqun Lisa Yin για να ανταποκριθεί στις απαιτήσεις του διαγωνισμού Advanced Encryption Standard (AES). Ήταν ένας

ιδιόκτητος αλγόριθμος, κατοχυρωμένος με δίπλωμα ευρεσιτεχνίας από την RSA Security. Το RC6 έχει μέγεθος μπλοκ 128 bit και υποστηρίζει μεγέθη κλειδιών 128, 192 και 256 bits έως και 2040 bits, αλλά, όπως το RC5, μπορεί να παραμετροποιηθεί για να υποστηρίξει μια μεγάλη ποικιλία λέξεων-μήκους, μεγέθους κλειδιού και αριθμός γύρων. Το RC6 είναι πολύ παρόμοιο με το RC5 σε δομή, χρησιμοποιώντας εξαρτώμενες από δεδομένα περιστροφές, αρθρωτή προσθήκη και λειτουργίες XOR. Στην πραγματικότητα, το RC6 μπορεί να θεωρηθεί ότι συνενώνει δύο παράλληλες διαδικασίες κρυπτογράφησης RC5, αν και η RC6 χρησιμοποιεί μια επιπλέον λειτουργία πολλαπλασιασμού που δεν υπάρχει στο RC5, προκειμένου να γίνει η περιστροφή εξαρτώμενη από κάθε bit σε μια λέξη και όχι μόνο τα λιγότερο σημαντικά κομμάτια.

Το XTEA (eXtended TEA) είναι ένα κρυπτογράφημα σχεδιασμένο για να διορθώνει τις αδυναμίες του TEA. Οι σχεδιαστές του κρυπτογράφου ήταν ο David Wheeler και ο Roger Needham του Cambridge Computer Laboratory και ο αλγόριθμος παρουσιάστηκε σε μια μη δημοσιευμένη τεχνική έκθεση το 1997. Όπως και ο TEA, το XTEA είναι ένας κρυπτογράφος Feistel σε μπλοκ 64 bit με κλειδί 128 bit και προτεινόμενο 64 γύρους. Πολλές διαφορές από το TEA είναι εμφανείς, συμπεριλαμβανομένου ενός κάπως πιο περίπλοκου κλειδιού και μιας αναδιάταξης των μετατοπίσεων, των XOR και των προσθηκών.

7.2.2 Παρουσίαση των δυο προσεγγίσεων

Η πολλαπλή κρυπτογράφηση αυξάνει την ασφάλεια του αλγορίθμου εφαρμόζοντάς την επανειλημμένα. Οι κρυπτογράφοι σε μπλοκ καταναλώνουν μικρή ισχύ, αλλά έχουν ασφαλεία και τρέχουν αρκετές φορές σε σειρά, επιτυγχάνοντας έτσι έναν ασφαλέστερο συνολικό αποτέλεσμα. Οι εφαρμοστές πρέπει να είναι σε θέση να κλιμακώσουν τον αλγόριθμο από μια bit-serial υλοποίηση σε μια εξαιρετικά παράλληλη υλοποίηση, ανάλογα με την επιθυμητή μέγιστη κατανάλωση ισχύος και ταχύτητα. Παρουσιάζονται δυο προσεγγίσεις: (α) Πολλαπλή κρυπτογράφηση με XTEA μετά από RC6, (β) Πολλαπλή κρυπτογράφηση με RC6 μετά από XTEA.

- **Πολλαπλή κρυπτογράφηση με XTEA μετά από RC6**

Η παρακάτω διαδικασία παρέχει όλες τις λεπτομέρειες για τη μετατροπή του plaintext σε ciphertext χρησιμοποιώντας τους αλγορίθμους RC6 και XTEA. Το XTEA εφαρμόζεται με το κρυπτογράφημα, το οποίο παράγεται μετά την εφαρμογή του αλγορίθμου RC6. Η ακόλουθη διαδικασία περιγράφεται παρακάτω :
-Το απλό κείμενο χωρίζεται μπλόκ των 128-bit.
-Επιλογή του μπλοκ των 128- bit ένα προς ένα και εφαρμογή του RC6 αλγόριθμου.
-Διαχωρισμός του κρυπτογράφηματος των 128- bit σε δύο μπλοκ των 64-bit.
-Εφαρμογή του αλγόριθμου XTEA ξεχωριστά σε κάθε μπλόκ των 64-bit.
-Συνδυασμός των μπλοκ κρυπτογράφησης των 64-bit σε 128-bit.
-Επανάληψη των βημάτων δυο έως πέντε μέχρι το τέλος του απλού κειμένου.

- **Πολλαπλή κρυπτογράφηση με RC6 μετά από XTEA**

Σε αυτή τη μέθοδο ο αλγόριθμος XTEA χρησιμοποιείται πρώτα. Τα βήματα για την μετατροπή του απλού κειμενου σε μορφή κρυπτογραφήματος εξηγούνται παρακάτω:
-Το απλό κείμενο χωρίζεται μπλόκ των 64-bit.
-Επιλογή δύο μπλοκ των 64-bit και εφαρμογή του αλγόριθμου XTEA χωριστά.
-Συνδυασμός των δύο κρυπτογραφήματων των 64-bit σε ενιαίο 128-bit μπλοκ.
-Εφαρμογή του αλγόριθμου RC6 με το κείμενο εισόδου ως κρυπτογράφημα του δεύτερου βήματος .
-Το αποτέλεσμα των 128-bit είναι το κρυπτογράφημα του απλού κειμένου.
-Επανάληψη των βημάτων δυο έως πέντε μέχρι το τέλος του απλού κειμένου

7.2.3 Ανάλυση των αποτελεσμάτων

Στην ανάλυση του συντελεστή συσχέτισης που πραγματοποιείται, παρατηρούμαι μεταξύ των δυαδικών ψηφίων του απλού κειμένου και των αντίστοιχων δυαδικών ψηφίων του κρυπτογραφήματος . Εάν ο συντελεστής συσχέτισης ισούται με το μηδέν, τότε το απλό κείμενο και το κείμενο κρυπτογράφησης είναι τελείως διαφορετικά. Αν ο συντελεστής συσχέτισης είναι ίσος με -1 τότε το κρυπτογράφημα είναι το αρνητικό του απλού κειμένου. Αν ο συντελεστής συσχέτισης είναι σε τέλεια συσχέτιση τότε το κείμενο κρυπτογράφησης και το απλό κείμενο είναι τα ίδια. Η ανάλυση πραγματοποιήθηκε με τρεις διαφορετικούς τύπους δεδομένων που περιέχουν μόνο γράμματα του αλφαβήτου, μόνο ψηφία και αλφαριθμητικό κείμενο.

Για να αποφευχθεί η διαρροή πληροφοριών που σχετιζονται με την ιδωτικότητα, είναι πλεονεκτικό εάν το cipherimage φέρει ελάχιστη ή καθόλου στατιστική ομοιότητα με το απλό κείμενο. Με την χρήση ιστογραμμάτων έγινε υπολογισμός και ανάλυση πολλών κρυπτογραφημένων και πρωτότυπων αρχείων που έχουν πολύ διαφορετικό περιεχόμενοκαι το κρυπτογράφημα όταν χρησιμοποιήσαμε αλφάβητα, ψηφία και αλφαριθμητικά Οι μεταβολές κατανέμονται εξίσου στον αλγόριθμο AES όταν συγκρίνουμε τους υπόλοιπους τρεις αλγόριθμους (RC6, DES και XTEA). Η πολυκρυπτογράφηση παρέχει το καλύτερο αποτέλεσμα όταν το XTEA χρησιμοποιήθει πρώτος. Το ιστόγραμμα του κρυπτογράφου για την πολυκρυπτογράφηση με το XTEA είναι πιο ομοιόμορφο, σημαντικά διαφορετικό από αυτό του απλού κειμένου και δεν φέρει στατιστική ομοιότητα με το απλό κείμενο. Στο κρυπτογραφημένο κειμένο έχουμε μεγαλύτερη ομοιομορφία και συνεπώς δεν παρέχει καμία ένδειξη για τη χρήση οποιασδήποτε στατιστικής επίθεσης στην προτεινόμενη διαδικασία πολυκρυπτογράφησης. Για να εκτιμηθεί η ποιότητα της διαδικασίας πολυκρυπτογράφησης, είναι απαραίτητο να μελετηθεί η εξέλιξη της εντροπίας. Η εντροπία επιτρέπει να έχουμε μια ιδέα της ανακατανομής των εικονοστοιχείων και του αριθμού που απαιτείται για τη μετάδοση από το δίκτυο.

Η ταχύτητα της διαδικασίας κρυπτογράφησης και αποκρυπτογράφησης υπολογίστηκε ξεχωριστά χρησιμοποιώντας αρχείο κειμένου με μέγεθος 24KB δηλαδή 1398 μπλοκ, καθένα από τα οποία έχει 128-bit. Οι περισσότεροι από τους αλγόριθμους παρήγαγαν λιγότερο χρόνο αποκρυπτογράφησης σε σύγκριση με την κρυπτογράφηση, ενώ παράλληλα ο ρυθμός μετατροπής της αποκρυπτογράφησης είναι μεγαλύτερος. Κατά την σύγκριση του ρυθμού μετατροπής μεταξύ πολυκρυπτογράφησης και κανονικής κρυπτογράφησης, η πολυκρυπτογράφηση παρήγαγε μικρότερο ποσοστό μετατροπής.

Η συνδυασμένη προσέγγιση των λειτουργιών στα μπλόκ των κρυπτογραφιών RC6 και XTEA θα βελτιώσει την αποδοτικότητα του ρυθμού μετατροπής. Πράγμα που μπορεί να προσφέρει τον καλύτερο αλγόριθμο με καλύτερη ασφάλεια των δεδομένων που χρησιμοποιούνται.

7.3 Σύγκριση των μεθόδων που παρουσιάστηκαν

Η χρήση της κρυπτογράφησης ώς μέθοδος κάλυψης τηα ασφάλειας των Μεγάλων Δεδομένων αποτελεί ένα σημαντικό βήμα ώστε να υπάρχει ασφαλής χρήση των προσωπικών δεδομένων και σεβασμός στην ιδιωτικότητά τους. Υπάρχουν πολύ τρόποι που μπορεί να χρησιμοποιηθεί η κρυπτογράφηση όπως αναφέραμε στην εισαγωγή που έγινε προηγουμένως ώστε να δούμε την έννοια και τα βασικά στοιχεία. Προχωρώντας έπειτα από έρευνα που έχει πραγματοποιήθει παρουσιάστηκαν δυο μέθοδοι που με την χρήση της κρυπτογράφησης παρέχουν ασφάλεια σε δεδομένα όπως τα δεδομένα μεγάλης κλίμακας. Η πρώτη μέθοδος απευθύνεται στην χρήση ευαίσθητων δεδομένων σε πλατφόρμα μεγάλων δεδομένων, η οποία βασίζεται σε ένα VMM καθώς και στο Heterogeneous Proxy Re-Encryption (H-PRE). Η δεύτερη μέθοδος στηρίζεται στην Πολυ-κρυπτογράφηση δεδομένων με την χρήση δυο προσεγγίσεων των RC6 και XTEA.

Στο κεφάλαιο αυτό θα προσπαθήσουμε να κάνουμε σύγκριση των δυο αυτών μεθόδων και των τεχνικών που χρησιμοποιήθηκαν ώστε να δούμε τα αποτελέσματα τους,

ώστε να μπορούμε να κρίνουμε εαν είναι θετικά ή αρνητικά. Η πρώτη μέθοδος είναι εμφανές οτι έχει μεγαλύτερη πολυπλοκότητα σαν εκτέλεση και διαδικασία καθώς χρησιμοποιούνται διαφορες τεχνικές για το απαιτούμενο αποτέλεσμα. Αρχικά, βασίζεται σε ένα VMM ώστε να εγγυάται την ασφαλή χρήση των δεδομένων στην πλατφόρμα του cloud από τον ιδιωτικό χώρο της διαδικασίας του χρήστη. Όπως βλέπουμε κάτι τέτοιο ή παρόμοιο δεν παρατηρείται στην δεύτερη μέθοδο που παρουσιάστηκε καθώς στηρίζεται στην εκτέλεση της διαδικασίας της πολυ-κρυπτογράφησης των δεδομένων και όχι στο περιβάλλον που θα μπορούσε να στηθεί γύρω από μια πλατφόρμα νέφους όπου οι χρήστες θα διαχειρίζονται τα δεδομένα τους.

Στην πρώτη μέθοδο παρατηρούμε ότι ο κάτοχος δεδομένων κρυπτογραφεί τα δεδομένα τοπικά, χρησιμοποιώντας πρώτα το πρότυπο Advanced Encryption Standard (AES), το οποίο είναι συμμετρικός αλγόριθμος κρυπτογράφησης για την κρυπτογράφηση των δεδομένων υποβολής και στη συνέχεια, χρησιμοποιείται ο αλγόριθμος Heterogeneous Proxy re-encryption (H-PRE) για την κρυπτογράφηση του συμμετρικού κλειδιού των δεδομένων. Αντίθετα, στην μέθοδο της Πολυ-κρυπτογράφησης χρησιμοποιούνται οι αλγόριθμοι RC6 και XTEA εναλλάξ ώστε να καλυφθούν όλες οι περιπτώσεις χρήσης τους. Το κοινό χαρακτηριστικό και τον δύο μεθόδων είναι ότι το πρότυπο Advanced Encryption Standard (AES) και ο RC6 αλγόριθμος είναι αλγόριθμοι συμμετρικού κλειδιού δηλαδή είναι αλγόριθμοι κρυπτογραφίας που χρησιμοποιούν τα ίδια κρυπτογραφικά κλειδιά τόσο για την κρυπτογράφηση του απλού κειμένου όσο και για την αποκρυπτογράφηση. Τα κλειδιά μπορεί να είναι ίδια ή μπορεί να υπάρχει ένας απλός μετασχηματισμός, στην πράξη αντιπροσωπεύουν το μυστικό μεταξύ δύο ή περισσοτέρων μερών που μπορούν να χρησιμοποιηθούν για τη διατήρηση ενός ιδιωτικού συνδέσμου πληροφοριών. Αυτή η απαίτηση και για τα δύο μέρη να έχουν πρόσβαση στο μυστικό κλειδί είναι ένα από τα κύρια μειονεκτήματα της συμμετρικής κρυπτογράφησης κλειδιών, σε σύγκριση με την κρυπτογράφηση δημόσιου κλειδιού (γνωστή και ως ασύμμετρη κρυπτογράφηση κλειδιών. Ο αλγόριθμος Heterogeneous Proxy re-encryption (H-PRE) είναι κρυπτοσύστημα το οποίο επιτρέπει σε τρίτους (proxy) να αλλάξουν ένα κρυπτογράφημα το οποίο έχει κρυπτογραφηθεί από ένα μέρος, έτσι ώστε να μπορεί να αποκρυπτογραφηθεί από άλλο.

Η διαδικασία ανάλυσης της πρώτης μεθόδου περιλαμβάνει πιο πολυσύνθετες διαδικασίες που συμπεριλαμβάνουν ξεχωριστά στοιχεία και μεταβλητές τα οποία αναφέρονται αναλυτικά. Επίσης, περιλαμβάνονται οι λειτουργίες της υποβολή, της αποθήκευσης και της εξαγωγής των ευαίσθητων δεδομένων του συστήματος που αποτελούν και τον κύριο κορμό της μεθόδου που παρουσιάζεται. Καθώς, γίνεται η χρήση του VMM για την ασφάλεια των ευαίσθητων δεδομένων παρουσιάζονται και τα στοιχεία αυτής της διαδικασίας τα οποία έχουν τρία στάδια την δημιουργία ενός αξιόπιστου περιβάλλοντος και καναλιών, την μεταφόρτωση και εξαγωγή δεδομένων και τέλος την εκτέλεση του προγράμματος. Στην μέθοδο της πολυ-κρυπτογράφησης έχουμε επίσης τα βασικά στοιχεία στον κεντρικό κορμό τα οποία παρουσιάζουν τα βήματα με την σειρά για την εκτέλεση της πολυ-κρυπτογράφησης τα οποία αλλάζουν και ανάλογα με την σειρά χρήσης των δυο αλγορίθμων καθώς βλέπουμε τον συνδιασμό τους.

Ολοκληρώνοντας την σύγκριση των δυό μεθόδων παρατηρούμε ότι στην πολυ-κρυπτογράφηση έχουμε ανάλυση των αποτελεσμάτων με την χρήση του συντελεστή συσχέτισης καθως και με την χρήση ιστογραμμάτων. Επίσης πραγματοποιείται ανάλυση που βασίζεται στις ταχύτητες κρυπτογράφησης και αποκρυπτογράφησης σε διάφορες περιπτώσεις. Γίνεται σύγκριση των αποτελεσμάτων και αναλυτικός σχολιασμός για το καλύτερο αποτέλεσμα. Αντίθετα, στην άλλη μέθοδο την οποία περιγράψαμε παρατηρούμε οτι ενώ υπάρχει εκτενής ανάλυση των διαδικασίων και των βημάτων που εκτελούνται δεν έχουμε κάποια παράθεση αποτελεσμάτων ή σύγκριση μεταξύ κάποιων στοιχείων.

Κεφάλαιο 8: Cloud Computing – Νέφος

8.1 Εισαγωγή στο Cloud Computing

Το Cloud computing, που αναφέρεται συχνά ως απλά το "σύννεφο", είναι κοινόχρηστα σύνολα διαμορφωμένων πόρων συστήματος υπολογιστών και υπηρεσιών ανώτερου επιπέδου που μπορούν να εξασφαλιστούν γρήγορα με ελάχιστη προσπάθεια διαχείρισης, συχνά μέσω του διαδικτύου. Το Cloud computing βασίζεται στην ανταλλαγή πόρων για την επίτευξη συνοχής και οικονομίας, παρόμοια με μια δημόσια υπηρεσία.

Τα σύννεφα τρίτων επιτρέπουν στους οργανισμούς να επικεντρωθούν στις βασικές τους δραστηριότητες αντί να δαπανήσουν πόρους για την υποδομή και τη συντήρηση υπολογιστών. Σημειώνεται ότι το cloud computing επιτρέπει στις εταιρείες να αποφεύγουν ή να ελαχιστοποιούν τα αρχικά κόστη υποδομής πληροφορικής καθώς και να κάνουν πιο γρήγορες τις εφαρμογές τους με βελτιωμένη διαχειρισιμότητα και λιγότερη συντήρηση, και επιτρέπει στις ομάδες τεχνολογίας πληροφορικής να προσαρμόζουν ταχύτερα τους πόρους για να ανταποκριθούν σε κυμαινόμενη και απρόβλεπτη ζήτηση.

