



**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ**

Διπλωματική Εργασία

**ΕΦΑΡΜΟΓΕΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ BIG DATA
ΣΤΗ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ**

του:

ΣΤΕΦΑΝΟΥ ΜΟΝΑΣΤΗΡΙΔΗ:

Αριθμός Μητρώου :mba23007

Επιβλέπον: ΣΟΥΜΠΕΝΙΩΤΗΣ ΔΗΜΗΤΡΙΟΣ

ΚΑΘΗΓΗΤΗΣ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

Υποβλήθηκε ως απαιτούμενο για την απόκτηση του μεταπτυχιακού
διπλώματος στη Διοίκηση Επιχειρήσεων

Ιανουάριος 2024

Περιεχόμενα

Πίνακας Εικονογραφήσεων	V
Πίνακας Πινάκων	V
Περίληψη	VI
Abstract.....	VII
Εισαγωγή	1
Ανασκόπηση της Βιβλιογραφίας.....	2
Κεφάλαιο 1 ^ο	3
1.1 Ορισμός Big Data	3
1.2 Τα Vs	6
1.3 Big Data Value Chain.....	8
1.3.1 Δημιουργία δεδομένων(Data Generation)	8
1.3.2 Συλλογή δεδομένων(Data Collection)	9
1.3.3 Μετάδοση δεδομένων(Data Transmission)	9
1.3.4 Προεπεξεργασία δεδομένων (Data Preprocessing)	9
1.3.5 Αποθήκευση δεδομένων	10
1.3.6 Ανάλυση δεδομένων	11
1.3.7 Λήψη αποφάσεων.....	12
Κεφάλαιο 2 ^ο	13
2. Προκλήσεις των Μεγάλων Δεδομένων	13
2.1 Πολυπλοκότητα των δεδομένων	13
2.2 Ετερογένεια δεδομένων	14
2.3 Ποιότητα δεδομένων.....	14
2.4 Επεκτασιμότητα.....	14
2.5 Αποθήκευση και διαχείριση.....	15
2.6 Ιδιωτικότητα και ασφάλεια	15
2.7 Προεπεξεργασία.....	15
2.8 Οπτικοποίηση.....	16
2.9 Ανάλυση.....	16
2.10 Ανάλυση σε πραγματικό χρόνο	17
Κεφάλαιο 3 ^ο	18
3.1 Data Analytics	18
3.2 Εργαλεία για τη συλλογή, προεπεξεργασία και ανάλυση μεγάλων δεδομένων.....	19
3.2.1 NoSQL	20
3.2.2 Hadoop.....	21

3.2.3 Hadoop Distributed File System (HDFS)	21
3.2.4 Hadoop MapReduce.....	21
3.3 Λογισμικά για το Χειρισμό Μεγάλων Δεδομένων.....	23
3.3.1 R.....	23
3.3.2 Python	23
3.3.3 Scala	24
3.3.4 Apache Spark	24
3.3.5 Apache Hive.....	25
3.3.6 Apache Pig	25
3.3.7 Amazon Elastic Compute Cloud (EC2).....	25
Κεφάλαιο 4 ^ο	26
4. Big Data & Νέες Τεχνολογίες	26
4.1 Big Data και Cloud Computing	26
4.2 Big Data και Internet of Things	27
4.3 Big Data και Artificial Intelligence.....	29
Κεφάλαιο 5ο	31
5. Big Data & Επιχειρηματικότητα	31
5.1 Οφέλη των Big Data στις επιχειρήσεις.....	32
5.1.1 Εξατομίκευση μάρκετινγκ	32
5.1.2 Καλύτερη τιμολόγηση	32
5.1.3 Μείωση κόστους.....	33
5.1.4 Βελτιωμένη εξυπηρέτηση πελατών	33
Κεφάλαιο 6ο	34
6. Τα Big Data σε κλάδους των επιχειρήσεων	34
6.1 Big Data στα Logistics.....	34
6.2 Χρηματοοικονομικές Υπηρεσίες	35
6.3 Big Data στις Υπηρεσίες Υγείας.....	36
6.4 Big Data στη Βιομηχανία.....	37
6.5 Big Data στη Γεωργία.....	38
6.6 Τηλεπικοινωνίες.....	39
6.7 Λιανικό Εμπόριο	40
6.8 Επιβολή του Νόμου	41
6.9 Τραπεζικό Σύστημα	41
6.10 Ενέργεια και υπηρεσίες κοινής ωφέλειας.....	42
6.11 Εκπαίδευση	42

Κεφάλαιο 7ο	43
7. Παραδείγματα χρήσης Big Data σε γνωστές επιχειρήσεις.....	43
Κεφάλαιο 8ο	45
8. Βήματα για την αξιοποίηση των Big Data από νέες επιχειρήσεις	45
8.1.Καθορισμός ξεκάθαρων στόχων.....	45
8.2.Διαχωρισμός Σημαντικών Δεδομένων.....	46
8.3.Διασφάλιση Ποιότητας Δεδομένων	46
8.4.Σωστή επισήμανση δεδομένων	46
8.5.Επιλογή Κατάλληλων Θέσεων Αποθήκευσης.....	47
8.6.Διαχείριση Κύκλου Ζωής Δεδομένων	47
8.7.Απλοποίηση Διαδικασιών Δημιουργίας Αντιγράφων Ασφαλείας.....	48
8.8. Εφαρμογή μέτρων ασφαλείας.....	49
8.9.Κλιμακόμενη Υποδομή.....	49
8.10.Αποτελεσματική Χρήση Εργαλείων	49
8.11.Τακτικοί έλεγχοι δεδομένων	50
Κεφάλαιο 9ο	50
9.Big Data στην Ευρώπη	50
Κεφάλαιο 10ο.....	51
10.Big Data στην Ελλάδα.....	51
Συμπεράσματα.....	54

Πίνακας Εικονογραφήσεων

Εικόνα 1. Big Data (Google)

Εικόνα 2. The 3V's of Big Data [Pedro César Tebaldi Gomes (2014)]

Εικόνα 3. Big Data Value Chain [Bhadani, A., Jothimani, D. (2016)]

Εικόνα 4. Εξέλιξη της Επιστήμης Δεδομένων (Google)

Εικόνα 5. Διαδικασίες Μεγάλων Δεδομένων [Bhadani, A., Jothimani, D. (2016)]

Εικόνα 6. Βασικά στοιχεία του υπολογιστικού νέφους [Chen (2014)]

Εικόνα 7. Απεικόνιση εξοπλισμού απόκτησης δεδομένων στο IoT [Chen (2014)]

Εικόνα 8. Εταιρείες Επεξεργασίας BigData (Firstmark)

Εικόνα 9. Logistics (Google)

Εικόνα 10. Big Data in Healthcare (Google)

Εικόνα 11. Big Data in Manufacturing Industry (Google)

Εικόνα 12. Big Data in Agriculture (Google)

Πίνακας Πινάκων

Πίνακας 1. Χαρακτηριστικά των Big Data (Big Data Vs)

Πίνακας 2. Οι προκλήσεις των Μεγάλων Δεδομένων [Ashabi. (2020)]

Πίνακας 3. Δυνατότητες & Τεχνολογίες των Big Data

Περίληψη

Το αντικείμενο της παρούσας εργασίας αφορά τη μελέτη των Μεγάλων Δεδομένων (Big Data) και των τεχνολογιών τους καθώς και την αναζήτηση του ρόλου τους στην σημερινή επιχειρηματικότητα. Λόγω της ραγδαίας εξέλιξης του διαδικτύου και της ψηφιοποίησης των διαδικασιών, τεράστιος όγκος δεδομένων δημιουργείται συνεχώς. Τα Big Data ορίζονται ως ένα τεράστιο μέγεθος μη δομημένων δεδομένων που παράγονται από ετερογενείς ομάδες εφαρμογών υψηλής απόδοσης που εκτείνονται από τα κοινωνικά δίκτυα έως τις επιστημονικές εφαρμογές υπολογιστών. Η ενσωμάτωση, η αξιοποίηση και εφαρμογή αυτών των Μεγάλων Δεδομένων στις επιχειρήσεις μπορούν να ενισχύσουν την αποτελεσματικότητα της παραγωγής τους καθώς και την ανταγωνιστικότητα τους σε πολλές πτυχές. Στη συγκεκριμένη εργασία αναλύθηκε ο όρος Big Data και ο τρόπος με τον οποίο αντλούμε αξία από τα δεδομένα. Στη συνέχεια αναφέρθηκαν οι προκλήσεις των μεγάλων δεδομένων καθώς και τα εργαλεία και τα λογισμικά που χρησιμοποιούμε για την ανάλυση τους. Μελετήθηκαν τα οφέλη των Μεγάλων Δεδομένων στις επιχειρήσεις και ο τρόπος εφαρμογής τους σε συγκεκριμένους επιχειρηματικούς κλάδους. Τέλος, παρατέθηκαν παραδείγματα εφαρμογής των Big Data σε μεγάλες παγκόσμιες επιχειρήσεις και αναλύθηκε η αξία που έχουν για την Ευρώπη και την Ελλάδα.

Abstract

The subject of this review is the study of Big Data and their technologies as well as the search for their role in today's entrepreneurship. Due to the rapid development of the internet and the digitization of processes, a huge amount of data is constantly being generated. Big Data is defined as an enormous amount of unstructured data produced by heterogeneous groups of high-performance applications spanning from social networks to scientific computer applications. The integration, utilization and application of Big Data in enterprises can enhance their production efficiency as well as their competitiveness in many aspects. In this specific review, the term Big Data and the way in which we derive value from the data were analyzed. Then the challenges of big data were mentioned as well as the tools and software we use to analyze them. The benefits of Big Data in business and how to apply it to specific business sectors were studied. Finally, examples of the application of Big Data in large global companies were listed and their value for Europe and Greece was analyzed.

Εισαγωγή

Με την ψηφιοποίηση των περισσότερων διαδικασιών, την εμφάνιση διαφορετικών πλατφορμών κοινωνικών δικτύων, ιστολογίων, την ανάπτυξη διαφορετικών ειδών αισθητήρων, την υιοθέτηση ψηφιακών συσκευών χειρός, φορητών συσκευών και την έκρηξη στη χρήση του Διαδικτύου, τεράστιος όγκος δεδομένων δημιουργείται σε συνεχή βάση. Κανείς δεν μπορεί να αρνηθεί ότι το Διαδίκτυο έχει αλλάξει τον τρόπο λειτουργίας των επιχειρήσεων, τη λειτουργία της κυβέρνησης, την εκπαίδευση και τον τρόπο ζωής των ανθρώπων σε όλο τον κόσμο. Σήμερα, αυτή η τάση βρίσκεται σε ένα μετασχηματιστικό στάδιο, όπου ο ρυθμός παραγωγής δεδομένων είναι πολύ υψηλός και ο τύπος των δεδομένων που παράγονται ξεπερνά τις δυνατότητες των υφιστάμενων τεχνικών αποθήκευσης δεδομένων. Δεν μπορούμε να αμφισβητήσουμε ότι αυτά τα δεδομένα φέρουν πολύ περισσότερες πληροφορίες από ποτέ, λόγω της εμφάνισης και της υιοθέτησης του Διαδικτύου. Τις τελευταίες δύο δεκαετίες, υπάρχει τεράστια αύξηση στα δεδομένα. Αυτή η τάση μπορεί να παρατηρηθεί σχεδόν σε κάθε τομέα. Σύμφωνα με έκθεση της International Data Corporation (IDC), μια ερευνητική εταιρεία ισχυρίζεται ότι μεταξύ 2012 και 2020, ο όγκος των πληροφοριών στο ψηφιακό σύμπαν θα αυξηθεί κατά 35 τρισεκατομμύρια gigabyte (1 gigabyte που ισοδυναμεί με 40 ντουλάπια αρχείων (τεσσάρων συρταριών) κειμένου. ή δύο μουσικά CD). Αυτός είναι στο ίδιο επίπεδο με τον αριθμό των αστεριών στο φυσικό σύμπαν! (Forsyth, 2012).

Στα μέσα της δεκαετίας του 2000, η εμφάνιση των μέσων κοινωνικής δικτύωσης, του cloud computing και της επεξεργαστικής ισχύος (μέσω πολυπύρηνων επεξεργαστών και GPU) συνέβαλαν στην άνοδο των μεγάλων δεδομένων (Manovich, 2011· Agneeswaran, 2012). Αυτό έχει δημιουργήσει νέα μονοπάτια για τη μελέτη της κοινωνικής και πολιτιστικής δυναμικής.

Η εξεύρεση λογικής από τα τεράστια δεδομένα μπορεί να βοηθήσει τον οργανισμό στη λήψη τεκμηριωμένων αποφάσεων και να προσφέρει ανταγωνιστικό πλεονέκτημα. Παλαιότερα, οι οργανισμοί χρησιμοποιούσαν συστήματα επεξεργασίας συναλλαγών που χρησιμοποιούσαν εγγενώς Συστήματα Διαχείρισης Βάσεων Δεδομένων Σχέσεων (RDBMS) και απλές τεχνικές ανάλυσης δεδομένων όπως η δομημένη γλώσσα ερωτημάτων (SQL) για τις καθημερινές τους λειτουργίες που τους βοήθησαν στη λήψη αποφάσεων και τον προγραμματισμό τους. Ωστόσο, λόγω της αύξησης του μεγέθους των δεδομένων, ειδικά της μη δομημένης μορφής δεδομένων (για παράδειγμα, κριτικές πελατών των σελίδων τους στο Facebook ή tweets), έχει καταστεί σχεδόν αδύνατη η επεξεργασία αυτών των δεδομένων με τις υπάρχουσες τεχνικές αποθήκευσης και απλά ερωτήματα.

Ανασκόπηση της Βιβλιογραφίας

Αυτή η διπλωματική εργασία παρέχει μια ολοκληρωμένη επισκόπηση της βασικής βιβλιογραφίας που αφορά τις εφαρμογές της ανάλυσης Big Data στην επιχειρηματικότητα. Καθώς η τεχνολογία εξελίσσεται, οι επιχειρήσεις ενθαρρύνονται να υιοθετήσουν προσεγγίσεις που βασίζονται σε μεγάλα δεδομένα (big data) για να αποκτήσουν ανταγωνιστικό πλεονέκτημα και να προωθήσουν την καινοτομία στο σύγχρονο επιχειρηματικό κόσμο. Οι γνώσεις που συγκεντρώθηκαν από αυτές τις πηγές θέτουν τις βάσεις για περαιτέρω έρευνα και εξερεύνηση στη δυναμική διασταύρωση των μεγάλων δεδομένων και της επιχειρηματικότητας.

Για την υλοποίηση αυτής της εργασίας συγκεντρώθηκαν δεδομένα για όλο το φάσμα των μεγάλων δεδομένων (Big Data). Συγκεντρώθηκαν δεδομένα για τον προσδιορισμό του όρου Big Data, τα οφέλη τους στην επιχειρηματικότητα, τις επιμέρους κατηγορίες τους καθώς και τους κινδύνους και τις προκλήσεις που μπορούμε να αντιμετωπίσουμε με αυτά.

Επιπρόσθετα, συγκέντρωσα πληροφορίες για τα εργαλεία ανάλυσης των Big Data, τα λογισμικά που χρησιμοποιούνται, τον τρόπο με τον οποίο τα αναλύουμε και τα χρησιμοποιούμε και τους τομείς στους οποίους μπορούμε να τα αξιοποιήσουμε.

Τέλος, αναλύονται σημαντικά παραδείγματα χρήσης των Big Data σε μεγάλες παγκόσμιες επιχειρήσεις ώστε να κατανοήσουμε τον τρόπο εφαρμογής τους στον επιχειρηματικό κόσμο και να πάρουμε ερεθίσματα για μελλοντική χρήση των μεγάλων δεδομένων σε νέες επιχειρήσεις καθώς και το πως η Ελλάδα και η Ευρώπη αντιμετωπίζουν σήμερα τα Μεγάλα Δεδομένα.

Οι πληροφορίες συγκεντρώθηκαν από επιστημονικά άρθρα, case studies και από το διαδίκτυο.

Κεφάλαιο 1^ο

1.1 Ορισμός Big Data

Ο όρος "Big Data" έχει γίνει πολύ δημοφιλής αφού χρησιμοποιείται από σχεδόν όλους, συμπεριλαμβανομένων ακαδημαϊκών και επιχειρηματιών. Με την πάροδο των χρόνων, πολλοί είναι οι ορισμοί που έχουν δοθεί για τα Μεγάλα Δεδομένα . Ο όρος Big Data αρχικά χρησιμοποιήθηκε για τον ορισμό μεγάλων δεδομένων. Ο όρος big data χρησιμοποιήθηκε για πρώτη φορά το 1997 από επιστήμονες της NASA καθώς τα δεδομένα ήταν τόσο μεγάλα που καθιστούσαν αδύνατη την αναπαράσταση τους γραφικά και την αποθήκευση τους στην κύρια μνήμη του υπολογιστή. Αυτό αποτέλεσε και ένα από τα πρώτα προβλήματα με τα μεγάλα δεδομένα. Αργότερα διάφορες επιχειρήσεις, συμπεριλαμβανομένης της Microsoft και της IBM όρισαν τα Big Data σε μεγάλου μεγέθους δεδομένα των οποίων η διαχείριση και επεξεργασία γίνεται από γενικούς υπολογιστές εντός ενός αποδεκτού πεδίου» (Chen et al., 2014). Μετά από αυτό, το 2011, το McKinsey Global Institute όρισε τα μεγάλα δεδομένα ως «σύνολα δεδομένων των οποίων το μέγεθος καθιστά τη διαχείριση, αποθήκευση και ανάλυση τους δύσκολη να πραγματοποιηθεί από τυπικά εργαλεία λογισμικού (Manyika et al., 2011). Επίσης, η International Data Corporation (IDC) ορίζει τις τεχνολογίες μεγάλων δεδομένων ως νέα γενιά τεχνολογιών που έχουν σχεδιαστεί για να εξάγουν οικονομική αξία από πολύ μεγάλους όγκους μεγάλης ποικιλίας δεδομένων, επιτρέποντας τη λήψη, την ανακάλυψη και διαχείριση υψηλής ταχύτητας» (Gantz και Reinsel, 2011). Οι ακαδημαϊκοί ορίζουν τα μεγάλα δεδομένα ως ένα τεράστιο μέγεθος μη δομημένων δεδομένων που παράγονται από ετερογενείς ομάδες εφαρμογών υψηλής απόδοσης που εκτείνονται από τα κοινωνικά δίκτυα έως τις επιστημονικές εφαρμογές υπολογιστών. Τα σύνολα δεδομένων κυμαίνονται από μερικές εκατοντάδες gigabyte έως zetabyte (Cuzzocrea et al., 2011; Qin et al., 2012; Agneeswaran, 2012).

Αν και τα μεγάλα δεδομένα έχουν οριστεί με διάφορες μορφές, κανένας ορισμός δεν είναι πλήρης. Λίγοι έχουν ορίσει τι κάνει, ενώ πολύ λίγοι έχουν επικεντρωθεί σε αυτό που είναι. Εκτός από τον ορισμό των μεγάλων δεδομένων, υπάρχει ανάγκη να κατανοήσουμε πώς να αξιοποιήσουμε καλύτερα αυτά τα δεδομένα για να αποκτήσουμε πολύτιμες πληροφορίες για τη λήψη αποφάσεων.

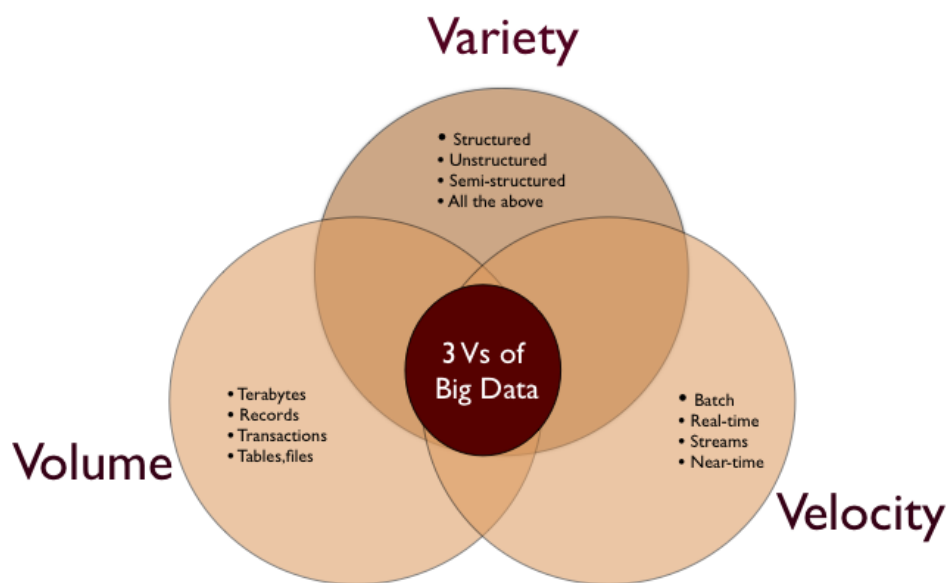
Τα τελευταία χρόνια, η ανάγκη διαχείρισης μεγάλων δεδομένων (Big Data) αυξάνεται λόγω της ανάγκης για μαζική παραγωγή. Τεχνολογικές εξελίξεις στον τομέα των τηλεπικοινωνιών και των ολοκληρωμένων κυκλωμάτων έχουν ως αποτέλεσμα τη δημιουργία μηχανισμών παρακολούθησης των λειτουργιών ενός οργανισμού σε ανεπτυγμένο επίπεδο. Λόγο αυτής της ψηφιοποίησης των παραγωγικών διαδικασιών μεγάλοι οργανισμοί και επιχειρήσεις κατάφεραν να παράγουν τεράστιο όγκο δεδομένων σε πολύ σύντομο χρονικό διάστημα. Η ανάλυση των δεδομένων αυτών παρέχει χρήσιμες πληροφορίες στις επιχειρήσεις και έχει ως αποτέλεσμα την βελτιστοποίηση των διαδικασιών παραγωγής των επιχειρήσεων.

