

**Τμήμα
Μηχανικών
Πληροφορικής τ.ε.**

Τεχνολογικό Εκπα. δευτικό Ίδρυμα
Δυτικής Ελλάδας

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιχειρηματική Ευφυΐα και Σύγχρονες Τάσεις σε
Μεγάλο Όγκο Δεδομένων

Φώτης Σοφούλης Α.Μ.2227

Επιβλέπων καθηγητής: Ιωάννης Τζήμας

Αντίρριο 2019

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή
Αντίρριο, Ημερομηνία

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Ονοματεπώνυμο, Υπογραφή
2. Ονοματεπώνυμο, Υπογραφή
3. Ονοματεπώνυμο, Υπογραφή

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Ιωάννη Τζήμα για την ώθηση του, να ασχοληθώ με το συγκεκριμένο αντικείμενο, καθώς και την καθοδήγηση του και τις συμβουλές του καθ' όλη τη διάρκεια των σπουδών μου.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για την στήριξη τους όλα αυτά τα χρόνια.

Περίληψη

Στην παρούσα πτυχιακή εργασία με θέμα “*Επιχειρηματική Ευφυΐα και Σύγχρονες Τάσεις σε Μεγάλο Όγκο Δεδομένων*”, παρουσιάζεται η σημασία και ο αντίκτυπος της επιχειρηματικής ευφυΐας, για την διαχείριση του τεράστιου όγκου δεδομένων των επιχειρήσεων, με στόχο την καλύτερη λήψη αποφάσεων, καθώς και οι νεότερες τάσεις σε αυτήν. Αρχικά, αναλύεται η έννοια της επιχειρηματικής ευφυΐας, και διατυπώνεται μία ιστορική αναδρομή της. Στη συνέχεια, αναφέρεται η σημασία και ο αντίκτυπός της στην σύγχρονη επιχείρηση, καθώς και η αξιοποίηση της σε πραγματικό χρόνο. Ύστερα, παρουσιάζεται η έννοια, η ιστορία και τα χαρακτηριστικά των μεγάλων δεδομένων, όπως επίσης και η χρησιμότητά τους. Επιπλέον, αναλύονται εργαλεία και τεχνολογίες, βασισμένα στο οικοσύστημα του Hadoop. Στη συνέχεια, παρουσιάζεται ο ορισμός και η εξέλιξη της εξόρυξης δεδομένων και επισημαίνονται οι τεχνικές εξόρυξης δεδομένων και η ενσωμάτωσή της στις επιχειρηματικές διαδικασίες. Μετά, αναλύονται οι σύγχρονες τάσεις στην επιχειρηματική ευφυΐα, με βάση ερευνητική μελέτη της BARC. Έπειτα, αναφέρονται μερικοί από τους προμηθευτές εμπορικών προϊόντων επιχειρηματικής ευφυΐας, καθώς και τα πλεονεκτήματα και τα μειονεκτήματά τους, με βάση το MagicQuadrant του 2018, της εταιρείας Gartner. Επίσης, γίνεται αναφορά σε ένα πλήθος προϊόντων επιχειρηματικής ευφυΐας ανοιχτού κώδικα, όπως και τα θετικά και τα αρνητικά των λογισμικών ανοιχτού κώδικα. Στο τέλος της εργασίας παρατίθενται συμπεράσματα και ένας συνοπτικός συγκριτικός πίνακας των προμηθευτών επιχειρηματικής ευφυΐας.

Λέξεις κλειδιά: επιχειρηματική ευφυΐα, μεγάλα δεδομένα, Hadoop, εξόρυξη δεδομένων

Abstract

This thesis on "Business Intelligence and Current Trends in Big Data" presents the importance and impact of business intelligence, in order to manage the vast amount of business data for better decision making, as well as current trends. Initially, the concept of business intelligence is analyzed, and a historical review of it is made. Next, the significance and impact on the modern business are proposed, as well as its use in real time scenarios. Then, the concept, history and features of Big Data, as well as its usefulness, are presented. In addition, tools and technologies are analyzed, based on the Hadoop ecosystem. Afterwards, the definition and evolution of data mining is presented, and furthermore, data mining techniques and their integration into business processes are highlighted. Then, current trends in business intelligence are analyzed, based on the BARC research study. Next, some of the vendors of commercial business intelligence products are listed, as well as their pros and cons based on the 2018 Magic Quadrant, by Gartner. Also, reference to a number of open source business intelligence products is made, as well as the positive and negative aspects of open source software. Finally, the conclusions are presented, as well as, a summary table of business intelligence vendors.

Key words: Business Intelligence, Big Data, Hadoop, Data Mining

Πρόλογος

Στη σημερινή εποχή του πελατοκεντρικού και ψηφιακού κόσμου, η αναζήτηση, ο έλεγχος, αλλά και η κατανόηση των δεδομένων αποτελούν βασικά στοιχεία για την καλύτερη λήψη αποφάσεων μίας επιχείρησης. Η λήψη αποφάσεων είναι μία από τις σημαντικότερες διαδικασίες των διοικητικών στελεχών μίας επιχείρησης, η οποία, τα τελευταία χρόνια, λόγω της υπερφόρτωσης πληροφοριών, καθίσταται ακόμα πιο δύσκολη.

Ο τεράστιος όγκος πληροφοριών που προέρχονται τόσο από προσωπικά δεδομένα, όπως όνομα, επίθετο, ηλικία, φύλλο, διεύθυνση και ιθαγένεια, όσο από επιχειρησιακά στοιχεία, όπως ηλεκτρονικές αγορές, προτιμήσεις προϊόντων, ερωτηματολόγια και διαφημιστικές καμπάνιες, αλλά και δεδομένα της ίδιας της επιχείρησης, καθιστά σαφές ότι η σύγχρονη επιχείρηση έχει στη διάθεση της μία εξαιρετικά μεγάλη πληθώρα και ποικιλία δεδομένων.