8.2 Χαρακτηριστικά του Cloud Computing

Ο ορισμός του cloud computing από το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας προσδιορίζει "πέντε βασικά χαρακτηριστικά":

Αυτοεξυπηρέτηση: Ένας καταναλωτής μπορεί μονομερώς να παρέχει δυνατότητες υπολογιστών, όπως διακομιστή και αποθήκευση δικτύου, αυτόματα χωρίς να απαιτείται ανθρώπινη αλληλεπίδραση με κάθε πάροχο υπηρεσιών.
Ευρεία πρόσβαση στο δίκτυο: Οι δυνατότητες είναι διαθέσιμες μέσω του δικτύου και υπάρχει προσπέλαση μέσω τυποποιημένων μηχανισμών που προωθούν τη χρήση από πλατφόρμες πελατών.
Συγκέντρωση πόρων: Οι υπολογιστικοί πόροι του παρόχου συγκεντρώνονται για να εξυπηρετούν πολλούς καταναλωτές χρησιμοποιώντας ένα μοντέλο πολλαπλών μισθωτών, με διαφορετικούς φυσικούς και εικονικούς πόρους δυναμικώς εκχωρημένους.
Ταχεία ελαστικότητα: Οι δυνατότητες μπορούν να απελευθερωθούν και να κλιμακώνονται ταχέως προς τα έξω και προς τα μέσα ανάλογα με τη ζήτηση. Στον καταναλωτή, οι διαθέσιμες δυνατότητες παροχής υπηρεσιών συχνά εμφανίζονται απεριόριστες και μπορούν να χρησιμοποιηθούν σε οποιαδήποτε ποσότητα ανά πάσα στιγμή.
Μετρούμενη υπηρεσία: Τα συστήματα Cloud ελέγχουν αυτόματα και βελτιστοποιούν τη χρήση των πόρων, αξιοποιώντας τη δυνατότητα μέτρησης σε κάποιο επίπεδο αφαίρεσης κατάλληλου για τον τύπο υπηρεσίας. Η χρήση των πόρων μπορεί να παρακολουθείται, να ελέγχεται και να αναφέρεται, παρέχοντας διαφάνεια τόσο για τον πάροχο όσο και για τον καταναλωτή της χρησιμοποιούμενης υπηρεσίας.

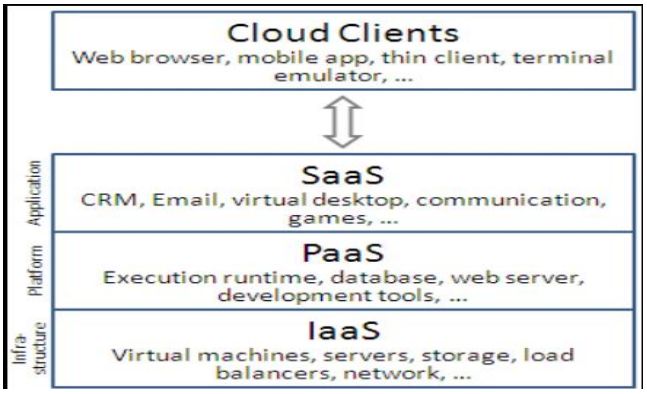
Παρόλα αυτά τα χαρακτηρίστηκα το Cloud computing παρουσιάζει και κάποια επιπλέον, ώστε να προσδιορίζεται καλύτερα και πιο επεξηγηματικά. Αυτά είναι τα εξής:

- Η ευελιξία των οργανισμών μπορεί να βελτιωθεί, καθώς το cloud computing μπορεί να αυξήσει την ευελιξία των χρηστών με την επαναφορά των πόρων, την προσθήκη ή την επέκταση των πόρων της τεχνολογικής υποδομής.
- Η ανεξαρτησία της συσκευής και της τοποθεσίας επιτρέπει στους χρήστες να έχουν πρόσβαση σε συστήματα χρησιμοποιώντας ένα πρόγραμμα περιήγησης ιστού, ανεξάρτητα από την τοποθεσία τους ή τη συσκευή που χρησιμοποιούν.
- Η συντήρηση των εφαρμογών του cloud computing είναι ευκολότερη, επειδή δεν χρειάζεται να εγκατασταθεί στον υπολογιστή του κάθε χρήστη και μπορεί να προσεγγιστεί από διαφορετικά μέρη.

- Οι μειώσεις κόστους απαιτούνται από τους παρόχους cloud. Αυτό μειώνει τα εμπόδια στην είσοδο, καθώς η υποδομή παρέχεται συνήθως από τρίτο μέρος και δεν χρειάζεται να αγοραστεί για ενίοτε ή σπάνια εντατικά καθήκοντα πληροφορικής. Επίσης, απαιτούνται λιγότερες δεξιότητες πληροφορικής για την υλοποίηση έργων που χρησιμοποιούν υπολογιστικά cloud.
- Η απόδοση παρακολουθείται από εμπειρογνώμονες της τεχνολογίας της πληροφορικής από τον πάροχο υπηρεσιών και οι συνεπείς και χαλαρά συζευγμένες αρχιτεκτονικές κατασκευάζονται χρησιμοποιώντας υπηρεσίες ιστού ως διεπαφή συστήματος.
- Η παραγωγικότητα μπορεί να αυξηθεί όταν πολλοί χρήστες μπορούν να δουλέψουν ταυτόχρονα στα ίδια δεδομένα, αντί να περιμένουν να αποθηκευτούν και να αποσταλούν μέσω ηλεκτρονικού ταχυδρομείου. Ο χρόνος μπορεί να αποθηκευτεί, καθώς οι πληροφορίες δεν χρειάζεται να εισαχθούν ξανά όταν τα πεδία αντιστοιχούν, ούτε και οι χρήστες πρέπει να εγκαταστήσουν αναβαθμίσεις λογισμικού εφαρμογών στον υπολογιστή τους.
- Η αξιοπιστία βελτιώνεται με τη χρήση πολλαπλών τοποθεσιών, γεγονός που καθιστά το cloud κατάλληλα σχεδιασμένο για την επιχειρησιακή συνέχεια και την αποκατάσταση καταστροφών.
- Η ασφάλεια μπορεί να βελτιωθεί εξαιτίας της συγκέντρωσης δεδομένων και των αυξημένων πόρων, αλλά οι ανησυχίες μπορούν να εξακολουθήσουν να αφορούν την απώλεια ελέγχου ορισμένων ευαίσθητων δεδομένων και την έλλειψη ασφάλειας για τους αποθηκευμένους πυρήνες. Η ασφάλεια είναι συχνά τόσο καλή, εν μέρει επειδή οι πάροχοι υπηρεσιών είναι σε θέση να αφιερώσουν πόρους για την επίλυση ζητημάτων ασφάλειας που πολλοί πελάτες δεν έχουν την οικονομική δυνατότητα να αντιμετωπίσουν ή για τους οποίους δεν διαθέτουν τις τεχνικές δεξιότητες που πρέπει να αντιμετωπίσουν. Ωστόσο, η πολυπλοκότητα της ασφάλειας αυξάνεται σημαντικά όταν τα δεδομένα διανέμονται σε μια ευρύτερη περιοχή ή σε έναν μεγαλύτερο αριθμό συσκευών. Επιπλέον, η πρόσβαση των χρηστών σε αρχεία καταγραφής ελέγχου ασφαλείας μπορεί να είναι δύσκολη ή αδύνατη. Οι ιδιωτικές εγκαταστάσεις cloud ενθαρρύνονται εν μέρει από την επιθυμία των χρηστών να διατηρούν τον έλεγχο της υποδομής και να αποφεύγουν να χάσουν τον έλεγχο της ασφάλειας των πληροφοριών.

8.3 Μοντέλα Υπηρεσιών (Service models)

Αν και η αρχιτεκτονική προσανατολισμένη στις υπηρεσίες υποστηρίζει το "everything as services", οι πάροχοι υπολογιστικού νέφους προσφέρουν τις "υπηρεσίες" τους σύμφωνα με διαφορετικά μοντέλα, εκ των οποίων τα τρία πρότυπα μοντέλα είναι Infrastructure as a Service (IaaS), Platform as a Service (PaaS) και Software as a Service (SaaS). Συχνά απεικονίζονται ως στρώματα σε μια στοίβα: υποδομή, πλατφόρμα και λογισμικό, αλλά αυτά δεν χρειάζεται να σχετίζονται. Για παράδειγμα, μπορούμε να παρέχουμε SaaS που εφαρμόζονται σε φυσικές μηχανές, χωρίς να χρησιμοποιούμε υποκείμενα στρώματα PaaS ή IaaS, και αντίστροφα μπορεί κανείς να τρέξει ένα πρόγραμμα στο IaaS και να το αποκτήσει απευθείας πρόσβαση χωρίς το SaaS.



8.3.1 Υποδομή ως υπηρεσία (Infrastructure as a Service -IaaS)

Η "Υποδομή ως υπηρεσία" (IaaS) αναφέρεται σε ηλεκτρονικές υπηρεσίες που παρέχουν API υψηλού επιπέδου που χρησιμοποιούνται για διαφορές λεπτομερειών χαμηλού επιπέδου της υποκείμενης υποδομής δικτύου όπως φυσικοί υπολογιστικοί πόροι, τοποθεσία, κλιμάκωση, ασφάλεια, , όπως το Xen, το Oracle VirtualBox, που εκτελούν τις εικονικές μηχανές ως φιλοξενούμενοι. Συγκεντρώσεις των hypervisors στο πλαίσιο του λειτουργικού συστήματος cloud μπορούν να υποστηρίξουν μεγάλο αριθμό εικονικών μηχανών και την ικανότητα κλιμάκωσης υπηρεσιών προς τα πάνω και προς τα κάτω σύμφωνα με τις διαφορετικές απαιτήσεις των πελατών. Τα νέφη IaaS προσφέρουν συχνά πρόσθετους πόρους, όπως βιβλιοθήκη δίσκων εικονικής μηχανής, αποθήκευση πρωτογενών μονάδων αποθήκευσης κλπ.

Ο ορισμός του cloud computing του NIST περιγράφει το IaaS ως "όπου ο καταναλωτής είναι σε θέση να αναπτύξει και να εκτελέσει αυθαίρετο λογισμικό, το οποίο μπορεί να περιλαμβάνει λειτουργικά συστήματα και εφαρμογές. Ο καταναλωτής δεν διαχειρίζεται ή ελέγχει την υποκείμενη υποδομή του cloud αλλά έχει τον έλεγχο των λειτουργικών συστημάτων, και αναπτύσσονται εφαρμογές και ενδεχομένως περιορισμένος έλεγχος επιλεγμένων στοιχείων δικτύωσης (π.χ. firewalls υποδοχής).

Οι πάροχοι IaaS-cloud παρέχουν αυτούς τους πόρους κατ' απαίτηση από τις μεγάλες δεξαμενές εξοπλισμού που είναι εγκατεστημένες σε κέντρα δεδομένων. Για ευζωνική σύνδεση, οι πελάτες μπορούν να χρησιμοποιήσουν είτε το Internet ή τα σύννεφα φορέων (αποκλειστικά εικονικά ιδιωτικά δίκτυα). Για να αναπτύξουν τις εφαρμογές τους, οι χρήστες cloud εγκαθιστούν εικόνες λειτουργικού συστήματος και το λογισμικό εφαρμογών τους στην υποδομή του cloud. Σε αυτό το μοντέλο, ο χρήστης cloud επιδιορθώνει και διατηρεί τα λειτουργικά συστήματα και το λογισμικό εφαρμογής. Οι πάροχοι νέφους συνήθως χρεώνουν τις υπηρεσίες IaaS σε βάση υπολογιστικής χρησιμότητας: το κόστος αντικατοπτρίζει το ποσό των πόρων που διατίθενται και καταναλώνονται.

8.3.2 Η πλατφόρμα ως υπηρεσία (Platform as a Service - PaaS)

Ο ορισμός του NIST για το cloud computing ορίζει την πλατφόρμα ως υπηρεσία ως εξής: «Η δυνατότητα που παρέχεται στον καταναλωτή είναι η ανάπτυξη στην υποδομή του cloud των εφαρμογών που δημιουργούνται από καταναλωτές ή αποκτώνται, οι οποίες δημιουργούνται χρησιμοποιώντας γλώσσες προγραμματισμού, βιβλιοθήκες, υπηρεσίες και εργαλεία. Ο καταναλωτής δεν διαχειρίζεται ή ελέγχει την υποκείμενη υποδομή του cloud, συμπεριλαμβανομένων των δικτύων, των διακομιστών, των λειτουργικών συστημάτων ή του χώρου αποθήκευσης, αλλά έχει τον έλεγχο των αναπτυγμένων εφαρμογών και ενδεχομένως των ρυθμίσεων διαμόρφωσης για το περιβάλλον φιλοξενίας εφαρμογών.»

Οι πωλητές PaaS προσφέρουν ένα περιβάλλον ανάπτυξης στους προγραμματιστές εφαρμογών. Ο πάροχος τυπικά αναπτύσσει ένα σύνολο εργαλείων και προτύπων για την ανάπτυξη και τα κανάλια διανομής και πληρωμής. Στα μοντέλα PaaS, οι πάροχοι cloud παρέχουν μια πλατφόρμα υπολογιστών, που συνήθως περιλαμβάνει λειτουργικό σύστημα,

περιβάλλον εκτέλεσης γλώσσας προγραμματισμού, βάση δεδομένων και διακομιστή ιστού. Οι προγραμματιστές εφαρμογών μπορούν να αναπτύξουν και να εκτελέσουν τις λύσεις λογισμικού τους σε μια πλατφόρμα νέφους χωρίς το κόστος και την πολυπλοκότητα της αγοράς και διαχείρισης των υποκείμενων στρώσεων υλικού και λογισμικού. Με ορισμένες εφαρμογές του PaaS, όπως το Microsoft Azure, την πλατφόρμα Oracle Cloud και το Google App Engine, η υποκείμενη κλίμακα υπολογιστών και αποθηκευτικών πόρων προσαρμόζεται αυτόματα στη ζήτηση εφαρμογών, ώστε ο χρήστης του cloud να μην χρειάζεται να διαθέσει τους πόρους με μη αυτόματο τρόπο. Ο τελευταίος έχει επίσης προταθεί από μια αρχιτεκτονική που στοχεύει στη διευκόλυνση του πραγματικού χρόνου σε περιβάλλοντα νέφους.

Ορισμένοι πάροχοι υπηρεσιών ολοκλήρωσης και διαχείρισης δεδομένων έχουν επίσης αγκαλιάσει εξειδικευμένες εφαρμογές της PaaS ως μοντέλα παράδοσης για λύσεις δεδομένων. Παραδείγματα περιλαμβάνουν το iPaaS (πλατφόρμα ενοποίησης ως υπηρεσία) και το dPaaS (πλατφόρμα δεδομένων ως υπηρεσία). Το iPaaS επιτρέπει στους πελάτες να αναπτύσσουν, να εκτελούν και να ρυθμίζουν τις ροές ενοποίησης. Στο πλαίσιο του μοντέλου ενσωμάτωσης του iPaaS, οι πελάτες προωθούν την ανάπτυξη και την ανάπτυξη ενοποιήσεων χωρίς να εγκαταστήσουν ή να διαχειριστούν οποιοδήποτε υλικό ή μεσαίο λογισμικό. Το dPaaS παρέχει προϊόντα ολοκλήρωσης και διαχείρισης δεδομένων ως μια πλήρως διαχειριζόμενη υπηρεσία. Σύμφωνα με το μοντέλο dPaaS, ο πάροχος PaaS, διαχειρίζεται την ανάπτυξη και εκτέλεση λύσεων δεδομένων, δημιουργώντας προσαρμοσμένες εφαρμογές δεδομένων για τον πελάτη. Οι χρήστες του dPaaS διατηρούν τη διαφάνεια και τον έλεγχο των δεδομένων μέσω εργαλείων οπτικοποίησης δεδομένων. Οι χρήστες της πλατφόρμας ως υπηρεσία (PaaS) δεν διαχειρίζονται ή δεν ελέγχουν την υποκείμενη υποδομή του cloud, συμπεριλαμβανομένων των δικτύων, των διακομιστών, των λειτουργικών συστημάτων ή του χώρου αποθήκευσης, αλλά έχουν τον έλεγχο των αναπτυσσόμενων εφαρμογών και ενδεχομένως των ρυθμίσεων διαμόρφωσης για το περιβάλλον φιλοξενίας εφαρμογών.

Ένα πρόσφατο εξειδικευμένο PaaS είναι το blockchain ως υπηρεσία (BaaS), που ορισμένοι προμηθευτές όπως το IBM Bluemix και το Oracle Cloud Platform έχουν ήδη συμπεριλάβει στην προσφορά PaaS.

8.3.3 Το λογισμικό ως υπηρεσία (Software as a Service - SaaS)

Ο ορισμός του NIST για το cloud computing ορίζει το λογισμικό ως υπηρεσία ως εξής: «Η δυνατότητα που παρέχεται στον καταναλωτή είναι να χρησιμοποιεί τις εφαρμογές του πάροχου που εκτελούνται σε μια υποδομή σύννεφο. Οι εφαρμογές είναι προσβάσιμες από διάφορες συσκευές είτε μέσω διεπαφής, όπως ενός προγράμματος περιήγησης ιστού (π.χ. ηλεκτρονικού ταχυδρομείου μέσω διαδικτύου) είτε μέσω διεπαφής προγράμματος. Ο καταναλωτής δεν διαχειρίζεται ή ελέγχει την υποκείμενη υποδομή του cloud, συμπεριλαμβανομένων των δικτύων, των διακομιστών, των λειτουργικών συστημάτων, των αποθηκευτικών χώρων ή ακόμη και των ατομικών δυνατοτήτων εφαρμογής, με την πιθανή εξαίρεση περιορισμένων ρυθμίσεων διαμόρφωσης εφαρμογών για συγκεκριμένους χρήστες.»

Στο μοντέλο λογισμικού ως υπηρεσία (SaaS), οι χρήστες αποκτούν πρόσβαση σε λογισμικό και βάσεις δεδομένων εφαρμογών. Οι παροχείς υπηρεσιών Cloud διαχειρίζονται την υποδομή και τις πλατφόρμες που εκτελούν τις εφαρμογές. Το SaaS αναφέρεται μερικές φορές ως "λογισμικό κατά παραγγελία" και συνήθως διατίθεται με χρέωση ανά χρήση ή με συνδρομή. Στο μοντέλο SaaS, οι πάροχοι νέφους εγκαθιστούν και λειτουργούν λογισμικό εφαρμογών στο cloud και οι χρήστες του cloud έχουν πρόσβαση στο λογισμικό. Οι χρήστες του Cloud δεν διαχειρίζονται την υποδομή του cloud και την πλατφόρμα όπου εκτελείται η εφαρμογή. Αυτό εξαλείφει την ανάγκη εγκατάστασης και λειτουργίας της εφαρμογής στους υπολογιστές του χρήστη, γεγονός που απλοποιεί τη συντήρηση και την υποστήριξη. Οι εφαρμογές cloud διαφέρουν από άλλες εφαρμογές στην κλιμάκωση τους, κάτι που μπορεί να επιτευχθεί με την κλωνοποίηση εργασιών σε πολλαπλές εικονικές μηχανές κατά το χρόνο εκτέλεσης για να ικανοποιηθεί η μεταβαλλόμενη ζήτηση εργασίας. Οι αντισταθμιστές φορτίου διανέμουν το έργο πάνω από το σύνολο των εικονικών μηχανών. Αυτή η διαδικασία είναι διαφανής για τον χρήστη νέφους, ο οποίος βλέπει μόνο ένα σημείο πρόσβασης.