Σε κάθε οικονομία, τομέα και οργανισμό τα ψηφιακά δεδομένα βρίσκονται πλέον παντού. Επιχειρήσεις αλλά και καταναλωτές έλκονται ολοένα και περισσότερο από τα big data λόγω του κέρδους που μπορούν να αποκομίσουν μέσω της ανάλυσης τους. Στη σημερινή εποχή η δυνατότητα συγκέντρωσης, αποθήκευσης αλλά και ανάλυσης των δεδομένων είναι πιο προσιτή και εφικτή από ποτέ. Παράλληλα η αύξηση των κινητών συσκευών έχει εκτοξεύσει τη δυνατότητα παραγωγής και πρόσβασης δεδομένων. Κάθε άνθρωπος σήμερα αποτελεί μια μεγάλη πηγή δεδομένων καθώς μέσω των ψηφιακών του προφίλ στα κοινωνικά δίκτυα οι πληροφορίες και τα δεδομένα που παράγονται είναι χιλιάδες. Η Google αναφέρει ότι το ψηφιακό υλικό που δημιουργείται από τους χρήστες κάθε δύο μέρες ισοδυναμεί με το συνολικό έντυπο υλικό που παρήγαγε η ανθρωπότητα από την αρχή της γραφής μέχρι το 2003.



Εικόνα 1. Big Data (Google)

Προκειμένου να εντοπίσουμε τις προκλήσεις των Big Data , είναι απαραίτητο να κατανοήσουμε τα βασικά τους χαρακτηριστικά. Υπάρχουν πολλοί ορισμοί για τα “Μεγάλα Δεδομένα” αλλά τα κύρια χαρακτηριστικά τους περιγράφονται ξεκάθαρα από τον Gartner (Gartner et al., 2018). Σύμφωνα με αυτόν τα “Μεγάλα Δεδομένα” είναι δεδομένα μεγάλου όγκου(Volume), υψηλής ταχύτητας(Velocity) και μεγάλης ποικιλίας(Variety) που απαιτούν νέες μεθόδους διαχείρισης ώστε να αξιοποιηθούν πλήρως και να βελτιστοποιήσουν τη λήψη αποφάσεων και το τελική διαδικασία. Κάθε ένα από αυτά τα τρία Vs έχει τις δικές του ιδιαιτερότητες ως προς την ανάλυση του.



Εικόνα 2. The 3V's of Big Data[Pedro César Tebaldi Gomes (2014)]

1.2 Τα Vs

Volume : Ο όγκος αναφέρεται στο μέγεθος των δεδομένων που δημιουργούνται και συλλέγονται. Αυξάνεται με ταχύτατο ρυθμό από terabyte σε petabyte (1024 terabytes) (Zikopoulos et al., 2012). Με την αύξηση της χωρητικότητας αποθήκευσης, ό,τι είναι αδύνατον να συλληφθεί και να αποθηκευτεί τώρα θα είναι δυνατό στο μέλλον. Η ταξινόμηση των μεγάλων δεδομένων με βάση τον όγκο είναι σχετική με τον τύπο των δεδομένων που παράγονται και τον χρόνο.

Variety : Η ποικιλία αναφέρεται σε διαφορετικούς τύπους δεδομένων που παράγονται και συγκεντρώνονται. Εκτείνονται πέρα από τα δομημένα δεδομένα και εμπίπτουν στις κατηγορίες ημιδομημένων και μη δομημένων δεδομένων (Zikopoulos et al., 2012) . Τα δεδομένα που μπορούν να οργανωθούν χρησιμοποιώντας ένα προκαθορισμένο μοντέλο δεδομένων είναι γνωστά ως δομημένα δεδομένα. Τα δεδομένα σε πίνακα στις σχεσιακές βάσεις δεδομένων και στο Excel είναι παραδείγματα δομημένων δεδομένων και αποτελούν μόνο το 5% όλων των υπαρχόντων δεδομένων (Cukier, 2010). Τα μη δομημένα δεδομένα δεν μπορούν να οργανωθούν χρησιμοποιώντας αυτά τα προκαθορισμένα μοντέλα και παραδείγματα περιλαμβάνουν βίντεο, κείμενο και ήχο. Τα ημι-δομημένα δεδομένα είναι εκείνα που έχουν δομή αλλά δεν ταιριάζουν στη σχεσιακή βάση δεδομένων. Τα ημι-δομημένα δεδομένα είναι οργανωμένα, γεγονός που καθιστά ευκολότερη την ανάλυση τους σε σύγκριση με τα μη δομημένα δεδομένα. Το JSON και το XML είναι παραδείγματα ημι-δομημένων δεδομένων.

Velocity : αναφέρεται στον ρυθμό παραγωγής δεδομένων. Οι παραδοσιακές αναλύσεις δεδομένων βασίζονται σε περιοδικές ενημερώσεις (καθημερινά, εβδομαδιαία ή μηνιαία). Με τον αυξανόμενο ρυθμό παραγωγής δεδομένων, τα μεγάλα δεδομένα θα πρέπει να υποβάλλονται σε επεξεργασία και να αναλύονται σε πραγματικό ή σχεδόν πραγματικό χρόνο για τη λήψη τεκμηριωμένων αποφάσεων. Ο ρόλος του χρόνου είναι πολύ κρίσιμος εδώ (Gandomi and Haider, 2015). Λίγοι τομείς, όπως το Λιανικό εμπόριο, οι Τηλεπικοινωνίες και τα Οικονομικά παράγουν δεδομένα υψηλής συχνότητας. Τα δεδομένα που δημιουργούνται μέσω εφαρμογών για κινητά, για παράδειγμα, δημογραφικά στοιχεία, γεωγραφική τοποθεσία και ιστορικό συναλλαγών, μπορούν να χρησιμοποιηθούν σε πραγματικό χρόνο για να προσφέρουν εξατομικευμένες υπηρεσίες στους πελάτες. Αυτό θα βοηθούσε στη διατήρηση των πελατών καθώς και στην αύξηση του επιπέδου εξυπηρέτησης.

Αργότερα, προστέθηκαν μερικές ακόμη διαστάσεις, οι οποίες απαριθμούνται παρακάτω:

Veracity: Επινοήθηκε από την IBM, και αναφέρεται στην αναξιοπιστία που σχετίζεται με τις πηγές δεδομένων (Gandomi and Haider, 2015). Για παράδειγμα, η ανάλυση συναισθήματος με χρήση δεδομένων μέσω κοινωνικής δικτύωσης (Twitter, Facebook, κ.λπ.) υπόκειται σε αβεβαιότητα. Υπάρχει ανάγκη διαφοροποίησης των αξιόπιστων δεδομένων από τα αβέβαια και ανακριβή δεδομένα καθώς και διαφοροποίησης στη διαχείριση τους.

Variability: Αναφέρεται στην μεταβλητότητα των δεδομένων. Η μεταβλητότητα και η πολυπλοκότητα προστέθηκαν ως πρόσθετες διαστάσεις από το SAS. Συχνά, η ασυνέπεια στη μεγάλη ταχύτητα δεδομένων οδηγεί σε διακύμανση του ρυθμού ροής δεδομένων, η οποία αναφέρεται ως μεταβλητότητα (Gandomi and Haider, 2015). Τα δεδομένα παράγονται από διάφορες πηγές και υπάρχει μια αυξανόμενη πολυπλοκότητα στη διαχείριση τους.

Visualization : Η απεικόνιση των “Μεγάλων Δεδομένων” είναι εξαιρετικά σημαντική καθώς μετατρέπει τα δεδομένα σε μορφή κατανοητή για τον αναγνώστη. Αποτελεί ένα από τα πιο δύσκολα κομμάτια της ανάλυσης δεδομένων καθώς πίσω από κάθε απεικόνιση υπάρχουν σύνθετες ιστορίες που είναι δύσκολο να αποτυπωθούν πλήρως.

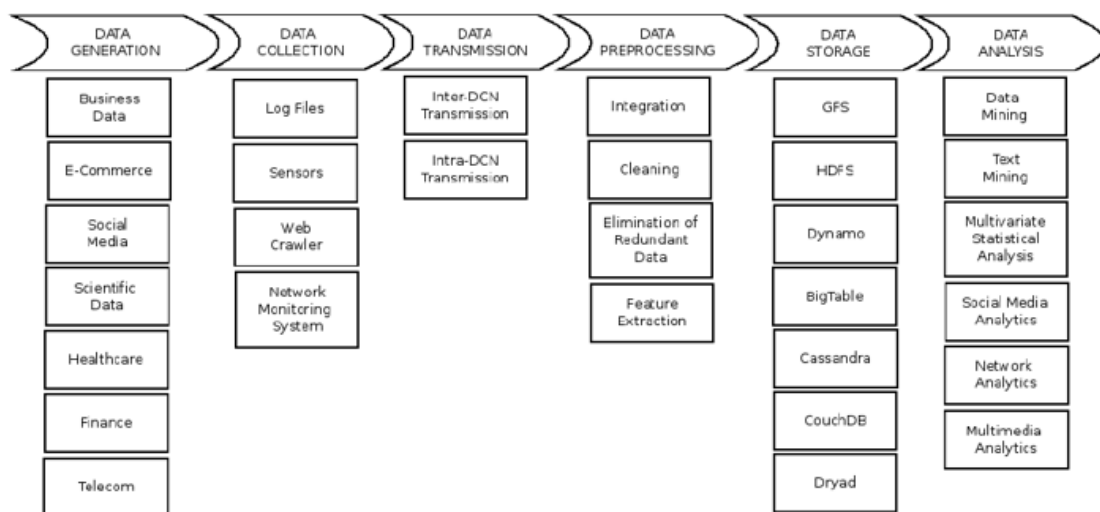
Πίνακας 1. Χαρακτηριστικά των Big Data (Big Data Vs)

Χαρακτηριστικά των Big Data	
Volume	Μέγεθος Δεδομένων
Velocity	Ταχύτητα Δεδομένων
Variety	Τύπος Δεδομένων
Veracity	Ποιότητα Δεδομένων
Validity	Αυθεντικότητα Δεδομένων
Value	Σημαντικότητα Δεδομένων
Volatility	Διάρκεια Χρησιμότητας Δεδομένων
Visualization	Απεικόνιση Δεδομένων
Vitality	Ταχύτητα Διάδοσης
Viscosity	Καθυστέρηση Εκδήλωσης
Variability	Διαφοροποίηση Δεδομένων
Venue	Πλατφόρμα Δεδομένων
Vocabulary	Ορολογία Δεδομένων

1.3 Big Data Value Chain

Η Αλυσίδα Αξίας (Value Chain), που προτάθηκε αρχικά από τον Porter το 1980, περιγράφει μια σειρά ενεργειών που αναλαμβάνει μια εταιρεία για να ενισχύσει την αξία του προϊόντος ή της υπηρεσίας τους καθ' όλη τη διαδικασία παράδοσης στους πελάτες. Ομοίως, η αλυσίδα αξίας δεδομένων (Big Data Value Chain) περιλαμβάνει μια δομή που διαχειρίζεται μια σειρά εργασιών με στόχο την άντληση αξίας από τα υπάρχοντα δεδομένα. Η διαδικασία αυτή μπορεί να χωριστεί σε επτά στάδια:

1. Παραγωγή δεδομένων
2. Συλλογή δεδομένων
3. Μετάδοση δεδομένων
4. Προ επεξεργασία δεδομένων
5. Αποθήκευση δεδομένων
6. Ανάλυση δεδομένων
7. Λήψη αποφάσεων



Εικόνα 3. Big Data Value Chain [Bhadani, A., Jothimani, D. (2016)]

1.3.1 Δημιουργία δεδομένων(Data Generation)

Το αρχικό βήμα στην αλυσίδα αξίας των μεγάλων δεδομένων περιλαμβάνει τη δημιουργία δεδομένων. Όπως περιεγράφηκε στην προηγούμενη ενότητα, τα δεδομένα προέρχονται από διάφορες πηγές, που περιλαμβάνουν αρχεία λεπτομερειών κλήσεων (CDR), ιστολόγια, tweets και σελίδες στο Facebook.

1.3.2 Συλλογή δεδομένων(Data Collection)

Σε αυτό το στάδιο, τα δεδομένα συλλέγονται από όλες τις πιθανές πηγές (Miller and Mork, 2013- Chen et al., 2014). Για παράδειγμα, για την πρόβλεψη της απομάκρυνσης πελατών στον κλάδο των τηλεπικοινωνιών, τα δεδομένα μπορούν να αντληθούν από τα CDR και τις απόψεις/καταγγελίες των πελατών σε ιστότοπους κοινωνικής δικτύωσης όπως το Twitter (με τη μορφή tweets) και το Facebook (που εκφράζονται στη σελίδα της εταιρείας στο Facebook). Οι μέθοδοι που χρησιμοποιούνται συνήθως περιλαμβάνουν αρχεία καταγραφής, αισθητήρες, web crawlers και λογισμικό παρακολούθησης δικτύου (Chen et al., 2014).

1.3.3 Μετάδοση δεδομένων(Data Transmission)

Μετά τη συλλογή δεδομένων, αυτά διαβιβάζονται σε μια υποδομή αποθήκευσης και επεξεργασίας δεδομένων για περαιτέρω ανάλυση. Η εν λόγω διαβίβαση πραγματοποιείται σε δύο φάσεις: Μετάδοση εντός του δικτύου δυναμικού κυκλώματος (DCN), η οποία περιλαμβάνει τη μεταφορά δεδομένων από την πηγή στο κέντρο δεδομένων, και μετάδοση εντός του δικτύου δυναμικού κυκλώματος (Intra-DCN), που διευκολύνει τη μεταφορά δεδομένων εντός του κέντρου δεδομένων. Εκτός από την αποθήκευση δεδομένων, το κέντρο δεδομένων βοηθά στη συλλογή, οργάνωση και διαχείριση των δεδομένων.

1.3.4 Προεπεξεργασία δεδομένων (Data Preprocessing)

Τα δεδομένα που συλλέγονται από διάφορες πηγές μπορεί να περιέχουν πλεονασμούς, θόρυβο και ασυνέπειες. Ως εκ τούτου, σε αυτή τη φάση, τα δεδομένα υποβάλλονται σε προεπεξεργασία για τη βελτίωση της ποιότητάς τους για την επακόλουθη ανάλυση. Το στάδιο αυτό συμβάλλει στη βελτίωση της ακρίβειας της ανάλυσης και στη μείωση των εξόδων αποθήκευσης.

Η προεπεξεργασία περιλαμβάνει τα ακόλουθα βήματα:

α) Ολοκλήρωση

Οι διάφορες πηγές δεδομένων συγχωνεύονται για να παρουσιάσουν μια ενιαία και συνεπή εικόνα. Συνήθως χρησιμοποιούνται παραδοσιακές μέθοδοι όπως η ομοσπονδία δεδομένων και η αποθήκευση δεδομένων, που περιλαμβάνουν τη διαδικασία εξαγωγής, μετασχηματισμού και φόρτωσης (ETL). Η δυναμική ενοποίηση δεδομένων μπορεί να επιτευχθεί με τη συγκέντρωση δεδομένων από διαφορετικές πηγές με τη χρήση μιας εικονικής βάσης δεδομένων.

β) Καθαρισμός

Τα δεδομένα ελέγχονται για την ακρίβεια, την πληρότητα και τη συνοχή τους. Η διαδικασία αυτή περιλαμβάνει τον εντοπισμό και τη διόρθωση σφαλμάτων, την τεκμηρίωση των τύπων και των παραδειγμάτων σφαλμάτων και την τροποποίηση των διαδικασιών εισαγωγής δεδομένων για την αποφυγή μελλοντικών σφαλμάτων.

γ) Εξάλειψη περιττών δεδομένων

Πολλά σύνολα δεδομένων περιέχουν πλεονάζοντα ή επαναλαμβανόμενα δεδομένα, γεγονός που οδηγεί σε πλεονασμό δεδομένων. Αυτά τα πλεονάζοντα δεδομένα αυξάνουν το κόστος αποθήκευσης, εισάγουν ασυνέπεια και επηρεάζουν την ποιότητα των δεδομένων. Για την αντιμετώπιση αυτού του ζητήματος χρησιμοποιούνται διάφορες μέθοδοι μείωσης δεδομένων, όπως το φιλτράρισμα και η συμπίεση τους. Ωστόσο, είναι ζωτικής σημασίας η διενέργεια ανάλυσης κόστους-οφέλους πριν από τη χρήση αυτών των τεχνικών λόγω της πιθανής αύξησης του υπολογιστικού κόστους.

1.3.5 Αποθήκευση δεδομένων

Τα συστήματα αποθήκευσης μεγάλων δεδομένων θα πρέπει να παρέχουν αξιόπιστο χώρο αποθήκευσης και ισχυρή πρόσβαση στα δεδομένα. Τα κατανομημένα συστήματα αποθήκευσης για μεγάλα δεδομένα θα πρέπει να λαμβάνουν υπόψη παράγοντες όπως η συνέπεια (C), η διαθεσιμότητα (A) και η ανοχή κατατμήσεων (P). Σύμφωνα με τη θεωρία CAP που προτάθηκε από τον Brewer (2000), τα κατανομημένα συστήματα αποθήκευσης θα μπορούσαν να

ικανοποιούν ταυτόχρονα δύο απαιτήσεις, δηλαδή είτε τη συνέπεια και τη διαθεσιμότητα είτε τη διαθεσιμότητα και την ανοχή κατατμήσεων είτε τη συνέπεια και την ανοχή κατατμήσεων, αλλά όχι όλες τις απαιτήσεις ταυτόχρονα (Gilbert and Lynch, 2002). Σημαντική έρευνα συνεχίζεται ακόμη στον τομέα του μηχανισμού αποθήκευσης μεγάλων δεδομένων. Μικρή πρόοδος σε αυτό το θέμα είναι το Google File System (GFS), το Dynamo, το BigTable, το Cassandra, το CouchDB και το Dryad.

1.3.6 Ανάλυση δεδομένων

Μετά τη συλλογή, μετατροπή και αποθήκευση των δεδομένων, η επόμενη διαδικασία είναι η αξιοποίηση των δεδομένων ή η ανάλυση των δεδομένων, η οποία απαριθμείται χρησιμοποιώντας τα ακόλουθα βήματα:

α) Καθορισμός μετρήσεων

Με βάση τα δεδομένα που έχουν συλλεχθεί και μετασχηματιστεί, ορίζεται ένα σύνολο μετρικών για ένα συγκεκριμένο πρόβλημα. Για παράδειγμα, για τον εντοπισμό ενός δυνητικού πελάτη που πρόκειται να εκπέσει, μπορεί να ληφθεί υπόψη ο αριθμός των φορών που επικοινωνήσε (είτε μέσω φωνητικής κλήσης, είτε μέσω tweets ή παραπόνων στη σελίδα του Facebook). (Miller και Mork, 2013).

β) Επιλογή της αρχιτεκτονικής με βάση τον τύπο ανάλυσης

Με βάση το χρόνο της ανάλυσης που πρόκειται να πραγματοποιηθεί, επιλέγεται η κατάλληλη αρχιτεκτονική. Η ανάλυση σε πραγματικό χρόνο χρησιμοποιείται στον τομέα όπου τα δεδομένα μεταβάλλονται συνεχώς και υπάρχει ανάγκη για ταχεία ανάλυση για τη λήψη μέτρων. Οι υπολογισμοί με βάση τη μνήμη και τα συστήματα παράλληλης επεξεργασίας είναι οι υπάρχουσες αρχιτεκτονικές. Η ανίχνευση απάτης στον τομέα του λιανικού εμπορίου και η απάτη στις τηλεπικοινωνίες αποτελούν παραδείγματα ανάλυσης σε πραγματικό χρόνο. Οι εφαρμογές που δεν απαιτούν υψηλό χρόνο απόκρισης πραγματοποιούνται με τη χρήση ανάλυσης εκτός σύνδεσης. Τα δεδομένα μπορούν να εξαχθούν, να αποθηκευτούν και να αναλύονται σχετικά αργότερα στο χρόνο. Η αρχιτεκτονική που χρησιμοποιείται γενικά είναι η πλατφόρμα Hadoop (Chen et al., 2014).

γ) Επιλογή των κατάλληλων αλγορίθμων και εργαλείων

Ένα από τα σημαντικότερα βήματα της ανάλυσης δεδομένων είναι η επιλογή των κατάλληλων τεχνικών για την ανάλυση δεδομένων. Λίγες παραδοσιακές τεχνικές ανάλυσης δεδομένων, όπως η ανάλυση συστάδων, η ανάλυση παλινδρόμησης και οι αλγόριθμοι εξόρυξης δεδομένων,

εξακολουθούν να ισχύουν για την ανάλυση μεγάλων δεδομένων. Η ανάλυση συστάδων είναι μια μη επιβλεπόμενη τεχνική που ομαδοποιεί τα αντικείμενα με βάση ορισμένα χαρακτηριστικά. Οι τεχνικές εξόρυξης δεδομένων βοηθούν στην εξαγωγή άγνωστων, κρυφών και χρήσιμων πληροφοριών από ένα τεράστιο σύνολο δεδομένων. Οι 10 ισχυρότεροι αλγόριθμοι εξόρυξης δεδομένων επιλέχθηκαν και συζητήθηκαν στο Wu et al. (2007). Διάφορα εργαλεία είναι διαθέσιμα για την ανάλυση δεδομένων, συμπεριλαμβανομένων λογισμικών ανοικτού κώδικα και εμπορικών λογισμικών.

δ) Οπτικοποίηση δεδομένων

Η ανάγκη για επιθεώρηση λεπτομερειών σε πολλαπλές κλίμακες και μικρολεπτομέρειες έδωσε το έναυσμα για την οπτικοποίηση δεδομένων. Οι οπτικές διεπαφές μαζί με τις στατιστικές αναλύσεις και το σχετικό πλαίσιο βοηθούν στον εντοπισμό μοτίβων σε μεγάλα δεδομένα με την πάροδο του χρόνου (Fisher et al., 2012). Η οπτική ανάλυση (VA) ορίζεται ως "η επιστήμη της αναλυτικής συλλογιστικής που διευκολύνεται από οπτικές διαδραστικές διεπαφές" (Thomas and Cook, 2005). Λίγα εργαλεία οπτικοποίησης είναι τα Tableau, QlikView, Spotfire, JMP, Jaspersoft, Visual Analytics, Centrifuge, Visual Mining και Board.

1.3.7 Λήψη αποφάσεων

Με βάση την ανάλυση και τα οπτικοποιημένα αποτελέσματα, οι υπεύθυνοι λήψης αποφάσεων μπορούν να αποφασίσουν εάν και πώς να επιβραβεύσουν μια θετική συμπεριφορά και να αλλάξουν μια αρνητική. Οι λεπτομέρειες ενός συγκεκριμένου προβλήματος μπορούν να αναλυθούν ώστε να κατανοηθούν οι αιτίες των προβλημάτων να ληφθούν τεκμηριωμένες αποφάσεις και να προγραμματιστούν οι απαραίτητες ενέργειες (Miller and Mork, 2013).