Με την κατάλληλη επεξεργασία και ανάλυση, τα δεδομένα αυτά, μπορούν να αποτελέσουν μία πολύτιμη πηγή πληροφόρησης, αναγκαία για τη βελτίωση της λήψης επιχειρησιακών και στρατηγικών αποφάσεων. Η παροχή αυξημένης πληροφόρησης οδηγεί στην καλύτερη κατανόηση του προβλήματος, και κατ' επέκταση στη μείωση της αβεβαιότητας και στον περιορισμό του ρίσκου. Ως αποτέλεσμα, αυξάνονται οι επιδόσεις της επιχείρησης και εξασφαλίζεται το ανταγωνιστικό πλεονέκτημα.

Η ανάγκη των επιχειρήσεων για βελτιωμένη πληροφόρηση και αναβάθμιση των διαδικασιών λήψης αποφάσεων, η διαθεσιμότητα μεγάλου όγκου και ποικιλομορφίας δεδομένων, και η ανάπτυξη νέων τεχνολογιών και μεθοδολογιών για την ανάλυση των δεδομένων, αποτέλεσαν τα μέσα για την ανάπτυξη συστημάτων επιχειρηματικής ευφυΐας.

Τις τελευταίες δεκαετίες, τα συστήματα επιχειρηματικής ευφυΐας βρίσκονται στο επίκεντρο του ενδιαφέροντος του επιχειρηματικού κόσμου. Με βάση μελέτες που πραγματοποιούν οίκοι ερευνών και συμβουλευτικών υπηρεσιών, όπως η Gartner και η BARC, η επιχειρηματική ευφυΐα κατατάσσεται στις κορυφαίες θέσεις των τεχνολογικών προτεραιοτήτων των επιχειρήσεων παγκοσμίως. Ως αποτέλεσμα, έχει αναπτυχθεί μια αγορά σχετικών συστημάτων και λογισμικού με κύκλο εργασιών της τάξης δεκάδων δισεκατομμυρίων, με κορυφαίες εταιρείες πληροφορικής, όπως η Microsoft, η IBM και η Oracle, αλλά και εξειδικευμένες και νεότερες επιχειρήσεις,

όπως η Qlik, η Tableau και η Sisense να προσφέρουν ανταγωνιστικές και προσιτές λύσεις στην αγορά.

Περιεχόμενα

Ευχαριστίες	3
Περίληψη	4
Abstract	5
Πρόλογος	6
Περιεχόμενα.....	8
Πίνακας Εικόνων	10
1. Ορισμός – Επιχειρηματική Ευφυΐα	11
2. Ιστορία της Επιχειρηματικής Ευφυΐας	12
3. Σημασία και αντίκτυπος της επιχειρηματικής ευφυΐας.....	13
4. Επιχειρηματική ευφυΐα σε πραγματικό χρόνο.....	16
5. Επιχειρηματική ευφυΐα και Μεγάλα δεδομένα	18
5.1. Τι είναι τα μεγάλα δεδομένα	18
5.2. Ιστορία των μεγάλων δεδομένων	19
5.3. Χαρακτηριστικά και κατηγορίες μεγάλων δεδομένων.....	20
5.3.1. Χαρακτηριστικά μεγάλων δεδομένων	20
5.3.2. Κατηγορίες Μεγάλων Δεδομένων	22
5.4. Εργαλεία και τεχνολογίες μεγάλων δεδομένων	24
5.5. Χρησιμότητα των μεγάλων δεδομένων	45
6. Επιχειρηματική ευφυΐα και εξόρυξη δεδομένων.....	48
6.1. Τί είναι η εξόρυξη δεδομένων	48
6.2. Ιστορία και εξέλιξη	49
6.3. Διαδικασία ανακάλυψης γνώσης και τεχνικές εξόρυξης δεδομένων.....	50
6.3.1. Διαδικασία ανακάλυψης γνώσης σε βάσεις δεδομένων.....	50
6.3.2. Τεχνικές εξόρυξης δεδομένων	52
6.4. Εξόρυξη δεδομένων στις επιχειρηματικές διαδικασίες	54
7. Σύγχρονες τάσεις στην επιχειρηματική ευφυΐα.....	56
7.1. Προβλέψεις της Gartner για την επιχειρηματική ευφυΐα.....	56
7.2. Οι κυριότερες σύγχρονες τάσεις στην επιχειρηματική ευφυΐα.....	57
7.2.1. Διαχείριση κύριων δεδομένων & ποιότητας δεδομένων	58
7.2.2. Ανακάλυψη & Οπτικοποίηση δεδομένων.....	58
7.2.3. Αυτοεξυπηρετούμενη επιχειρηματική ευφυΐα.....	59

7.2.4.	Διακυβέρνηση δεδομένων	60
7.2.5.	Ανάλυση πραγματικού χρόνου.....	60
8.	Προμηθευτές και εμπορικά προϊόντα επιχειρηματικής ευφυΐας.....	61
8.1.	Microsoft.....	61
8.1.1.	Ιστορία	61
8.1.2.	Προϊόντα.....	62
8.1.3.	Δυνατά σημεία	64
8.1.4.	Αδυναμίες.....	65
8.2.	IBM.....	67
8.2.1.	Ιστορία	67
8.2.2.	Προϊόντα.....	67
8.2.3.	Δυνατά σημεία	69
8.2.4.	Αδυναμίες.....	71
8.3.	Qlik.....	72
8.3.1.	Ιστορία	72
8.3.2.	Προϊόντα.....	72
8.3.3.	Δυνατά σημεία	74
8.3.4.	Αδυναμίες.....	75
8.4.	Tableau	77
8.4.1.	Ιστορία	77
8.4.2.	Προϊόντα.....	77
8.4.3.	Δυνατά σημεία	79
8.4.4.	Αδυναμίες.....	80
8.5.	Sisense	82
8.5.1.	Ιστορία	82
8.5.2.	Προϊόντα.....	82
8.5.3.	Δυνατά σημεία	83
8.5.4.	Αδυναμίες.....	85
9.	Επιχειρηματική Ευφυΐα & Λογισμικό Ανοιχτού Κώδικα.....	86
9.1.	Τί είναι το λογισμικό ανοιχτού κώδικα.....	86
9.2.	Πλεονεκτήματα	86
9.3.	Μειονεκτήματα.....	87
9.4.	Προμηθευτές επιχειρηματικής ευφυΐας ανοιχτού κώδικα.....	88