Το μοντέλο τιμολόγησης για τις εφαρμογές SaaS είναι συνήθως μια μηνιαία ή ετήσια κατ 'αποκοπή αμοιβή ανά χρήστη, ώστε οι τιμές να είναι κλιμακωτές και να ρυθμίζονται αν οι χρήστες προστεθούν ή αφαιρεθούν σε οποιοδήποτε σημείο. Οι υποστηρικτές ισχυρίζονται ότι η SaaS δίνει στην επιχείρηση τη δυνατότητα να μειώσει το λειτουργικό κόστος της τεχνολογίας της πληροφορικής με εξωτερική ανάθεση συντήρησης υλικού και λογισμικού και υποστήριξη στον πάροχο νέφους. Αυτό επιτρέπει στην επιχείρηση να ανακατανείμει το κόστος των λειτουργιών πληροφορικής μακριά από τις δαπάνες υλικού / λογισμικού και από τα έξοδα προσωπικού, προς την επίτευξη άλλων στόχων. Επιπλέον, με εφαρμογές που φιλοξενούνται κεντρικά, οι ενημερώσεις μπορούν να απελευθερωθούν χωρίς την ανάγκη για τους χρήστες να εγκαταστήσουν νέο λογισμικό. Ένα μειονέκτημα του SaaS έρχεται με την αποθήκευση των δεδομένων των χρηστών στον εξυπηρετητή του παροχέα cloud. Ως εκ τούτου, ενδέχεται να υπάρχει μη εξουσιοδοτημένη πρόσβαση στα δεδομένα.

8.4 Μοντέλα ανάπτυξης (Deployment Models)

- **Ιδιωτικό Νέφος**

Το ιδιωτικό νέφος είναι υποδομή cloud που λειτουργεί αποκλειστικά για έναν μόνο οργανισμό, είτε διαχειρίζεται εσωτερικά είτε από τρίτο μέρος, και φιλοξενείται είτε εσωτερικά είτε εξωτερικά. Η ανάληψη ενός ιδιωτικού έργου cloud απαιτεί σημαντική δέσμευση για την εικονικοποίηση του επιχειρηματικού περιβάλλοντος και απαιτεί από τον οργανισμό να επανεκτιμήσει τις αποφάσεις σχετικά με τους υπάρχοντες πόρους. Μπορεί να βελτιώσει τις επιχειρήσεις, αλλά κάθε βήμα στο σχέδιο εγείρει ζητήματα ασφάλειας που πρέπει να αντιμετωπιστούν για την αποφυγή σοβαρών τρωτών σημείων. Τα κέντρα δεδομένων αυτοεξυπηρέτησης είναι εν γένει εντατικά κεφάλαια. Έχουν ένα σημαντικό φυσικό αποτύπωμα, που απαιτεί κατανομές του χώρου, του υλικού και των περιβαλλοντικών ελέγχων. Έχουν προσελκύσει επικρίσεις, διότι οι χρήστες "πρέπει να αγοράσουν, να κατασκευάσουν και να διαχειριστούν" και έτσι δεν επωφελούνται από τη χειρότερη διαχείριση, ουσιαστικά "λείπει το οικονομικό μοντέλο που καθιστά το cloud computing μια τόσο ενδιαφέρουσα ιδέα".

- **Δημόσιο Νέφος**

Ένα σύννεφο ονομάζεται "δημόσιο νέφος" όταν οι υπηρεσίες εκχωρούνται μέσω δικτύου που είναι ανοιχτό για δημόσια χρήση. Οι δημόσιες υπηρεσίες cloud ενδέχεται να είναι δωρεάν. Από τεχνικής απόψεως μπορεί να υπάρχει ελάχιστη ή καμία διαφορά μεταξύ δημόσιας και ιδιωτικής αρχιτεκτονικής cloud, ωστόσο η ασφάλεια μπορεί να διαφέρει αισθητά όσον αφορά τις υπηρεσίες (εφαρμογές, αποθήκευση και άλλους πόρους) που παρέχονται από πάροχο υπηρεσιών για ένα κοινό και όταν η επικοινωνία πραγματοποιείται μέσω ενός μη αξιόπιστου δικτύου. Σε γενικές γραμμές, οι δημόσιοι πάροχοι υπηρεσιών cloud όπως το Amazon Web Services (AWS), η Oracle, η Microsoft και η Google κατέχουν και λειτουργούν την υποδομή στο κέντρο δεδομένων τους και η πρόσβαση γίνεται γενικά μέσω του διαδικτύου. Η AWS, η Oracle, η Microsoft και η Google προσφέρουν επίσης υπηρεσίες άμεσης σύνδεσης με την ονομασία "AWS Direct Connect", "Oracle FastConnect", "Azure ExpressRoute" και "Cloud Interconnect" αντίστοιχα.

- **Υβριδικό Νέφος**

Το υβριδικό νέφος είναι μια σύνθεση δύο ή περισσότερων νεφών (ιδιωτικών, κοινοτικών ή δημόσιων) που παραμένουν ξεχωριστές οντότητες αλλά συνδέονται μεταξύ τους, προσφέροντας τα οφέλη από τα μοντέλα πολλαπλής ανάπτυξης. Το υβριδικό νέφος μπορεί επίσης να σημαίνει τη δυνατότητα σύνδεσης της συνεγκατάστασης, των διαχειριζόμενων ή και των αποκλειστικών υπηρεσιών με τους πόρους του cloud. Η Gartner

ορίζει μια υπηρεσία υβριδικού cloud ως υπηρεσία cloud computing που αποτελείται από έναν συνδυασμό ιδιωτικών, δημόσιων και κοινοτικών υπηρεσιών cloud από διάφορους παρόχους. Μια υπηρεσία υβριδικών νεφών διασχίζει τα όρια απομόνωσης και παρόχων, ώστε να μην μπορεί να τοποθετηθεί απλά σε μια κατηγορία ιδιωτικών, δημόσιων ή κοινοτικών υπηρεσιών cloud. Επιτρέπει την επέκταση είτε της χωρητικότητας είτε της ικανότητας μιας υπηρεσίας νέφους, με τη συνάθροιση, την ολοκλήρωση ή την προσαρμογή με μια άλλη υπηρεσία νέφους.

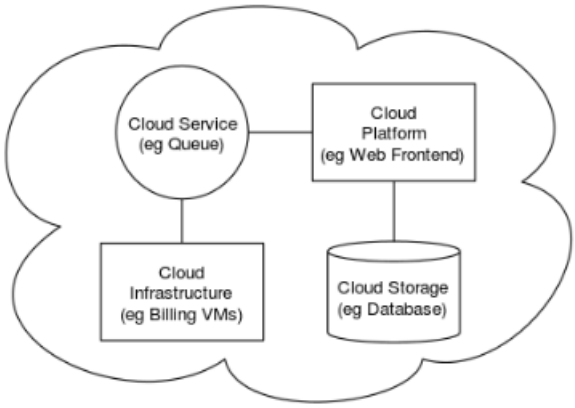
Υπάρχουν διάφορες περιπτώσεις χρήσης για τη σύνθεση υβριδικού νέφους. Για παράδειγμα, ένας οργανισμός μπορεί να αποθηκεύει ευαίσθητα δεδομένα πελάτη στο σπίτι σε μια ιδιωτική εφαρμογή, αλλά να διασυνδέει την εφαρμογή αυτή με μια εφαρμογή επιχειρηματικής ευφυΐας που παρέχεται σε ένα δημόσιο νέφος ως υπηρεσία λογισμικού. Αυτό το παράδειγμα υβριδικού νέφους επεκτείνει τις δυνατότητες της επιχείρησης να παρέχει μια συγκεκριμένη επιχειρηματική υπηρεσία μέσω της προσθήκης εξωτερικά διαθέσιμων δημόσιων υπηρεσιών cloud. Η υιοθέτηση του υβριδικού σύννεφου εξαρτάται από πολλούς παράγοντες όπως η ασφάλεια των δεδομένων και οι απαιτήσεις συμμόρφωσης, το επίπεδο ελέγχου που απαιτείται για τα δεδομένα και οι εφαρμογές που χρησιμοποιεί μια οργάνωση.

Ένα άλλο παράδειγμα υβριδικού νέφους είναι εκείνο όπου οι οργανισμοί πληροφορικής χρησιμοποιούν δημόσιους πόρους υπολογιστικού cloud για να καλύψουν τις προσωρινές ανάγκες χωρητικότητας που δεν μπορούν να καλύψουν με το ιδιωτικό νέφος. Αυτή η δυνατότητα προσφέρει στα υβριδικά νέφοι να χρησιμοποιούν εκρήξεις νεφών για κλιμάκωση. Το cloud bursting είναι ένα μοντέλο ανάπτυξης εφαρμογών στο οποίο μια εφαρμογή τρέχει σε ένα ιδιωτικό cloud ή κέντρο δεδομένων και "εκτοξεύεται" σε ένα δημόσιο νέφος όταν αυξάνεται η ζήτηση για υπολογιστική χωρητικότητα. Η έκρηξη του νέφους επιτρέπει στα κέντρα δεδομένων να δημιουργήσουν μια εσωτερική υποδομή πληροφορικής που να υποστηρίζει τους μέσους φόρτους εργασίας και να χρησιμοποιεί τους νέους πόρους από δημόσια ή ιδιωτικά σύννεφα κατά τη διάρκεια των αιχμών στις απαιτήσεις επεξεργασίας. Το εξειδικευμένο μοντέλο του υβριδικού νέφους, το οποίο είναι χτισμένο πάνω σε ετερογενές υλικό, ονομάζεται "Cross-Platform Hybrid Cloud".

8.5 Αρχιτεκτονική

Η αρχιτεκτονική συστημάτων λογισμικού που εμπλέκονται στην παροχή του cloud computing, συνήθως περιλαμβάνει πολλαπλές συνιστώσες νέφους που επικοινωνούν μεταξύ τους μέσω ενός χαλαρού μηχανισμού ζεύξης όπως μια ουρά μηνυμάτων. Η ελαστική πρόβλεψη συνεπάγεται ευφυΐα στη χρήση σφιχτής ή χαλαρής σύζευξης όπως εφαρμόζεται σε μηχανισμούς όπως αυτοί και άλλοι.

Η τεχνολογία νεφών είναι η εφαρμογή των κλάδων της μηχανικής στο cloud computing. Εισάγει μια συστηματική προσέγγιση στις ανησυχίες υψηλού επιπέδου σχετικά με την εμπορευματοποίηση, την τυποποίηση και τη διακυβέρνηση στη σύλληψη, ανάπτυξη, λειτουργία και συντήρηση των συστημάτων υπολογιστικού νέφους. Πρόκειται για μια πολυεπιστημονική μέθοδο που περιλαμβάνει συμβολές από διάφορους τομείς, όπως συστήματα, λογισμικό, ιστό, επιδόσεις, πληροφορίες, ασφάλεια, πλατφόρμα, κίνδυνος και ποιοτική μηχανική.



8.6 Ασφάλεια και προστασία της ιδιωτικότητας

Το Cloud computing δημιουργεί ανησυχίες για την προστασία της ιδιωτικότητας, επειδή ο πάροχος υπηρεσιών μπορεί να έχει πρόσβαση σε δεδομένα που βρίσκονται στο νέφος ανά πάσα στιγμή. Θα μπορούσε τυχαία ή σκόπιμα να μεταβάλει ή να διαγράψει πληροφορίες. Πολλοί πάροχοι νέφους μπορούν να μοιράζονται πληροφορίες με τρίτους, εάν αυτό είναι απαραίτητο για λόγους νόμου και τάξης χωρίς ένταλμα. Αυτό επιτρέπεται στις πολιτικές απορρήτου τους, τις οποίες πρέπει να συμφωνούν οι χρήστες πριν αρχίσουν να χρησιμοποιούν υπηρεσίες cloud. Οι λύσεις για την προστασία της ιδιωτικής ζωής περιλαμβάνουν την πολιτική και τη νομοθεσία, καθώς και τις επιλογές των τελικών χρηστών για τον τρόπο αποθήκευσης των δεδομένων. Οι χρήστες μπορούν να κρυπτογραφήσουν τα δεδομένα που επεξεργάζονται ή αποθηκεύονται στο νέφος για να αποτρέψουν την ανεπίτρεπτη πρόσβαση.

Σύμφωνα με τη Cloud Security Alliance, οι τρεις πρώτες απειλές στο νέφος είναι η απώλεια δεδομένων και διαρροή και αποτυχία υλικού, που αντιστοιχεί σε 29%, 25% και 10% όλων των διακοπών ασφαλείας cloud αντίστοιχα. Μαζί, αυτά αποτελούν κοινές ευπάθειες τεχνολογίας. Σε μια πλατφόρμα παροχής νέφους που μοιράζονται διάφοροι χρήστες, ενδέχεται να υπάρχει πιθανότητα οι πληροφορίες που ανήκουν σε διαφορετικούς πελάτες να βρίσκονται στον ίδιο διακομιστή δεδομένων. Επιπλέον, ο Eugene Schultz, επικεφαλής της τεχνολογίας στο Emagined Security, δήλωσε ότι οι χάκερ ξοδεύουν σημαντικό χρόνο και προσπάθεια αναζητώντας τρόπους να διεισδύσουν στο νέφος. Επειδή τα δεδομένα από εκατοντάδες ή χιλιάδες εταιρείες μπορούν να αποθηκευτούν σε μεγάλους διακομιστές, οι χάκερ μπορούν θεωρητικά να αποκτήσουν τον έλεγχο των τεράστιων όγκων πληροφοριών μέσω μιας ενιαίας επίθεσης, μιας διαδικασίας που ονομάζεται "hyperjacking". Ορισμένα παραδείγματα περιλαμβάνουν την παραβίαση ασφαλείας Dropbox και τη διαρροή iCloud

2014. Το Dropbox είχε παραβιαστεί τον Οκτώβριο του 2014, έχοντας πάνω από 7 εκατομμύρια κωδικούς πρόσβασης χρηστών κλεμμένους από χάκερ, σε μια προσπάθεια να αποκτήσει χρηματική αξία από την Bitcoins (BTC). Έχοντας αυτούς τους κωδικούς πρόσβασης, είναι σε θέση να διαβάσουν τα ιδιωτικά δεδομένα καθώς και να αναπροσαρμόσουν αυτά τα δεδομένα από μηχανές αναζήτησης, καθιστώντας τις πληροφορίες δημόσιες.

Υπάρχει το πρόβλημα της νόμιμης ιδιοκτησίας των δεδομένων καθώς πολλές συμφωνίες περί Όρων Παροχής Υπηρεσιών σιωπούν σχετικά με το ζήτημα της ιδιοκτησίας. Ο φυσικός έλεγχος του εξοπλισμού ηλεκτρονικών υπολογιστών (ιδιωτικό νέφος) είναι πιο ασφαλής από την απενεργοποίηση του εξοπλισμού και υπό τον έλεγχο κάποιου άλλου (δημόσιο νέφος). Αυτό παρέχει μεγάλο κίνητρο στους δημόσιους παρόχους υπηρεσιών cloud computing να δώσουν προτεραιότητα στην οικοδόμηση και διατήρηση ισχυρής διαχείρισης ασφαλών υπηρεσιών. Υπάρχει ο κίνδυνος οι τελικοί χρήστες να μην κατανοούν τα θέματα που εμπλέκονται κατά την πρόσβασή σε μια υπηρεσία νέφους, αυτό είναι σημαντικό αφού το cloud computing γίνεται δημοφιλές και απαιτείται για να λειτουργούν ορισμένες υπηρεσίες, όπως για παράδειγμα για έναν ευφυή προσωπικό βοηθό (το Siri της Apple ή το Google Now). Βασικά, το ιδιωτικό νέφος θεωρείται πιο ασφαλές με υψηλότερα επίπεδα ελέγχου για τον ιδιοκτήτη, ωστόσο το δημόσιο νέφος φαίνεται να είναι πιο ευέλικτο και απαιτεί λιγότερες επενδύσεις χρόνου και χρήματος από τον χρήστη.

Κεφάλαιο 9: CloudAnalyst

9.1 Εισαγωγή

Με την πρόοδο του Cloud, υπάρχουν νέες δυνατότητες του τρόπου με τον οποίο μπορούν να χτιστούν οι εφαρμογές στο διαδίκτυο. Από τη μία πλευρά υπάρχουν οι πάροχοι υπηρεσιών cloud οι οποίοι είναι πρόθυμοι να προσφέρουν μεγάλης κλίμακας υπολογιστική υποδομή σε φθηνότερη τιμή, εξαλείφοντας το υψηλό αρχικό κόστος δημιουργίας περιβάλλοντος ανάπτυξης εφαρμογών και παρέχοντας τις υπηρεσίες υποδομής σε πολύ ευέλικτο τρόπο. Από την άλλη πλευρά, υπάρχουν μεγάλα συστήματα λογισμικού όπως οι ιστότοποι κοινωνικής δικτύωσης και οι εφαρμογές ηλεκτρονικού εμπορίου που κερδίζουν σήμερα δημοτικότητα, οι οποίες μπορούν να επωφεληθούν σε μεγάλο βαθμό από τη χρήση τέτοιων υπηρεσιών cloud για την ελαχιστοποίηση του κόστους και τη βελτίωση της ποιότητας των υπηρεσιών στους τελικούς χρήστες. Αλλά όταν συνδυάζουμε αυτούς τους δύο άξονες, υπάρχουν διάφοροι παράγοντες που θα επηρεάσουν το καθαρό όφελος, όπως η γεωγραφική κατανομή των βάσεων χρηστών, η διαθέσιμη υποδομή του διαδικτύου εντός αυτών των γεωγραφικών περιοχών, η δυναμική φύση των προτύπων χρήσης της βάσης χρηστών πόσο καλά οι υπηρεσίες cloud μπορούν να προσαρμοστούν ή να αναμορφωθούν δυναμικά, κλπ.

Η διεξαγωγή εμπεριστατωμένης μελέτης σχετικά με αυτό το γενικό πρόβλημα στον πραγματικό κόσμο θα είναι εξαιρετικά δύσκολη και η καλύτερη προσέγγιση για τη μελέτη ενός τόσο δυναμικού και μαζικά κατανεμημένου περιβάλλοντος είναι μέσω της προσομοίωσης. Έχουν γίνει πολλές μελέτες που χρησιμοποιούν τεχνικές προσομοίωσης για τη διερεύνηση της συμπεριφοράς κατανεμημένων συστημάτων μεγάλης κλίμακας, όπως τα έργα GridSim και CloudSim.