Έχοντας συζητήσει για το πώς μπορεί να εξαχθεί αξία από τα μεγάλα δεδομένα, μια βιομηχανία, ανεξαρτήτως τομέα, θα πρέπει να εξετάσει τρία κριτήρια πριν από την εφαρμογή της ανάλυσης μεγάλων δεδομένων:

1. μπορούν να ληφθούν χρήσιμες πληροφορίες επιπλέον αυτών που λαμβάνονται από τα υπάρχοντα συστήματα;
2. θα υπάρξει βελτίωση στην ακρίβεια των πληροφοριών που λαμβάνονται με τη χρήση της ανάλυσης μεγάλων δεδομένων;

3. η εφαρμογή της ανάλυσης μεγάλων δεδομένων θα βοηθήσει στη βελτίωση της έγκαιρης ανταπόκρισης; (Villars et al., 2011).

Κεφάλαιο 2^ο

2. Προκλήσεις των Μεγάλων Δεδομένων

Τα μεγάλα δεδομένα παρουσιάζουν σημαντικές προκλήσεις, οι οποίες μπορούν να κατηγοριοποιηθούν σε τρεις κύριες ομάδες με βάση τον κύκλο ζωής των δεδομένων - προκλήσεις δεδομένων, προκλήσεις επεξεργασίας και προκλήσεις διαχείρισης (U.Sivarajah et al., 2017).

- Οι προκλήσεις δεδομένων αφορούν τα εγγενή χαρακτηριστικά των ίδιων των δεδομένων, όπως ο όγκος, η ποικιλία και η ταχύτητά τους.
- Οι προκλήσεις επεξεργασίας αφορούν τον καθορισμό προσεγγίσεων για τη σύλληψη, το μετασχηματισμό, την ενσωμάτωση, την επιλογή κατάλληλων μοντέλων για ανάλυση και την παρουσίαση των ευρημάτων.
- Οι προκλήσεις διαχείρισης περιλαμβάνουν ανησυχίες σχετικά με την ασφάλεια, την προστασία της ιδιωτικής ζωής, τη διακυβέρνηση και τις ηθικές προοπτικές.

2.1 Πολυπλοκότητα των δεδομένων

Τα τυπικά χαρακτηριστικά των μεγάλων δεδομένων, συμπεριλαμβανομένης της ταξινόμησης σε τύπους και τάσεις, των περίπλοκων σχέσεων και του ποικίλου φάσματος ποιότητας των δεδομένων, συμβάλλουν στην εγγενή πολυπλοκότητά τους. Οι εγγενείς περιπλοκές των μεγάλων δεδομένων, όπως οι περίπλοκοι τύποι, οι πολύπλοκες δομές και οι εμπλεκόμενες τάσεις, δημιουργούν πρόσθετες προκλήσεις όσον αφορά την αναγνώριση, την απεικόνιση, την κατανόηση και τον υπολογισμό. Κατά συνέπεια, παρατηρείται αξιοσημείωτη αύξηση της υπολογιστικής πολυπλοκότητας σε σύγκριση με τα παραδοσιακά μοντέλα υπολογισμού που βασίζονται σε ολοκληρωμένα δεδομένα. Οι συμβατικές εργασίες ανάλυσης και εξόρυξης δεδομένων, όπως η ανάκτηση, η ανακάλυψη θεμάτων, οι σημασιολογικές αναλύσεις και οι αναλύσεις συναισθήματος, γίνονται εξαιρετικά χρονοβόρες όταν πρόκειται για μεγάλα δεδομένα (X.Jin et al., 2015).

2.2 Ετερογένεια δεδομένων

Μια άλλη ουσιαστική πρόκληση που αντιμετωπίζουν οι ερευνητές αφορά την ενσωμάτωση δεδομένων από διαφορετικές πηγές για τη μεγιστοποίηση της χρησιμότητάς τους. Ο σημαντικός όγκος δεδομένων που παράγεται από τα μέσα κοινωνικής δικτύωσης, τα ιστολόγια και τους ιστότοπους παρουσιάζει διαφορές στη μορφή, τη σημασιολογία και την πηγή δεδομένων για κάθε ξεχωριστή προέλευση. Οι δομές δεδομένων από αυτές τις πηγές κυμαίνονται από εξαιρετικά οργανωμένα δεδομένα, όπως οι βάσεις δεδομένων, έως αδόμητα δεδομένα με τη μορφή ετερογενών αναφορών (P Murali K et al .,2016).

2.3 Ποιότητα δεδομένων

Η αξιοπιστία των μεγάλων δεδομένων για τη λήψη αποφάσεων εξαρτάται σημαντικά από την ποιότητα των δεδομένων, η οποία αξιολογείται με βάση τέσσερα βασικά κριτήρια (N.T.Tariq et all., 2015):

- Πληρότητα: Διασφάλιση ότι όλα τα σχετικά δεδομένα είναι διαθέσιμα, όπως τα στοιχεία του πωλητή, όπως το όνομα, η διεύθυνση και ο τραπεζικός λογαριασμός.
- Ακρίβεια: Επαλήθευση ότι τα δεδομένα δεν περιέχουν λανθασμένη ορθογραφία, τυπογραφικά λάθη, ανακρίβειες και συντομεύσεις.
- Διαθεσιμότητα: Διασφάλιση ότι τα δεδομένα είναι εύκολα προσβάσιμα όταν χρειάζονται και ότι είναι εύκολος ο εντοπισμός τους.
- Επικαιρότητα: Επιβεβαίωση ότι οι πληροφορίες είναι επίκαιρες και άμεσα διαθέσιμες για να βοηθήσουν στη διαδικασία λήψης αποφάσεων.

2.4 Επεκτασιμότητα

Στο πεδίο των μεγάλων δεδομένων, η επεκτασιμότητα αποτελεί αξιοσημείωτη πρόκληση, ιδίως κατά τη φάση της ανάλυσης. Οι αυξητικές στρατηγικές παρουσιάζουν εξαιρετικά χαρακτηριστικά κλιμάκωσης όταν έχουν να κάνουν με την ανάλυση μεγάλων συνόλων δεδομένων. Αυτό είναι ζωτικής σημασίας επειδή το μέγεθος των δεδομένων επεκτείνεται με ρυθμό ταχύτερο από την ταχύτητα της CPU. Κατά συνέπεια, η αντιμετώπιση αυτής της πρόκλησης οδήγησε στην εξέλιξη του παράλληλου υπολογισμού, ο οποίος είναι απαραίτητος

για εφαρμογές πραγματικού χρόνου, όπως τα κοινωνικά δίκτυα, η πλοήγηση, τα οικονομικά, η επικαιρότητα και οι αναζητήσεις στον ιστό (R.S.K.Althaf et al., 2018).

2.5 Αποθήκευση και διαχείριση

Ο σημαντικός όγκος δεδομένων που παράγεται καθημερινά καθιστά αναγκαία την αποθήκευση για αναλυτικούς σκοπούς και τη συμμόρφωση με τους κανονισμούς και τις συμφωνίες επιπέδου υπηρεσιών για την ασφάλεια και τη διατήρηση των δεδομένων. Στον τομέα των μεγάλων δεδομένων, η αποθήκευση και η διαχείριση αναδεικνύονται ως δύο κεντρικές προκλήσεις (Agrawal et al., 2016). Η ικανότητα κλιμάκωσης των συσκευών αποθήκευσης για την προσαρμογή στην αύξηση των δεδομένων, την ενίσχυση του χρόνου πρόσβασης και τη βελτίωση των ρυθμών μεταφοράς αποτελεί σημαντική πρόκληση. Οι παράγοντες αυτοί επηρεάζουν σημαντικά τη συνολική απόδοση της αποθήκευσης και της διαχείρισης δεδομένων.

2.6 Ιδιωτικότητα και ασφάλεια

Το 2015, μια ολοκληρωμένη ανασκόπηση κατέληξε στο συμπέρασμα ότι οι συμβατικές εφαρμογές αντιμετωπίζουν ακραίες προκλήσεις κατά την αποθήκευση και ανάλυση μεγάλων δεδομένων. Επιπλέον, προκύπτουν ανησυχίες για την προστασία της ιδιωτικής ζωής και την ασφάλεια, με πιθανά τρωτά σημεία στα συστήματα κρυπτογράφησης, τα τείχη προστασίας, τα δικαιώματα πρόσβασης και την ασφάλεια επιπέδου μεταφοράς. Η προέλευση των δεδομένων μπορεί να είναι ασαφής και ακόμη και ανώνυμα δεδομένα μπορούν να επαναπροσδιοριστούν. Η έρευνα υπογράμμισε ότι τα ζητήματα ιδιωτικότητας και ασφάλειας που σχετίζονται με τα μεγάλα δεδομένα χρήζουν περαιτέρω διερεύνησης. Παρά την αυξημένη εστίαση στην έρευνα και την ανάπτυξη, μια πρόσφατη ανάλυση το 2018 αποκάλυψε ότι η προστασία της ιδιωτικής ζωής και η ασφάλεια παραμένουν σημαντικές προκλήσεις για τα Μεγάλα Δεδομένα και την εφαρμογή τους (Milan et al., 2018).

2.7 Προεπεξεργασία

Η προεπεξεργασία στέκεται ως κρίσιμο συστατικό της Εξόρυξης Δεδομένων. Αυτό οφείλεται κυρίως στο γεγονός ότι οι βάσεις δεδομένων του πραγματικού κόσμου επηρεάζονται σημαντικά από θόρυβο, ελλιπείς τιμές, αντικρουόμενα και περιττά δεδομένα. Η

προεπεξεργασία περιλαμβάνει ένα σύνολο στρατηγικών, συμπεριλαμβανομένης της επιλογής χαρακτηριστικών, του χειρισμού ελαττωματικών δεδομένων, της διαχείρισης ανισόρροπης μάθησης κ.λπ. Τα μέτρα αυτά εφαρμόζονται πριν από τη χρήση τεχνικών εξόρυξης δεδομένων και αποτελούν βασική πτυχή στην περίπλοκη διαδικασία της ανακάλυψης γνώσης από δεδομένα (Garcia et al.,2016).

2.8 Οπτικοποίηση

Μια ουσιαστική πτυχή της ανάλυσης μεγάλων δεδομένων περιλαμβάνει την αποτελεσματική οπτικοποίηση των αποτελεσμάτων της εξέτασης των δεδομένων. Δεδομένου του τεράστιου μεγέθους των δεδομένων, η δημιουργία φιλικών προς τον χρήστη οπτικοποιήσεων καθίσταται εξαιρετικά δύσκολο έργο. Ο πρωταρχικός στόχος της οπτικοποίησης δεδομένων είναι η πληρέστερη μεταφορά πληροφοριών, χρησιμοποιώντας διάφορες τεχνικές από τη θεωρία γραφημάτων. Η γραφική απεικόνιση δημιουργεί έναν κρίσιμο σύνδεσμο μεταξύ των δεδομένων και της ακριβούς ερμηνείας τους. Οι διαδικτυακές αγορές όπως η Amazon και η E-cone, με την εκτεταμένη βάση χρηστών και τα δισεκατομμύρια προϊόντα τους, παράγουν σημαντικά δεδομένα. Για να το αντιμετωπίσουν αυτό, ορισμένοι οργανισμοί χρησιμοποιούν ένα εργαλείο που ονομάζεται Tableau για την οπτικοποίηση μεγάλων δεδομένων. Το Tableau διαθέτει την ικανότητα να μεταφράζει μεγάλους όγκους πολύπλοκων δεδομένων σε διαισθητικές οπτικές απεικονίσεις, βοηθώντας:

- Ερμηνεία της συνάφειας αναζήτησης
- Παρακολούθηση των πιο πρόσφατων ανατροφοδοτήσεων των πελατών
- Ανάλυση του συναισθήματος των πελατών.

2.9 Ανάλυση

Η ανάλυση μεγάλων δεδομένων περιλαμβάνει την εξέταση μεγάλων συνόλων δεδομένων, αποκαλύπτοντας κρυφές τάσεις, ασαφείς σχέσεις και άλλες πολύτιμες πληροφορίες για τη βελτίωση της λήψης αποφάσεων. Η διενέργεια τέτοιων αναλύσεων σε εκτεταμένες ποσότητες αδόμητων, ημιδομημένων και δομημένων δεδομένων απαιτεί υψηλό επίπεδο τεχνικής εμπειρογνωμοσύνης. Επιπλέον, παραμένει ασαφές πώς μια βέλτιστη αρχιτεκτονική για

τεχνικές ανάλυσης θα πρέπει να χειρίζεται αποτελεσματικά ιστορικές πληροφορίες και δεδομένα πραγματικού χρόνου ταυτόχρονα (Mishra et al.,2015).

2.10 Ανάλυση σε πραγματικό χρόνο

Οι υπηρεσίες ανάλυσης σε πραγματικό χρόνο αποσκοπούν στον εντοπισμό των βαθύτερων αιτιών των επιχειρησιακών και λειτουργικών προβλημάτων και εξαιρέσεων. Η επεξεργασία της ανάλυσης σε πραγματικό χρόνο μπορεί να περιλαμβάνει μεμονωμένες ή πολλαπλές ολοκληρωμένες αναλυτικές υπηρεσίες, οι οποίες βασίζονται σε πληροφορίες που λαμβάνονται σε πραγματικό χρόνο από τη διαδικασία ανακάλυψης καταστάσεων και άλλα αποθηκευμένα εκτός σύνδεσης μεγάλα δεδομένα, όπως χάρτες, προηγούμενες συναλλαγές, καταστάσεις και αποφάσεις. Ο χειρισμός των αποθηκευμένων μεγάλων δεδομένων τυπικά δημιουργεί προκλήσεις (Mohamed et al.,2014). Αυτές οι υπηρεσίες απαιτούν την ανάπτυξη γρήγορων αλγορίθμων που παρέχουν εναλλακτικές επιλογές εντός περιορισμένου χρονικού πλαισίου. Αυτές οι εναλλακτικές λύσεις μπορεί να είναι είτε βέλτιστες είτε ημι-βέλτιστες λόγω των περιορισμών του χρόνου και των διαθέσιμων πόρων.

Οι προκλήσεις των μεγάλων δεδομένων που αναφέρθηκαν παραπάνω συνοψίζονται στον πίνακα II.

Πίνακας 2. Οι προκλήσεις των Μεγάλων Δεδομένων [Ashabi. (2020)]

Category	Big Data Challenges
Data	Complexity of Data
	Heterogeneity of Data
	Quality of Data
	Scalability of Data
Management	Storage
	Management
	Security
	Privacy
Analysis process	Preprocessing
	Visualization
	Analytics
	Real-time analysis

Κεφάλαιο 3^ο

3. Τεχνολογίες των Big Data

3.1 Data Analytics

Τα Big Data Analytics είναι η επιστήμη της εξέτασης ή της ανάλυσης μεγάλων συνόλων δεδομένων με μια ποικιλία τύπων δεδομένων, δηλαδή, δομημένα, ημι-δομημένα ή μη δομημένα δεδομένα, τα οποία μπορεί να είναι δεδομένα ροής ή δέσμης δεδομένων. Τα Data analytics επιτρέπουν τη λήψη καλύτερων αποφάσεων, την εύρεση νέων επιχειρηματικών ευκαιριών, τον ανταγωνισμό ενάντι επιχειρηματικών αντιπάλων, τη βελτίωση της απόδοσης και της αποτελεσματικότητας και τη μείωση του κόστους χρησιμοποιώντας προηγμένες τεχνικές ανάλυσης δεδομένων. Τα Big Data analytics επικεντρώνονται στην εξαγωγή σημαντικών πληροφοριών χρησιμοποιώντας αποτελεσματικούς αλγόριθμους στα δεδομένα που συλλαμβάνονται για την επεξεργασία, την ανάλυση και την οπτικοποίηση τους. Αυτό περιλαμβάνει τη διαμόρφωση του αποτελεσματικού αλγορίθμου και συστήματος για την ενσωμάτωση δεδομένων, αναλύοντας τις πληροφορίες που παράγονται έτσι ώστε να δημιουργηθούν άμεσες επιχειρηματικές λύσεις. Για παράδειγμα, στο ηλεκτρονικό εμπόριο, η ανάλυση του τεράστιου όγκου δεδομένων που παράγεται από διαδικτυακές συναλλαγές είναι το κλειδί για την ενίσχυση της αντίληψης των εμπόρων για τη συμπεριφορά των πελατών και των προτύπων αγορών για τη λήψη κατάλληλων επιχειρηματικών αποφάσεων. Ομοίως, σε πολλές σελίδες του Facebook εμφανίζονται κατάλληλες διαφημίσεις, μέσω της ανάλυσης αναρτήσεων στο Facebook, φωτογραφιών και ούτω καθεξής. Όταν χρησιμοποιείτε πιστωτικές κάρτες, οι πάροχοι πιστωτικών καρτών χρησιμοποιούν έναν έλεγχο ανίχνευσης απάτης για να επιβεβαιώσουν ότι η συναλλαγή είναι νόμιμη. Η βαθμολογία πίστωσης των πελατών αναλύεται από χρηματοπιστωτικά ιδρύματα για να προβλέψει εάν ο αιτών θα χρεοκοπήσει. Συνοψίζοντας, το αντίκτυπο και η σημασία των analytics έχει φτάσει σε μεγάλο βαθμό με περισσότερα δεδομένα να συλλέγονται συνεχώς. Τα Analytics θα συνεχίσουν να αυξάνονται έως ότου υπάρξει κάποιο στρατηγικό αντίκτυπο στην αντίληψη των κρυφών πληροφοριών από τα δεδομένα.

Τα analytics χωρίζονται σε 4 κατηγορίες. Αυτές οι κατηγορίες είναι οι εξής:

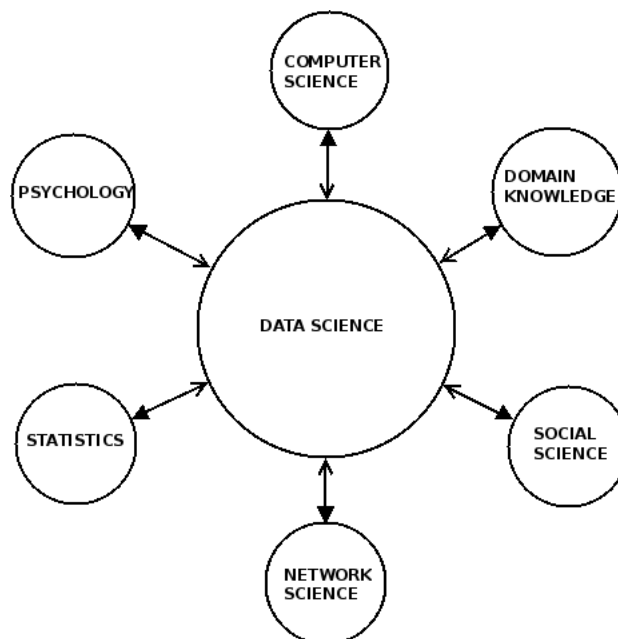
- Descriptive Analytics: Πληροφορίες για το παρελθόν.
- Diagnostic Analytics: Κατανόηση του τι συμβαίνει και γιατί συμβαίνει.
- Predictive Analytics: Κατανόηση του μέλλοντος.
- Prescriptive Analytics: Συμβουλές για πιθανά αποτελέσματα.

3.2 Εργαλεία για τη συλλογή, προεπεξεργασία και ανάλυση μεγάλων δεδομένων

Η παρούσα ενότητα παρέχει μια επισκόπηση των τεχνολογιών που υιοθετούνται για την ανάλυση μεγάλων δεδομένων.

Η προσιτή τιμή της υψηλής υπολογιστικής χωρητικότητας επέτρεψε στους ερευνητές να εμβαθύνουν στις δυνατότητες των μεγάλων δεδομένων στο πλαίσιο της επιστήμης των δεδομένων. Η συλλογή δεδομένων, η οποία δεν περιορίζεται πλέον στην απλή τήρηση αρχείων, περιλαμβάνει πλέον την εξερεύνηση μέσω ευφών συστημάτων για την αποκάλυψη απρόβλεπτων γνώσεων. Χωρίς την παρέμβαση εμπειρογνομώνων, η διατύπωση και η τελειοποίηση ενδιαφέροντων ερωτημάτων ενδυναμώνουν αυτά τα συστήματα να ανακαλύπτουν αυτόνομα νέες πληροφορίες (Dhar, 2013).

Οι τεχνολογικές εξελίξεις στον τομέα αυτό συμβάλλουν στην ανάλυση της οικοδόμησης σεναρίων, βρίσκοντας εφαρμογές σε διάφορους τομείς όπως η αστρονομία, η υγειονομική περίθαλψη και οι τηλεπικοινωνίες. Ωστόσο, παρά τα οφέλη της, η ανάλυση μεγάλων δεδομένων αντιμετωπίζει προκλήσεις και περιορισμούς, με τα ζητήματα ασφάλειας και προστασίας της ιδιωτικής ζωής να δημιουργούν σημαντικές ανησυχίες στους ερευνητές. Οι εξελίξεις αυτές, μαζί με τις σχετικές προκλήσεις, έχουν δώσει το έναυσμα για τη δημιουργία ενός αναπτυσσόμενου διεπιστημονικού τομέα, γνωστού ως επιστήμη των δεδομένων. Ο τομέας αυτός αντλεί γνώσεις από διάφορα γνωστικά αντικείμενα, όπως η ψυχολογία, η στατιστική, τα οικονομικά, οι κοινωνικές επιστήμες, η επιστήμη των δικτύων και η επιστήμη των υπολογιστών (βλ. Σχήμα 3).



Εικόνα 4.Εξέλιξη της Επιστήμης Δεδομένων (Google)

Διάφορες ψηφιακές συσκευές, όπως κινητές συσκευές, κάμερες, φορητές συσκευές και έξυπνα ρολόγια, μαζί με εφαρμογές, χρησιμεύουν ως εργαλεία για τη συλλογή δεδομένων. Αυτά τα εργαλεία παράγουν εκτεταμένα δεδομένα με τη μορφή αρχείων καταγραφής, κειμένου, ηχογραφήσεων φωνής, εικόνων και βίντεο. Για τον αποτελεσματικό χειρισμό και την ανάλυση αυτών των ποικίλων δεδομένων, οι ερευνητές αναπτύσσουν νέες τεχνικές που βελτιώνουν την αναπαράσταση των αδόμητων δεδομένων. Η προσέγγιση αυτή καθίσταται ιδιαίτερα σημαντική στο πλαίσιο των μεγάλων δεδομένων, επιτρέποντας την εξαγωγή πολύτιμων πληροφοριών που μπορεί να μην είχαν προβλεφθεί προηγουμένως.