9.4.1.	Jaspersoft Community	88
9.4.2.	Birt	89
9.4.3.	Pentaho	91
10.	Σύνοψη	93
	Βιβλιογραφία.....	98

Πίνακας Εικόνων

Εικόνα 1	Το φάσμα των επιχειρηματικών πληροφοριών	15
Εικόνα 2	Τα 3V των Μεγάλων Δεδομένων	21
Εικόνα 3	Τύποι Μεγάλων Δεδομένων.....	24
Εικόνα 4	Το οικοσύστημα του Hadoop	27
Εικόνα 5	Η αρχιτεκτονική του HDFS.....	28
Εικόνα 6	Η αρχιτεκτονική του YARN	30
Εικόνα 7	Η εσωτερική λειτουργία του Hive.....	36
Εικόνα 8	Οι κύριες φάσεις της ανακάλυψης γνώσης	51
Εικόνα 9	Ο κύριος πίνακας ελέγχου του Power BI.....	63
Εικόνα 10	Ο βασικός πίνακας ελέγχου του Cognos Analytics	68
Εικόνα 11	Ο κύριος πίνακας ελέγχου του Watson Analytics.....	69
Εικόνα 12	Ο πίνακας ελέγχου του QlikView	73
Εικόνα 13	Η αρχική οθόνη του Tableau	78
Εικόνα 14	Η αρχική οθόνη του Sisense	83
Εικόνα 15	Η κύρια οθόνη του JasperSoft	89
Εικόνα 16	Η κύρια οθόνη του BIRT	91
Εικόνα 17	Ο βασικός πίνακας ελέγχου του Pentaho	92

1. Ορισμός – Επιχειρηματική Ευφυΐα

Με βάση την TechTarget ο ορισμός της Επιχειρηματικής Ευφυΐας είναι, *“Είναι μία τεχνολογικά-καθοδηγούμενη διαδικασία για την ανάλυση δεδομένων και την παρουσίαση πληροφοριών, οι οποίες μπορούν να βοηθήσουν στελέχη μίας επιχείρησης να λάβουν πληροφορημένες επιχειρηματικές αποφάσεις”*

Περαιτέρω αναφέρει ότι, η επιχειρηματική ευφυΐα περιλαμβάνει μια ευρεία ποικιλία από εργαλεία, εφαρμογές και μεθοδολογίες που επιτρέπουν σε οργανισμούς τη συλλογή και την αποθήκευση των δεδομένων τους, από εσωτερικές ή εξωτερικές πηγές, την προετοιμασία των δεδομένων αυτών προς ανάλυση, την ανάπτυξη ερωτημάτων πάνω στα δεδομένα ώστε να είναι δυνατή η δημιουργία αναφορών και απεικονίσεων δεδομένων. Με τα αναλυτικά εξαγόμενα, η διαδικασία της λήψης αποφάσεων για τα στελέχη της επιχείρησης είναι πιο αποτελεσματική, καθώς έχουν μία πιο καθαρή εικόνα στα δεδομένα.

Σύμφωνα με την Wikipedia, η επιχειρηματική ευφυΐα είναι *“τεχνολογίες που έχουν την δυνατότητα να ανταπεξέλθουν σε μεγάλο όγκο, δομημένων ή μη-δομημένων δεδομένων για την αναγνώριση και ανάπτυξη ή εκ νέου δημιουργία στρατηγικών επιχειρησιακών αποφάσεων. Στοχεύουν ώστε η αναπαράσταση των δεδομένων να είναι απλή και κατανοητή, ώστε η υλοποίηση μίας στρατηγικής με βάση τα δεδομένα αυτά να δώσει το ανταγωνιστικό πλεονέκτημα στην επιχείρηση.”*

Το Wikipedia, όπως και άλλοι φορείς που έχουν βοηθήσει με την πρόσδοση του ορισμού αυτού, έχουν δώσει ιδιαίτερη σημασία στον όρο «Μεγάλα Δεδομένα»(BigData). Ο συνεχώς αυξανόμενος αριθμός των δομημένων και μη-δομημένων δεδομένων, τόσο σε όγκο και πολυπλοκότητα, όσο και στην ταχύτητα εισαγωγής των δεδομένων αυτών, στις αποθήκες δεδομένων των επιχειρήσεων, καθιστά την αποθήκευση, επεξεργασία και αναπαράσταση τους αδύνατη από παραδοσιακά εργαλεία εξόρυξης γνώσης. Λόγο των παραπάνω, το μεγαλύτερο ποσοστό των επιχειρήσεων μεταβαίνει σε ένα μοντέλο ανάπτυξης και υλοποίησης το οποίο να μπορεί να υποστηρίξει «Μεγάλα Δεδομένα» και να μπορεί να εκλάβει χρήσιμες στρατηγικές πληροφορίες από αυτά μέσω εργαλείων επιχειρηματικής ευφυΐας.

2. Ιστορία της Επιχειρηματικής Ευφυΐας

Η παλαιότερη χρήση του όρου «Επιχειρηματική Ευφυΐα» ήταν από τον Richard Millar Devens στο βιβλίο του με τίτλο “Cyclopaedia of Commercial and Business Anecdotes” το 1865. Στο βιβλίο αυτό, χρησιμοποίησε τον όρο «επιχειρηματική ευφυΐα» για να περιγράψει πως ένας τραπεζίτης με το όνομα Sir Henry Furnese αποκόμισε κέρδος, αξιοποιώντας τις πληροφορίες από το περιβάλλον του, νωρίτερα από τους αντιπάλους του. Η ικανότητα του στην συλλογή και επεξεργασία, όπως και στην λήψη κατάλληλων αποφάσεων με βάση τις εισερχόμενες πληροφορίες του, είναι τα βασικά θεμέλια της επιχειρηματικής ευφυΐας.

Ύστερα, το 1958, σε ένα άρθρο της IBM με τίτλο «A Business Intelligence System», ο ερευνητής της IBM Hans Peter Luhn χρησιμοποίησε μια πιο αυθαίρετη μορφή του όρου επιχειρηματική ευφυΐα με βάση τον ορισμό της νοημοσύνης, από το λεξικό Webster, η οποία ήταν «Η ικανότητα της αντίληψης των αλληλεξαρτήσεων, των παρόντων γεγονότων, με τρόπο τέτοιο ώστε να οδηγεί προς ένα επιθυμητό στόχο».