Θα παρουσιαστεί μια έρευνα που θα ασχολείται με την επέκταση αυτών των τεχνικών για να μελετήσει τη συμπεριφορά μεγάλης κλίμακας εφαρμογών διαδικτύου σε περιβάλλον cloud και με την χρήση ενός εργαλείου το CloudAnalyst το οποίο μπορεί να χρησιμοποιηθεί για την προσομοίωση αυτού του τύπου μεγάλων κλιμακωτών εφαρμογών μαζί με μια νέα προσέγγιση για τέτοιες μελέτες. Με τον τρόπο αυτό μπορούμε να ερευνήσουμε ένα υπάρχει ικανοποιητική μετάδοση δεδομένων, πράγμα που έχει σαν αντίκτυπο και ένα ικανοποιητικό επίπεδο ασφάλειας. Καθώς, μπορούμε και να κρίνουμε πόσα χρήματα χρειάζεται να δαπανηθούν για την μελέτη της διασφάλισης των δεδομένων.

9.1.1 Ορολογία και συντομογραφίες

Data Transmission Latency	"Λανθάνουσα μετάδοση" σημαίνει καθυστέρηση στο δίκτυο (με βάση τη γεωγραφική απόσταση, τη λειτουργία του δικτύου εξοπλισμό κ.λπ.) μεταξύ δύο σημείων
Data Transfer Time	Ο χρόνος μεταφοράς δεδομένων είναι ο χρόνος που απαιτείται για μια δεδομένη ποσότητα δεδομένων να μεταφέρονται από ένα σημείο στο άλλο. Αυτό θεωρείται ότι είναι ισοδύναμο με το διαθέσιμο εύρος ζώνης διαιρούμενο με το μέγεθος του ανά μονάδα δεδομένων.
Response Time	Ο χρόνος που απαιτείται από μια εφαρμογή στο διαδίκτυο ορίζεται ως η ώρα διάστημα μεταξύ της αποστολής της αίτησης και της λήψης απάντησης
VM	Εικονική μηχανή
VMM	Παρακολούθηση εικονικής μηχανής

9.2 Χαρακτηριστικά του προγράμματος προσομοίωσης

Υπάρχουν πολλά ιδιαίτερα επιθυμητά χαρακτηριστικά ενός εργαλείου σαν αυτό που θα χρησιμοποιηθεί :

- *Ευκολία χρήσης* : Η ευκολία δημιουργίας και εκτέλεσης ενός πειράματος προσομοίωσης είναι το κύριο σημείο της ύπαρξης ενός εργαλείου προσομοίωσης. Ο προσομοιωτής πρέπει να παρέχει ένα εύκολο στη χρήση γραφικό περιβάλλον χρήστη το οποίο είναι διαισθητικό αλλά πλήρες.
- *Δυνατότητα καθορισμού προσομοίωσης με υψηλό βαθμό ευελιξίας* : Ίσως το πιο σημαντικό χαρακτηριστικό είναι το επίπεδο της διαμορφωσιμότητας που μπορεί να προσφέρει το εργαλείο. Μια προσομοίωση, σαν μια εφαρμογή στο διαδίκτυο εξαρτάται από πολλές παραμέτρους, επομένως είναι σημαντικό να μπορεί να γίνει εισαγωγή και αλλαγή σε αυτές τις παραμέτρους γρήγορα και εύκολα.
- *Γραφική έξοδος* : Μια εικόνα λέγεται ότι αξίζει χίλιες λέξεις. Το γραφικό αποτέλεσμα με τη μορφή πινάκων και γραφημάτων είναι ιδιαίτερα επιθυμητό ώστε να παρουσιάζονται τα στατιστικά που συλλέγονται κατά τη διάρκεια της προσομοίωσης. Αυτή η αποτελεσματική παρουσίαση βοηθά στην αναγνώριση των σημαντικών προτύπων των παραμέτρων εξόδου και βοηθάει στις συγκρίσεις μεταξύ των σχετικών παραμέτρων.
- *Επαναληψιμότητα* : Η επαναληψιμότητα των πειραμάτων είναι μια πολύ σημαντική απαίτηση ενός προσομοιωτή. Το ίδιο πείραμα με τις ίδιες παραμέτρους θα πρέπει να παράγει παρόμοια αποτελέσματα κάθε φορά που εκτελείται η προσομοίωση. Διαφορετικά, η προσομοίωση γίνεται απλώς μια τυχαία ακολουθία γεγονότων και όχι ένα ελεγχόμενο πείραμα.
- *Ευκολία επέκτασης* : Όπως ήδη αναφέρθηκε, η προσομοίωση είναι ένα πολύπλοκο έργο και ένα σύνολο παραμέτρων εισόδου μπορεί να επιτευχθεί σε λίγες προσπάθειες. Επομένως, αναμένεται να υπάρχει εξέλιξη συνεχώς και όχι ένα πρόγραμμα που γράφεται μία για πάντα και στη συνέχεια χρησιμοποιείται συνεχώς. Επομένως, η αρχιτεκτονική προσομοίωσης θα πρέπει να υποστηρίζει τις επεκτάσεις με ελάχιστη προσπάθεια με κατάλληλα πλαίσια.

9.3 Χρησιμοποιώντας το CloudAnalyst

Το CloudAnalyst είναι εξοπλισμένο με ένα ολοκληρωμένο GUI ενσωματωμένο στην Java Swing. Αυτή η ενότητα περιγράφει σύντομα τις οθόνες και πώς να χρησιμοποιηθούν για να γίνουν οι ρυθμίσεις και η εκτέλεση μιας προσομοίωσης.

9.3.1 Δημιουργία προσομοίωσης

Για να ρυθμιστεί μια προσομοίωση θα πρέπει να εκτελεστούν τα παρακάτω βήματα:

1. Καθορισμός βάσεων χρηστών - Χρησιμοποιώντας οντότητες βάσης χρηστών γίνεται ορισμός των χρηστών της εφαρμογής, της γεωγραφικής τους κατανομής και άλλων ιδιοτήτων όπως η συχνότητα χρήσης και το πρότυπο χρήσης όπως οι ώρες αιχμής. Αυτό γίνεται στην κύρια καρτέλα της οθόνης Configure Simulation.
2. Καθορισμός κέντρων δεδομένων - Χρησιμοποιώντας την καρτέλα Data Centers στην οθόνη Configuration γίνεται ορισμός των κέντρων δεδομένων που θα χρησιμοποιηθούν στην προσομοίωση.
3. Κατανομή Εικονικών Μηχανών για την εφαρμογή στα Κέντρα Δεδομένων - Μόλις δημιουργηθούν τα κέντρα δεδομένων, πρέπει να οριστούν εικονικές μηχανές σε αυτά για την

προσομοίωση εφαρμογής χρησιμοποιώντας την κύρια καρτέλα στο Configuration. Μπορεί να γίνει ορισμός πολλών τύπων εικονικών μηχανών στο ίδιο κέντρο δεδομένων κατά τη διάρκεια αυτού του βήματος.
4. Έλεγχος και προσαρμογή των προηγμένων παραμέτρων στην καρτέλα Advanced στην Configuration Screen.
5. Αναθεώρηση και προσαρμογή των χρόνων καθυστέρησης δικτύου και του εύρους ζώνης στο Internet Characteristics.

9.3.2 Οθόνες προσομοιωτή

9.3.2.1 Κύρια οθόνη με Simulation Panel

Όταν ξεκινήσει το CloudAnalyst, η πρώτη οθόνη που εμφανίζεται είναι η κύρια οθόνη. Έχει το πίνακα προσομοίωσης με χάρτη του κόσμου στα δεξιά και τον κύριο πίνακα ελέγχου στα αριστερά. Όπως αναφέρθηκε, το CloudAnalyst διαιρεί τον κόσμο σε 6 περιοχές που συμπίπτουν κατά προσέγγιση με τις 6 κύριες ηπείρους. Οι τοποθεσίες όλων των στοιχείων της προσομοίωσης αναγνωρίζονται μόνο από την περιοχή για λόγους απλότητας (δηλαδή δεν υπάρχουν συντεταγμένες x-y, όλες οι οντότητες εντός της περιοχής είναι παρόμοιες για ειδικές γεωγραφικές παραμέτρους.)

Οι επιλογές του πίνακα ελέγχου είναι:

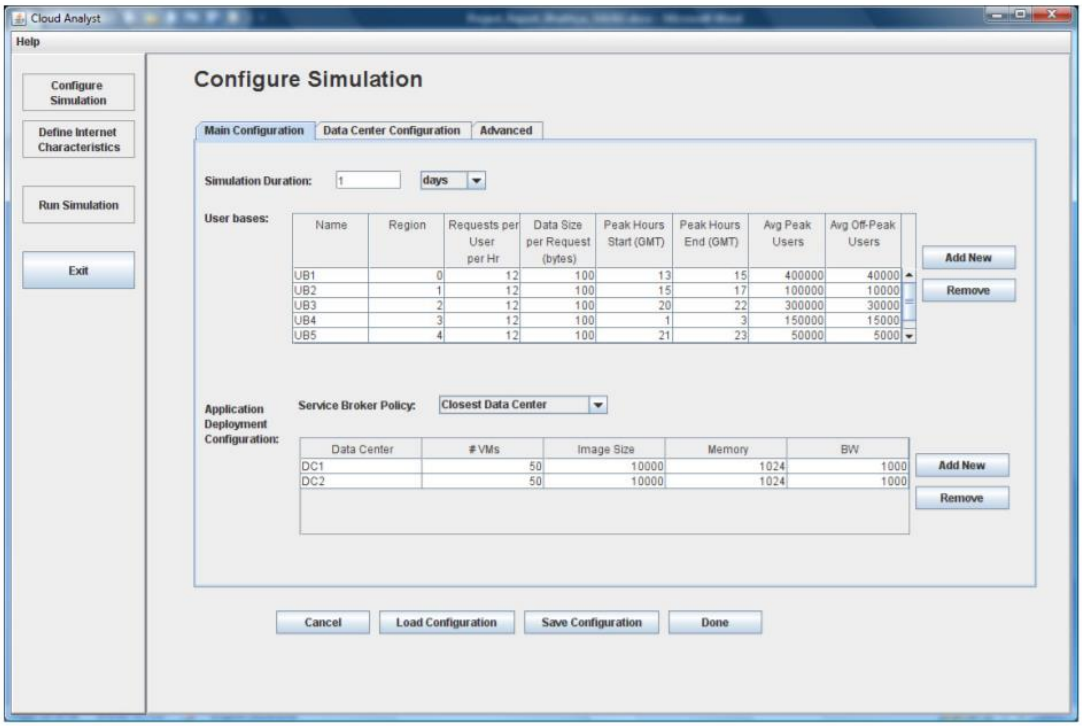
- Configure Simulation
- Define Internet Characteristics
- Run Simulation
- Exit

9.3.2.2 Οθόνη Configure Simulation

Η οθόνη Configure Simulation περιλαμβάνει τρεις καρτέλες.

- *Main Tab*

Οι επιλογές διαμόρφωσης στην κύρια καρτέλα είναι οι εξής:
1. Simulation time - η διάρκεια της προσομοίωσης που μπορεί να δοθεί σε λεπτά, ώρες ή ημέρες
2. User Bases Table - Αυτός είναι ένας πίνακας που απαριθμεί όλες τις βάσεις χρηστών στην προσομοίωση. Κάθε βάση χρηστών έχει τα ακόλουθα διαμορφώσιμα πεδία, που αντιπροσωπεύονται από μια μόνο γραμμή στον πίνακα : Name, Region, Requests per user per hour, Data size per request, Peak hours, Average users during peak hours, Average users during off-peak hours.
3. Application Deployment Configuration - Αυτός ο πίνακας αναφέρει πόσες εικονικές μηχανές έχουν κατανεμηθεί για την εφαρμογή σε κάθε κέντρο δεδομένων από την καρτέλα Data Centers, μαζί με τις λεπτομέρειες μιας εικονικής μηχανής. Τα πεδία είναι: Data Center, Number of VMs, Image Size, Memory, BW.



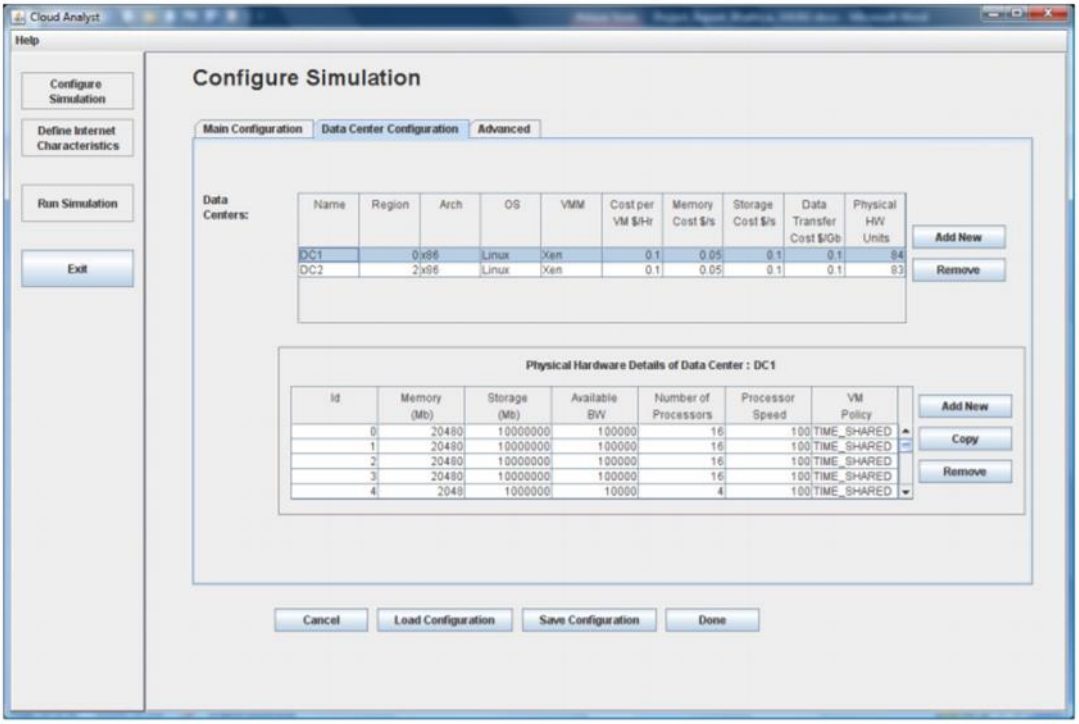
4. Service Broker Policy - Αυτή η αναπτυσσόμενη λίστα επιτρέπει την επιλογή μεταξύ των κέντρων δεδομένων που αποφασίζουν ποιο κέντρο δεδομένων θα πρέπει να λαμβάνει κίνηση από ποια βάση χρηστών.

Το κουμπί αποθήκευσης επιτρέπει να αποθηκευτεί η διαμόρφωση που δημιουργήθηκε ως αρχείο. Τα αρχεία προσομοίωσης αποθηκεύονται με επέκταση .sim.

- *Data Center Tab*

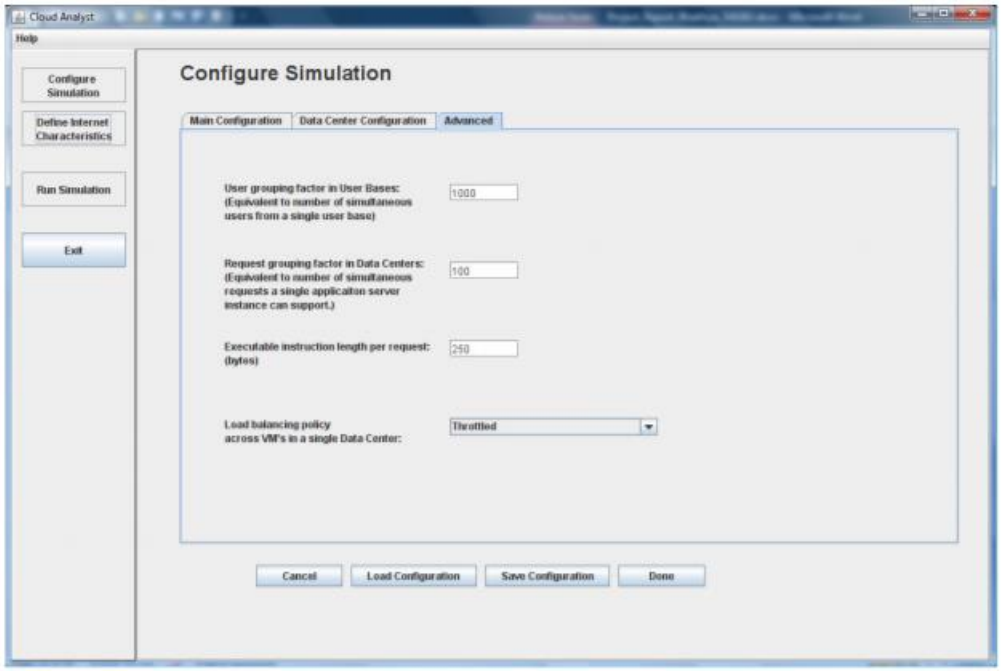
Η καρτέλα κέντρου δεδομένων επιτρέπει τον ορισμό της διαμόρφωσης ενός κέντρου δεδομένων. Ο πίνακας στην κορυφή παραθέτει τα κέντρα δεδομένων και χρησιμοποιώντας τα κουμπιά "Προσθαφαίρεση" μπορείτε να προσθέσετε ή να αφαιρέσετε τα κέντρα δεδομένων στη διαμόρφωση. Τα πεδία παραμέτρων είναι: Name, Region, Architecture, Operating

System, Virtual Machine Monitor (VMM), Cost per VM Hour, Cost per 1Mb Memory Hour, Storage cost per Gb, Data Transfer cost per Gb, Number of servers.
Όταν γίνεται η επιλογή ενός κέντρου δεδομένων από αυτόν τον πίνακα, θα εμφανιστεί κάτω από αυτόν ένας δεύτερος πίνακας με τις λεπτομέρειες των μηχανών διακομιστή στο κέντρο δεδομένων. Οι παράμετροι για κάθε μηχανή μπορούν να δοθούν σύμφωνα με τα διαθέσιμα πεδία.