3.2.1 NoSQL

Η συμβατική μέθοδος για τη διαχείριση δομημένων δεδομένων περιλαμβάνει τη χρήση Σχεσιακών Συστημάτων Διαχείρισης Βάσεων Δεδομένων (RDBMS), συμπεριλαμβανομένων των Δομημένων Γλωσσών Ερωτήσεων (SQL). Τα RDBMS χρησιμοποιούν μια σχεσιακή βάση δεδομένων και ένα σχήμα για την αποθήκευση και την ανάκτηση δεδομένων, χρησιμοποιώντας συχνά μια αποθήκη δεδομένων για μεγάλα σύνολα δεδομένων. Η ευρέως χρησιμοποιούμενη SQL διευκολύνει τα ερωτήματα στη βάση δεδομένων, με τα δεδομένα να αποθηκεύονται σε μια αποθήκη δεδομένων μέσω διαστασιοποιημένων και κανονικοποιημένων προσεγγίσεων (Bakshi, 2012). Η προσέγγιση διαστάσεων περιλαμβάνει τη διαίρεση των δεδομένων σε πίνακες γεγονότων και διαστάσεων, ενώ η κανονικοποιημένη προσέγγιση συνεπάγεται τη διαίρεση των δεδομένων σε οντότητες, δημιουργώντας πολλαπλούς πίνακες σε μια σχεσιακή βάση δεδομένων.

Ωστόσο, λόγω των περιορισμών της Ατομικότητας, Συνέπειας, Απομόνωσης και Ανθεκτικότητας (ACID), η κλιμάκωση μεγάλου όγκου δεδομένων καθίσταται δύσκολη στο πλαίσιο των RDBMS. Επιπλέον, τα RDBMS δυσκολεύονται με ημιδομημένα και αδόμητα δεδομένα (Qin et al., 2012- Mukherjee et al., 2012- Zikopoulos et al., 2012), οδηγώντας στην εμφάνιση των NoSQL βάσεων δεδομένων.

Οι βάσεις δεδομένων NoSQL ειδικεύονται στην αποθήκευση και τη διαχείριση μη δομημένων δεδομένων και συχνά αναφέρονται ως βάσεις δεδομένων "χωρίς σχήμα". Επιτρέπουν τη γρήγορη αναβάθμιση της δομής χωρίς επανεγγραφή πινάκων και υποστηρίζουν διάφορα μοντέλα, όπως αποθήκευση εγγράφων, αποθήκευση κλειδιών-τιμών, BigTable και βάσεις δεδομένων γράφων. Οι NoSQL υιοθετούν ένα πιο χαλαρό μοντέλο συνέπειας σε σύγκριση με τις παραδοσιακές βάσεις δεδομένων και οι λειτουργίες διαχείρισης και αποθήκευσης των

δεδομένων τους είναι ξεχωριστές, επιτρέποντας κλιμακούμενες λύσεις δεδομένων. Παραδείγματα βάσεων δεδομένων NoSQL περιλαμβάνουν τις HBase, MongoDB και Dynamo.

3.2.2 Hadoop

Η έναρξη και η υλοποίηση του έργου ανοικτού κώδικα Apache Hadoop πραγματοποιήθηκε το 2005, αντλώντας έμπνευση από το σύστημα αρχείων της Google και το προγραμματιστικό παράδειγμα του MapReduce (Prekopcsák et al., 2011- Bakshi, 2012- Minelli et al., 2013).

3.2.3 Hadoop Distributed File System (HDFS)

Το HDFS χρησιμεύει ως ένα ανεκτικό σε σφάλματα, κλιμακούμενο και εκτενώς παραμετροποιήσιμο σύστημα κατανεμημένης αποθήκευσης σε ένα σύμπλεγμα Hadoop. Διαχωρίζει τα δεδομένα εντός της συστάδας Hadoop σε τμήματα, διανέμοντάς τα σε διάφορους διακομιστές. Κάθε διακομιστής αποθηκεύει ένα μικρό τμήμα του πλήρους συνόλου δεδομένων.

3.2.4 Hadoop MapReduce

Το MapReduce λειτουργεί ως ένα πλαίσιο λογισμικού που έχει σχεδιαστεί για την κατανεμημένη επεξεργασία εκτεταμένων συνόλων δεδομένων με αξιόπιστο και ανεκτικό σε σφάλματα τρόπο. Η διαδικασία περιλαμβάνει δύο διακριτές φάσεις:

Φάση χαρτογράφησης (Map Phase): Κατά τη διάρκεια αυτής της φάσης, ο συνολικός φόρτος εργασίας διαιρείται σε μικρότερους επιμέρους φόρτους εργασίας, με εργασίες που ανατίθενται σε χαρτογράφους. Κάθε χαρτογράφος επεξεργάζεται ένα συγκεκριμένο μοναδιαίο μπλοκ δεδομένων, δημιουργώντας έναν ταξινομημένο κατάλογο ζευγών (key,value). Αυτό το αποτέλεσμα, γνωστό ως έξοδος του χαρτογράφου, υφίσταται ανακάτεμα και περνά στην επόμενη φάση.

Φάση μείωσης (Reduce phase): Σε αυτό το στάδιο, το αποτέλεσμα της χαρτογράφησης αναλύεται και συγχωνεύεται για την παραγωγή της τελικής εξόδου, η οποία στη συνέχεια εγγράφεται στο κατανεμημένο σύστημα αρχείων Hadoop (HDFS) στη συστάδα.

Ο Πίνακας 3 παρέχει μια σύνοψη των δυνατοτήτων που σχετίζονται με τα μεγάλα δεδομένα και τις κύριες διαθέσιμες τεχνολογίες (Sun and Heller, 2012).

Πίνακας 3. Δυνατότητες & Τεχνολογίες των Big Data

Big Data Capability	Primary Technology	Features
Δυνατότητα αποθήκευσης και διαχείρισης	Hadoop Distributed File System (HDFS)	Κατανεμημένο σύστημα αρχείων ανοικτού κώδικα. Τρέχει σε υλικό βασικών προϊόντων υψηλής απόδοσης, Ιδιαίτερα κλιμακούμενη αποθήκευση και αυτόματη αντιγραφή δεδομένων.
Δυνατότητα βάσης δεδομένων	Oracle NoSQL	Δυναμική και ευέλικτη σχεδίαση σχήματος, υψηλή κλιμάκωση πολλαπλών κόμβων, πολλαπλών κέντρων δεδομένων, ανοχή σε σφάλματα, λειτουργίες ACID, βάση δεδομένων ζεύγους key-value υψηλής απόδοσης.
	Apache HBase	Αυτόματη υποστήριξη failover μεταξύ διακομιστών Region. Αυτόματη και παραμετροποιήσιμη κατανομή πινάκων.
	Apache Cassandra	Δυνατότητα ανοχής σφαλμάτων για κάθε κόμβο. Ευρετήρια στηλών με την απόδοση των ενημερώσεων με δομημένο αρχείο καταγραφής και ενσωματωμένη προσωρινή αποθήκευση δεδομένων.
	Apache Hive	Εκτέλεση ερωτημάτων μέσω MapReduce, Χρήση γλώσσας HiveQL που μοιάζει με SQL.Εύκολη διαδικασία ETL είτε από HDFS είτε από Apache HBase.
Ικανότητα επεξεργασίας	MapReduce	Κατανομή του φόρτου εργασίας δεδομένων σε χιλιάδες κόμβους. Διάσπαση του προβλήματος σε μικρότερα υποπροβλήματα.
	Apache Hadoop	Ιδιαίτερα προσαρμόσιμη υποδομή, Ιδιαίτερα κλιμακούμενη παράλληλη επεξεργασία παρτίδων.Ανοχή σε σφάλματα.
Δυνατότητα ενσωμάτωσης δεδομένων	Oracle big data connectors, Oracle data integrator	Εξάγει αποτελέσματα MapReduce σε RDBMS, Hadoop και άλλους στόχους. Περιλαμβάνει γραφική διεπαφή χρήστη

Παρά τα πλεονεκτήματα του Hadoop έναντι των RDBMS, αντιμετωπίζει αρκετούς περιορισμούς, όπως περιγράφει η ParAccel το 2012:

Αντιγραφή δεδομένων: Το Hadoop Distributed File System (HDFS) έχει ως αποτέλεσμα τη δημιουργία πολλαπλών αντιγράφων δεδομένων.

Περιορισμένη υποστήριξη SQL: Το Hadoop παρέχει μόνο περιορισμένη υποστήριξη SQL και στερείται βασικών λειτουργιών, όπως οι υποερωτήσεις και οι αναλύσεις "group by".

Αναποτελεσματική εκτέλεση: Η απουσία βελτιστοποιητή ερωτημάτων οδηγεί σε αναποτελεσματικά σχέδια εκτέλεσης με βάση το κόστος, με αποτέλεσμα την ανάγκη για μεγαλύτερες συστάδες σε σύγκριση με παρόμοιες βάσεις δεδομένων.

Απαιτητικό πλαίσιο: Το πλαίσιο MapReduce δεν ευνοεί τον αποτελεσματικό χειρισμό σύνθετης μετασχηματιστικής λογικής.

Απαιτήσεις δεξιοτήτων: Η επιτυχής εφαρμογή του Hadoop απαιτεί γνώσεις αλγορίθμων και δεξιότητες στην ανάπτυξη κατανεμημένων MapReduce.

3.3 Λογισμικά για το Χειρισμό Μεγάλων Δεδομένων

Πολυάριθμα εργαλεία διευκολύνουν τους επιστήμονες δεδομένων στην επεξεργασία και την ανάλυση δεδομένων. Μια πληθώρα νέων γλωσσών, πλαισίων και τεχνολογιών αποθήκευσης δεδομένων έχουν εμφανιστεί για την υποστήριξη της διαχείρισης μεγάλων δεδομένων.

3.3.1 R

Γλώσσα στατιστικών υπολογισμών ανοικτού κώδικα που διαδραματίζει σημαντικό ρόλο προσφέροντας ένα ευρύ φάσμα στατιστικών και γραφικών τεχνικών για την εξαγωγή συμπερασμάτων από τα δεδομένα. Διαθέτει μια αποτελεσματική δυνατότητα χειρισμού και αποθήκευσης δεδομένων, υποστηρίζοντας διανυσματικές λειτουργίες με μια σειρά από τελεστές που ενισχύουν την ταχύτητα επεξεργασίας. Παρουσιάζοντας τα χαρακτηριστικά μιας τυπικής γλώσσας προγραμματισμού, η R υποστηρίζει ορίσματα υπό όρους, βρόχους και συναρτήσεις που ορίζονται από τον χρήστη. Το Comprehensive R Archive Network (CRAN) υποστηρίζει την R με ένα ευρύ φάσμα πακέτων. Συμβατό με πλατφόρμες Windows, Linux και Mac, το R συνοδεύεται από ισχυρή τεκμηρίωση για κάθε πακέτο. Ειδικότερα, υπερέχει σε αλγορίθμους επεξεργασίας δεδομένων, εξόρυξης δεδομένων και μηχανικής μάθησης, επιδεικνύοντας αξιόπαινη υποστήριξη για ανάγνωση και εγγραφή σε κατανομημένα περιβάλλοντα, καθιστώντας την κατάλληλη για το χειρισμό μεγάλων δεδομένων. Ωστόσο, οι προκλήσεις περιλαμβάνουν τη διαχείριση της μνήμης, την ταχύτητα και την αποδοτικότητα.

3.3.2 Python

Χαρακτηρίζεται ως γλώσσα ανοικτού κώδικα και έχει συμβατότητά με τις πλατφόρμες Windows, Linux και Mac. Διαθέτει μια εκτεταμένη συλλογή πακέτων που συνεισφέρει η κοινότητα ή τρίτοι προγραμματιστές. Αξιοσημείωτες ενότητες όπως οι NumPy, Scikit και Pandas είναι καθοριστικής σημασίας για τη μηχανική μάθηση και την εξόρυξη δεδομένων, διευκολύνοντας εργασίες όπως η προεπεξεργασία δεδομένων, ο υπολογισμός και η μοντελοποίηση.

Το NumPy χρησιμεύει ως το θεμελιώδες πακέτο για τον επιστημονικό υπολογισμό, ενισχύοντας την Python με υποστήριξη για ευμεγέθεις, πολυδιάστατους πίνακες.

Το Scikit περιλαμβάνει μια ποικιλία αλγορίθμων για ταξινόμηση, παλινδρόμηση, ομαδοποίηση, μείωση διαστάσεων, επιλογή χαρακτηριστικών, προεπεξεργασία και επιλογή μοντέλου.

Το Pandas παίζει καθοριστικό ρόλο στη συγκέντρωση δεδομένων και την προετοιμασία για την επακόλουθη ανάλυση δεδομένων και τη μοντελοποίηση. Επιπλέον, υπερέχει στην ανάλυση γράφων μέσω της βιβλιοθήκης NetworkX και υποστηρίζει την ανάλυση κειμένου και την επεξεργασία φυσικής γλώσσας με το nltk. Η Python εγκωμιάζεται για τη φιλική προς το χρήστη φύση της, καθιστώντας την ιδανική για γρήγορη και αποτελεσματική ανάλυση προβλημάτων. Επιπλέον, ενσωματώνεται απρόσκοπτα με τη Spark μέσω της βιβλιοθήκης pyspark.

3.3.3 Scala

"Scalable Language", είναι μια αντικειμενοστραφής γλώσσα προγραμματισμού όπου κάθε λειτουργία και αντικείμενο λειτουργεί ως κλήση μεθόδου, ευθυγραμμισμένη με τις αρχές των αντικειμενοστραφών γλωσσών. Για την εκτέλεσή της απαιτείται το περιβάλλον της Java Virtual Machine. Η Spark, ένα πλαίσιο υπολογισμού συστάδων που λειτουργεί στη μνήμη, υλοποιείται στη Scala. Η Scala κερδίζει ολοένα και περισσότερο έδαφος ως εργαλείο προγραμματισμού για την αντιμετώπιση των προκλήσεων που σχετίζονται με τα μεγάλα δεδομένα.

3.3.4 Apache Spark

Είναι μια τεχνολογία υπολογιστών υψηλής ταχύτητας που έχει σχεδιαστεί για επεξεργασία σε συστοιχίες. Χρησιμοποιεί το Hadoop για αποθήκευση και ενσωματώνει τη δυνατότητα διαχείρισης συστάδων του. Η πλατφόρμα προσφέρει ενσωματωμένα APIs για Java, Scala και Python, ενώ πιο πρόσφατα έχει επεκτείνει την υποστήριξή της και στην R. Με 80 υψηλού επιπέδου τελεστές για διαδραστικές επερωτήσεις, η Spark βασίζεται στο πλαίσιο Resilient Distributed Data (RDD) για υπολογισμούς στη μνήμη. Αυτό το πλαίσιο κατανέμει το πλαίσιο δεδομένων σε μικρότερα τμήματα σε διαφορετικές μηχανές, ενισχύοντας την ταχύτητα υπολογισμού. Επιπλέον, το Spark υποστηρίζει λειτουργίες Map και Reduce για την επεξεργασία δεδομένων, μαζί με δυνατότητες για SQL, ροή δεδομένων, αλγόριθμους επεξεργασίας γράφων και αλγόριθμους μηχανικής μάθησης. Αν και είναι προσβάσιμο μέσω των Python, Java και R, το Spark παρουσιάζει ισχυρή υποστήριξη για τη Scala και επί του παρόντος επιδεικνύει μεγαλύτερη σταθερότητα. Επιτρέπει επίσης τη βαθιά μάθηση μέσω της ενσωμάτωσης της Spark στο H2O.

3.3.5 Apache Hive

Πλατφόρμα ανοικτού κώδικα που έχει σχεδιαστεί για την αναζήτηση και διαχείριση εκτεταμένων συνόλων δεδομένων που βρίσκονται σε κατανομημένα συστήματα αποθήκευσης, όπως το HDFS. Αναφέρεται ως HiveQL και λειτουργεί παρόμοια με την SQL. Για την επεξεργασία των ερωτημάτων, χρησιμοποιεί MapReduce και προσφέρει στους προγραμματιστές την ευελιξία να ενσωματώνουν προσαρμοσμένους κώδικες mapper και reducer όταν η έκφραση της επιθυμητής λογικής μέσω της HiveQL αποτελεί πρόκληση.

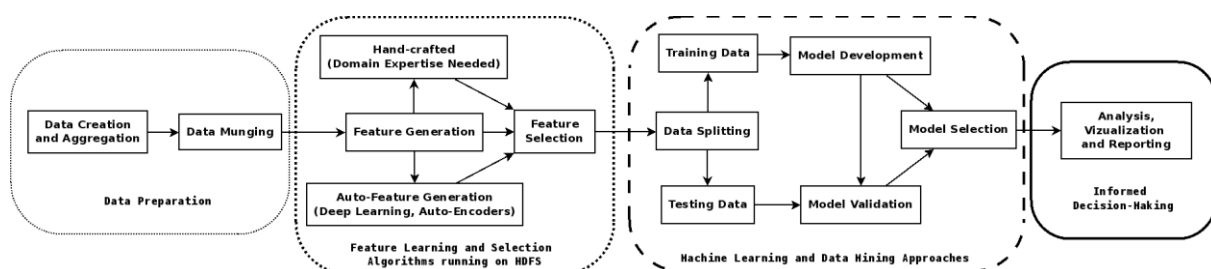
3.3.6 Apache Pig

Πλατφόρμα που επιτρέπει στους αναλυτές να αναλύουν μεγάλα σύνολα δεδομένων. Χρησιμοποιεί μια γλώσσα προγραμματισμού υψηλού επιπέδου, γνωστή ως Pig Latin, για τη δημιουργία προγραμμάτων MapReduce, βασισμένη στο Hadoop για την αποθήκευση δεδομένων. Ο κώδικας Pig Latin μπορεί να επεκταθεί χρησιμοποιώντας User-Defined Functions γραμμένες σε γλώσσες όπως η Java ή η Python. Η Pig είναι κατάλληλη για ουσιαστικό παραλληλισμό, επιτρέποντας τον αποτελεσματικό χειρισμό πολύ μεγάλων συνόλων δεδομένων.

3.3.7 Amazon Elastic Compute Cloud (EC2)

Είναι μια διαδικτυακή υπηρεσία cloud που παρέχει υπολογιστική χωρητικότητα. Προσφέρει πλήρη έλεγχο των υπολογιστικών πόρων, επιτρέποντας στους προγραμματιστές να εκτελούν υπολογισμούς στο υπολογιστικό περιβάλλον που προτιμούν. Λειτουργώντας σε ένα μοντέλο pay-as-you-go, το EC2 έχει γίνει μια από τις πιο επιτυχημένες πλατφόρμες υπολογιστικού νέφους.

Διάφορα άλλα λογισμικά που υποστηρίζουν μεγάλα δεδομένα είναι τα MongoDB, BlinkDB, Tachyon, Cassandra, CouchDB, Clojure, Tableau, Splunk και άλλα.



Εικόνα 5. Διαδικασίες Μεγάλων Δεδομένων [Bhadani, A., Jothimani, D. (2016)]

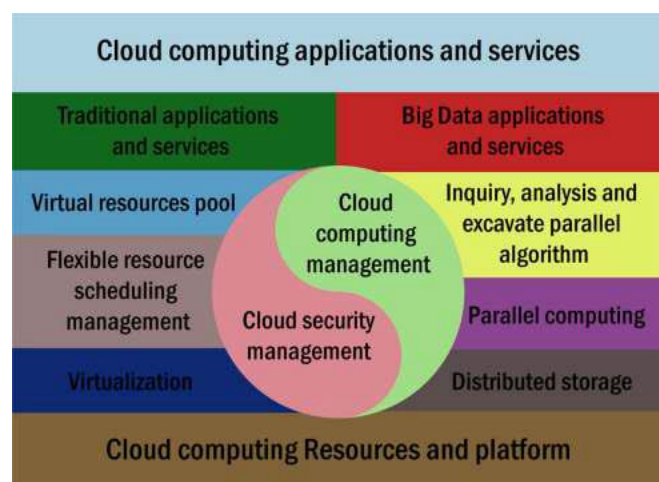
Κεφάλαιο 4^ο

4. Big Data & Νέες Τεχνολογίες

4.1 Big Data και Cloud Computing

Το cloud computing και τα μεγάλα δεδομένα είναι στενά συνδεδεμένα. Τα μεγάλα δεδομένα, τα οποία υποβάλλονται σε λειτουργίες έντασης υπολογισμού, ασκούν σημαντική πίεση στις δυνατότητες αποθήκευσης ενός συστήματος cloud. Ο πρωταρχικός στόχος του υπολογιστικού νέφους είναι η αποτελεσματική χρήση εκτεταμένων πόρων υπολογιστών και αποθήκευσης μέσω κεντρικής διαχείρισης, καλύπτοντας τις περίπλοκες υπολογιστικές ανάγκες των εφαρμογών μεγάλων δεδομένων. Η εξέλιξη του υπολογιστικού νέφους προσφέρει λύσεις τόσο για προκλήσεις αποθήκευσης όσο και επεξεργασίας που δημιουργούνται από τα μεγάλα δεδομένα. Αντίθετα, η έλευση των μεγάλων δεδομένων επιταχύνει επίσης την πρόοδο του υπολογιστικού νέφους.

Παρά τις πολυάριθμες αλληλεπικαλυπτόμενες τεχνολογίες μεταξύ του υπολογιστικού νέφους και των μεγάλων δεδομένων, διαφέρουν σε δύο σημαντικές πτυχές. Πρώτον, διαφέρουν εννοιολογικά, με το cloud computing να αναδιαμορφώνει την αρχιτεκτονική IT και τα μεγάλα δεδομένα να επηρεάζουν τη λήψη επιχειρηματικών αποφάσεων. Ωστόσο, τα μεγάλα δεδομένα βασίζονται στο cloud computing ως τη θεμελιώδη υποδομή του για απρόσκοπτη λειτουργία. Δεύτερον, οι πελάτες-στόχοι τους ποικίλλουν, με το cloud computing να εξυπηρετεί τους Chief Information Officers (CIO) ως προηγμένη λύση πληροφορικής, ενώ τα μεγάλα δεδομένα στοχεύουν στους Chief Executive Officers (CEOs) που εστιάζουν στην ενίσχυση των επιχειρηματικών λειτουργιών.