Τα συστήματα υποστήριξης αποφάσεων (Σ.Υ.Α), τα οποία ξεκίνησαν από το 1960 και εξελίχθηκαν μέχρι και τα μέσα του 1980, αποτελούν την αρχική ιδέα για την σημερινή υλοποίηση της επιχειρηματικής ευφυΐας. Με βάση το βιβλίο του Peter G. W. Keen και Michael S. Scott Morton με τίτλο “Decision Support Systems: A Research Perspective” που εκδόθηκε το 1978, η έννοια των συστημάτων υποστήριξης αποφάσεων προέκυψε, κυρίως από τις θεωρητικές έρευνες στο πεδίο της λήψης επιχειρησιακών αποφάσεων από το Ινστιτούτο Τεχνολογίας του Carnegie, στα τέλη του 1950 και τις αρχές του 1960, αλλά και τις υλοποιήσεις που πήραν μέρος στα μέσα του 1960. Μέσω των συστημάτων υποστήριξης αποφάσεων, κύριοι τομείς της σημερινής τεχνολογίας πληροφοριών, όπως αποθήκες δεδομένων (Data Warehouses), πληροφοριακά συστήματα διοίκησης (Executive Information Systems), συστήματα On-Line Analytical Processing (OLAP), αλλά και η επιχειρηματική ευφυΐα έγιναν το επίκεντρο ενδιαφέροντος στα τέλη του 1980.

Το 1989, ο Howard Dresner, ο οποίος ήταν αργότερα αναλυτής της Gartner, πρότεινε την «Επιχειρηματική Ευφυΐα» ως ένα όρο “ομπρέλα”, που θα συμπεριλάμβανε ιδέες και μεθόδους για την βελτίωση της λήψης επιχειρησιακών αποφάσεων με την χρήση συστημάτων υποστήριξης βασισμένα σε δεδομένα. Ο όρος αυτός διαδόθηκε και χρησιμοποιείται μέχρι και σήμερα, μετά τα τέλη του 1990.

3. Σημασία και αντίκτυπος της επιχειρηματικής ευφυΐας

Από τις αρχές του 21^{ου} αιώνα, η ανάγκη των επιχειρήσεων για ψηφιοποίηση, τόσο των προϊόντων τους, όσο και του εσωτερικού διοικητικού περιβάλλοντος, έφερε πολλές δυνατότητες και ευκαιρίες, όπως και προβλήματα. Για την αντιμετώπιση αυτών των νέων προκλήσεων είναι αναγκαία η εξέλιξη και αναβάθμιση των πρακτικών του διοικητικού τμήματος των επιχειρήσεων και συγκεκριμένα την βελτιστοποίηση της διαδικασίας λήψης αποφάσεων. Για να είναι δυνατό αυτό, απαιτείται βαθιά κατανόηση της πληροφορίας που εισάγεται στην επιχείρηση, είτε από την ίδια, είτε από το περιβάλλον της, καθώς και η πληροφορία αυτή να είναι έγκαιρη και ουσιαστική.

Η αξία της επιχειρηματικής ευφυΐας, όπως και η σημασία της, αυξάνεται συνεχώς καθώς οι ανερχόμενες προκλήσεις των επιχειρήσεων, όπως η εκθετική αύξηση των δεδομένων τους σε ταχύ ρυθμό ή νόμοι και κανονισμοί που επηρεάζουν τον τρόπο και το χρονικό διάστημα αποθήκευσης και επεξεργασίας των δεδομένων αυτών, γίνονται ολοένα και πιο δύσκολες στην διαχείριση και την αντιμετώπισή τους. Γι' αυτό το λόγο, η ανάγκη των επιχειρησιακών εργαλείων να έχουν την δυνατότητα να ανταπεξέλθουν σε αυτές τις προκλήσεις, με την μέγιστη δυνατή απόδοση, φέρνει την επιχειρηματική ευφυΐα στο προσκήνιο.

Μερικές από τις λύσεις που προσφέρει η επιχειρηματική ευφυΐα είναι :

Απόκτηση γνώσης για την συμπεριφορά των καταναλωτών

Ένα από τα βασικά πλεονεκτήματα υλοποίησης επιχειρηματικής ευφυΐας, σε μία επιχείρηση, είναι η ενίσχυση της ικανότητας της, στην ανάλυση των σύγχρονων τάσεων αγορών των καταναλωτών της. Μέσα από την κατανόηση των τάσεων και χρησιμοποιώντας τις εξαγόμενες πληροφορίες για την ανάπτυξη προϊόντων που πληρούν τις ανάγκες των καταναλωτών, μπορεί η επιχείρηση να αυξήσει το κέρδος της.