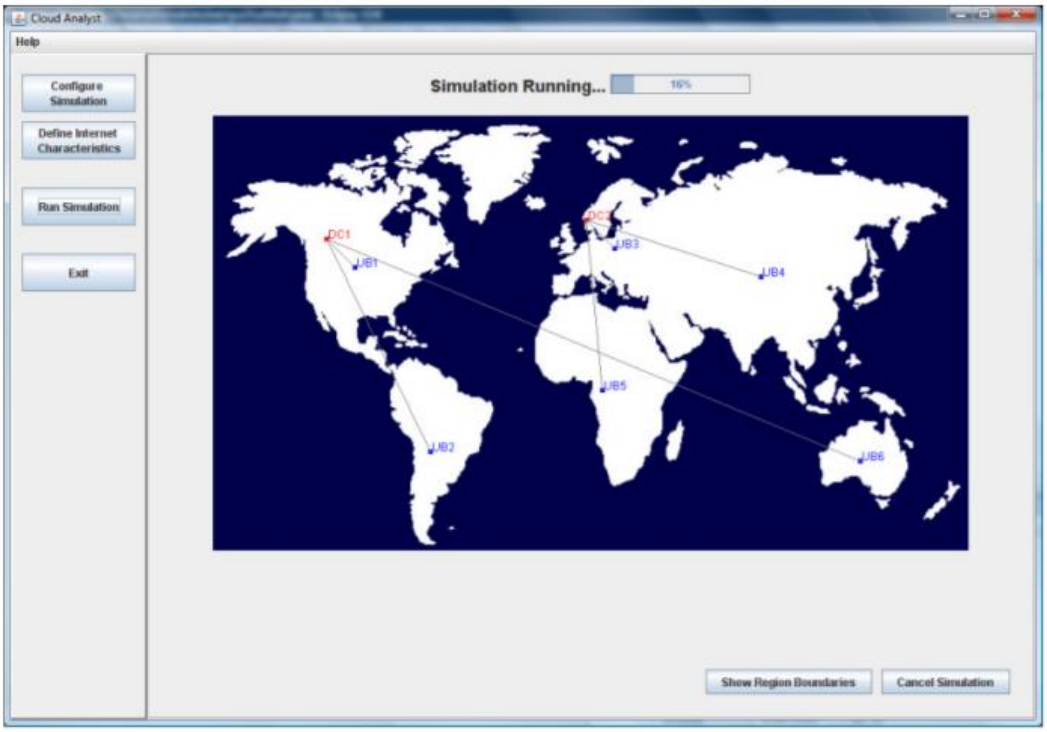
- *Advanced Tab*

Η καρτέλα για προχωρημένους περιέχει κάποιες σημαντικές παραμέτρους που ισχύουν για την ολοκλήρωση της προσομοίωσης. Αυτές οι παράμετροι είναι : User Grouping Factor (in User Bases), Request Factor Grouping (in Data Centers), Executable instruction length (in bytes) και Load balancing policy.



9.3.3 Εκτέλεση προσομοίωσης

Μόλις οι παραπάνω οθόνες έχουν χρησιμοποιηθεί για να δημιουργηθεί επιτυχώς μια διαμόρφωση προσομοίωσης, ο χρήστης πρέπει να επιστρέψει στην κύρια οθόνη και να εκτελεστεί η προσομοίωση επιλέγοντας το "Run Simulation" από τον πίνακα ελέγχου. Αυτό θα ξεκινήσει την προσομοίωση και η γραμμή προόδου στην κορυφή του πίνακα προσομοίωσης δείχνει το ποσοστό ολοκλήρωσης της προσομοίωσης. Η οθόνη προσομοίωσης θα εμφανίσει ένα απλό κινούμενο σχέδιο που θα δείχνει ποιες βάσεις χρηστών στέλνουν μηνύματα στα οποία τα κέντρα δεδομένων. Μια προσομοίωση μπορεί να ακυρωθεί πριν από την ολοκλήρωση της εκτέλεσης, χρησιμοποιώντας το κουμπί ακύρωσης στην κάτω δεξιά γωνία. Μπορεί να χρειαστεί λίγος χρόνος αφού πατηθεί το κουμπί ακύρωσης για να σταματήσει η προσομοίωση καθώς θα συνεχίσει να συγκεντρώνει τα δεδομένα προσομοίωσης των αιτημάτων που είχαν δημιουργηθεί πριν την ακύρωση αλλά δεν είχαν ολοκληρωθεί.



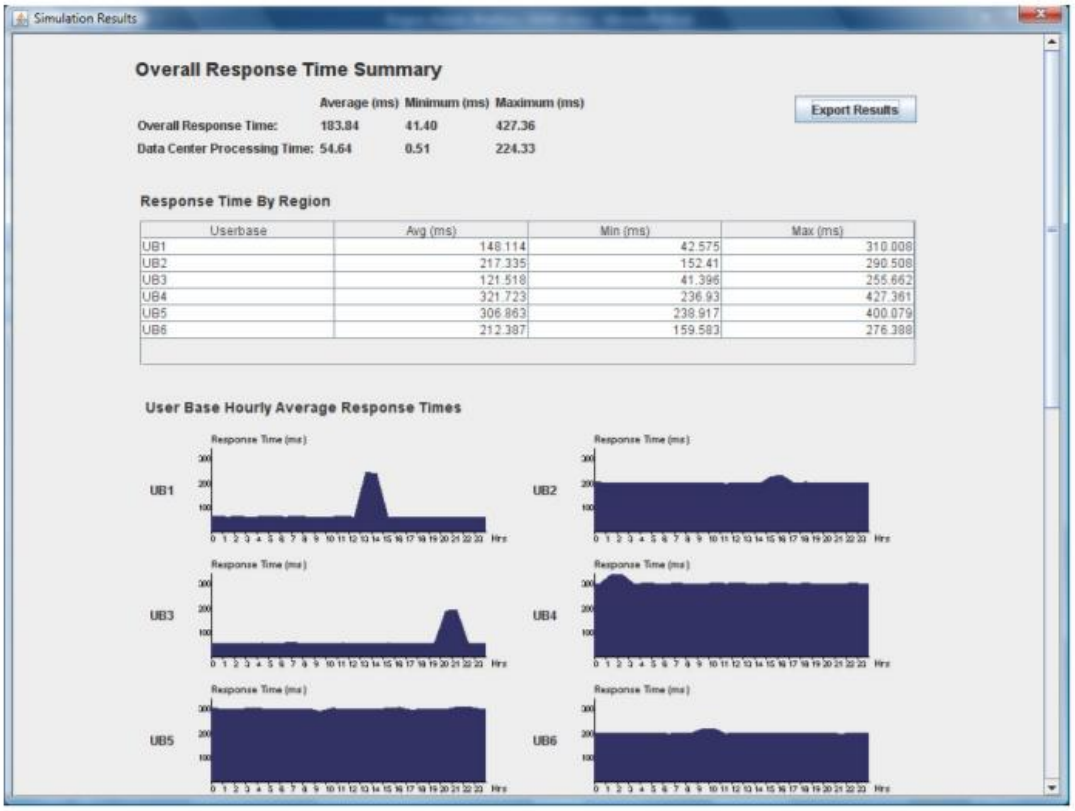
9.3.4 Οθόνη αποτελεσμάτων

Μόλις ολοκληρωθεί η προσομοίωση, οι κύριοι χρόνοι απόκρισης θα εμφανιστούν στον πίνακα προσομοίωσης δίπλα σε κάθε βάση χρηστών. Τα λεπτομερή αποτελέσματα μπορούν να προβληθούν κάνοντας κλικ στο κουμπί " View Detailed Results " που εμφανίζεται στη δεξιά κάτω γωνία της οθόνης μετά την ολοκλήρωση της προσομοίωσης.

Στην οθόνη αποτελεσμάτων θα εμφανιστούν τα δεδομένα που συλλέγονται από την προσομοίωση. Αυτό περιλαμβάνει:

1. Συνολική περίληψη χρόνου απόκρισης (για όλες τις βάσεις χρηστών)
2. Χρόνος απόκρισης από τη βάση χρηστών σε μορφή πίνακα
3. Χρόνος απόκρισης από τη βάση χρηστών σε γραφική μορφή κατανεμημένη στις 24 ώρες της ημέρας.
4. Χρόνος εξυπηρέτησης από κάθε κέντρο δεδομένων σε μορφή πίνακα

5. Χρόνος εξυπηρέτησης από το κέντρο δεδομένων σε γραφική μορφή, κατανεμημένη σε 24 ώρες της ημέρας.
6. Φόρτωση κέντρου δεδομένων (αριθμός αιτημάτων που εξυπηρετούνται) σε γραφική μορφή που κατανεμήθηκε σε 24 ώρες της ημέρας
7. Στοιχεία κόστους



9.4 Προσομοίωση μιας εφαρμογής μεγάλης κλίμακας Internet που εκτελείται στο Cloud

Ένας τυπικός μεγάλος κλιμακωτός τύπος εφαρμογής στο διαδίκτυο σήμερα που θα μπορούσε να ωφεληθεί από το Cloud είναι εφαρμογές κοινωνικής δικτύωσης. Π.χ. Το Facebook, ένας από τους πιο δημοφιλείς ιστότοπους κοινωνικής δικτύωσης, έχει πάνω από 2 δισεκατομμύρια εγγεγραμμένους χρήστες παγκοσμίως. Στις η κατά προσέγγιση διανομή της βάσης χρηστών Facebook σε ολόκληρο τον πλανήτη δίνεται στον παρακάτω πίνακα :

Περιοχή	**CloudAnalyst Id Περιοχής**	**Χρήστες**
Βόρεια Αμερική	0	320 εκατ.
Νότια Αμερική	1	325 εκατ.
Ευρώπη	2	340 εκατ.
Ασία	3	820 εκατ.
Αφρική	4	180 εκατ.
Ωκεανία	5	20 εκατ.

Για την προσομοίωση μας χρησιμοποιούμε ένα παρόμοιο σύστημα στο 1/10 της κλίμακας του Facebook.

9.4.1 Διαμόρφωση προσομοίωσης

Ορίζουμε 6 βάσεις χρηστών που αντιπροσωπεύουν τις παραπάνω 6 περιοχές με τις ακόλουθες παραμέτρους:

Χρήστης Βάσης	**Περιοχή**	**Ζώνη ώρας**	**Ώρες αιχμής (Τοπική ώρα)**	**Ώρες αιχμής (GMT)**	**Συνδεδεμένοι χρήστες στη διάρκεια των ωρών αιχμής**	**Συνδεδεμένοι χρήστες στη διάρκεια εκτός αιχμής ωρών**
UB1	0	GMT - 6.00	7.00 - 9.00pm	13:00-15:00	500.000	50.000
UB2	1	GMT - 4.00	7.00 - 9.00pm	15:00-17:00	150.000	15.000
UB3	2	GMT + 1.00	7.00 - 9.00pm	20:00-22:00	300.000	30.000
UB4	3	GMT + 6.00	7.00 - 9.00pm	01:00-03:00	200.000	20.000
UB5	4	GMT + 2.00	7.00 - 9.00pm	21:00-23:00	50.000	5.000
UB6	5	GMT+ 10.00	7.00 - 9.00pm	09:00-11:00	80.000	8.000

Για λόγους απλότητας, η εφαρμογή γίνεται τα βράδια μετά την εργασία για περίπου 2 ώρες. Ας υποθέσουμε ότι το 5% των εγγεγραμμένων χρηστών θα είναι online κατά τη διάρκεια του χρόνου αιχμής ταυτόχρονα και μόνο το 1/10 κατά τις ώρες εκτός αιχμής. Ας υποθέσουμε ότι κάθε χρήστης υποβάλλει ένα νέο αίτημα κάθε 5 λεπτά όταν γίνεται online.

Από την άποψη του κόστους φιλοξενίας το υποθετικό σχέδιο είναι:

Κόστος ανά VM ανά ώρα (1024Mb, 100MIPS)	$0,10
Κόστος ανά 1Gb μεταφοράς δεδομένων (from/to Internet)	$0,10

Άλλες παράμετροι δίνονται στον παρακάτω πίνακα:

Παράμετροι	**Τιμές**
VM Image Size	10000
VM Memory	1024Mb
VM Bandwidth	1000
Data Center – Architecture	X86
Data Center – OS	Linux
Data Center – VMM	Xen
Data Center – Number of Machines	20
Data Center – Memory per Machine	2048Mb
Data Center – Storage per machine	100000
Data Center – Available BW per Machine	10000
Data Center – Number of processors per machine	4
Data Center – Processor speed	100MIPS
Data Center – VM Policy	Time Shared
User Grouping Factor	1000
Request Grouping Factor	100
Executable Instruction Length	250

Delay Matrix values (in milliseconds):

Region/Region	0	1	2	3	4	5
0	25	100	150	250	250	100
1	100	25	250	500	350	200
2	150	250	25	150	150	200
3	250	500	150	25	500	500
4	250	350	150	500	25	500
5	100	200	200	500	500	25

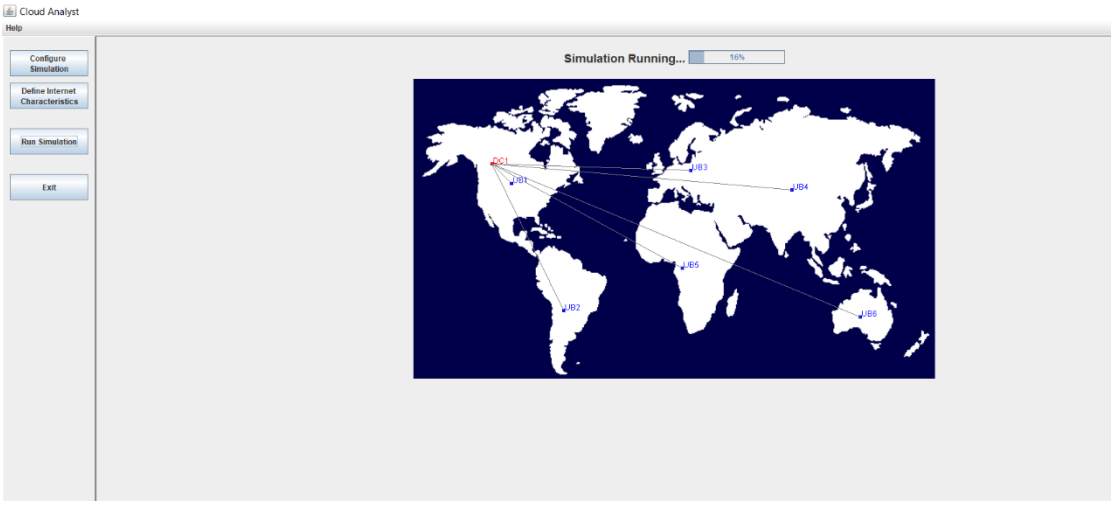
Bandwidth Matrix values (in Mbps):

Region/Region	0	1	2	3	4	5
0	2.000	1.000	1.000	250	1.000	1.000
1	1.000	800	1.000	500	1.000	1.000
2	1.000	1.000	2.500	150	1.000	1.000
3	1.000	1000	1.000	25	1.000	1.000
4	1.000	1.000	1.000	500	500	1.000
5	1.000	1.000	1.000	500	1.000	2.000

Τώρα ας προσπαθήσουμε να προσομοιώσουμε αυτή την εφαρμογή με το CloudAnalyst και να παρατηρήσουμε τη συμπεριφορά.

9.5 Senario 1 - Απλή εφαρμογή στο Web που φιλοξενείται σε ένα ενιαίο κέντρο δεδομένων

Όπως και με τις περισσότερες πραγματικές εφαρμογές Web, ας υποθέσουμε αρχικά ότι η εφαρμογή αναπτύσσεται σε ένα μία θέση, στην περιοχή 0 (Βόρεια Αμερική).



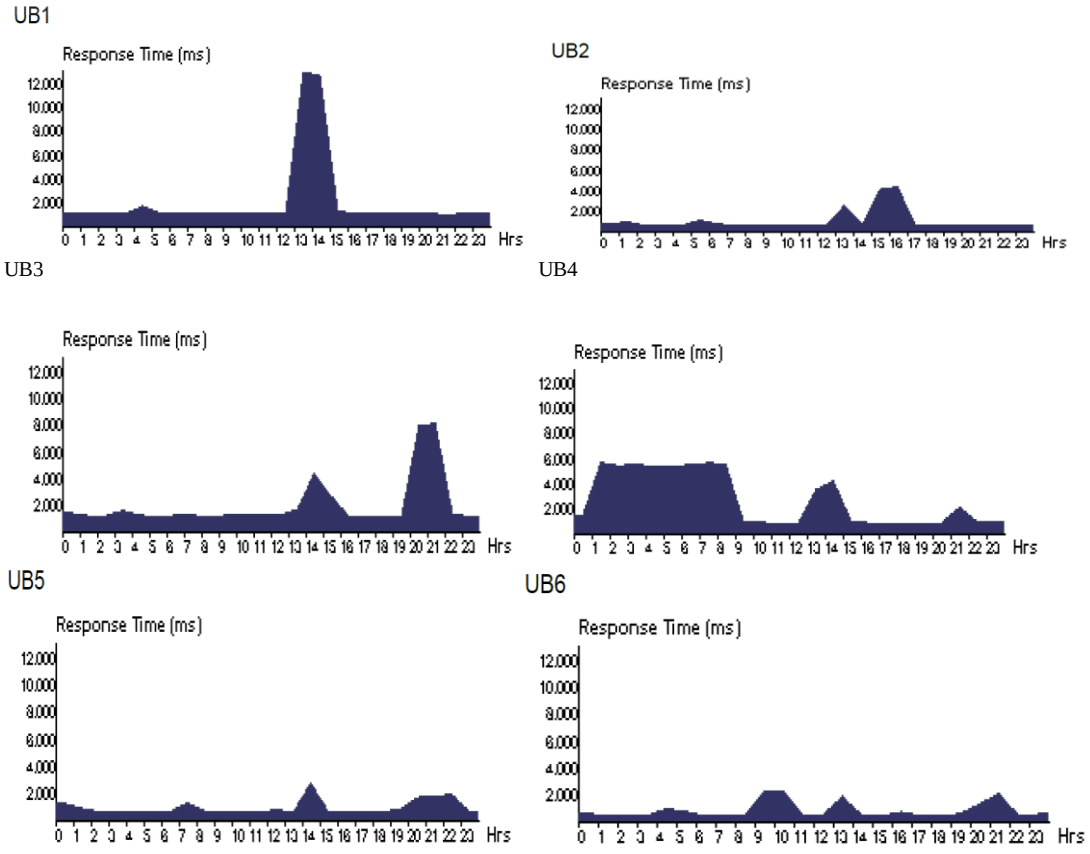
Υποθέτοντας ότι η εφαρμογή αναπτύσσεται σε 50 εικονικές μηχανές (με 1024Mb μνήμης σε κάθε VM που εκτελείται σε φυσικούς επεξεργαστές με ταχύτητα 100MIPS), ακολουθεί η έξοδος προσομοίωσης.