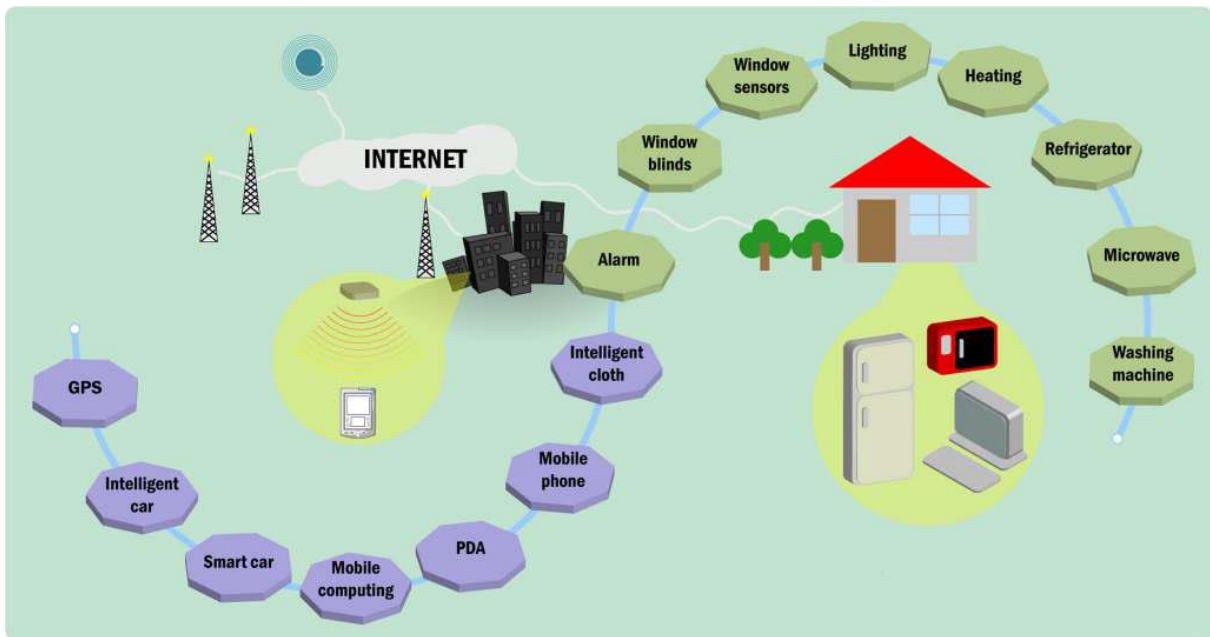
Εικόνα 6. Βασικά στοιχεία του υπολογιστικού νέφους [Chen (2014)]

Καθώς οι υπεύθυνοι λήψης αποφάσεων αντιμετωπίζουν άμεση πίεση ανταγωνισμού στην αγορά, πρέπει να ξεπεράσουν τους αντιπάλους τους με πιο ανταγωνιστικούς τρόπους. Η ενοποίηση των μεγάλων δεδομένων και του υπολογιστικού νέφους συνεχίζει να ενισχύεται, με το cloud computing να παρέχει πόρους σε επίπεδο συστήματος παρόμοιους με υπολογιστές και λειτουργικά συστήματα, και μεγάλα δεδομένα που λειτουργούν στο ανώτερο επίπεδο που υποστηρίζονται από το cloud computing, προσφέροντας λειτουργίες παρόμοιες με βάσεις δεδομένων και αποτελεσματική ικανότητα επεξεργασίας δεδομένων . Σύμφωνα με τον Κίσινγκερ, Πρόεδρο της EMC, η εφαρμογή μεγάλων δεδομένων απαιτεί μια βάση στο cloud computing.

Η εξέλιξη των μεγάλων δεδομένων πηγάζει από τις αυξανόμενες απαιτήσεις των εφαρμογών, ενώ το cloud computing έχει εξελιχθεί από εικονικοποιημένες τεχνολογίες. Κατά συνέπεια, το cloud computing όχι μόνο παρέχει δυνατότητες υπολογισμού και επεξεργασίας για μεγάλα δεδομένα, αλλά χρησιμεύει και ως υπηρεσία από μόνο του. Σε κάποιο βαθμό, οι εξελίξεις στο cloud computing προωθούν περαιτέρω την ανάπτυξη μεγάλων δεδομένων, δημιουργώντας μια αμοιβαία επωφελή σχέση μεταξύ των δύο τεχνολογιών.

4.2 Big Data και Internet of Things

Στο παράδειγμα του Διαδικτύου των Πραγμάτων (IoT), ένα εκτεταμένο δίκτυο αισθητήρων είναι ενσωματωμένο σε διάφορες συσκευές και μηχανές του πραγματικού κόσμου. Αυτοί οι αισθητήρες, που αναπτύσσονται σε διαφορετικά πεδία, συλλέγουν μια σειρά τύπων δεδομένων, συμπεριλαμβανομένων περιβαλλοντικών, γεωγραφικών, αστρονομικών και λογιστικών δεδομένων. Οι κινητές συσκευές, τα συστήματα μεταφοράς, οι δημόσιες εγκαταστάσεις και οι οικιακές συσκευές μπορούν όλα να χρησιμεύσουν ως συσκευές απόκτησης δεδομένων εντός του IoT, όπως απεικονίζεται στην Εικ. 7.



Εικόνα 7.Απεικόνιση εξοπλισμού απόκτησης δεδομένων στο IoT [Chen (2014)]

Τα μεγάλα δεδομένα που παράγονται από το IoT διαθέτουν ξεχωριστά χαρακτηριστικά σε σύγκριση με τα γενικά μεγάλα δεδομένα λόγω των διαφορετικών τύπων δεδομένων που συλλέγονται. Τα κλασικά χαρακτηριστικά περιλαμβάνουν την ετερογένεια, την ποικιλία, την αδόμητη φύση, τον θόρυβο και τον υψηλό πλεονασμό. Αν και τα τρέχοντα δεδομένα IoT δεν κυριαρχούν στα μεγάλα δεδομένα, οι προβλέψεις της HP προβλέπουν ότι έως το 2030, η ποσότητα των αισθητήρων θα φτάσει το ένα τρισεκατομμύριο, καθιστώντας τα δεδομένα IoT το πιο σημαντικό συστατικό των μεγάλων δεδομένων. Η αναφορά της Intel υπογραμμίζει τρία χαρακτηριστικά των μεγάλων δεδομένων στο IoT που ευθυγραμμίζονται με το παράδειγμα των μεγάλων δεδομένων: (i) την αφθονία των τερματικών που παράγουν τεράστια δεδομένα. (ii) ημιδομημένη ή αδόμητη φύση των δεδομένων που δημιουργούνται από το IoT. και (iii) τη χρησιμότητα των δεδομένων IoT μόνο όταν αναλύονται.

Επί του παρόντος, η ικανότητα επεξεργασίας δεδομένων του IoT υστερεί σε σχέση με τα δεδομένα που συλλέγονται, γεγονός που καθιστά αναγκαία την επείγουσα υιοθέτηση τεχνολογιών μεγάλων δεδομένων για την προώθηση της ανάπτυξης του IoT. Πολλοί φορείς εκμετάλλευσης IoT αναγνωρίζουν τη σημασία των μεγάλων δεδομένων, αναγνωρίζοντας ότι η επιτυχία του IoT βασίζεται στην αποτελεσματική ενσωμάτωση μεγάλων δεδομένων και υπολογιστικού νέφους. Η ευρεία εξάπλωση του IoT είναι έτοιμη να εισαγάγει πολλές πόλεις στην εποχή των μεγάλων δεδομένων.

Είναι ευρέως αποδεκτό ότι αυτές οι δύο τεχνολογίες είναι αλληλεξαρτώμενες και πρέπει να αναπτυχθούν από κοινού. Από τη μία πλευρά, η εκτεταμένη ανάπτυξη του IoT οδηγεί σε σημαντική ανάπτυξη τόσο στην ποσότητα όσο και στην ποικιλία των δεδομένων, δημιουργώντας ευκαιρίες για εφαρμογή και ανάπτυξη μεγάλων δεδομένων. Από την άλλη πλευρά, η εφαρμογή της τεχνολογίας μεγάλων δεδομένων στο IoT επιταχύνει τις ερευνητικές προόδους και τα επιχειρηματικά μοντέλα στο τοπίο του IoT.

4.3 Big Data και Artificial Intelligence

Τα μεγάλα δεδομένα και η τεχνητή νοημοσύνη έχουν μια συνεργιστική σχέση. Η τεχνητή νοημοσύνη απαιτεί μια τεράστια κλίμακα δεδομένων για την εκμάθηση και τη βελτίωση των διαδικασιών λήψης αποφάσεων και η ανάλυση μεγάλων δεδομένων αξιοποιεί την τεχνητή νοημοσύνη για καλύτερη ανάλυση δεδομένων. Με αυτήν τη σύγκλιση, μπορείτε να αξιοποιήσετε πιο εύκολα προηγμένες δυνατότητες ανάλυσης, όπως επαυξημένη ή πρόβλεψη αναλυτικών στοιχείων και πιο αποτελεσματικά να εμφανίσετε χρήσιμες πληροφορίες από τα τεράστια αποθέματα δεδομένων σας. Με τα αναλυτικά στοιχεία που υποστηρίζονται από τεχνητή νοημοσύνη μεγάλων δεδομένων, μπορείτε να ενδυναμώσετε τους χρήστες σας με τα διαισθητικά εργαλεία και τις ισχυρές τεχνολογίες που χρειάζονται για να εξάγουν insights υψηλής αξίας από δεδομένα, ενισχύοντας τον αλφαριθμητισμό δεδομένων σε ολόκληρο τον οργανισμό σας, ενώ αποκομίζετε τα οφέλη του να γίνετε ένας οργανισμός που βασίζεται πραγματικά στα δεδομένα.

Συνδυάζοντας μεγάλα δεδομένα και τεχνολογία ανάλυσης AI, οι εταιρείες μπορούν να βελτιώσουν την απόδοση και την αποδοτικότητα των επιχειρήσεων με:

- Πρόβλεψη και αξιοποίηση των τάσεων της αναδυόμενης βιομηχανίας και της αγοράς.
- Ανάλυση της συμπεριφοράς των καταναλωτών και αυτοματοποίηση τμηματοποίησης πελατών

- Εξατομίκευση και βελτιστοποίηση της απόδοσης των καμπανιών ψηφιακού μάρκετινγκ
- Χρησιμοποιώντας έξυπνα συστήματα υποστήριξης αποφάσεων που τροφοδοτούνται από μεγάλα δεδομένα, ΑΙ και προγνωστικά αναλυτικά στοιχεία

ΑΙ Big Data Analytics

Η τεχνητή νοημοσύνη μπορεί να βοηθήσει τους χρήστες σε όλες τις φάσεις του μεγάλου κύκλου δεδομένων ή στις διαδικασίες που εμπλέκονται στη συγκέντρωση, αποθήκευση και ανάκτηση διαφορετικών τύπων δεδομένων από διάφορες πηγές. Αυτές περιλαμβάνουν τη διαχείριση δεδομένων, τη διαχείριση προτύπων, τη διαχείριση περιβάλλοντος, τη διαχείριση αποφάσεων, τη διαχείριση ενεργειών, τη διαχείριση στόχων και τη διαχείριση κινδύνου.

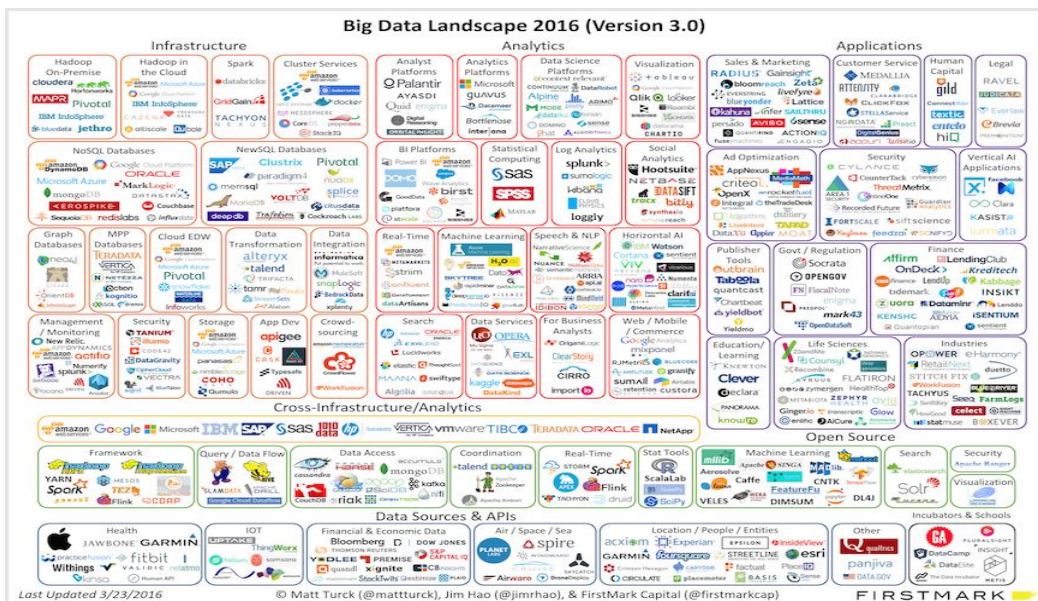
Η τεχνητή νοημοσύνη μπορεί να αναγνωρίσει τύπους δεδομένων, να βρει πιθανές συνδέσεις μεταξύ συνόλων δεδομένων και να αναγνωρίσει τη γνώση χρησιμοποιώντας επεξεργασία φυσικής γλώσσας. Μπορεί να χρησιμοποιηθεί για την αυτοματοποίηση και την επιτάχυνση των εργασιών προετοιμασίας δεδομένων, συμπεριλαμβανομένης της δημιουργίας μοντέλων δεδομένων, και για να βοηθήσει στην εξερεύνηση δεδομένων. Επιπρόσθετα, μπορεί να μάθει κοινά πρότυπα ανθρώπινων σφαλμάτων, εντοπίζοντας και επιλύοντας πιθανές ατέλειες στις πληροφορίες. Επίσης, μπορεί να μάθει παρακολουθώντας πώς ο χρήστης αλληλεπιδρά με ένα πρόγραμμα αναλυτικών στοιχείων, εμφανίζοντας γρήγορα απροσδόκητες πληροφορίες από τεράστια σύνολα δεδομένων. Τελός, μπορεί να ειδοποιεί τους χρήστες για ανωμαλίες ή απροσδόκητα μοτίβα στα δεδομένα, παρακολουθώντας ενεργά συμβάντα και εντοπίζοντας πιθανές απειλές από αρχεία καταγραφής συστήματος ή δεδομένα κοινωνικής δικτύωσης.

Κεφάλαιο 5ο

5. Big Data & Επιχειρηματικότητα

Ο ρόλος των Big Data στην τέταρτη βιομηχανική επανάσταση θα είναι πολύ σημαντικός (H.Kagermann, 2013). Μέχρι σήμερα, οι επιχειρήσεις χρησιμοποιούσαν σχεσιακές βάσεις δεδομένων για την επεξεργασία και τη διαχείριση των πληροφοριών. Οι μέθοδοι και οι τεχνικές που χρησιμοποιούνταν στηρίζονταν σε απλά ερωτηματολόγια, αναζητήσεις, διαδικτυακή επεξεργασία συναλλαγών και διαδραστικών οπτικοποιήσεων. Από τις αρχές του 21ου αιώνα το διαδίκτυο βοήθησε τις επιχειρήσεις να αλληλοεπιδρούν άμεσα με τους πελάτες τους ενώ αργότερα η εμφάνιση των Big Data προέτρεψε τις επιχειρήσεις στην εφαρμογή τους για την βέλτιστη αξιοποίηση των συνόλων δεδομένων. Με την εφαρμογή των Big Data οι επιχειρήσεις μπορούν να ενισχύσουν την αποτελεσματικότητα της παραγωγής τους καθώς και την ανταγωνιστικότητά τους σε πολλές πτυχές.

Τα μεγάλα δεδομένα παρέχουν μεγάλες δυνατότητες στις επιχειρήσεις στο να δημιουργήσουν νέες επιχειρήσεις, στην ανάπτυξη νέων προϊόντων και υπηρεσιών καθώς και στη βελτίωση των επιχειρηματικών λειτουργιών (Lee I, 2017). Η ανάλυση των Big Data μπορεί να συμβάλει στην εξοικονόμηση κόστους, στη καλύτερη λήψη αποφάσεων και τελικά στην υψηλότερη ποιότητα προϊόντων και υπηρεσιών (Davenport, 2014). Στοιχευμένες διαφημίσεις, τέλεια προσαρμοσμένες στους καταναλωτές καθώς και άρθρα με ειδήσεις ανάλογες των προτιμήσεων τους είναι μερικές από τις επιπτώσεις των ‘Μεγάλων Δεδομένων’.



Εικόνα 8: Εταιρείες Επεξεργασίας BigData (Firstmark)

5.1 Οφέλη των Big Data στις επιχειρήσεις

5.1.1 Εξατομίκευση μάρκετινγκ

Με την εκμετάλλευση μεγάλων δεδομένων από πολλαπλές πηγές, εταιρείες μπορεί να προσφέρουν εξατομικευμένες προτάσεις προϊόντων, υπηρεσιών και άλλες προωθητικές προσφορές. Μεγάλοι έμποροι λιανικής όπως τα Macy's και τα Target χρησιμοποιούν Big Data για να αναλύσουν τις προτιμήσεις και τα συναισθήματα των καταναλωτών και να βελτιώσουν την εμπειρία αγοράς τους. Καινοτόμες εταιρείες τεχνολογίας έχουν ήδη αρχίσει να χρησιμοποιούν δεδομένα από τα μέσα κοινωνικής δικτύωσης για την αξιολόγηση του πιστωτικού κινδύνου και των χρηματοδοτικών αναγκών των πιθανών πελατών τους έτσι ώστε να τους παρέχουν νέου τύπου προϊόντα προσαρμοσμένα στις ανάγκες τους. Επίσης, οι τράπεζες αναλύουν μεγάλα δεδομένα για την αύξηση των εσόδων, την διατήρηση των πελατών και την καλύτερη εξυπηρέτηση των πελατών τους (Lee I, 2017).

5.1.2 Καλύτερη τιμολόγηση

Με την σωστή αξιοποίηση μεγάλων δεδομένων που συλλέγονται από πελάτες οι επιχειρήσεις έχουν τη δυνατότητα να τιμολογούν με μεγαλύτερη ακρίβεια με αποτέλεσμα να εισπράττουν καλύτερες ανταμοιβές (Baker, Kiewell, & Winkler, 2014). Συγκεκριμένα το Ebay χρησιμοποιεί τεχνολογίες ανοιχτού κώδικα και ανάλυσης δεδομένων για τη βελτιστοποίηση των τιμών και της ικανοποίηση των πελατών. Για παράδειγμα, για να επιτύχει την υψηλότερη δυνατή τιμή, συνήθως σε προϊόντα δημοπρασιών, το eBay εξετάζει όλα τα δεδομένα που σχετίζονται με τα προϊόντα που έχουν πουληθεί στο παρελθόν (π.χ. σχέση μεταξύ της ποιότητας των ειδών δημοπρασίας και της προσφοράς, τιμές) και προτείνει τρόπους μεγιστοποίησης των κερδών σε πωλητές.

5.1.3 Μείωση κόστους

Η ανάλυση των Big Data μειώνει το λειτουργικό κόστος για πολλές εταιρείες. Σύμφωνα με την Accenture (2016), οι εταιρείες που χρησιμοποιούν ανάλυση δεδομένων στις λειτουργίες τους έχουν ταχύτερο και πιο αποτελεσματικό χρόνο αντίδρασης σε ζητήματα της εφοδιαστικής αλυσίδας σε σχέση με εταιρείες που κάνουν ανάλυση δεδομένων σε ad-hoc (47% έναντι 18%). Η ανάλυση μεγάλων δεδομένων οδηγεί σε πιο σωστές προβλέψεις ζήτησης, πιο αποτελεσματική δρομολόγηση με παρακολούθηση σε πραγματικό χρόνο κατά τις αποστολές, και πολύ βελτιστοποιημένη διαχείριση δικτύου διανομής (House, 2014). Τα μεγάλα δεδομένα έχουν επίσης οδηγήσει σε τεράστια μείωση του κόστους στον κλάδο του λιανικού εμπορίου. Η Tesco, μια ευρωπαϊκή εταιρεία με σούπερ μάρκετ, αναλύει τα δεδομένα από τη χρήση ψυγείων με αποτέλεσμα τη μείωση του κόστους ενέργειας κατά περίπου 25 εκατομμύρια δολάρια ετησίως. Για τη βελτιστοποίηση της θερμοκρασίας των ψυγείων, όλα τα ψυγεία Tesco στην Ιρλανδία είναι εξοπλισμένα με αισθητήρες που παρακολουθούν την θερμοκρασία κάθε 3 δευτερόλεπτα.

5.1.4 Βελτιωμένη εξυπηρέτηση πελατών

Η ανάλυση μεγάλων δεδομένων μπορεί να ενσωματώσει δεδομένα από πολλά κανάλια επικοινωνίας (π.χ. τηλέφωνο, email, μηνύματα) και να βοηθήσει το τμήμα εξυπηρέτησης πελατών στην καλύτερη κατανόηση των προβλημάτων του πελάτη με αποτέλεσμα την ταχύτερη αντιμετώπιση προβλημάτων. Επίσης με την ανάλυση των συναλλαγών και των δραστηριοτήτων σε πραγματικό χρόνο, μπορούν ανιχνευθούν δόλιες δραστηριότητες και να ειδοποιηθούν οι πελάτες άμεσα. Η Hertz, μια εταιρεία ενοικίασης αυτοκινήτων στις ΗΠΑ, χρησιμοποιεί τα Big Data για τη βελτίωση της ικανοποίησης των πελατών. Η Hertz συγκεντρώνει δεδομένα για τους πελάτες της από email, μηνύματα και διαδικτυακές έρευνες με σκοπό την καλύτερη λειτουργικότητα της εταιρείας. Για παράδειγμα, η Hertz ανακάλυψε καθυστερήσεις επιστροφής σημειώθηκαν συγκεκριμένες ώρες μιας ημέρας σε ένα γραφείο στη Φιλαδέλφεια και μπόρεσε να αυξήσει το προσωπικό κατά τις ώρες αιχμής για να βεβαιωθεί ότι τέτοια προβλήματα δεν θα επαναληφθούν (IBM, 2010).