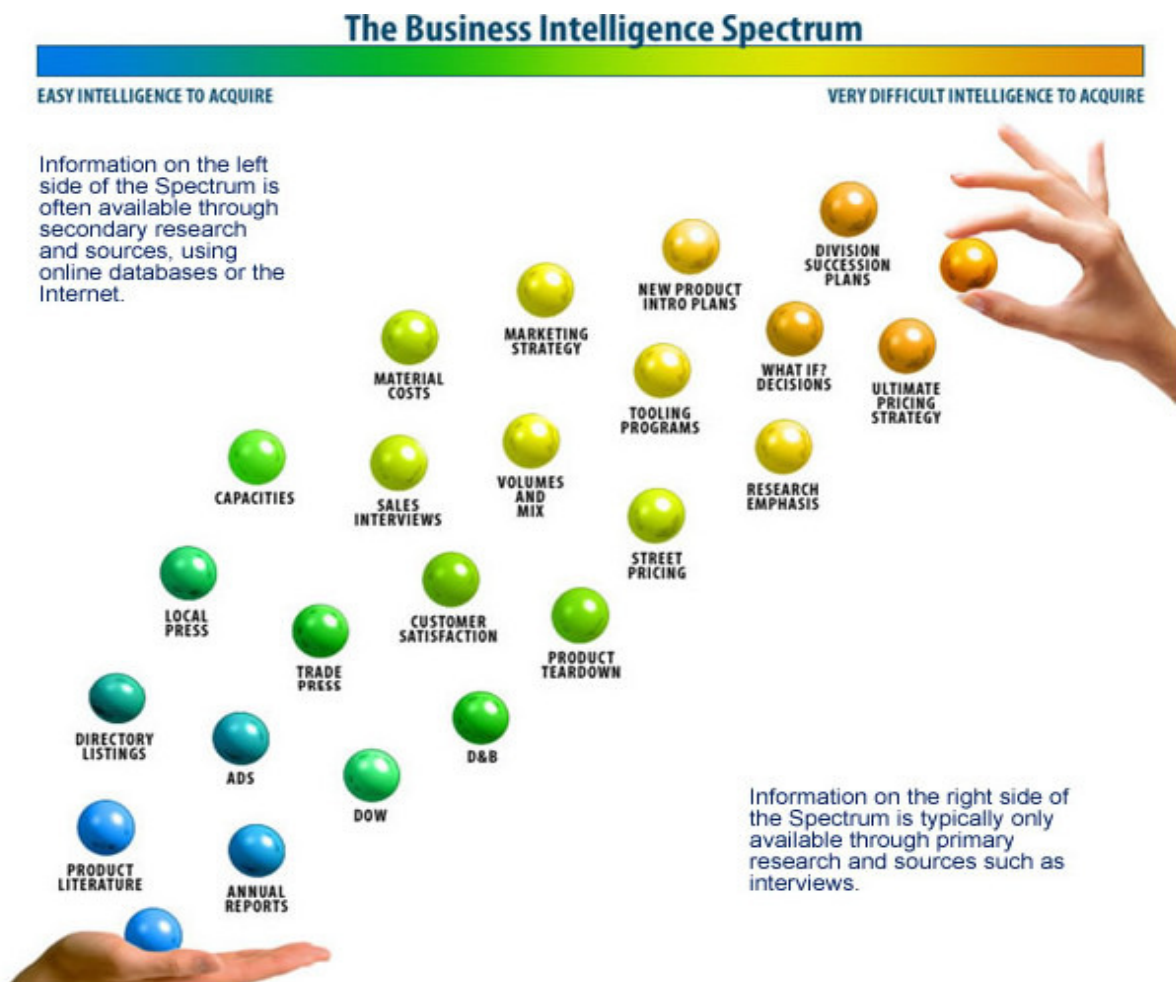
Βελτίωση της αντίληψης

Ο πλήρης έλεγχος των διάφορων σημαντικών διεργασιών μίας επιχείρησης, ήταν και θα είναι στο επίκεντρο της προσοχής της. Για την επίτευξη αυτού του στόχου όμως, χρειάζεται καλή αντίληψη για τον εντοπισμό των διεργασιών ή ολόκληρων τομέων που χρειάζονται βελτίωση. Η προσφορά της επιχειρηματικής ευφυΐας, σε αυτό το κομμάτι, είναι αναλυτικές αναφορές και ευκατανόητες παρουσιάσεις, από τις οποίες μπορούν πολύ πιο εύκολα να εντοπιστούν

προβλήματα ή και βελτιώσεις σε κάποια διεργασία. Με τη βελτίωση της αντίληψης, μειώνονται οι κίνδυνοι και οι απρόοπτες ζημιές, εξοικονομείται σημαντικός χρόνος και βελτιώνεται η παραγωγικότητα.

Μετατροπή δεδομένων σε χρήσιμη πληροφορία

Μία ιδιαίτερα σημαντική διαδικασία είναι η λήψη στρατηγικών αποφάσεων, που στη βέλτιστη περίπτωση αποφέρουν κέρδος. Η επίτευξη ίδιων ή ανάλογων διαδικασιών απαιτεί την μετατροπή του τεράστιου όγκου των δεδομένων σε χρήσιμη πληροφορία. Κατά συνέπεια, η αναγνώριση τάσεων και μοτίβων στα δεδομένα της επιχείρησης, γίνεται πιο εύκολη και πραγματοποιούνται συνδέσεις των διάφορων διεργασιών της επιχείρησης, οι οποίες αρχικά φαίνονται άσχετες μεταξύ τους. Η διαδικασία της μετατροπής βοηθάει στην καλύτερη κατανόηση των διάφορων διεργασιών και την πιθανότητα σύνδεσης τους, και στην ενίσχυση της ικανότητας της επιχείρησης στον εντοπισμό καλύτερων ευκαιριών για την ίδια.



Εικόνα 1 Το φάσμα των επιχειρηματικών πληροφοριών

Βελτιστοποίηση απόδοσης

Ο κύριος λόγος για επένδυση σε συστήματα επιχειρηματικής ευφυίας, ο οποίος είναι αλληλένδετος με τα παραπάνω, είναι η βελτιστοποίηση της απόδοσης. Η ευκολία στην εξαγωγή και ανταλλαγή πληροφοριών εντός μία επιχείρησης, είναι πιθανό να μειώσει την επικάλυψη ρόλων / καθηκόντων και να βελτιώσει την ακρίβεια και τη χρησιμότητα των δεδομένων που παράγονται από διάφορα τμήματα.

Ανταγωνιστική ευφυία

Ο όρος «ανταγωνιστική ευφυία» συγγέεται πολλές φορές με τον όρο επιχειρηματική ευφυία, όμως η επιχειρηματική ευφυία χρησιμοποιεί διαδικασίες και εφαρμογές, κυρίως για την ανάλυση και επεξεργασία εσωτερικών δεδομένων, ενώ η ανταγωνιστική ευφυία ασχολείται κυρίως με τη συλλογή και ανάλυση της πληροφορίας με στόχο την απόκτηση στρατηγικού πλεονεκτήματος από τους αντιπάλους της επιχείρησης. Με μία πιο ευρεία κατανόηση, η επιχειρηματική ευφυία μπορεί να περιλαμβάνει την ανταγωνιστική ευφυία. Επομένως, η επιχειρηματική ευφυία μπορεί να χρησιμοποιηθεί για την λήψη στρατηγικών αποφάσεων, σε σχέση με τις κινήσεις των ανταγωνιστών, καθώς και το σχεδιασμό μελλοντικών αποφάσεων.