Overall Response Time:

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	4962,25	225,82	29799,47
Data Center processing time:	4667,22	50,65	29711,68

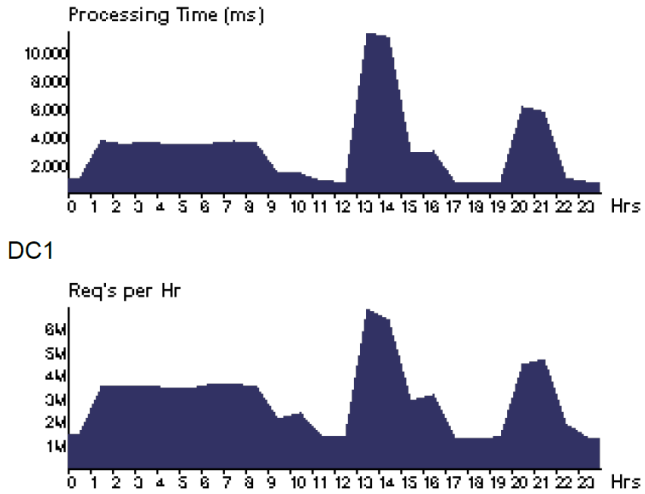
Πρέπει να λάβουμε υπόψη ότι αυτοί οι αριθμοί βασίζονται σε όλες τις παραμέτρους που αναφέρονται σε παραπάνω ενότητα και συνεπώς μπορεί να μην είναι ρεαλιστικές. Αλλά χρησιμεύουν ως μια καλή βάση για τη σύγκριση μεταξύ διαφόρων σεναρίων.

Οι χρόνοι απόκρισης κάθε βάσης χρήστη απεικονίζονται γραφικά ως εξής:



Οι αιχμές στους χρόνους απόκρισης μπορούν να φανούν σαφώς κατά τη διάρκεια της περιόδου αιχμής δύο ωρών και μπορεί να παρατηρηθεί πως τα φορτία αιχμής μιας βάσης χρηστών θα μπορούσαν να επηρεάσουν και άλλες βάσεις χρηστών. Για παράδειγμα, το UB1 έχει την κορυφή μεταξύ 13: 00-15: 00 GMT και όλες οι άλλες βάσεις χρηστών έχουν μικρές αιχμές κατά την ίδια περίοδο. Ωστόσο, ο αντίκτυπος ήταν μικρότερος, καθώς ο αριθμός των αιτημάτων που προέκυψαν από αυτές τις περιοχές κατά τη διάρκεια αυτής της περιόδου είναι μικρότερος.

Κατά τη διάρκεια αυτής της περιόδου 24 ωρών, ο μέσος χρόνος που απαιτείται από το κέντρο δεδομένων για την επεξεργασία ενός αιτήματος και ο αριθμός των αιτημάτων που υποβάλλονται σε επεξεργασία έχει ως εξής:



Όπως αναμένεται, αυτά τα δύο γραφήματα αντανακλούν το ένα το άλλο στενά και το πρώτο είναι πολύ κοντά σε μια υπέρθεση όλων των συγκεκριμένων γραφικών περιοχών παραπάνω. Ο συνολικός χρόνος επεξεργασίας στατιστικών στοιχείων για το κέντρο δεδομένων σε αριθμούς έχει ως εξής:

Data Center Request Servicing Times:

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	4667,22	50,65	29711,68

Έτσι, συνολικά, οι χρήστες μπορούν να περιμένουν χρόνο απόκρισης περίπου 5 δευτερολέπτων, αλλά κατά τη διάρκεια της ημέρας υπάρχουν αρκετές χρονικές περιόδους, όταν ο πραγματικός χρόνος απόκρισης μπορεί να αναμένεται να αυξηθεί πολύ υψηλότερα.

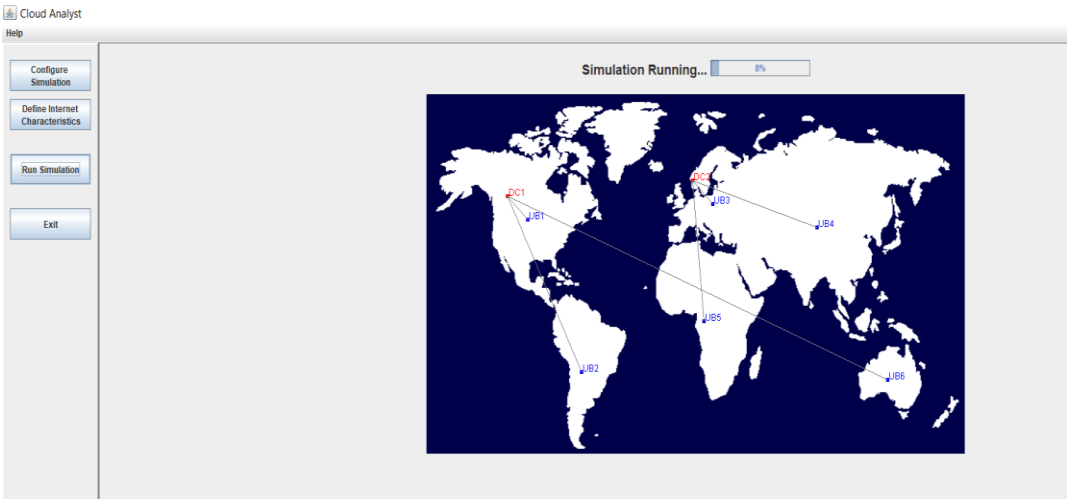
Ένας άλλος σημαντικός παράγοντας για τον ιδιοκτήτη της εφαρμογής θα ήταν το πόσο θα κόστιζε η λειτουργία αυτής της εφαρμογής για μια ημέρα. Το CloudAnalyst υπολογίζει ότι το ποσό αυτό είναι:

Data Center	VM Cost $	Data Transfer Cost $	Total $
DC1	120,05	707,68	827,73

9.6 Senario 2 - Εφαρμογή στο Web που φιλοξενείται σε πολλαπλά κέντρα δεδομένων στον κόσμο

Όταν οι εφαρμογές αυξάνονται σε δημοτικότητα στο διαδίκτυο, η πιο κοινή προσέγγιση για τη βελτίωση της ποιότητας των υπηρεσιών είναι η ανάπτυξη της εφαρμογής σε διάφορες τοποθεσίες σε όλο τον κόσμο. Για το δεύτερο σενάριο, διατηρώντας ταυτόχρονα τις βάσεις των χρηστών, προσθέτουμε ένα ακόμα κέντρο δεδομένων, στην περιοχή 2 (Ευρώπη). Για να διατηρηθεί το κόστος το ίδιο, οι 50 εικονικές μηχανές κατανέμονται 25 σε κάθε κέντρο δεδομένων.

9.6.1 Περίπτωση 1: Δύο κέντρα δεδομένων με 25 VM στο κάθε ένα

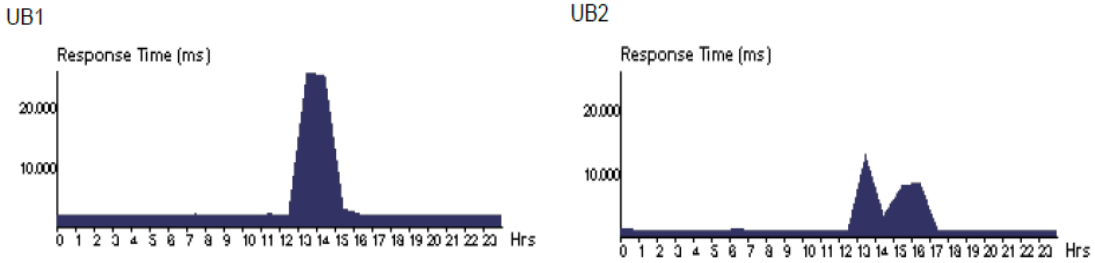


Στη συνέχεια, τα αποτελέσματα προσομοίωσης έχουν ως εξής.

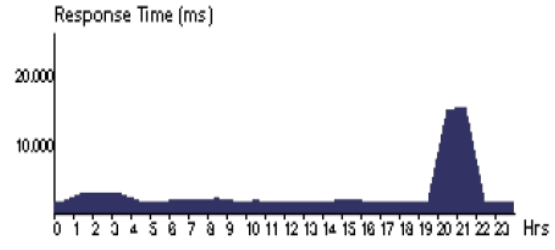
Overall response time and request processing times:

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	9306,01	170,35	59544,99
Data Center processing time:	9116,07	50,65	59458,75

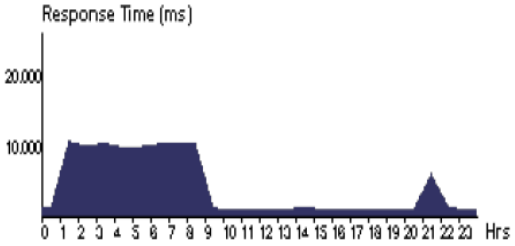
User Base Hourly Response Times:

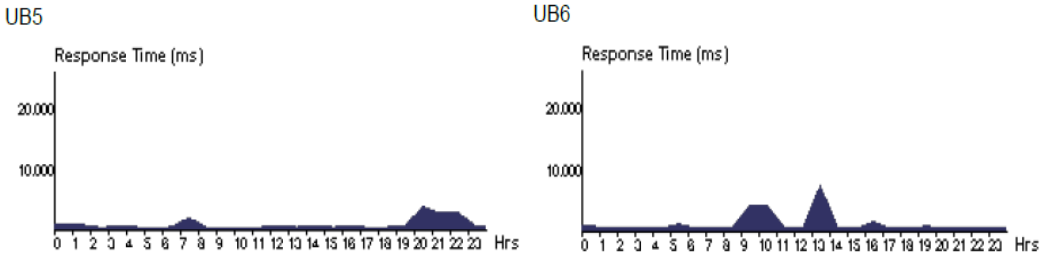


UB3



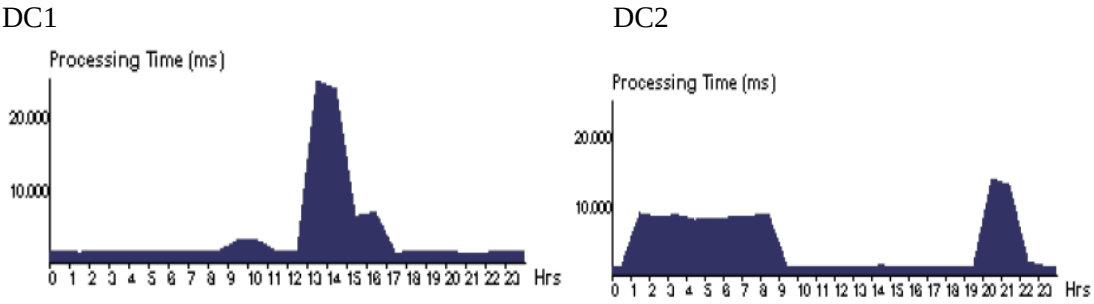
UB4



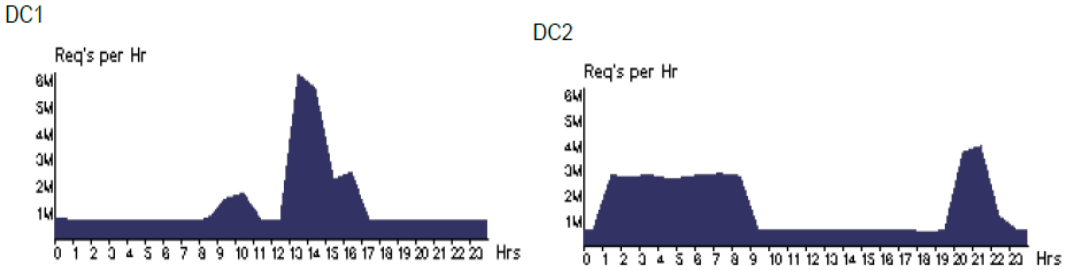


Το αποτέλεσμα μπορεί να μην είναι ακριβώς αυτό που αναμενόταν, φέρνοντας την υπηρεσία πιο κοντά στους χρήστες. Ο χρόνος απόκρισης έχει σχεδόν διπλασιαστεί. Το μοτίβο της κατανομής του χρόνου απόκρισης δεν άλλαξε πολύ, με την κύρια διαφορά να είναι ο αριθμός των μικρότερων κορυφών που μειώνονται, καθώς όλη η κίνηση δεν κατευθύνεται στο ίδιο κέντρο δεδομένων αυτή τη φορά.

Data Center Hourly Average Processing Times:



Data Center Hourly Loading:



Το συνολικό κόστος και τα επιμέρους :

Data Center	VM Cost $	Data Transfer Cost $	Total $
Total	**120,05**	**707,68**	**827,73**
DC1	60,03	384,11	444,14
DC2	60,03	323,57	383,59

Που είναι ακριβώς το ίδιο όπως πριν (όπως αναμενόταν).

Υπάρχουν δύο λόγοι για τους φτωχότερους χρόνους απόκρισης:
1. Τα κέντρα δεδομένων είναι υπερφορτωμένα κατά τους χρόνους αιχμής, όπως είναι εμφανές από το γεγονός ότι ο χρόνος επεξεργασίας αυξάνεται.
2. Η μείωση του αριθμού των εικονικών μηχανών κατά το ήμισυ κατά τη διάρκεια καθενός από αυτά τα φορτία αιχμής επιμηκύνει αποτελεσματικά τον χρόνο που απαιτείται για την επεξεργασία κατά περίπου διπλάσια σε κάθε κέντρο δεδομένων.

Γενικά, το πρώτο σημείο παραπάνω σημαίνει ότι η διαθέσιμη χωρητικότητα επεξεργασίας δεν επαρκεί. Στην ιδανική περίπτωση, το σημαντικό στοιχείο του συνολικού χρόνου απόκρισης θα πρέπει να είναι οι καθυστερήσεις του δικτύου, όπως φαίνεται να συνέβη όταν καταγράφηκε ο ελάχιστος χρόνος απόκρισης 170,35ms όταν ο ελάχιστος χρόνος επεξεργασίας ήταν 50,65ms. Ένας βασικός λόγος για αυτή τη βαριά φόρτωση είναι η χαμηλή βαθμολογία MIPS που επιλέχθηκε για τις παραμέτρους προσομοίωσης και την περιορισμένη μνήμη ανά VM. Αλλά επειδή αυτές οι ακραίες συνθήκες φωτίζουν τα πρότυπα δραστηριότητας, συνεχίζουμε να χρησιμοποιούμε αυτές τις ίδιες παραμέτρους και για τα επόμενα μερικά πειράματα.

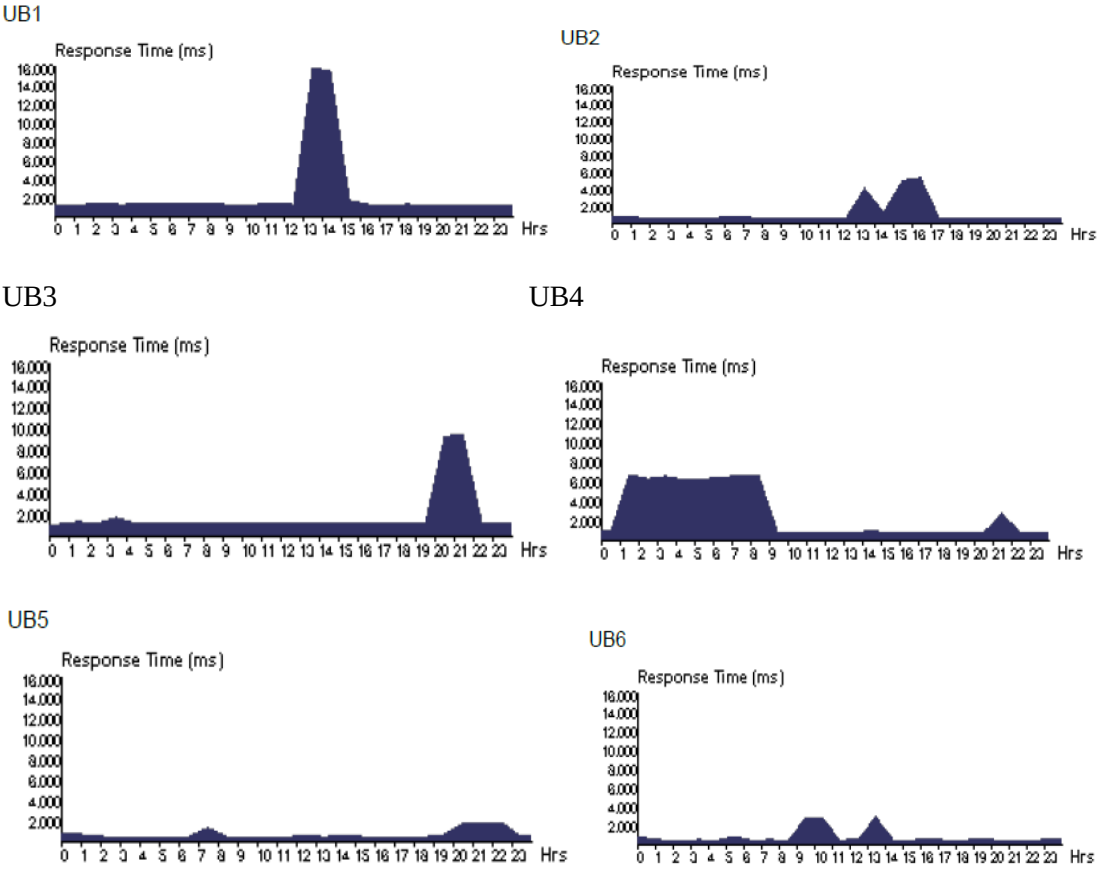
9.6.2 Περίπτωση 2: Δύο κέντρα δεδομένων με 50 VM στο καθένα

Με την αύξηση του αριθμού των εικονικών μηχανών σε κάθε κέντρο δεδομένων στα 50 ακόλουθα αποτελέσματα.