Κεφάλαιο 6ο

6. Τα Big Data σε κλάδους των επιχειρήσεων

6.1 Big Data στα Logistics



Εικόνα 9. Logistics (Google)

Ένα από τα μεγαλύτερα οφέλη που προκύπτουν από την εφαρμογή μεγάλων δεδομένων σε σύγχρονους οργανισμούς και ιδιαίτερα στον τομέα των Logistics είναι η ικανότητά διεξαγωγής αναλύσεων υψηλής απόδοσης. Ο T. Davenport αναφέρει ως νέες ευκαιρίες από τη χρήση δεδομένων « την μείωση κόστους, βελτίωση αποφάσεων, βελτιώσεις σε προϊόντα και υπηρεσίες και αναφέρει επίσης ότι αποφάσεις διαφόρων τύπων μπορούν να βελτιωθούν με την αξιοποίηση των μεγάλων δεδομένων» (Davenport, 2013). Τέτοιες λύσεις μπορούν να βελτιώσουν την παραγωγικότητα και την ανταγωνιστικότητα(Kott, 2015) των επιχειρήσεων και να δημιουργήσουν τεράστια οφέλη για τους καταναλωτές(Chen , 2014). Επίσης, η χρήση των Big Data μπορεί να συμβάλει στον μετασχηματισμό των λειτουργικών διαδικασιών στα Logistics σε τομείς όπως η διανομή, τα αποθέματα, οι προμήθειες, η ανάπτυξη προϊόντων, κατασκευή, μάρκετινγκ, πωλήσεις και ανθρώπινο δυναμικό(Schmarzo, 2013).

Ο L. Columbus δηλώνει ότι «τα μεγάλα δεδομένα φέρνουν επανάσταση στον τρόπο με τον οποίο τα δίκτυα προμηθευτών σχηματίζονται, αναπτύσσονται, πολλαπλασιάζονται σε νέες αγορές και ωριμάζουν με την πάροδο του χρόνου. Οι συναλλαγές δεν είναι ο μόνος στόχος, η δημιουργία δικτύων ανταλλαγής γνώσεων βασίζεται στις γνώσεις που αποκτήθηκαν από την

ανάλυση μεγάλων δεδομένων» . Ισχυρίζεται επίσης ότι «μεγάλα δεδομένα και προηγμένα αναλυτικά στοιχεία ενσωματώνονται σε εργαλεία βελτιστοποίησης, πρόβλεψη ζήτησης, ολοκληρωμένες επιχειρήσεις σχεδιασμός και συνεργασία προμηθευτών και ανάλυση κινδύνου με επιταχυνόμενο ρυθμό»(Columbus, 2016) και υπογραμμίζει το γεγονός ότι «το 64% των στελεχών της εφοδιαστικής αλυσίδας εξετάζει την ανάλυση μεγάλων δεδομένων ως μια ανατρεπτική και σημαντική τεχνολογία, που θέτει τα θεμέλια για μακροπρόθεσμη αλλαγή διαχείριση στους οργανισμούς τους» (Columbus, 2016). Η βελτίωση της διαχείρισης των αποθεμάτων μπορεί να επιτευχθεί συνδυάζοντας πολλαπλά σύνολα δεδομένων όπως ιστορικά πωλήσεων, καιρικές προβλέψεις και εποχιακούς κύκλους πωλήσεων» (Manyika, 2011). Η ανάλυση των δεδομένων μεταφορών μπορεί να βελτιώσει την παραγωγικότητα με βελτιστοποίηση της απόδοσης καυσίμου, της συντήρησης και της δρομολόγηση οχημάτων. Η βέλτιστη αξιοποίηση των πόρων είναι ένα βασικό ανταγωνιστικό πλεονέκτημα για τις εταιρείες logistics.

6.2 Χρηματοοικονομικές Υπηρεσίες

Οι χρηματοοικονομικές υπηρεσίες είναι καλά παραδείγματα για το πώς η τεχνολογία μπορεί να βοηθήσει στην επεξεργασία μεγάλων ποσοτήτων δεδομένων. Τα μεγάλα δεδομένα μπορούν να χρησιμοποιηθούν για τη μείωση των ρυθμών ανατροπής, προκειμένου να προβλεφθούν πιθανές εξόδους πελατών. Μπορεί επίσης να χρησιμοποιηθεί για σύσφιξη των σχέσεων με τους πελάτες και την προσαρμογή των υπηρεσιών. Για παράδειγμα, η κατανόηση του πώς οι πελάτες χρησιμοποιούν τις πιστωτικές κάρτες και του τι είδος δανείων χρειάζονται, βοηθούν τις τράπεζες να δημιουργήσουν νέων δεικναιστικά προϊόντα που ανταποκρίνονται στις ανάγκες των πελατών. Ο Sarrocco (F. Sarrocco, 2016) υποστηρίζει ότι σε αντίθεση με τα παραδοσιακά συστήματα επιχειρηματικής ευφυΐας, οι νέες τεχνικές και τεχνολογίες που χρησιμοποιούν μεγάλα δεδομένα επιτρέπουν στους Chief Financial Officers (CFOs) να αποκτούν χρήσιμες πληροφορίες με πολύ χαμηλότερο κόστος.

6.3 Big Data στις Υπηρεσίες Υγείας

Η Ψηφιακή Υγεία είναι πολύ σημαντική για τον κλάδο της υγείας, καθώς και για την κοινωνία γενικότερα, και η πρόσβαση σε αυτό είναι δυνατή μόνο με τη χρήση των μεγάλων δεδομένων. Η Ψηφιακή Υγεία απαιτεί τη χρήση πληροφοριακών συστημάτων ώστε να συλλέγει, να συγκεντρώνει και να επεξεργάζεται δομημένα και μη δομημένα δεδομένα που συνδέονται με τον τομέα της υγείας, δημιουργώντας έτσι κλινικές πληροφορίες που ενισχύουν την ιατρική ακριβείας. Τα μεγάλα δεδομένα μπορούν να χρησιμοποιηθούν για την πρόληψη επιδημιών παρακολουθώντας πληθυσμούς στα κοινωνικά δίκτυα. Σε συνδυασμό με τη χρήση της στατιστικής ανάλυσης, μπορεί να επιτευχθεί η έγκαιρη πρόβλεψη εκδήλωσης επιδημίας, δίνοντας στα ιδρύματα υγείας χρόνο να προσαρμοστούν στις ξαφνικές αυξήσεις της ζήτησης για φροντίδα και φαρμακευτική αγωγή. Παρά την ύπαρξη αρκετών μελετών που αναφέρουν την επιτυχή χρήση των Μεγάλων Δεδομένων στις Υπηρεσίες Υγείας, ορισμένα από τα εμπόδια παραμένουν. Ο Kruse (C.Kruse, 2016) αναφέρει ως εμπόδια ζητήματα δομής δεδομένων, ασφάλειας, τυποποίησης δεδομένων, αποθήκευσης και μεταφορών, και διαχειριστικές δεξιότητες όπως η διακυβέρνηση δεδομένων. Στην πραγματικότητα, η ασφάλεια και το απόρρητο των Μεγάλων Δεδομένων εξακολουθεί να είναι το σημαντικότερο εμπόδιο.



Εικόνα 10. Big Data in Healthcare (Google)

6.4 Big Data στη Βιομηχανία

Τα τελευταία χρόνια οι βιομηχανίες μπόρεσαν να μειώσουν τα απόβλητα και τη μεταβλητότητα στις διαδικασίες παραγωγής τους με την εισαγωγή νέων πρακτικών διαχείρισης. Εκτός από την εξοικονόμηση κόστους, αυτές οι πρακτικές παρείχαν σημαντικές βελτιώσεις στην ποιότητα του τελικού προϊόντος. Τα μεγάλα δεδομένα μπορούν να χρησιμοποιηθούν με πολλούς τρόπους. Ένας από αυτούς είναι η ενσωμάτωση τους στην αλυσίδα εφοδιασμού για την επίτευξη καλύτερης κατάστασης στην προμήθεια πρώτων υλών και την έναρξη παραγγελιών σε πραγματικό χρόνο ανάλογα με τις ανάγκες των πελατών. Επιπλέον, μπορούν να χρησιμοποιηθούν στην ανάλυση ιστορικών δεδομένων διεργασίας για την ταυτοποίηση προτύπων και σχέσεων μεταξύ των διαφόρων εσωτερικών διεργασιών. Ο Agrawal (Agrawal, 2016) ανέφερε ότι ένας από τους τομείς που θα επωφεληθούν περισσότερο από την ανάπτυξη της αγοράς μεγάλων δεδομένων είναι η βιομηχανία, με έσοδα που προβλέπεται να φτάσουν τα 39 δισεκατομμύρια δολάρια έως το 2019. Ο Delgado (Delgado, 2017) αναφέρει ότι οι εταιρείες μπορούν να εγκαταστήσουν ηλεκτρονικούς αισθητήρες στις γραμμές συναρμολόγησης ώστε να επιτρέπουν τη χρήση των παραγόμενων δεδομένων για τη βελτίωση της ποιότητας και της ασφάλειας της διαδικασίας παραγωγής. Η έξυπνη κατασκευή εμφανίζεται ως ένα νέο παράδειγμα στην ανάπτυξη του μεταποιητικού τομέα, που σκοπεύει να εκμεταλλεύονται τις τεχνολογίες της πληροφορίας για να αυξήσουν τον βαθμό αυτονομίας και ευελιξίας της παραγωγικής διαδικασίας.



Εικόνα 11. Big Data in Manufacturing Industry (Google)

6.5 Big Data στη Γεωργία

Η γεωργία ακριβείας έχει αποκτήσει εξέχουσα θέση σε πολλές χώρες, όπως οι ΗΠΑ, η Βραζιλία ή η Αυστραλία. Η χρήση μεγάλων δεδομένων και οι τεχνολογίες δημιουργούν μια νέα αγροτική επανάσταση, που εγγυάται αύξηση της παραγωγικότητας και της παραγωγής χωρίς την ανάγκη να αυξηθεί η έκταση. Η απόφαση για το είδος του σπόρου που θα φυτευτεί, πώς θα φυτευτεί, πώς θα γονιμοποιηθεί, πώς θα χρησιμοποιηθεί, οι αποφάσεις θρέψης, ο χρόνος συγκομιδής, η αποθήκευση, η μεταφορά και η εμπορία που πρέπει να ληφθούν από τον αγρότη μπορούν ήδη να υπολογιστούν με τη βοήθεια της τεχνολογία. Ο Wolfert (Wolfert, 2017) αναφέρει ότι τα μεγάλα δεδομένα αναμένεται να έχουν μεγάλο αντίκτυπο στην έξυπνη γεωργία και σε ολόκληρη την εφοδιαστική αλυσίδα. Αισθητήρες μπορούν να χρησιμοποιηθούν για τη συλλογή δεδομένων σχετικά με τις συνθήκες του εδάφους, τις ποικιλίες ζιζανίων και τον καιρό. Τέτοιες πληροφορίες βοηθούν στην ελαχιστοποίηση του κινδύνου και στη σωστή και έγκαιρη λήψη αποφάσεων.



Εικόνα 12. Big Data in Agriculture (Google)

6.6 Τηλεπικοινωνίες

Οι πάροχοι υπηρεσιών κινητής τηλεφωνίας (ΠΥΚ) αντιμετωπίζουν συνήθως προκλήσεις όπως η χαμηλή αποδοχή των υπηρεσιών κινητής τηλεφωνίας και οι δυσκολίες στη διαχείριση της αποχώρησης των πελατών. Η απόκτηση νέων πελατών είναι πιο δαπανηρή από τη διατήρηση των υφιστάμενων, και η εμπειρία των πελατών διαδραματίζει κρίσιμο ρόλο στην προώθηση της αφοσίωσης και τη δημιουργία εσόδων (Soares, 2012a,b). Για τη βελτίωση της εμπειρίας των πελατών, οι ΠΥΠ εξετάζουν εξονυχιστικά διάφορους παράγοντες, συμπεριλαμβανομένων δημογραφικών πληροφοριών (όπως το φύλο, η ηλικία, η οικογενειακή κατάσταση και οι γλωσσικές προτιμήσεις), τις προτιμήσεις των πελατών, τη δομή του νοικοκυριού και τα στοιχεία χρήσης (αρχεία λεπτομερειών κλήσεων, χρήση του διαδικτύου και υπηρεσίες προστιθέμενης αξίας). Με την ανάλυση αυτών των παραγόντων, οι MSP δημιουργούν μοντέλα προτιμήσεων των πελατών για την παροχή εξατομικευμένων και σχετικών υπηρεσιών, μια στρατηγική γνωστή ως στοχευμένο μάρκετινγκ. Η προσέγγιση αυτή αποσκοπεί στην ενίσχυση της υιοθέτησης των υπηρεσιών κινητής τηλεφωνίας, στη μείωση των ποσοστών αποχώρησης και, τελικά, στην αύξηση των εσόδων των MSP.

Χαρακτηριστικό παράδειγμα αποτελεί η Ufone, ένας πάροχος υπηρεσιών κινητής τηλεφωνίας με έδρα το Πακιστάν, ο οποίος μείωσε με επιτυχία τα ποσοστά αποχώρησης εφαρμόζοντας ακριβές μάρκετινγκ εξατομικευμένων προσφορών στους πελάτες του (Utsler, 2013). Η εταιρεία χρησιμοποιεί δεδομένα Call Detail Record (CDR) για τον εντοπισμό μοτίβων κλήσεων, επιτρέποντάς της να προσφέρει διαφορετικά πακέτα σε μεμονωμένους πελάτες. Στη συνέχεια, οι υπηρεσίες αυτές προωθούνται στους πελάτες μέσω κλήσεων ή γραπτών μηνυμάτων και οι απαντήσεις τους καταγράφονται σχολαστικά για μετέπειτα ανάλυση.

Οι εταιρείες τηλεπικοινωνιών αντιμετωπίζουν ενεργά την πρόκληση της απάτης στις τηλεπικοινωνίες. Τα παραδοσιακά συστήματα διαχείρισης απάτης συχνά δυσκολεύονται να εντοπίσουν άμεσα νέες μορφές απάτης, επιτρέποντας σε αυτούς που την διαπράττουν να αλλάξουν την τακτική τους πριν από την ανίχνευση. Για να ξεπεράσουν αυτές τις ελλείψεις, οι πάροχοι υπηρεσιών κινητής τηλεφωνίας (MSP) στρέφονται στην ανάλυση δεδομένων σε πραγματικό χρόνο για να ελαχιστοποιήσουν τις απώλειες που οφείλονται στην απάτη. Η Mobileum Inc., ένας διακεκριμένος πάροχος λύσεων τηλεπικοινωνιακής ανάλυσης, βρίσκεται

στην πρώτη γραμμή της ανάπτυξης ενός συστήματος ανίχνευσης απάτης σε πραγματικό χρόνο που χρησιμοποιεί προγνωστική ανάλυση και μηχανική μάθηση (Ray, 2015).

Η αναδύομενη τάση στις τηλεπικοινωνίες, γνωστή ως Network Analytics, επιτρέπει στους MSPs να παρακολουθούν την ταχύτητα του δικτύου και να επιβλέπουν ολόκληρο το δίκτυο. Η δυνατότητα αυτή διευκολύνει την ταχεία επίλυση προβλημάτων δικτύου, οδηγώντας σε βελτιωμένη ποιότητα υπηρεσιών και βελτιωμένη εμπειρία πελατών μέσα σε λίγα λεπτά. Με την εξάπλωση των smartphones, η ανάλυση δεδομένων θέσης και συμπεριφοράς σε πραγματικό χρόνο ανοίγει δυνατότητες για την παροχή υπηρεσιών με βάση την τοποθεσία ή υπηρεσιών με βάση το πλαίσιο κατόπιν αιτήματος του πελάτη. Η προσέγγιση αυτή αναμένεται να ενισχύσει την υιοθέτηση των υπηρεσιών κινητής τηλεφωνίας.

6.7 Λιανικό Εμπόριο

Η εξέλιξη του ηλεκτρονικού εμπορίου, των ηλεκτρονικών συναλλαγών, των συζητήσεων στα μέσα κοινωνικής δικτύωσης και των πρόσφατων αλληλεπιδράσεων με βάση την τοποθεσία μέσω των smart phones συμβάλλουν στην ποσότητα και την ποιότητα των δεδομένων για την προσαρμογή βάσει δεδομένων στον τομέα του λιανικού εμπορίου (Brown et al., 2011). Τα μεγάλα καταστήματα λιανικής πώλησης αναπτύσσουν CCTV όχι μόνο για την παρακολούθηση περιπτώσεων κλοπής αλλά και για την παρακολούθηση της κίνησης των πελατών (Villars et al., 2011). Αυτό τους επιτρέπει να παρατηρούν τα δημογραφικά στοιχεία των πελατών, το φύλο και την αγοραστική συμπεριφορά τις καθημερινές και τα Σαββατοκύριακα. Η ανάλυση των αγοραστικών προτύπων των πελατών με τη χρήση της ανάλυσης καλαθιού αγοράς, μιας ευρέως χρησιμοποιούμενης τεχνικής εξόρυξης δεδομένων που εισήγαγαν οι Agrawal και Srikanth το 1994, βοηθά τους λιανοπωλητές να κατηγοριοποιήσουν τα είδη. Για παράδειγμα, εάν ένας πελάτης αγοράζει ψωμί και γάλα, μπορεί επίσης να αγοράζει μαρμελάδα, επηρεάζοντας τις αποφάσεις σχετικά με την τοποθέτηση προϊόντων και την τιμολόγηση (Brown et al., 2011-Villars et al., 2011).

Στο σύγχρονο λιανικό εμπόριο, οι επιχειρήσεις ηλεκτρονικού εμπορίου χρησιμοποιούν την ανάλυση καλαθιού αγοράς και τα συστήματα συστάσεων για την τμηματοποίηση και τη στόχευση των πελατών. Συλλέγουν δεδομένα ροής κλικ, παρατηρούν τη συμπεριφορά και προσφέρουν συστάσεις προϊόντων σε πραγματικό χρόνο. Οι αναλύσεις διαδραματίζουν κρίσιμο ρόλο στο να βοηθούν τις εταιρείες λιανικής στη διαχείριση των αποθεμάτων τους. Η

Stage Stores, που λειτουργεί υπό την επωνυμία Stage Stores Inc. σε περισσότερες από 40 αμερικανικές πολιτείες, είναι ένα παράδειγμα μάρκας που χρησιμοποιεί την ανάλυση για να προβλέψει παραγγελίες ενδυμάτων διαφόρων μεγεθών για διαφορετικές γεωγραφικές περιοχές (Meek, 2015).

6.8 Επιβολή του Νόμου

Οι αρχές επιβολής του νόμου χρησιμοποιούν διάφορες μεθόδους για να προβλέψουν την πιθανή τοποθεσία μελλοντικών εγκλημάτων, χρησιμοποιώντας δεδομένα του παρελθόντος σχετικά με τον τύπο του εγκλήματος, την τοποθεσία και τον χρόνο, καθώς και ενσωματώνοντας πληροφορίες από τα μέσα κοινωνικής δικτύωσης, την παρακολούθηση με μη επανδρωμένα αεροσκάφη και την παρακολούθηση μέσω smartphone. Ερευνητές του Πανεπιστημίου Rutgers ανέπτυξαν μια εφαρμογή με την ονομασία RTM Dx, η οποία χρησιμοποιείται από αστυνομικά τμήματα στο Ιλινόις, το Τέξας, την Αριζόνα, το Νιου Τζέρσεϊ, το Μιζούρι και το Κολοράντο για την προληπτική πρόληψη εγκλημάτων. Η εφαρμογή αυτή επιτρέπει στις αρχές επιβολής του νόμου να αξιολογούν τη χωρική συσχέτιση μεταξύ των τοποθεσιών εγκλημάτων και των περιβαλλοντικών χαρακτηριστικών, βοηθώντας στην πρόληψη του εγκλήματος (Mor, 2014).

Έχει εισαχθεί μια νέα τεχνολογία γνωστή ως facial analytics, η οποία επιτρέπει την ανάλυση εικόνων ατόμων χωρίς να θίγεται η ιδιωτική τους ζωή. Η ανάλυση προσώπου χρησιμοποιείται ιδιαίτερα στην ανίχνευση της παιδικής πορνογραφίας, μειώνοντας σημαντικά τον χρόνο που απαιτείται για τη χειροκίνητη εξέταση. Ο εντοπισμός της παιδικής πορνογραφίας περιλαμβάνει την ενσωμάτωση διαφόρων τεχνολογιών, όπως η Artemis και η PhotoDNA, οι οποίες συγκρίνουν κατακερματισμούς αρχείων και εικόνων με υπάρχουσες βάσεις δεδομένων για να προσδιορίσουν αν το θέμα είναι ενήλικος ή παιδί. Επιπλέον, η τεχνολογία αυτή είναι ικανή να αναγνωρίζει πορνογραφία που βασίζεται σε κινούμενα σχέδια (Ricanek and Boehnen, 2012).

6.9 Τραπεζικό Σύστημα

Η ανάλυση της πιστοληπτικής ικανότητας των πελατών περιλαμβάνει την εξέταση δημογραφικών πληροφοριών, δεδομένων συμπεριφοράς και αρχείων οικονομικής απασχόλησης. Η στρατηγική των διασταυρούμενων πωλήσεων μπορεί να εφαρμοστεί σε αυτό το πλαίσιο για την εστίαση σε συγκεκριμένα τμήματα πελατών, λαμβάνοντας υπόψη τις προηγούμενες αγοραστικές τους συνήθειες, τις δημογραφικές πληροφορίες, την ανάλυση

συναισθήματος και τα δεδομένα από τα συστήματα διαχείρισης πελατειακών σχέσεων (CRM) (Forsyth, 2012- Coumaros et al., 2014).

6.10 Ενέργεια και υπηρεσίες κοινής ωφέλειας

Οι έξυπνοι μετρητές χρησιμοποιούνται για τη μέτρηση της κατανάλωσης νερού, φυσικού αερίου και ηλεκτρικής ενέργειας σε ωριαία διαστήματα. Η διαδικασία αυτή παράγει σημαντικό όγκο δεδομένων, τα οποία στη συνέχεια αναλύονται για την τροποποίηση των προτύπων χρήσης ενέργειας (Brown et al., 2011). Η ανάλυση σε πραγματικό χρόνο αποκαλύπτει τα πρότυπα κατανάλωσης ενέργειας, τις περιπτώσεις κλοπής ηλεκτρικής ενέργειας και τις διακυμάνσεις των τιμών.