Λόγω των παραπάνω, η επιχειρηματική ευφυία έχει ένα μεγάλο αντίκτυπο στη σημερινή και την μελλοντική αγορά. Σύμφωνα με την εταιρεία Gartner, τα παγκόσμια έσοδα της επιχειρηματικής ευφυίας και της επιχειρηματικής ανάλυσης πρόκειται να φθάσουν τα \$22.8 δισεκατομμύρια, μέχρι το τέλος του 2020. Στη σχετική έκθεση των ετών 2015(«GartnerCIOLeadershipForum 2015»), 2016 (“GartnerCIOLeadershipForum 2016”) και 2017(«InsightsFromthe 2017 CIOAgendaReport: SeizetheDigitalEcosystemOpportunity»), η επιχειρηματική ευφυία βρίσκεται στην πρώτη θέση του καταλόγου των τεχνολογικών προτεραιοτήτων.

4. Επιχειρηματική ευφυία σε πραγματικό χρόνο

Τα τελευταία χρόνια, καθώς η πληροφορική ενσωματώνεται όλο και πιο πολύ στον πυρήνα της επιχείρησης και ειδικά με την άνοδο του διαδικτύου και των τεχνολογιών παγκοσμίου ιστού, έχει αυξηθεί και η ανάγκη για άμεση και έγκαιρη λήψη στρατηγικών αποφάσεων. Η επιχειρηματική ευφυία πραγματικού χρόνου προήλθε από αυτή την ανάγκη, με σκοπό όχι μόνο την ανάλυση, αλλά και την παρουσίαση των δεδομένων σε πραγματικό χρόνο, ούτως ώστε να επωφεληθούν επιχειρήσεις και εταιρείες που βασίζονται από τα άμεσα δεδομένα τους.

Τις προηγούμενες δεκαετίες, η εξόρυξη γνώσης από τα ιστορικά δεδομένα μίας εταιρείας, τα οποία αποθηκεύονταν σε μεγάλες αποθήκες δεδομένων, ήταν ο προκαθορισμένος τρόπος ανάλυσης των δεδομένων για τους περισσότερους οργανισμούς. Η τεχνολογία των κύβων δεδομένων (DataCubes) διεύρυνε περαιτέρω την αποδοχή αυτού του τύπου ανάλυσης δεδομένων. Δυστυχώς, λόγω της συσσωρευμένης προσέγγισης της ενσωμάτωσης δεδομένων που είχαν οι κύβοι δεδομένων, συνήθως, υπήρχε μία σημαντική καθυστέρηση για τη λήψη της πληροφορίας από τις αποθήκες δεδομένων και την μετέπειτα επεξεργασία και ανάλυση της, με αποτέλεσμα να μειώνεται δραστικά η αξία της.

Καθώς το διαδίκτυο και οι τεχνολογίες παγκόσμιου ιστού αποκτούσαν μεγαλύτερη βαρύτητα από το ευρύ κοινό των επιχειρήσεων, αλλά και των πελατών, κατά συνέπεια και ο ρόλος των αποθηκών δεδομένων, και κατ' επέκταση της επιχειρηματικής ευφυΐας, άλλαξε σημαντικά από ήσσονος σημασίας τεχνολογικής συνιστώσα σε ένα κρίσιμο στοιχείο στην υποδομή πληροφοριών της επιχείρησης. Ως αποτέλεσμα, η ανάγκη για συνεχείς βελτιστοποιήσεις βάσει ανάλυσης επιδόσεων σε πραγματικό χρόνο έχει γίνει απαραίτητη. Αυτές οι απαιτήσεις δικαίωσαν την επανεξέταση των παραδοσιακών αρχιτεκτονικών των αποθηκών δεδομένων και της επιχειρηματικής ευφυΐας, που ήταν προσανατολισμένες σε συσσωρευμένη ενσωμάτωση και ενημέρωση των δεδομένων, και πλέον δεν αποτελούν για το μεγαλύτερο ποσοστό των επιχειρήσεων μία καλή και αξιόπιστη λύση.

Παρόλο που το τεχνολογικό τοπίο για την επιχειρησιακή ευφυΐα εξακολουθεί να εξελίσσεται, υπάρχει μια αναδυόμενη συζήτηση μεταξύ των ερευνητών βάσεων δεδομένων και των επαγγελματιών σχετικά με την ανάγκη για επιχειρηματική ευφυΐα σε πραγματικό χρόνο. Το επιχείρημα που διατύπωσαν οι υποστηρικτές είναι ότι δεν αρκεί να παρασχεθούν οι επιχειρηματικές πληροφορίες στους υπεύθυνους λήψης αποφάσεων, αλλά πρέπει επίσης να γίνει με "έγκαιρο" τρόπο. Αυτή η έννοια της επικαιρότητας έχει οδηγήσει στην ευρεία χρήση του όρου επιχειρησιακή ευφυΐα σε πραγματικό χρόνο.

Βάσει πολλών ερευνών, συμπεριλαμβανομένων αυτών τις Gartner και Forester, αναφέρουν ότι οι περισσότερες από τις επιχειρήσεις ενδιαφέρονται να επενδύσουν σε συστήματα επιχειρηματικής ευφυΐας. Πρέπει να σημειωθεί ότι παρά τις σημαντικές επενδύσεις στον προγραμματισμό επιχειρηματικών πόρων (ERP), τη διαχείριση της εφοδιαστικής αλυσίδας (SCM) και τη διαχείριση πελατειακών σχέσεων (CRM) κατά την τελευταία δεκαετία οι

επιχειρήσεις αγωνίζονται για να επιτύχουν κάποιο ανταγωνιστικό πλεονέκτημα. Αυτό μπορεί να οφείλεται στις πληροφορίες που συλλαμβάνονται ή δεν έχουν ληφθεί από αυτά τα συστήματα. Κάθε εταιρεία θα πρέπει να έχει ως στόχο την άμεση και έγκαιρη πρόσβαση στην πληροφορία. Ως εκ τούτου, οι επιχειρήσεις πρέπει να στηρίζουν την ανάλυση και την εφαρμογή των πληροφοριών που συλλέγονται προκειμένου να λάβουν κατάλληλες επιχειρησιακές αποφάσεις. Η υλοποίηση ευφύστερων επιχειρησιακών διαδικασιών είναι εκεί όπου η επιχειρηματική ευφύα έχει αντίκτυπο και επηρεάζει το συνολικό κέρδος, αλλά και ωφελεί οποιαδήποτε επιχείρηση.