Overall response time and processing times:

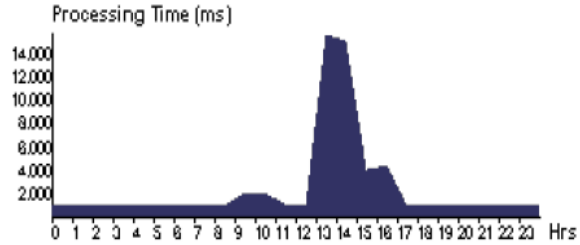
	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	5900,65	167,96	37299,24
Data Center processing time:	5721,44	50,60	37210,64

User Base Hourly Response Times:

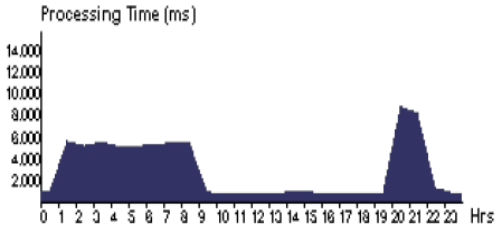


Η διαδικασία φόρτωσης του κέντρου δεδομένων για αυτό το σενάριο είναι ως εξής:

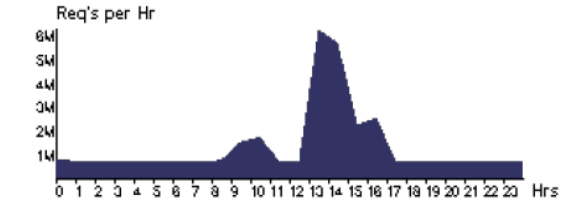
DC1



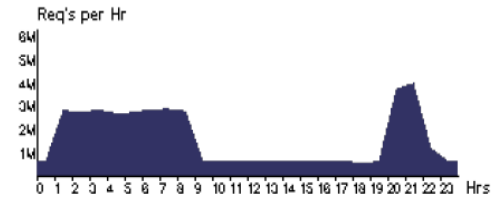
DC2



DC1



DC2



Data Center	VM Cost $	Data Transfer Cost $	Total $
Total	**192,08**	**707,68**	**899,76**
DC1	96,04	384,11	480,15
DC2	96,04	323,57	419,61

9.6.3 Περίπτωση 3: Δύο κέντρα δεδομένων με 50 VM το καθένα και φορτίο κατανομής κατά τη διάρκεια των ωρών αιχμής

Όπως φαίνεται από τα αποτελέσματα της προηγούμενης ενότητας, είναι σαφές ότι τα βαριά φορτία εμφανίζονται στα δύο κέντρα δεδομένων σε διαφορετικές χρονικές περιόδους. Τι θα συνέβαινε, αν κάποιο μέρος του φορτίου σε οποιαδήποτε χρονική στιγμή διοχετεύεται από το πιο φορτωμένο κέντρο δεδομένων στο μικρότερο φορτωμένο κέντρο δεδομένων; Αυτό το σενάριο μπορεί να αναλυθεί χρησιμοποιώντας την πολιτική μεσάζων υπηρεσιών βελτιστοποιημένου χρόνου απόκρισης στο CloudAnalyst και τα αποτελέσματα που λαμβάνονται είναι τα εξής

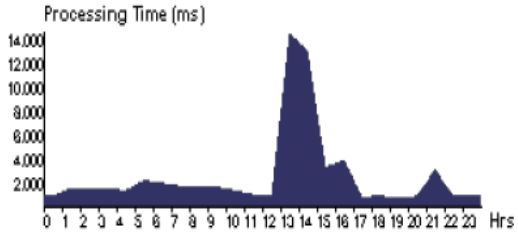
Overall response time and processing times:

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	5034,85	168,65	37297,71
Data Center processing time:	4816,43	50,10	37213,93

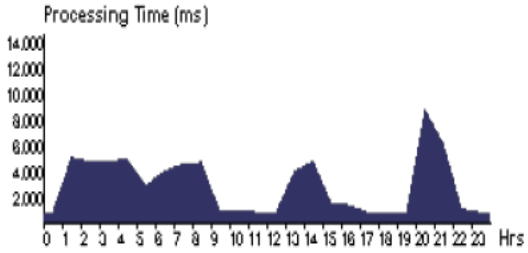
Η κατανομή χρόνου απόκρισης σε περιφέρειες δεν δείχνει καμία σημαντική διαφορά σε σχέση με την προηγούμενη περίπτωση, αλλά τα μοντέλα φόρτωσης κέντρου δεδομένων εμφανίζουν σημαντικές αλλαγές.

Data Center Hourly Average Processing Times:

DC1

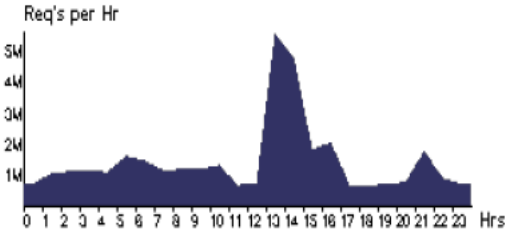


DC2

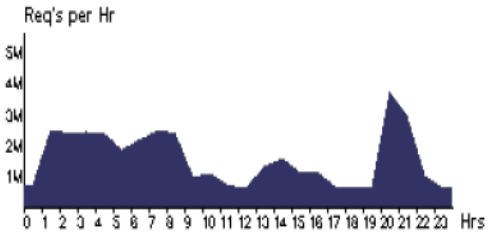


Data Center Hourly Loading:

DC1



DC2



Όπως φαίνεται στα γραφήματα, κάποια ποσότητα του μέγιστου φορτίου έχει μεταφερθεί στο μικρότερο φορτωμένο κέντρο δεδομένων. Το ποσοστό βελτίωσης που επιτυγχάνεται από την κατανομή φορτίου εξαρτάται σε μεγάλο βαθμό από τον αλγόριθμο που χρησιμοποιείται για την κατανομή φορτίου.

9.6.4 Περίπτωση 4: Εφαρμογή στο Web που φιλοξενείται σε 3 κέντρα δεδομένων με 50VM το καθένα

Σε αυτή την εφαρμογή, θα προσθέσουμε ένα τρίτο κέντρο δεδομένων στην περιοχή 3 (Ασία) με κατανομή επιπλέον 50 VM και θα παρατηρήσουμε πως συμπεριφέρεται η προσομοίωση. Χρησιμοποιούμε την Optimise Response Time, όπως στην περίπτωση των 2 κέντρων δεδομένων.

Overall response time and processing times:

	Avg (ms)	**Min (ms)**	**Max (ms)**
Overall response time:	4368,29	163,94	37301,50
Data Center DC1 processing time:	5653,18	50,10	37210,37
Data Center DC2 processing time:	3077,28	50,19	21227,88
Data Center DC3 processing time:	2726,36	50,46	18369,22

Δεν υπάρχουν εκπλήξεις σε αυτά τα αποτελέσματα. ο συνολικός χρόνος απόκρισης έχει βελτιωθεί. Το DC1 δίνει το χειρότερο μέσο χρόνο απόκρισης καθώς αντιμετωπίζει το υψηλότερο φορτίο.

9.6.5 Περίπτωση 5: Η εφαρμογή Web φιλοξενείται σε 3 κέντρα δεδομένων με 75, 50, 25 VM στο καθένα

Αφού, ολοκληρωθεί η ρύθμιση του αριθμού των VM σε DC1 - 75, DC2 - 50 και DC3 - 25 δίνει τα ακόλουθα.

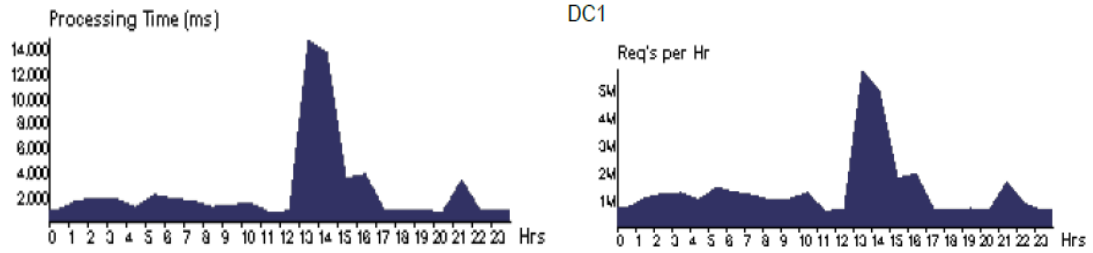
Overall response time and processing times:

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	4849,16	166,78	37301,50
Data Center DC1 processing time:	5790,10	50,00	37210,37
Data Center DC2 processing time:	3187,81	50,28	21227,88
Data Center DC3 processing time:	4354,46	50,33	28858,42

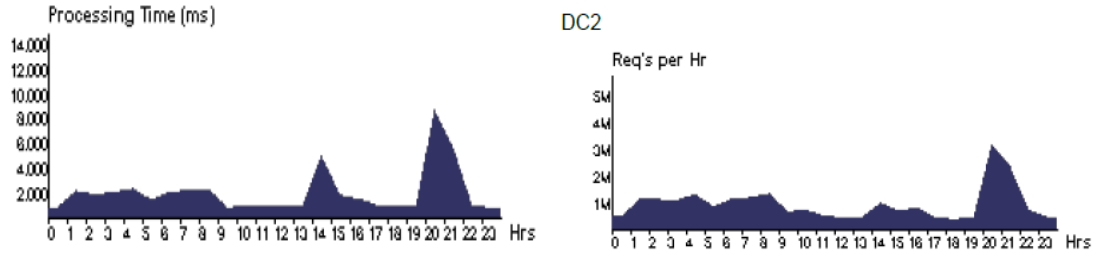
Τώρα ο χρόνος απόκρισης βελτιώνεται περαιτέρω, αλλά το πιο σημαντικό αποτέλεσμα είναι ότι ο μέσος χρόνος επεξεργασίας σε κάθε κέντρο δεδομένων είναι αρκετά παρόμοιος. Αυτό αποδεικνύεται σαφέστερα από τα διαγράμματα χρόνου επεξεργασίας του κέντρου δεδομένων.

Data Center Hourly Average Processing Times Data Center Hourly Loading

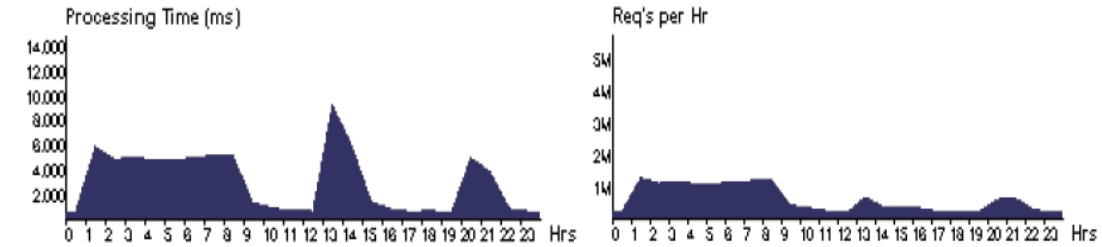
DC1



DC2



DC3



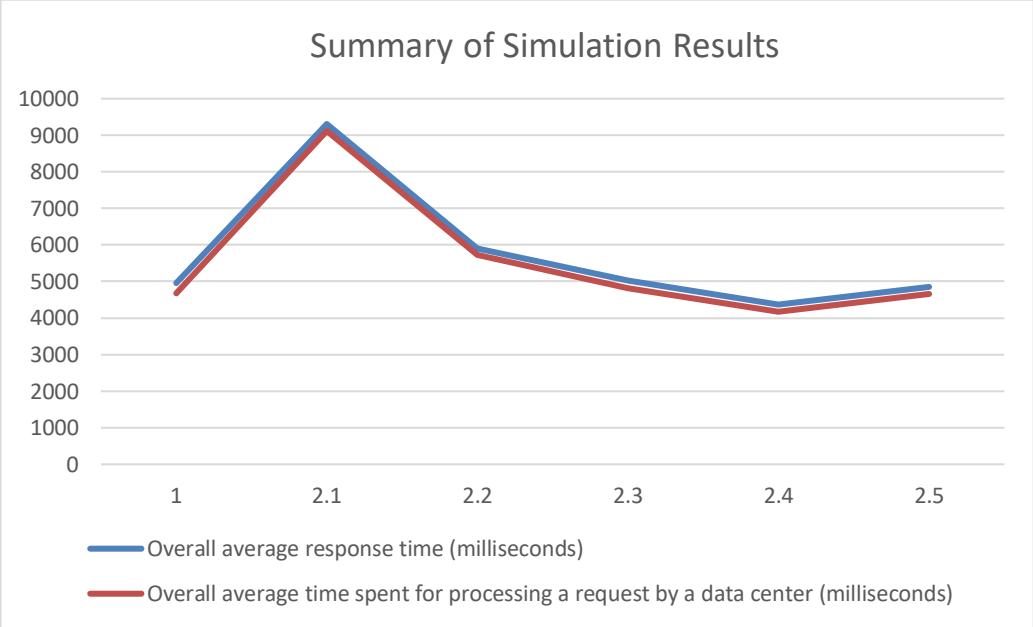
Από τα γραφήματα μπορεί να φανεί ότι παρόλο που τα διαφορετικά κέντρα δεδομένων αντιμετωπίζουν αιχμή κυκλοφορίας σε διαφορετικές χρονικές περιόδους και ο πραγματικός αριθμός αιτήσεων που δημιουργούνται κατά τις περιόδους αυτές ποικίλλει σημαντικά, ο λόγος 75-50-25 των VMs κατάφερε να διατηρήσει τη μέγιστη επεξεργασία χρόνο γύρω από το σήμα των 6 δευτερολέπτων και στα τρία κέντρα δεδομένων.

9.7 Σύνοψη αποτελεσμάτων προσομοίωσης

Έτσι, μετά τη σειρά των πειραμάτων, η περίληψη αποτελεσμάτων για τους χρόνους απόκρισης μπορεί να εκφραστεί ως εξής:

Σενάριο	Περιγραφή Σεναρίου	Overall average response time (milliseconds)	Overall average time spent for processing a request by a data center (milliseconds)
1	1 data center with 50 VMs	4962,25	4667,22
2.1	2 data centers with 25 VMs each	9306,01	9116,07
2.2	2 data centers with 50 VMs each	5900,65	5721,44
2.3	2 data centers with 50 VMs each with peak load sharing	5034,85	4816,43
2.4	3 data centers with 50 VMs each with peak load sharing and throttling	4368,29	4171,38
2.5	3 data centers with 75, 50, 25 VMs, with peak load sharing and throttling	4849,16	4655,12

Γραφικά μπορεί να παρουσιαστεί ως εξής:



9.8 Κύριες Παρατηρήσεις από τα Αποτελέσματα

Παρακάτω βλέπουμε τις κύριες παρατηρήσεις με βάση τα αποτελέσματα που βγήκαν:

- Η μεταφορά της υπηρεσίας πιο κοντά στους χρήστες βελτιώνει την ποιότητα της υπηρεσίας (χρόνος απόκρισης σε αυτή την περίπτωση)
- Η ποιότητα της υπηρεσίας μπορεί να βελτιωθεί περαιτέρω με την εφαρμογή εξισορρόπησης φορτίου σε επίπεδο εφαρμογής σε κέντρα δεδομένων και σε επίπεδο εικονικών μηχανών στα κέντρα δεδομένων. Ωστόσο, τα επίπεδα βελτίωσης που επιτυγχάνονται εξαρτώνται σε μεγάλο βαθμό από τους αλγόριθμους εξισορρόπησης φορτίου που χρησιμοποιούνται.
- Για να είναι αποτελεσματικές οι βελτιώσεις αυτές, απαιτείται επαρκής χωρητικότητα στα κέντρα δεδομένων για να ικανοποιηθεί η αιχμή της ζήτησης.
- Από την άλλη πλευρά, αν η μέγιστη χωρητικότητα κατανέμεται καθ' όλη τη διάρκεια, θα υπάρχει ένα σημαντικό μέρος του χρόνου κατά τον οποίο δεν χρησιμοποιείται πλήρως αυτή η ικανότητα. Αυτό δεν είναι οικονομικό.

Έτσι συνολικά, αυτά τα αποτελέσματα φαίνεται να προτείνουν μια άλλη προσέγγιση που θα μπορούσε να βελτιστοποιήσει την απόδοση, διατηρώντας παράλληλα το κόστος.

Κεφάλαιο 10: Συμπεράσματα – Μελλοντική Έρευνα

Ένα στοιχείο που συνδέεται άμεσα με την ανάπτυξη της τεχνολογίας και την εξέλιξη της είναι η έρευνα, καθώς αποτελεί συνδετικό κρίκο της σύλληψης μιας ιδέας με την υλοποίηση της. Τα μεγάλα δεδομένα βρίσκονται σε έκρηξη τα τελευταία χρόνια καθώς συνδέονται ολοένα και περισσότερο με την τεχνολογία. Νέες ιδέες αναπτύσσονται συνεχώς και υλοποιούνται ώστε να υπάρχει εξέλιξη και αντιμετώπιση τυχόν προβλημάτων που υπάρχουν. Η ασφάλεια των μεγάλων δεδομένων αποτελεί την βάση αυτής της έρευνα καθώς κάθε χρόνο, η προστασία ιδιωτικών και εμπιστευτικών πληροφοριών κερδίζει όλο και περισσότερη προσοχή, διότι η ασφάλεια είναι η πιο υψηλή προτεραιότητα. Σήμερα, οι οργανισμοί συλλέγουν, επεξεργάζονται και αποθηκεύουν τεράστιες ποσότητες πληροφοριών, όπου κυρίως σκοπός είναι η διασφάλιση της ασφάλειάς τους. Η έλλειψη ασφάλειας των δεδομένων, όσον αφορά τα μεγάλα δεδομένα μπορεί να οδηγήσει σε μεγάλες οικονομικές απώλειες για μια εταιρεία, λόγω κακής ασφάλειας. Ιδιαίτερα, σε εφαρμογές που στηρίζονται πάνω στο νέφος οι κίνδυνοι είναι μεγαλύτεροι καθώς μπορεί να χαθεί ή να κλαπεί τεράστιος όγκος δεδομένων των χρηστών και των επιχειρήσεων.

Για την ασφάλεια των μεγάλων δεδομένων έγινε έρευνα σχετικά με τις τεχνικές και τις μεθόδους πρόληψης για τυχόν απώλειες. Έπειτα, έγινε η σύγκριση μεταξύ δυο μεθόδων ασφάλειας δεδομένων, η οποία σχετίζεται άμεσα με την κρυπτογράφηση. Η κρυπτογράφηση αποτελεί βασική μέθοδο ασφάλειας των δεδομένων. Μέσα από αυτόν τον τρόπο παρουσιάστηκαν οι διαφορές και οι ομοιότητες μεταξύ των δυο προτεινόμενων μεθόδων και πως λειτουργεί η κάθε ως προς την προστασία. Επίσης, εξετάστηκε και το cloud αφού σχετίζεται άμεσα με την αποθήκευση δεδομένων τα όποια είναι σημαντικά.

Με την ταχεία πρόοδο των τεχνολογιών Cloud, υπάρχει μια νέα ανάγκη για εργαλεία για τη μελέτη και να αναλύσει τα οφέλη της τεχνολογίας και τον καλύτερο τρόπο εφαρμογής της τεχνολογίας σε μεγάλης κλίμακας εφαρμογών. Οι υπηρεσίες τύπου "νέφους" είναι η κοινωνική δικτύωση. Το CloudAnalyst είναι ένα νέο εργαλείο που αναπτύχθηκε στα πλαίσια της προσομοίωσης, για αυτό και χρησιμοποιήθηκε για να μοντελοποιήσει μια τυπική εφαρμογή τύπου κοινωνικής δικτύωσης με υψηλή χρήση.

Επιπλέον, η προσομοίωση μας οδήγησε σε πολλές νέες ιδέες που μπορούν να χρησιμοποιηθούν για τη βελτίωση της ποιότητας, και με δυνατότητα εισαγωγής δυναμικής διαμορφωσιμότητας μέσω ενός global Cloud Service Broker, αυξάνοντας ή μειώνοντας το μέγεθος της εφαρμογής σε διαφορετικά τοποθεσίες ανάλογα με το φορτίο. Η προσομοίωση μιας μεγάλης κλίμακας εφαρμογής στο Διαδίκτυο είναι μια πολύπλοκη εργασία, που το CloudAnalyst δεν είναι μια ολοκληρωμένη λύση σε όλες αυτές τις ανάγκες προσομοίωσης. Αυτή τη στιγμή είναι το πρώτο βήμα ενός εργαλείου και μια προσέγγιση για τη μελέτη αυτού του είδους εφαρμογών με προσομοίωση καθώς αναμένεται να εξελιχθεί με την πάροδο του χρόνου βελτιώνοντας την ποιότητα του ανάλυση στην πορεία.