6.11 Εκπαίδευση

Με την εισαγωγή ηλεκτρονικών ενοτήτων μαθημάτων, η αξιολόγηση των ακαδημαϊκών επιδόσεων σε πραγματικό χρόνο έχει καταστεί εφικτή. Η δυνατότητα αυτή επιτρέπει τη συνεχή παρακολούθηση των επιδόσεων των φοιτητών μετά από κάθε ενότητα, παρέχοντας άμεση ανατροφοδότηση σχετικά με τα μαθησιακά τους πρότυπα. Επιπλέον, δίνει τη δυνατότητα στους καθηγητές να αξιολογούν τις μεθόδους διδασκαλίας τους και να προβαίνουν σε προσαρμογές με βάση τις επιδόσεις και τις απαιτήσεις των μαθητών. Μπορεί να επιτευχθεί η πρόβλεψη των τάσεων εγκατάλειψης, ο εντοπισμός των μαθητών που χρήζουν ιδιαίτερης προσοχής και ο εντοπισμός εκείνων που είναι ικανοί να χειριστούν δύσκολες εργασίες (West, 2012). Οι Beck και Mostow (2008) διερεύνησαν την αναγνωστική κατανόηση των μαθητών με τη χρήση λογισμικού ευφυούς διδασκαλίας, σημειώνοντας σημαντική μείωση των αναγνωστικών λαθών όταν οι μαθητές επανέρχονταν σε παλιές ιστορίες σε σύγκριση με νέες.

Κεφάλαιο 7ο

7. Παραδείγματα χρήσης Big Data σε γνωστές επιχειρήσεις

7.1 UPS

Η UPS δεν είναι άγνωστη στα μεγάλα δεδομένα, αφού έχει αρχίσει να καταγράφει και να παρακολουθεί μια ποικιλία πακέτων κινήσεων και συναλλαγών ήδη από τη δεκαετία του 1980. Η εταιρεία παρακολουθεί τώρα δεδομένα για 16,3 εκατομμύρια πακέτα ανά ημέρα για 8,8 εκατομμύρια πελάτες, με μέσο όρο 39,5 εκατομμύρια αιτήματα παρακολούθησης από πελάτες ανά ημέρα. Η εταιρεία αποθηκεύει πάνω από 16 petabytes δεδομένων. Ωστόσο, μεγάλο μέρος από τα μεγάλα δεδομένα που απέκτησε πρόσφατα προέρχεται από τηλεματικούς αισθητήρες σε πάνω από 46.000 οχήματα. Τα δεδομένα για τα φορτηγά μεταφορών της UPS περιλαμβάνουν την ταχύτητα, την κατεύθυνσή τους, πέδηση και απόδοση αμαξοστοιχίας κίνησης. Τα δεδομένα δεν χρησιμοποιούνται μόνο για την παρακολούθηση της καθημερινής απόδοσης, αλλά και για τη δημιουργία ενός σημαντικού επανασχεδιασμό των δομών διαδρομής των οδηγών UPS. Αυτή η πρωτοβουλία, που ονομάζεται ORION είναι αναμφισβήτητα η μεγαλύτερη έρευνα λειτουργιών του κόσμου. Βασίζεται σε μεγάλο αριθμό δεδομένων διαδικτυακού χάρτη και τελικά επαναδιαμορφώνει τις παραλαβές και αποβιβάσεις του οδηγού σε πραγματικό χρόνο. Το έργο έχει ήδη οδηγήσει σε εξοικονόμηση περισσότερων από 8,4 εκατομμύρια γαλόνια καυσίμου το 2011, κόβοντας 85 εκατομμύρια μίλια από τις καθημερινές διαδρομές. Η UPS εκτιμά ότι εξοικονομώντας μόνο ένα ημερήσιο μίλι ανά οδηγό εξοικονομεί στην εταιρεία 30 εκατομμύρια δολάρια. Η εταιρεία προσπαθεί επίσης να χρησιμοποιήσει δεδομένα και αναλυτικά στοιχεία για τη βελτιστοποίηση της αποτελεσματικότητας των 2000 πτήσεων αεροσκαφών της την ημέρα.

7.2 AMAZON

Η Amazon είναι πρωτοπόρος στη συλλογή, αποθήκευση, επεξεργασία και ανάλυση προσωπικών πληροφοριών από κάθε πελάτη ως μέσο προσδιορισμού του τρόπου με τον οποίο οι πελάτες ξοδεύουν τα χρήματά τους. Η εταιρεία χρησιμοποιεί predictive analytics για στοχευμένο μάρκετινγκ που τη βοηθά να αυξήσει την ικανοποίηση των πελατών και να κερδίσει την αφοσίωση σε αντάλλαγμα.

Παρακάτω παρουσιάζονται οι τρόποι με τους οποίους χρησιμοποιεί τα Big Data :

1. Supply Chain Optimization

Η Amazon θέλει να εκπληρώσει τις παραγγελίες γρήγορα και για να πετύχει αυτό το κατόρθωμα, η εταιρεία συνδέεται με τους κατασκευαστές και παρακολουθεί τα αποθέματά τους. Η Amazon αναλύει τα διαθέσιμα δεδομένα και εντοπίζει την πλησιέστερη αποθήκη σε έναν πελάτη/πωλητή για να μειώσει τα έξοδα αποστολής. Επιπλέον, η θεωρία γραφημάτων βοηθά στον καθορισμό του καλύτερου προγράμματος παράδοσης, διαδρομής και ομαδοποίησης προϊόντων, γεγονός που μειώνει περαιτέρω τα έξοδα αποστολής.

2. Price Optimization

Οι τιμές καθορίζονται ανάλογα με τη δραστηριότητά του χρήστη στον ιστότοπο, τις τιμές των ανταγωνιστών, τη διαθεσιμότητα προϊόντων, τις προτιμήσεις ειδών, το ιστορικό παραγγελιών, το αναμενόμενο περιθώριο κέρδους και άλλους παράγοντες. Οι τιμές των προϊόντων αλλάζουν γενικά κάθε 10 λεπτά καθώς ενημερώνονται και αναλύονται τα big data. Ως αποτέλεσμα, η Amazon προσφέρει συνήθως εκπτώσεις σε είδη με τις μεγαλύτερες πωλήσεις και κερδίζει μεγαλύτερα κέρδη σε λιγότερο δημοφιλή είδη. Αυτό ωφέλησε την εταιρεία στην αύξηση του ετήσιου εισοδήματός της κατά 143% από το 2016 έως το 2019, σύμφωνα με άρθρο.

3. Παρακίνηση πελατών

Οι προτάσεις προϊόντων της Amazon είναι η πιο οικεία εφαρμογή μεγάλων δεδομένων στους χρήστες της ενώ χρησιμοποιείται επίσης για τη συλλογή πληροφοριών. Στη συνέχεια λειτουργεί παρουσιάζοντας στους χρήστες σχετικά αντικείμενα με βάση τα πράγματα που βρίσκονται ήδη στα καλάθι τους και τα προϊόντα που έχουν αγοράσει προηγουμένως. Στο σημείο που η Amazon μπορεί να προσελκύσει αποτελεσματικά τους πελάτες με προσαρμοσμένες επιλογές και να τους αναγκάσει να ξοδέψουν

περισσότερα, τα οφέλη του οργανισμού αυξάνονται και οι καταναλωτές αναγνωρίζουν ότι στην Amazon μπορούν να αγοράσουν οτιδήποτε χρειάζονται.

Κεφάλαιο 8°

8. Βήματα για την αξιοποίηση των Big Data από νέες επιχειρήσεις

Οι βέλτιστες πρακτικές μεγάλων δεδομένων, ειδικά προσαρμοσμένες για αποτελεσματική διαχείριση, διαδραματίζουν καθοριστικό ρόλο στη μετατροπή των ακατέργαστων δεδομένων σε χρήσιμες πληροφορίες. Οι οργανισμοί διαχειρίζονται μεγάλα δεδομένα για να λάβουν καλύτερες αποφάσεις, να δημιουργήσουν νέα προϊόντα και υπηρεσίες και να μειώσουν το κόστος. Καθώς οι επιχειρήσεις συνεχίζουν να συγκεντρώνουν τεράστιους όγκους δεδομένων από διαφορετικές πηγές, η ανάγκη για ολοκληρωμένες κατευθυντήριες γραμμές γίνεται όλο και πιο επιτακτική.

Η υιοθέτηση βέλτιστων πρακτικών μεγάλων δεδομένων και η παραμονή μπροστά στις τελευταίες τάσεις είναι απαραίτητη για τους οργανισμούς που στοχεύουν να ευδοκιμήσουν στο σημερινό τοπίο με επίκεντρο τα δεδομένα. Με την εκθετική αύξηση του όγκου δεδομένων, τις ποικίλες πηγές δεδομένων και την εξέλιξη των τεχνολογιών, τέτοιες πρακτικές προσφέρουν στρατηγικό πλεονέκτημα.

Οι βέλτιστες πρακτικές για μεγάλα δεδομένα είναι οι ακόλουθες:

8.1.Καθορισμός ξεκάθαρων στόχων

Ο καθορισμός σαφών στόχων είναι το θεμελιώδες βήμα σε κάθε επιτυχημένη πρωτοβουλία μεγάλων δεδομένων. Οι οργανισμοί πρέπει να διατυπώνουν συγκεκριμένους και μετρήσιμους επιχειρηματικούς στόχους που στοχεύουν να αντιμετωπίσουν οι προσπάθειές τους για μεγάλα δεδομένα.

Είτε οι στόχοι επικεντρώνονται στη βελτίωση της λειτουργικής αποτελεσματικότητας, την υπέρβαση των σχετικών προκλήσεων, τη βελτίωση των εμπειριών των πελατών ή την απόκτηση ανταγωνιστικού πλεονεκτήματος, ένας καλά καθορισμένος οδικός χάρτης διασφαλίζει ότι η στρατηγική ευθυγραμμίζεται απρόσκοπτα με ευρύτερους οργανωτικούς στόχους.

Αυτή η σαφήνεια όχι μόνο καθοδηγεί την επιλογή των κατάλληλων τεχνολογιών και πηγών δεδομένων, αλλά παρέχει επίσης ένα σημείο αναφοράς για τη μέτρηση της επιτυχίας της πρωτοβουλίας.

8.2. Διαχωρισμός Σημαντικών Δεδομένων

Το πρώτο βήμα στη χρήση μεγάλων δεδομένων είναι να γνωρίζουμε ποια δεδομένα είναι σημαντικά και τι όχι. Αυτό θα σας επιτρέψει να επιλέξετε τα καλύτερα δεδομένα για την επιχείρησή σας. Πρέπει να είστε συγκεκριμένοι σχετικά με το είδος των πληροφοριών που θέλετε, τις απαντήσεις που αναζητάτε και τον χρόνο που έχετε διαθέσιμο.

8.3. Διασφάλιση Ποιότητας Δεδομένων

Η διασφάλιση δεδομένων υψηλής ποιότητας είναι ένα σημαντικό ζήτημα στην ψηφιακή εποχή και θεωρείται σημαντική πτυχή των χαρακτηριστικών μεγάλων δεδομένων. Τα δεδομένα στα οποία βασίζομαστε μπορούν να έχουν τεράστιο αντίκτυπο στη ζωή και τις επιχειρήσεις μας, επομένως είναι σημαντικό να διασφαλίσουμε ότι είναι υψηλής ποιότητας. Από την υγειονομική περίθαλψη έως τα μέσα κοινωνικής δικτύωσης, η ποιότητα των δεδομένων πρέπει να παρακολουθείται και να ελέγχεται για ακρίβεια.

- Τα δεδομένα είναι ακριβή, πλήρη και ενημερωμένα.
- Τα δεδομένα είναι επίκαιρα, δεδομένου ότι δημιουργήθηκαν όχι περισσότερο από 60 ημέρες νωρίτερα ή αργότερα από την ημερομηνία τελευταίας τροποποίησης.
- Τα δεδομένα είναι πλήρη, που σημαίνει ότι περιλαμβάνουν όλες τις εγγραφές για μια συγκεκριμένη μεταβλητή.
- Τα σχετικά δεδομένα συνδέονται με την κατάλληλη μεταβλητή στη βάση δεδομένων.

8.4. Σωστή επισήμανση δεδομένων

Με την αυξανόμενη δημοτικότητα της επισήμανσης μεγάλων δεδομένων, είναι επιτακτική ανάγκη να κατανοήσουμε τη σημασία διαφορετικών τύπων συνόλων δεδομένων. Ένα καλό σύνολο δεδομένων είναι αυτό που έχει επισημανθεί σωστά και μπορεί να χρησιμοποιηθεί για διάφορους σκοπούς.

- Τοποθετήστε κατάλληλα τα μεγάλα δεδομένα σας, ώστε να είναι εύκολα κατανοητά και να ταξινομούνται αργότερα.

- Σκεφτείτε πώς θα επισημάνετε τα δεδομένα σας προτού αρχίσετε να τα συλλέγετε (δηλαδή ποιες είναι οι κατηγορίες;)
- Χρησιμοποιήστε συνεπείς ετικέτες που είναι κατανοητές από όλους.
- Κρατήστε τις ετικέτες όσο το δυνατόν πιο σύντομες.
- Χρησιμοποιήστε ένα υπόμνημα ή έναν πίνακα περιεχομένων για να δείξετε τι σημαίνει κάθε ετικέτα.

8.5.Επιλογή Κατάλληλων Θέσεων Αποθήκευσης

Η επιλογή κατάλληλων τοποθεσιών αποθήκευσης είναι μία από τις κορυφαίες βέλτιστες πρακτικές μεγάλων δεδομένων με εκτεταμένες συνέπειες για την αποτελεσματικότητα και την αποτελεσματικότητα της διαχείρισης δεδομένων. Ο τεράστιος όγκος και η ποικιλία δεδομένων που παράγονται από οργανισμούς απαιτούν προσεκτική εξέταση των λύσεων αποθήκευσης.

Η αποθήκευση δεδομένων αποτελεί σημαντική ευθύνη του σύγχρονου επιχειρηματικού κόσμου. Με τόσες πολλές παραβιάσεις δεδομένων που συμβαίνουν τα τελευταία χρόνια, είναι πιο σημαντικό από ποτέ να γνωρίζετε πού να αποθηκεύσετε τα δεδομένα σας. Η αποθήκευση στο cloud είναι μια πολύ δημοφιλής επιλογή για εταιρείες, καθώς είναι προσβάσιμη από πολλές συσκευές και μπορεί να αυξηθεί ή να μειωθεί ανάλογα με τις ανάγκες.

- Διατηρήστε τα δεδομένα σας σε κοντινή απόσταση από το σημείο όπου χρησιμοποιούνται και αναφέρονται.
- Τοποθετήστε τα μεταδεδομένα σας (πληροφορίες για τα δεδομένα σας) με το φυσικό αρχείο που περιέχει τα πραγματικά ακατέργαστα δεδομένα.
- Αποθηκεύστε τα δεδομένα σας σε πολλές φυσικές τοποθεσίες για πλεονασμό.
- Αποθηκεύστε πλήρη αντίγραφα ασφαλείας των δεδομένων και των μεταδεδομένων σας σε ξεχωριστό σύστημα από το μηχάνημα που περιέχει τα πρωτογενή δεδομένα.
- Κατανοήστε αποτελεσματικά τα στοιχεία των μεγάλων δεδομένων.

8.6.Διαχείριση Κύκλου Ζωής Δεδομένων

Η Διαχείριση Κύκλου Ζωής Δεδομένων είναι μια θεμελιώδης πτυχή των βέλτιστων πρακτικών μεγάλων δεδομένων, δίνοντας έμφαση στην ανάγκη για μια ολοκληρωμένη στρατηγική που

καλύπτει ολόκληρη τη διάρκεια ζωής των δεδομένων μέσα σε έναν οργανισμό. Αυτή η προσέγγιση περιλαμβάνει την ενορχήστρωση της συλλογής, αποθήκευσης, επεξεργασίας, ανάλυσης και, όταν είναι απαραίτητο, την αρχειοθέτηση ή τη διαγραφή δεδομένων.

Με την προσεκτική διαχείριση του κύκλου ζωής των δεδομένων, οι οργανισμοί μπορούν να βελτιστοποιήσουν την κατανομή πόρων, να διατηρήσουν την ποιότητα των δεδομένων και να συμμορφωθούν με τις κανονιστικές απαιτήσεις. Αυτή η πρακτική διασφαλίζει ότι τα δεδομένα παραμένουν σχετικά, ακριβή και προσβάσιμα καθ' όλη τη διάρκεια της διαδρομής τους, επιτρέποντας στις επιχειρήσεις να αντλούν διαρκή αξία από τα στοιχεία των πληροφοριών τους.

8.7.Απλοποίηση Διαδικασιών Δημιουργίας Αντιγράφων Ασφαλείας

Ένα από τα πιο σημαντικά καθήκοντα όταν ασχολείστε με μεγάλα δεδομένα είναι να έχετε ένα αντίγραφο ασφαλείας. Αυτό συμβαίνει επειδή μπορεί να χαθεί ή να καταστραφεί και σε αυτήν την εποχή αυξημένων απειλών για την ασφάλεια στον κυβερνοχώρο, τα δεδομένα σας πρέπει να δημιουργηθούν με ασφάλεια. Ο ευκολότερος τρόπος για να το κάνετε αυτό είναι να χρησιμοποιήσετε έναν πάροχο αποθήκευσης cloud.

Το Amazon, το Dropbox, το Google Drive και το Microsoft OneDrive είναι όλες δημοφιλείς επιλογές. Εναλλακτικά, μπορείτε να δημιουργήσετε αντίγραφα ασφαλείας των δεδομένων σας σε εξωτερικό σκληρό δίσκο ή CD/DVD.

- Να δημιουργείτε συχνά αντίγραφα ασφαλείας ολόκληρου του συστήματός σας
- Χρησιμοποιήστε δωρεάν λογισμικό για να δημιουργήσετε αντίγραφα ασφαλείας των αρχείων σας
- Δημιουργήστε αυτοματοποιημένα αντίγραφα ασφαλείας για τα δεδομένα και τα μεταδεδομένα σας
- Αποθηκεύστε αντίγραφα ασφαλείας εκτός τοποθεσίας
- Χρησιμοποιήστε υπηρεσίες απομακρυσμένης αποθήκευσης όπως το DropBox για να αποθηκεύσετε επιπλέον αντίγραφα ασφαλείας των αρχείων σας
- Πραγματοποιήστε τακτικά αντίγραφα ασφαλείας δεδομένων και ελέγχους ακεραιότητας στο αντίγραφο ασφαλείας

– Αποφύγετε τη χρήση μαγνητικών μέσων για την αποθήκευση των δεδομένων σας, καθώς μπορεί να υποβαθμιστούν με την πάροδο του χρόνου

8.8. Εφαρμογή μέτρων ασφαλείας

Η ασφάλεια των δεδομένων έχει γίνει βασικό μέλημα πολλών επιχειρήσεων και ιδιωτών. Οι παραβιάσεις μεγάλων δεδομένων μπορεί να οδηγήσουν σε κλοπή ταυτότητας, απώλεια εμπορικών μυστικών και πολλά άλλα. Εάν μια εταιρεία έχει ευαίσθητες πληροφορίες, θα πρέπει να είναι κρυπτογραφημένες και αποθηκευμένες σε μια απομονωμένη μονάδα δίσκου.

Η κρυπτογράφηση διασφαλίζει ότι μόνο το άτομο με το κλειδί αποκρυπτογράφησης έχει πρόσβαση στα μεγάλα δεδομένα. Η κύρια διαφορά μεταξύ της ασφάλειας που βασίζεται σε σύννεφο και της φυσικής ασφάλειας είναι ότι με την πρώτη, τα δεδομένα δεν αποθηκεύονται στη δική του συσκευή, αλλά σε έναν διακομιστή που βασίζεται σε σύννεφο. Αυτό σημαίνει μικρότερη χωρητικότητα αποθήκευσης για την εργασία κάποιου και μεγαλύτερη εξάρτηση από τον πάροχο υπηρεσιών.

8.9.Κλιμακόμενη Υποδομή

Μια επεκτάσιμη υποδομή διασφαλίζει ότι το τεχνολογικό θεμέλιο ενός οργανισμού μπορεί αβίαστα να προσαρμοστεί στις συνεχώς αυξανόμενες απαιτήσεις των μεγάλων δεδομένων. Αξιοποιώντας επεκτάσιμες λύσεις, όπως πλατφόρμες που βασίζονται σε σύννεφο, οι οργανισμοί μπορούν να επεκτείνουν ευέλικτα τις υπολογιστικές και αποθηκευτικές τους δυνατότητες καθώς εξελίσσονται οι απαιτήσεις δεδομένων.

Η επεκτασιμότητα δίνει στις επιχειρήσεις τη δυνατότητα να αντιμετωπίζουν αποτελεσματικά τις αυξήσεις της εισροής δεδομένων, είτε λόγω εποχιακών διακυμάνσεων είτε απρόβλεπτων αυξήσεων, χωρίς συμβιβασμούς στην απόδοση ή χωρίς περιττό κόστος υποδομής.

8.10.Αποτελεσματική Χρήση Εργαλείων

Τα οφέλη των μεγάλων δεδομένων στην ψηφιακή εποχή δεν μπορούν να παραβλεφθούν. Είναι ενσωματωμένο στην αρχιτεκτονική πολλών πραγμάτων. Η ανάλυση δεδομένων μπορεί να σας

βοηθήσει να ξεπεράσετε τον ανταγωνισμό και να κάνετε τη ζωή σας πιο εύκολη. Απλά πρέπει να ξέρετε τι να κάνετε με αυτό και πώς να το ερμηνεύσετε.

Η χρήση εργαλείων ΒΙ μπορεί να σας κάνει να αναλύσετε τα σύνολα δεδομένων που έχετε για να λάβετε περισσότερες πληροφορίες σχετικά με τη συμπεριφορά των πελατών, την απόδοση του προϊόντος και άλλα σχετικά σημαντικά θέματα.

8.11. Τακτικοί έλεγχοι δεδομένων

Η διεξαγωγή συστηματικών αναθεωρήσεων των διαδικασιών και των συστημάτων δεδομένων επιτρέπει στους οργανισμούς να εντοπίζουν και να διορθώνουν πιθανές αποκλίσεις, ασυνέπειες ή τρωτά σημεία που ενδέχεται να θέσουν σε κίνδυνο την ακεραιότητα των δεδομένων.

Εφαρμόζοντας ένα πρόγραμμα τακτικών ελέγχων, οι επιχειρήσεις μπορούν να διατηρήσουν την ποιότητα των στοιχείων ενεργητικού τους, μειώνοντας τον κίνδυνο σφαλμάτων που θα μπορούσαν να επηρεάσουν αρνητικά τις διαδικασίες λήψης αποφάσεων. Αυτή η πρακτική βοηθά επίσης στις προσπάθειες συμμόρφωσης διασφαλίζοντας ότι οι πρακτικές ευθυγραμμίζονται με τους σχετικούς κανονισμούς και τα πρότυπα του κλάδου.