5. Επιχειρηματική ευφύα και Μεγάλα δεδομένα

5.1. Τι είναι τα μεγάλα δεδομένα

Σύμφωνα με το Wikipediata *«μεγάλα δεδομένα είναι σύνολα δεδομένων, τα οποία είναι τόσο μεγάλα και πολύπλοκα που τα παραδοσιακά λογισμικά επεξεργασίας και ανάλυσης δεδομένων είναι ανεπαρκή στην αντιμετώπιση τους»*. Αντίστοιχα, ο ορισμός των μεγάλων δεδομένων που χρησιμοποιείται από την Gartner είναι *«πληροφοριακά στοιχεία υψηλού όγκου, υψηλής ταχύτητας και συνήθως μεγάλης ποικιλίας, τα οποία απαιτούν οικονομικά αποδοτικές και καινοτόμες μορφές επεξεργασίας πληροφοριών που επιτρέπουν βελτιωμένη γνώση, λήψη αποφάσεων και αυτοματοποίηση διαδικασιών.»*

Ένας πιο πρόσφατος ορισμός του 2018 δηλώνει ότι, *«τα μεγάλα δεδομένα είναι εκεί όπου απαιτούνται παράλληλα εργαλεία πληροφορικής για τη διαχείριση δεδομένων»* και σημειώνει, *«Αυτό αντιπροσωπεύει μια ξεχωριστή και σαφώς καθορισμένη αλλαγή στην επιστήμη των υπολογιστών που χρησιμοποιείται μέσω θεωριών παράλληλου προγραμματισμού και απώλειες ορισμένων εγγυήσεων και ικανότητες που δημιουργούνται από το σχεσιακό μοντέλο του Edgar Frank "Ted" Codd.»*

Η αυξανόμενη ωριμότητα της έννοιας περιγράφει με μεγαλύτερη ακρίβεια τη διαφορά μεταξύ των "μεγάλων δεδομένων" και της "επιχειρηματικής ευφύας":

- Η επιχειρηματική ευφύα χρησιμοποιεί περιγραφικά στατιστικά στοιχεία με δεδομένα υψηλής πυκνότητας πληροφοριών για να υπολογίσει το κέρδος, να εντοπίσει τις τάσεις κλπ.
- Τα μεγάλα δεδομένα χρησιμοποιούν επαγωγικές στατιστικές και έννοιες από τον μη-γραμμικό προσδιορισμό του συστήματος έως τη συναγωγή νόμων (παλινδρομήσεις, μη γραμμικές σχέσεις και αιτιακές συνέπειες) από μεγάλα σύνολα δεδομένων με χαμηλή

πυκνότητα πληροφοριών για να αποκαλύψουν σχέσεις και εξαρτήσεις ή για να εκτελέσουν προβλέψεις αποτελεσμάτων και συμπεριφορών.

Οι μεγάλες προκλήσεις στον τομέα των δεδομένων περιλαμβάνουν τη συλλογή δεδομένων, την αποθήκευση δεδομένων, την ανάλυση δεδομένων, την αναζήτηση, την κοινή χρήση, τη μεταφορά, την οπτικοποίηση, την ερώτηση, την ενημέρωση, την προστασία προσωπικών δεδομένων και την προέλευση δεδομένων.

5.2.Ιστορία των μεγάλων δεδομένων

Ενώ ο όρος "μεγάλα δεδομένα" είναι σχετικά νέος, η πράξη συλλογής και αποθήκευσης μεγάλου όγκου πληροφοριών για ενδεχόμενη ανάλυση είναι παλιά. Στην πραγματικότητα, τα πρώτα αρχεία της χρήσης δεδομένων για την παρακολούθηση και τον έλεγχο των επιχειρήσεων χρονολογούνται από 7.000 χρόνια πριν, όταν η λογιστική εισήχθη στη Μεσοποταμία για να μπορέσουν να καταγράψουν την ανάπτυξη των καλλιεργειών και των αγελάων.

Η ιδέα των μεγάλων δεδομένων απέκτησε μεγάλη σημασία στις αρχές της δεκαετίας του 2000, όταν ο αναλυτής της βιομηχανίας Doug Laney διατύπωσε τον ορισμό των μεγάλων δεδομένων, όπως και για τα κυριότερα χαρακτηριστικά τους, γνωστά και ως 3Vs, Volume, Velocity, Variety(Όγκος, Ταχύτητα, Ποικιλία).

Ύστερα, το 2005, ο Roger Mougallas από την O'Reilly Media εφάρμοσε για πρώτη φορά τον όρο μεγάλα δεδομένα, μόνο ένα χρόνο μετά τη δημιουργία του όρου Web 2.0. Αναφέρεται σε ένα μεγάλο σύνολο δεδομένων που είναι σχεδόν αδύνατο να διαχειριστεί και να επεξεργαστεί χρησιμοποιώντας παραδοσιακά εργαλεία επιχειρηματικής ευφυΐας. Το 2005 είναι επίσης η χρονιά που το Hadoop δημιουργήθηκε από τους Doug Cutting and Mike Cafarella, και είναι εμπνευσμένο πάνω στο MapReduce της Google. Σκοπός του ήταν η ευρετηρίαση ολόκληρου του Παγκόσμιου Ιστού και σήμερα το Hadoop, το οποίο είναι ανοιχτού κώδικα, χρησιμοποιείται από πολλούς οργανισμούς για να ανταπεξέλθουν σε τεράστια ποσά δεδομένων.

Καθώς αρχίζουν να εμφανίζονται όλο και περισσότερα κοινωνικά δίκτυα και το Web 2.0 αναπτύσσεται δραματικά, όλο και περισσότερα δεδομένα δημιουργούνται σε καθημερινή βάση. Οι καινοτόμες νεοσύστατες επιχειρήσεις αρχίζουν σιγά-σιγά να "σκάβουν" σε αυτό το τεράστιο όγκο δεδομένων, όπως επίσης οι κυβερνήσεις επιχειρούν να δουλεύουν όλο και περισσότερο σε έργα μεγάλων δεδομένων.