Σχεδόν όλα τα ζητήματα ασφάλειας δεδομένων προκαλούνται από την έλλειψη αποτελεσματικών μέτρων που παρέχονται από λογισμικό προστασίας από ιούς και τείχη προστασίας. Αυτά τα συστήματα αναπτύχθηκαν για να προστατεύσουν το περιορισμένο πεδίο των πληροφοριών που αποθηκεύονται, αλλά τα Big Data ξεπερνούν τις κλασσικές και απλές μεθόδους αποθήκευσης. Οι προτάσεις για μελλοντική έρευνα κατευθύνονται στους τρόπους βελτίωσης της ασφάλειας των Big Data που στηρίζονται στην επέκταση της προστασίας από ιούς. Υπάρχουν πολλοί τρόποι προστάσιας, που προσφέρουν ποικίλες λύσεις, παρέχουν καλύτερη άμυνα κατά των απειλών για την ασφάλεια των Big Data. Ορισμένες πρόσθετες συστάσεις για την ενίσχυση της ασφάλειας των Μεγάλων Δεδομένων επικεντρώνονται στην ασφάλεια των εφαρμογών, αντί της ασφάλειας των συσκευών, στην απομονώση των συσκευών και διακομιστών που περιέχουν κρίσιμα δεδομένα και στην εισαγωγή πληροφορίων ασφαλείας και διαχείριση συμβάντων σε πραγματικό χρόνο.

Βιβλιογραφία

Alexandru Adrian Tole (2013), Big Data Challenges, Database Systems Journal vol. IV, no. 3/2013.

Andreas P. Plageras, Christos Stergiou, George Kokkonis, Kostas E. Psannis, Yutaka Ishibashi, Byung-Gyu Kim, and Brij Gupta, Efficient Large-scale Medical Data (eHealth Big Data) Analytics in Internet of Things, International Workshop on Internet of Things and Smart Services in the 19th IEEE Conference on Business Informatics, Thessaloniki, Greece 24-26 July, 2017.

Andreu-Perez, J., Poon, C., Merrifield, R., Wong, S. and Yang, G, (2015), Big Data for Health, IEEE *Journal of Biomedical and Health Informatics*, 19(4), pp.1193-1208.

Evangelos Balasas, Kostas E. Psannis, and Manos Roumeliotis, Performance Evaluation of Routing Protocols for BIG Data applications, Springer Proceedings in Business and Economics, 2018.

Bagheri. H, and Abdullah Shaltooki.A (2015), Big Data: Challenges, Opportunities and Cloud Based Solutions, International Journal of Electrical and Computer Engineering (IJECE), Vol. 5, No. 2, pp. 340~343.

Bertino, E. (2015), Big Data - Security and Privacy, 2015 *IEEE International Congress on Big Data*.

Bi, Z. and Cochran, D. (2017), *Big data analytics with applications*.

Bi, S., Zhang, R., Ding, Z. and Cui, S. (2015), Wireless communications in the era of big data, *IEEE Communications Magazine*, 53(10), pp.190-199.

Bou-Harb, E. Husak, M., Debbabi M. and Assi C. (2017). Big Data Sanitization and Cyber Situational Awareness: A Network Telescope Perspective. *IEEE Transactions on Big Data*, pp.1-1.

Boyd, d. and Crawford, K. (2012), CRITICAL QUESTIONS FOR BIG DATA, *Information, Communication & Society*, 15(5), pp.662-679**.**

C. Stergiou, Kostas. E. Psannis, A. P. Plageras, Y. Ishibashi, B.-G. Kim, Algorithms for efficient digital media transmission over IoT and cloud networking, Journal of Multimedia Information System, vol. 5, no. 1, pp. 27-34, March 2018.

Cardenas. A., Manadhata, P. and Rajan, S. (2013), Big Data Analytics for Security, *IEEE Security & Privacy*, 11(6), pp.74-76.

Chen, M., Mao, S. and Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), pp.171-209.

Christos Stergiou, Kostas E. Psannis, Andreas P. Plageras, Theofanis Xifilidis, and B.B. Gupta, Security and Privacy of Big Data for Social Networking Services in Cloud, in Proceedings of IEEE conference on Computer Communications (IEEE INFOCOM 2018), Workshop on CCSNA: Cloud Computing Systems, Networks, and Applications, 15-20 April 2018, Honolulu, HI, USA.

Christos Stergiou, Andreas P. Plageras, Kostas E. Psannis, and Brij B. Gupta, Secure Machine Learning Scenario from Big Data in Cloud Computing via Internet of Things, Book chapter of Multimedia Information Systems, Springer, September 2018.

Christos Stergiou, Kostas E. Psannis, B.B. Gupta, and Yutaka Ishibashi, Security, Privacy & Efficiency of Sustainable Cloud Computing for Big Data & IoT, Sustainable Computing, Informatics and Systems, Elsevier, June 2018.

Codd, E. (1970), A relational model of data for large shared data banks, *Communications of the ACM*, 13(6), pp.377-387.

Costa.F, (2014), Big data in biomedicine, *Drug Discovery Today*, 19(4), pp.433-440.

Cuzzocrea, A., Saccà, D. and Ullman, J. (2013), Big data. *Proceedings of the 17th International Database Engineering & Applications Symposium on - IDEAS '13*.

Demchenko, Y., Grosso, P., de Laat, C. and Membrey, P. (2013), Addressing big data issues in Scientific Data Infrastructure, *2013 International Conference on Collaboration Technologies and Systems (CTS)*.

Dolev S., FlorissiP., Gudes, E., Sharma, S. and Singer, I. (2017). A Survey on Geographically Distributed Big-Data Processing using MapReduce. *IEEE Transactions on Big Data*, pp.1-1.

Dubey, A. and Srivastava, S. (2017), *A Major Threat to Big Data*.

Evangelos Balasas, Kostas Psannis, and Manos Roumeliotis, Performance Evaluation of Routing Protocols for BIG Data application, 6th International Symposium & 28th National Conference on Operational Research, "OR in the digital era - ICT challenges", Thessaloniki, University of Macedonia, Geece, June 8-10, 2017.

Emmanuel, W. (2009). Impact of Multiencryption in Data Security. *International Journal of Computer Theory and Engineering*, pp.571-576.

Fan, J., Han, F. and Liu, H. (2014), Challenges of Big Data analysis, *National Science Review*, 1(2), pp.293-314.

González-Bailón, S. (2013), Social science in the era of big data, Policy *& Internet*, 5(2), pp.147-160.

Goudarzi, M. (2017). Heterogeneous Architectures for Big Data Batch Processing in MapReduce Paradigm. *IEEE Transactions on Big Data*, pp.1-1.

Gupta, A., Verma, A., Kalra, P. and Kumar, L, (2014), Big Data: A security compliance model, *2014 Conference on IT in Business, Industry and Government (CSIBIG)*.

Gupta.B, Arachchilage. N, and Psannis.K, (2017), Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommunication Systems*.

Han Hu, Yonggang Wen, Tat-Seng Chua and Xuelong Li (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access. 2, pp.652-687.

Hashem, I., Yaqoob, I., Anuar, N., Mokhtar, S., Gani, A. and Ullah Khan, S. (2015), The rise of "big data" on cloud computing: Review and open research issues, *Information Systems*, 47, pp.98-115.

He Y., Yu F., Zhao N., Yin H., Yao, H. and Qiu, R. (2016). Big Data Analytics in Mobile Cellular Networks. *IEEE Access* 4, pp.1985-1996.

Hu, C., Li, W., Cheng, X., Yu, J., Wang, S. and Bie, R. (2017) A Secure and Verifiable Access Control Scheme for Big Data Storage in Clouds. *IEEE Transactions on Big Data*. pp.1-1.

Ji, C., Li, Y., Qiu, W., Awada, U. and Li, K. (2012), Big Data Processing in Cloud Computing Environments, *2012 12th International Symposium on Pervasive Systems, Algorithms and Networks*.

Jin. X., Wah, B., Cheng, X. and Wang, Y, (2015), Significance and Challenges of Big Data Research, *Big Data Research*, 2(2), pp.59-64.

Katal, A., Wazid, M. and Goudar, R. (2013), Big data: Issues, challenges, tools and Good practices. *2013 Sixth International Conference on Contemporary Computing (IC3)*.

Kim. S, Kim.N and Chung.T, (2013), Attribute Relationship Evaluation Methodology for Big Data Security. *2013 International Conference on IT Convergence and Security (ICITCS)*.

Kim. G, Trimi.S and Chung.J, (2014), Big-data applications in the government sector, *Communications of the ACM*, 57(3), pp.78-85.

Kitchin, R. and Lauriault,T. (2014), Small data in the era of big data, *GeoJournal*, 80(4), pp.463-475.

Kostas E, Psannis, Christos Stergiou, and BB Gupta, Advanced Media-based Smart Big Data on Intelligent Cloud Systems IEEE Transactions on Sustainable Computing, 2018 (Date of Publication: 21 March 2018)

Kshetri,N.(2014), Big data′ s impact on privacy, security and consumer welfare, *Telecommunications Policy*, 38(11), pp.1134-1145.

K.U, J. and M.David, J. (2014), Issues, Challenges and Solutions: Big Data Mining, *Computer Science & Information Technology (CS & IT)*.

Kupwade Patil. H, and Seshadri.R, (2014), Big Data Security and Privacy Issues in Healthcare, *2014 IEEE International Congress on Big Data*.

Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan and Yong Ren (2014). Information Security in Big Data: Privacy and Data Mining. *IEEE Access*. 2, pp.1149-1176.

Li. Y, Gai.K, Ming.Z, Zhao. H, and Qiu.M, (2016), Intercrossed Access Controls for Secure Financial Services on Multimedia Big Data in Cloud Systems, *ACM Transactions on Multimedia Computing, Communications, and Applications*, 12(4s), pp.1-18.

Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G. and Guo, S. (2016). Protection of Big Data Privacy, *IEEE Access*, 4, pp.1821-1834.

Mahmood, T. and Afzal, U. (2013), Security Analytics: Big Data Analytics for cybersecurity: A review of trends, techniques and tools, *2013 2nd National Conference on Information Assurance (NCIA)*.

Mahrt, M. and Scharkow, M. (2017), *The Value of Big Data in Digital Media Research*.

Manogaran. G, Thota.C and Kumar. M, (2016), MetaCloudDataStorage Architecture for Big Data Security In Cloud Computing, *Procedia Computer Science*, 87, pp.128-133.

Matturdi. B, Zhou. X, Li.S and Lin.F, (2014), Big Data security and privacy: A review, *China Communications*, 11(14), pp.135-145.

Meeker, W. and Hong, Y. (2017), Reliability *Meets Big Data: Opportunities and Challenges*.

Papadopoulos.N and Psannis. K, (2016), Sequential Multiple LSB methods and real-time data hiding: variations for Visual Cryptography ciphers, *Journal of Real-Time Image Processing*.

Philip Chen.C, and Zhang.C, (2014), Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, pp.314-347.

Plageras, A., Psannis, K., Gupta, B., Stergiou, C., Kim, B. and Ishibashi, Y. (2017). Solutions for inter-connectivity and security in a smart hospital building. *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*.

Plageras, A., Stergiou, C., Kokkonis, G., Psannis, K., Ishibashi, Y., Kim, B. and Gupta, B. (2017). Efficient Large-scale Medical Data (eHealth Big Data) Analytics in Internet of Things. *2017 IEEE 19th Conference on Business Informatics (CBI)*.

Plageras, A. and Psannis, K. (2017). Algorithms for Big Data Delivery over the Internet of Things. *2017 IEEE 19th Conference on Business Informatics (CBI)*.

Plageras, A., Psannis, K., Ishibashi, Y. and Kim, B. (2016). IoT-based surveillance system for ubiquitous healthcare. *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*.

Psannis, K., Stergiou, C. and Gupta, B. (2018). Advanced Media-based Smart Big Data on Intelligent Cloud Systems. *IEEE Transactions on Sustainable Computing*, pp.1-1.

,Rekha A. (2014). A SURVEY ON ENCRYPTION ALGORITHMS FOR DATA SECURITY. *International Journal of Research in Engineering and Technology*, 03(12), pp.131-134.

Rubinstein, I. (2017), *Big Data: The End of Privacy or a New Beginning?*.

Sagiroglu, S. and Sinanc, D. (2013), Big data: A review, 2013 International Conference on Collaboration Technologies and Systems (CTS).

Sapountzi.Aand Psannis. K, (2016), Social networking data analysis tools & challenges,*Future Generation Computer Systems*.

Sedayao, J., Bhardwaj, R. and Gorade, N. (2014), Making Big Data, Privacy, and Anonymization Work Together in the Enterprise: Experiences and Issues, *2014 IEEE International Congress on Big Data*.

Sivinski.G, Okuliar.A and Kjolbye.L, (2017), Is big data a big deal? A competition law approach to big data, *European Competition Journal*, pp.1-29.

Smith, M., Szongott, C., Henne, B. and von Voigt, G. (2012), Big data privacy issues in public social media, *2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST)*.

Stergiou, C., Psannis, K., Plageras, A., Kokkonis, G. and Ishibashi, Y. (2017). Architecture for security monitoring in IoT environments. *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*.

Stergiou.C, and Psannis. K, (2017), Efficient and secure BIG data delivery in Cloud Computing, Multimedia *Tools and Applications*.

Stergiou, C., Psannis, K., Gupta, B. and Ishibashi, Y. (2018). Security, privacy & efficiency of sustainable Cloud Computing for Big Data & IoT. *Sustainable Computing: Informatics and Systems*.

Stergiou.C, Psannis. K, Kim. B and Gupta.B, (2017), *Secure integration of IoT and Cloud Computing*.

Stergiou, C. and Psannis, K. (2016). Recent advances delivered by Mobile Cloud Computing and Internet of Things for Big Data applications: a survey. *International Journal of Network Management*, 27(3), p.e1930.

Strang.K, and Sun. Z, (2017), Big Data Paradigm: What is the Status of Privacy and Security?, Annals *of Data Science*, 4(1), pp.1-17.

Su, Z., Xu, Q. and Qi, Q. (2016), Big data in mobile social networks: a QoE-oriented framework, IEEE *Network*, 30(1), pp.52-57.

Tang, M., Alazab, M. and Luo, Y. (2017). Big Data for Cybersecurity: Vulnerability Disclosure Trends and Dependencies. *IEEE Transactions on Big Data*, pp.1-1.

Tankard, C. (2012), Big data security, *Network Security*, 2012(7), pp.5-8.

Tankard, C. (2017), Encryption as the cornerstone of big data security, *Network Security*, 2017(3), pp.5-7.

Wang S., Bonomi L., Dai, W., Chen, F., Cheung, C., Bloss, C., Cheng, S. and Jiang, X. (2016). Big Data Privacy in Biomedical Research. *IEEE Transactions on Big Data*, pp.1-1.

V. Memos and Kostas E. Psannis, Ένα Νέο Πρότυπο Ασφαλείας Βασισμένο στο Cloud Computing για Αποτελεσματική Ανίχνευση Απειλών, 2nd Student Conf. of the Dept. of Applied Informatics, University of Macedonia (FSTEP 2015), Dec. 2015.

V. Memos and Kostas E. Psannis, A New Methodology based on Cloud Computing for Efficient Virus Detection, New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, Lecture Notes in Electrical Engineering Volume 312, , pp 37-47. 2015.

Vasileios A. Memos and Kostas E. Psannis, Encryption algorithm for efficient transmission of HEVC media, Journal of Real-Time Image Processing, , Volume 12, Issue 2, pp 473–482, August 2016.

Wang, K., Xu, C., Zhang, Y., Guo, S. and Zomaya, A. (2017). Robust Big Data Analytics for Electricity Price Forecasting in the Smart Grid. *IEEE Transactions on Big Data*. pp.1-1.

Wang X., Yang, L., Liu, H. and Deen, M. (2017). A Big Data-as-a-Service Framework: State-of-the-art and Perspectives. *IEEE Transactions on Big Data*, pp.1-1.

΄Williamson, B. (2017), Who owns educational theory? Big data, algorithms and the expert power of education data science, *E-Learning and Digital Media*, p.204275301773123.

Wu J., Ota K., Dong, M., Li, J. and Wang, H. (2016). Big Data Analysis based Security Situational Awareness for Smart Grid. *IEEE Transactions on Big Data*, pp.1-1.

Wu X., Wu T., Khan, M Ni, Q. and Dou, W. (2017). Game Theory Based Correlated Privacy Preserving Analysis in Big Data. *IEEE Transactions on Big Data*. pp.1-1.

Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding (2014), Data mining with big data, *IEEE Transactions on Knowledge and Data Engineering*, 26(1), pp.97-107.

Xinhua Dong, Ruixuan Li, Heng He, Wanwan Zhou, Zhengyuan Xue and Hao Wu (2015). Secure sensitive data sharing on a big data platform. *Tsinghua Science and Technology*, 20(1), pp.72-80.

Xue-Wen Chen and Xiaotong Lin (2014), Big Data Deep Learning: Challenges and Perspectives, *IEEE Access*, 2, pp.514-525.

Yang.C, Liu.J, Chen.S and Lu. H, (2017), Implementation of a Big Data Accessing and Processing Platform for Medical Records in Cloud, *Journal of Medical Systems*, 41(10).

Ye, H., Cheng, X., Yuan, M., Xu, L., Gao, J. and Cheng, C. (2016). A survey of security and privacy in big data, *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*.

Yin, C., Zhang, S., Xi, J. and Wang, J. (2016), An improved anonymity model for big data security based on clustering algorithm. *Concurrency and Computation: Practice and Experience*, 29(7), p.e 3902.

Χαριτίδης Αλέξανδρος και Ψάννης Κωνσταντίνος, Υπολογιστική Νέφους Για Κινητά Τερματικά σε Δίκτυα 4ης Γενιάς: Apps4Thess, (in Greek), Φοιτητικό Συνέδριο Διοικητικής Επιστήμης και Τεχνολογίας στις 13 Μαΐου 2014.

http://bigdata-madesimple.com/research-papers-that-changed-the-world-of-big-data/
https://mapr.com/products/mapr-converged-data-platform/
https://www.ibm.com/analytics/hadoop/big-data-analytics
https://www.cloudera.com/products.html
https://www.1010data.com/products/insights-platform/
https://www.sap.com/products/hana.html
https://www.oracle.com/big-data/products.html
https://hortonworks.com/products/data-platforms/hdp/
https://www.datastax.com/products/datastax-enterprise
https://www.informatica.com/products/big-data.html?utm_source=PredictiveAnalyticsToday&utm_medium=Review&utm_campaign=PAT#fbid=58rEu6Ap82K
https://kognitio.com/?utm_source=PredictiveAnalyticsToday&utm_medium=Review&utm_campaign=PAT
https://www.internetworldstats.com/stats3.htm