Κεφάλαιο 9^ο

9. Big Data στην Ευρώπη

Στην Ευρώπη, η σημασία των μεγάλων δεδομένων έχει ξεπεράσει την απλή τεχνολογική καινοτομία, αναδεικνύοντας ως στρατηγική επιταγή που στηρίζει τις εξελίξεις σε διάφορους τομείς. Η Ευρώπη αναγνωρίζει ολοένα και περισσότερο το μετασχηματιστικό δυναμικό της αξιοποίησης τεράστιων και πολύπλοκων συνόλων δεδομένων για την προώθηση της τεκμηριωμένης λήψης αποφάσεων, την τόνωση της οικονομικής ανάπτυξης και την αντιμετώπιση των κοινωνικών προκλήσεων.

Καθώς τα ευρωπαϊκά έθνη περιηγούνται σε ένα ψηφιακό τοπίο, τα μεγάλα δεδομένα γίνονται βασικός άξονας για βιομηχανίες που κυμαίνονται από την υγειονομική περίθαλψη και τη χρηματοδότηση έως τη μεταποίηση και τις δημόσιες υπηρεσίες. Η στρατηγική κατεύθυνση της Ευρώπης προς τα μεγάλα δεδομένα σηματοδοτεί μια συλλογική στροφή προς τη διακυβέρνηση, την καινοτομία και την ανταγωνιστικότητα με γνώμονα τα δεδομένα. Η

υιοθέτηση αυτού του νέου μοντέλου επιτρέπει στις ευρωπαϊκές χώρες να βελτιστοποιήσουν την κατανομή των πόρων, να ενισχύσουν τη λειτουργική αποτελεσματικότητα και να προωθήσουν τα οικοσυστήματα καινοτομίας.

Επιπλέον, καθώς η Ευρώπη κατευθύνει την πορεία της, δίνεται αυξανόμενη έμφαση στη διασφάλιση ηθικών και υπεύθυνων πρακτικών δεδομένων, ευθυγραμμιζόμενες με ισχυρά ρυθμιστικά πλαίσια, όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR), για τη διασφάλιση της ιδιωτικής ζωής του ατόμου, ενώ παράλληλα ξεκλειδώνει το πλήρες δυναμικό των μεγάλων δεδομένων. για κοινωνικό όφελος.

Η Ευρωπαϊκή Επιτροπή τοποθετεί πολύ υψηλά στην ιεράρχηση των στόχων την ανάλυση και αξιοποίηση των δεδομένων, στο δρόμο για τον ψηφιακό μετασχηματισμό της μέχρι το 2030. Βάσει των ψηφιακών στόχων που έχει θέσει η ΕΕ απαιτείται έως το 2030 να έχουν υιοθετήσει τα big data πάνω από το 75% των ευρωπαϊκών εταιρειών.

Οι επιχειρήσεις σε ολόκληρη την ΕΕ προσαρμόζονται συνεχώς στις νέες τεχνολογίες για τη συλλογή, αποθήκευση και ανάλυση δεδομένων. Το 2020, το 14% των επιχειρήσεων πραγματοποίησε ανάλυση big data. Όπως επισημαίνει έκθεση της Κομισιόν, αυτό τις βοήθησε να παράγουν αποτελέσματα σχεδόν σε χρόνο ή σε πραγματικό χρόνο από δεδομένα που προέρχονται από διαφορετικούς τύπους μορφής.

Οι μεγάλες εταιρείες έχουν τη μερίδα του λέοντος στην επεξεργασία big data (με 34% αυτών να τα χρησιμοποιούν), ενώ οι μικρομεσαίες επιχειρήσεις έχουν ακόμη περιθώρια βελτίωσης για να εκμεταλλευτούν όλα τα οφέλη των big data, καθώς μόλις το 14% εξ αυτών τα χρησιμοποιούν.

Κεφάλαιο 10°

10.Big Data στην Ελλάδα

Στην Ελλάδα, η σημασία των μεγάλων δεδομένων αναγνωρίζεται ολοένα και περισσότερο ως καταλύτης για την οικονομική ανάπτυξη, την καινοτομία και τις βελτιωμένες δημόσιες υπηρεσίες. Η χώρα υιοθετεί μια νέα κατεύθυνση προς τα μεγάλα δεδομένα, αναγνωρίζοντας τις δυνατότητές της να αυξήσει την αποτελεσματικότητα και να παρέχει στρατηγικές γνώσεις σε διάφορους τομείς. Η εξελισσόμενη εστίαση της Ελλάδας στα μεγάλα δεδομένα

ευθυγραμμίζεται με τις παγκόσμιες τάσεις, αντανακλώντας τη δέσμευση για ψηφιακό μετασχηματισμό και λήψη αποφάσεων βάσει δεδομένων.

Για τον σκοπό αυτό, το υπουργείο Ψηφιακής Διακυβέρνησης έχει προκηρύξει διαγωνισμό για την υλοποίηση ενός κεντρικού κόμβου διαχείρισης και ανάλυσης μεγάλων δεδομένων, όπου θα μπορέσει να συγκεντρωθεί και να ομογενοποιηθεί η πληροφορία που όλα τα συστήματα του Δημοσίου προκειμένου να καταγράφεται και να παρακολουθείται η αποδοτικότητα υπηρεσιών και φορέων.

Το έργο, θα έχει προϋπολογισμό 22,4 εκατ. ευρώ, με τον κόμβο που θα δημιουργηθεί να συλλέγει τα δεδομένα του από συστήματα που υποστηρίζουν εσωτερικές διαδικασίες του Δημοσίου, όπως το Ολοκληρωμένο Πληροφοριακό Σύστημα Δημοσιονομικής Πολιτικής του ΓΛΚ (Gov ERP), το Σύστημα Διαχείρισης Ανθρώπινου Δυναμικού (ΣΔΑΔ) και το Σύστημα Μισθοδοσίας (Ενιαία Αρχή Πληρωμών).

Θα συγκεντρώνει επίσης δεδομένα από συστήματα υποστήριξης πολιτών, όπως το Μητρώο Πολιτών, η πλατφόρμα ηλεκτρονικής συνταγογράφησης, το Εθνικό Κτηματολόγιο και το Εθνικό Μητρώο, αλλά και επιλεγμένοι ιστότοποι και μέσα κοινωνικής δικτύωσης όπου πραγματοποιείται δημόσιος διάλογος για τις δράσεις και τις μεταρρυθμίσεις του πλαισίου «Ελλάδα 2.0».

Στόχος του έργου θα είναι η παροχή πληροφόρησης τόσο προς κυβερνητικούς χρήστες (Κεντρική Κυβέρνηση, Δημόσιες Υπηρεσίες, κλπ.) όσο και προς εξωτερικούς χρήστες (π.χ. πολίτες, ανεξάρτητες αρχές, θεσμοί, κλπ.), ώστε να υποστηρίζεται επαρκώς ο κυβερνητικός στρατηγικός σχεδιασμός, η διαχείριση κρίσεων, η παρακολούθηση της ποιότητας των υπηρεσιών προς τους πολίτες, ο σχεδιασμός νέων υπηρεσιών και γενικότερα η υποστήριξη της διαδικασίας παροχής ενημέρωσης/ πληροφόρησης και της λήψης αποφάσεων.

Οι αναφορές και οι αναλύσεις που θα παρέχονται από τον Κεντρικό Κόμβο Διαχείρισης και Ανάλυσης Πολυδιάστατων Δεδομένων Μεγάλου Όγκου κατηγοριοποιούνται σε τέσσερις βασικούς πυλώνες (Πυλώνες Διαχείρισης & Ανάλυσης Δεδομένων):

1. Οικονομία & Ανάπτυξη
2. Κοινωνικές Δομές & Υπηρεσίες
3. Περιβάλλον & Υποδομές
4. Ψηφιακός μετασχηματισμός & κοινή γνώμη

Ακολουθούν μερικές εταιρείες που είναι γνωστές για την εξειδίκευσή τους στην ανάλυση δεδομένων, τα μεγάλα δεδομένα και τις σχετικές τεχνολογίες στην Ελλάδα:

- Deloitte
- EY
- Accenture Greece
- Intrasoft International
- UniSystems
- Intelen Inc
- Qualco
- Advantis S.A

Συμπεράσματα

Η ραγδαία εξέλιξη της τεχνολογίας, κυρίως στην εφαρμογή και διαχείριση μεγάλου όγκου δεδομένων, κατέστησαν τα Big Data ως αναπόσπαστο κομμάτι της επιχειρηματικότητας της εποχής μας. Καθώς αναλογιζόμαστε τα ευρήματα που παρουσιάζονται σε αυτή τη διατριβή, γίνεται προφανές ότι το τοπίο της επιχειρηματικότητας εξελίσσεται ως απάντηση στις ραγδαίες εξελίξεις στις τεχνολογίες Big Data. Οι επιχειρηματίες αναγνωρίζουν όλο και περισσότερο την ανάγκη να αξιοποιήσουν τη δύναμη των γνώσεων που βασίζονται σε δεδομένα για την πλοήγηση σε περίπλοκες δυναμικές της αγοράς, τον εντοπισμό ευκαιριών και τον μετριασμό των κινδύνων αποτελεσματικά.

Ωστόσο, το ταξίδι προς την πλήρη αξιοποίηση των δυνατοτήτων των Μεγάλων Δεδομένων στην επιχειρηματικότητα είναι συνεχές και αρκετοί δρόμοι για μελλοντική έρευνα αξίζουν προσοχή. Πρώτα και κύρια, οι ηθικοί προβληματισμοί σχετικά με τη χρήση Big Data σε επιχειρηματικές επιχειρήσεις απαιτούν μια πιο προσεκτική εξέταση. Οι μελλοντικές μελέτες θα πρέπει να εμβαθύνουν στις ηθικές συνέπειες της συλλογής, αποθήκευσης και ανάλυσης δεδομένων, με έμφαση στη διασφάλιση υπεύθυνων και διαφανών πρακτικών.

Επιπλέον, η έρευνα σχετικά με την προγνωστική ανάλυση της επιχειρηματικής επιτυχίας έχει μεγάλες δυνατότητες. Η ανάπτυξη μοντέλων που μπορούν να προβλέψουν την πιθανότητα επιτυχίας για νέα εγχειρήματα με βάση ιστορικά δεδομένα και βασικούς παράγοντες επιτυχίας θα μπορούσε να είναι καθοριστική στη διαμόρφωση στρατηγικής λήψης αποφάσεων τόσο για επιχειρηματίες όσο και για επενδυτές.

Επιπλέον, ο ρόλος της Τεχνητής Νοημοσύνης (AI) στη λήψη επιχειρηματικών αποφάσεων, οι διακλαδικές εφαρμογές Big Data και ο μακροπρόθεσμος αντίκτυπος στα επιχειρηματικά οικοσυστήματα αντιπροσωπεύουν πλούσιους τομείς για περαιτέρω εξερεύνηση. Αυτοί οι τομείς υπόσχονται να αποκαλύψουν νέες διαστάσεις της δυναμικής αλληλεπίδρασης μεταξύ των μεγάλων δεδομένων και της επιχειρηματικότητας, προσφέροντας πολύτιμη γνώση τόσο για τον ακαδημαϊκό κόσμο όσο και για τους επαγγελματίες του κλάδου.

Καθώς κοιτάζουμε προς το μέλλον, είναι επιτακτική ανάγκη οι ερευνητές, οι επιχειρηματίες και οι υπεύθυνοι χάραξης πολιτικής να συνεργαστούν για να ξεκλειδώσουν πλήρως τις δυνατότητες των Big Data, ανοίγοντας το δρόμο για καινοτόμες και βιώσιμες επιχειρηματικές πρακτικές.

Κατάλογος Αναφορών

Accenture. (2016). Big data analytics in supply chain: Hype or here to stay? Available at <https://www.accenture.com>

Agneeswaran, V., 2012. Big-data - Theoretical, engineering and analytics perspective. In: Srinivasa, S., Bhatnagar, V. (Eds.), Big Data Analytics. Vol. 7678 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 8-15.

Almeida, F. (2018). Big data: Concept, potentialities and vulnerabilities. *Emerging Science Journal*, 2(1), 1–10. <https://doi.org/10.28991/esj-2018-01123>

Ashabi, Ardavan & Sahibuddin, Shamsul & Haghghi, Mehdi. (2020). Big Data: Current Challenges and Future Scope. 131-134. 10.1109/ISCAIE47305.2020.9108826.

Assuncao, M., Bianchi, S., Buyya, R., Assunção, M. D., Calheiros, R. N., & Netto, M. A. S. (n.d.). *Big Data computing and clouds: Trends and future directions Workflow as a Service in Cloud Computing Environment: Scheduling and Resource Provisioning Techniques View project design methods for finance View project Big Data Computing and Clouds: Trends and Future Directions*. <https://www.researchgate.net/publication/265129513>

Baker, W., Kiewell, D., & Winkler, G. (2014). Using big data to make better pricing decisions. McKinsey. Retrieved from <http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions>

Bakshi, K., 2012. Considerations for big data: Architecture and approach. In: 2012 IEEE Aerospace Conference, Big Sky Montana. pp. 1-7.

Beck, J., Mostow, J., 2008. How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. In: Proceedings of the 9th International Conference on Intelligent Tutoring Systems. pp. 353-362.

Bhadani, A., Jothimani, D. (2016), Big data: Challenges, opportunities and realities, In Singh, M.K., & Kumar, D.G. (Eds.), Effective Big Data Management and Opportunities for Implementation (pp. 1-24), Pennsylvania, USA, IGI Global

BME Master thesis- Mousouleas Ioannis. (n.d.).

Brewer, E. A., 2000. Towards robust distributed systems (abstract). In: Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing (PODC '00). ACM, New York, NY, USA, p. 7.

Brown, B., Chui, M., Manyika, J., October 2011. Are you ready for the era of big data? McKinsey Quarterly, <http://www.mckinsey.com/insights/strategy/are-you-ready-for-the-era-of-big-data>, Accessed on: Jan 20, 2015.

Brzozowska, A., Ziora, L., Sałek, R., & Wisniewska-Salek, A. (2018, September 29). *The Possibilities of Big Data Solutions Application in Logistics*. <https://doi.org/10.26649/musci.2016.044>

C. Kruse, R. Goswamy, Y. Raval, and S. Marawi, “Challenges and Opportunities of Big Data in Health Care: A Systematic

Chen, H., Chiang, R. H. L., Storey, V. C., Lindner, C. H., & Robinson, J. M. (2012). Business Intelligence and Analytics: From Big Data to Big Impact Quarterly-Business Intelligence and Analytics: From Big Data to Big Impact. In *Source: MIS Quarterly* (Vol. 36, Issue 4).

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209. <https://doi.org/10.1007/s11036-013-0489-0>

COLUMBUS, L.: Ten Ways Big Data Is Revolutionizing Supply Chain Management, Forbes.com LLC™ 2016, available online at: <http://onforb.es/1CzLEg2>.

Coumaros, J., de Roys, S., Chretien, L., Buvat, J., KVJ, S., Clerk, V., Auliard, O., 2014. Big data alchemy: How can banks maximize the value of their customer data? Capgemini Consulting White Paper <https://www.capgemini.com/resources/big-data-customer-analytics-in-banks>, Accessed on: March 20, 2015.

Cukier, K. (2010) Data, Data Everywhere: A Special Report on Managing Information. *The Economist*, 394, 3-5.

Davenport, T. H. (2014). How strategists use “big data” to support internal business decisions, discovery and production. *Strategy and Leadership*, 42(4), 45–50. <https://doi.org/10.1108/SL-05-2014-0034>

Davenport, T. H., & Dyché, J. (2013). *Big Data in Big Companies*.

Dezi, L., Santoro, G., Gabteni, H., & Pellicelli, A. C. (2018). The role of big data in shaping ambidextrous business process management: Case studies from the service industry. *Business Process Management Journal*, 24(5), 1163–1175. <https://doi.org/10.1108/BPMJ-07-2017-0215>

Dhar V., 2013. Data science and prediction. *Communications of ACM* 56(12), 64–73.

F. Sarrocco, V. Morabito, and G. Meyer, Exploring the Next Generation Financial Services: The Big Data Revolution, 2016, available at https://www.accenture.com/t20170314T051509_w_/nl-en/acnmedia/PDF/20/Accenture-Next-Generation-Financial.pdf.

Fisher, D., DeLine, R., Czerwinski, M., Drucker, S., 2012. Interactions with big data analytics. *Interactions* 19 (3), 50-59.

Forsyth, C., January 2012. For big data analytics there's no such thing as too big. Cisco White paper http://www.cisco.com/en/US/solutions/ns340/ns517/ns224/big_data_wp.pdf, Accessed on Feb 20, 2015.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Gantz J, Reinsel D (2011) Extracting value from chaos. IDC iView, pp 1–12

Gilbert, S., Lynch, N., 2002. Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. *SIGACT News* 33 (2), 51-59.

House, J. (2014, November 20). Big data analytics = Key to successful 2015 supply chain strategy. ModusLink. Retrieved from <https://www.moduslink.com/big-data-analytics-key-successful-2015-supply-chain-strategy/>

<http://www.paracel.com/resources/Whitepapers/Hadoop-Limitations-for-Big-Data-ParAccel-Whitepaper.pdf>

IBM. (2010). How big data is giving Hertz a big advantage. Retrieved from <https://www-01.ibm.com/software/ebusiness/jstart/portfolio/hertzCaseStudy.pdf>

Institute of Electrical and Electronics Engineers. (n.d.). *ISCAIE 2020 : IEEE 10th Symposium on Computer Applications and Industrial Electronics : Malaysia, 18-19 April 2020*.

Jin, X., Wah, B. W., Cheng, X., Wang, Y., 2015. Significance and challenges of big data research. *Big Data Research* 2 (2), 59-64.

.Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293–303. <https://doi.org/10.1016/j.bushor.2017.01.004>

Manovich, L., 2011. Trending: the promises and the challenges of big social data. In: Gold, M. K. (Ed.), *Debates in the Digital Humanities*. The University of Minnesota Press, Minneapolis, MN. Available at: http://www.manovich.net/DOCS/Manovich_trending_paper.pdf, Accessed on: 15 July 2015

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H., 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Meek, T., February 2015. Big data in retail: How to win with predictive analytics. *Forbes* <http://www.forbes.com/sites/netapp/2015/02/18/big-data-in-retail/>, Accessed on: May 27, 2015.

Miller, H. G., Mork, P., 2013. From data to decisions: A value chain for big data. *IT Professional* 15 (1), 57-59.

Minelli, M., Chambers, M., Dhiraj, A., 2013. *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, 1st Edition. Wiley Publishing, Wiley CIO Series.

Mor, Y., 2014. Big data and law enforcement: Was „minority report“ right? *Wired*, <http://www.wired.com/2014/03/big-data-law-enforcement-minority-report-right/>, Accessed on: June 20, 2015

Mukherjee, A., Pal, A., Misra, P., 2012. Data analytics in ubiquitous sensor-based health information systems. In: *6th International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST)*. pp. 193-198.

- N. Mohamed and J. Al-Jaroodi, "Real-time big data analytics: Applications and challenges," Proc. 2014 Int. Conf. High Perform. Comput. Simulation, HPCS 2014, no. October, pp. 305–310, 2014.
- N. T. Tariq RS, "Big Data Challenges," Comput. Eng. Inf. Technol., vol. 04, no. 03, 2015.
of-the-manufacturing-industry/.
- P. Murali K, M. Salehi Amini, K. Jayasimha R., Y. Xie, and V. Raghavan, "Massive Data Analysis: Tasks, Tools, Applications, and Challenges," Big Data Anal. Methods Appl., pp. 1–276, 2016.
- ParAccel, 2012. Hadoops limitations for big data analytics. ParAccel White Paper.
- Parashar, M., Jaypee Institute of Information Technology University, University of Florida. College of Engineering, Institute of Electrical and Electronics Engineers. Delhi Section, & Institute of Electrical and Electronics Engineers. (n.d.). *2013 sixth International Conference on Contemporary Computing (IC3-2013) : 8-10 August 2013, Jaypee Institute of Information Technology, Noida, India.*
- Patgiri, R., & Ahmed, A. (2016). *Big Data: The V's of the Game Changer Paradigm.* <https://doi.org/10.1109/HPCC-SmartCity-DSS.2016.8>
- Porter, M., 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors.* The Free Press, New York.
- R. Agrawal and C. Nyamful, "Challenges of big data storage and management," Glob. J. Inf. Technol., vol. 6, no. 1, 2016.
- R. Delgado, Big Data's Transformation of the Manufacturing Industry, 2017, available at <http://data-informed.com/big-datastransformation->
- R. Milan, K. Kumar Pandey, and D. Shukla, "Security and Privacy Challenges in Big Data Environment," in National Conference on "Data Analytics, Machine Learning and Security" 15 – 16 February 2018, 2018, no. February, pp. 315–325.
- R. S. K. Althaf, R. K. Sai, and R. K. Girija, "Challenging tools on Research Issues in Big Data Analytics," Int. J. Eng. Dev. Res., vol. 6, no. 1, pp. 637–644, 2018.

Review”, *JMIR Medical Informatics*, vol. 4, no. 4, 2016. doi: 10.2196/medinform.5359.

S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, “Big data preprocessing: methods and prospects,” *Big Data Anal.*, vol. 1, no. 1, p. 9, 2016.

S. Mishra, V. Dhote, G. S. Prajapati, and J. P. Shukla, “Challenges in Big Data Application: A Review,” *Int. J. Comput. Appl.*, vol. 121, no. 19, pp. 42–46, 2015.

Sun, Z., Hupman, A. C., & Abbas, A. E. (2021). The value of information for price dependent demand. *European Journal of Operational Research*, 288(2), 511–522. <https://doi.org/10.1016/j.ejor.2020.05.057>

U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, “Critical analysis of Big Data challenges and analytical methods,” *J. Bus. Res.*, vol. 70, pp. 263–286, 2017.

V. Agrawal, *The Impact of Big Data and Analytics on Manufacturing*, 2016, available at <https://tech.co/big-data-analyticsmanufacturing->

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A review. In *Agricultural Systems* (Vol. 153, pp. 69–80). Elsevier Ltd. <https://doi.org/10.1016/j.agsy.2017.01.023>

Zikopoulos P, Eaton C et al (2011) *Understanding big data: analytics for enterprise class hadoop and streaming data*. McGraw- Hill Osborne Media