5.3.Χαρακτηριστικά και κατηγορίες μεγάλων δεδομένων

5.3.1. Χαρακτηριστικά μεγάλων δεδομένων

Όπως προαναφέρθηκε, τα μεγάλα δεδομένα απαρτίζονται από τα παρακάτω χαρακτηριστικά :

- Όγκος(Volume)

Το ίδιο το όνομα, «Μεγάλα Δεδομένα», σχετίζεται με ένα μέγεθος που θεωρείται τεράστιο. Το μέγεθος των δεδομένων διαδραματίζει πολύ σημαντικό ρόλο στον καθορισμό της αξίας και της ενδεχόμενης γνώσης που μπορούν να προσφέρουν τα δεδομένα. Επίσης, εάν συγκεκριμένα δεδομένα μπορούν πραγματικά να θεωρηθούν ως μεγάλα δεδομένα ή όχι, εξαρτάται από τον όγκο των δεδομένων. Ως εκ τούτου, ο όρος «Όγκος» είναι ένα χαρακτηριστικό το οποίο πρέπει να ληφθεί υπόψιν κατά την εξέταση των «μεγάλων δεδομένων».

Η συγκέντρωση αυτού του τεράστιου όγκου γίνεται από οργανισμούς, οι οποίοι συλλέγουν δεδομένα από διάφορες πηγές, συμπεριλαμβανομένων επιχειρηματικών συναλλαγών, κοινωνικών μέσων και πληροφοριών από δεδομένα αισθητήρων ή μηχανών σε μηχανήματα.

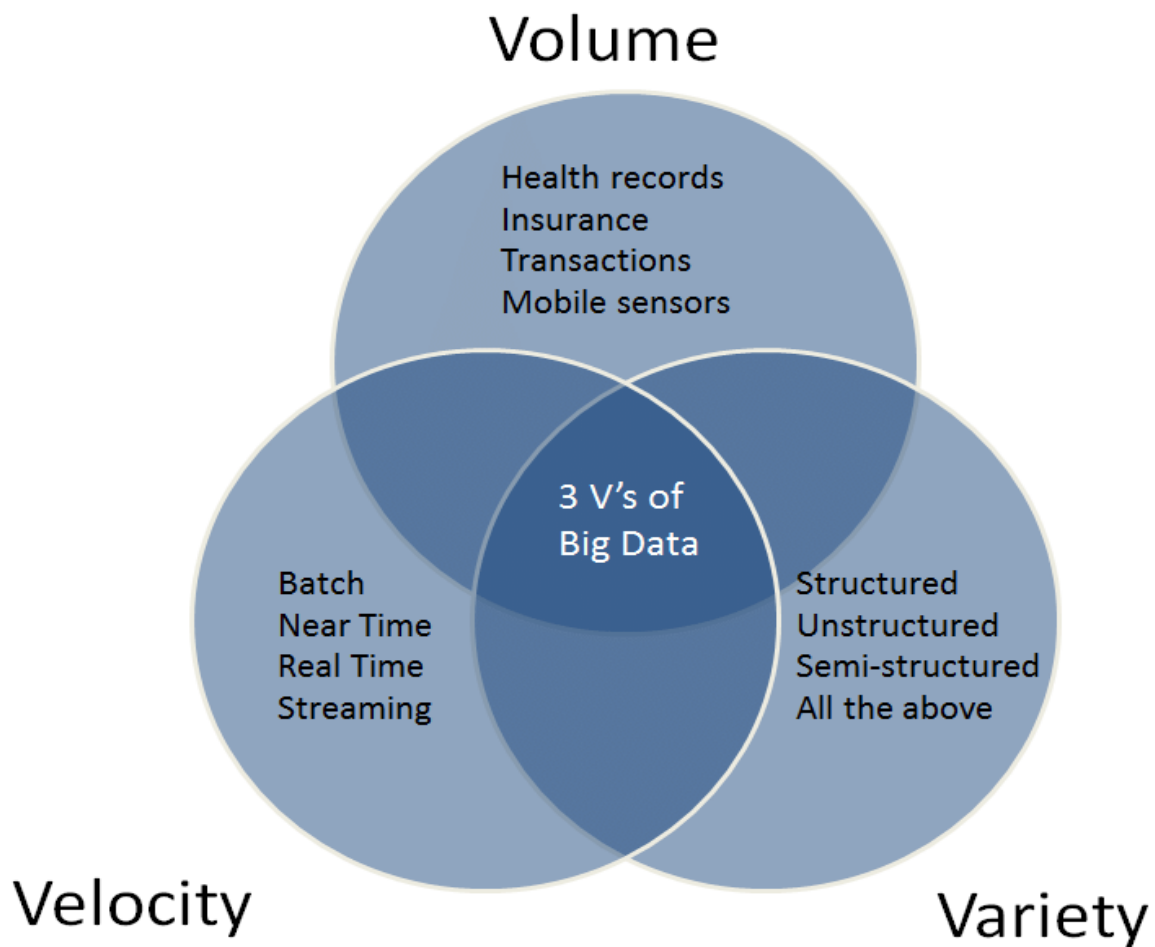
- Ταχύτητα(Velocity)

Ο όρος «ταχύτητα» αναφέρεται στην ταχύτητα παραγωγής δεδομένων, δηλαδή πόσο γρήγορα παράγονται και υποβάλλονται σε επεξεργασία ώστε να ικανοποιούνται οι απαιτήσεις και να καθορίζονται οι πραγματικές δυνατότητες των δεδομένα.

Τα δεδομένα αυτά ρέουν με πρωτοφανή ταχύτητα και πρέπει να αντιμετωπίζονται έγκαιρα. Οι ετικέτες RFID, οι αισθητήρες και οι έξυπνοι μετρητές οδηγούν στην ανάγκη αντιμετώπισης της ροής δεδομένων σε σχεδόν πραγματικό χρόνο.

- Ποικιλία(Variety)

Η ποικιλία αναφέρεται σε ετερογενείς πηγές και στη φύση των δεδομένων, δομημένων και αδόμητων. Κατά τα προηγούμενα χρόνια, τα υπολογιστικά φύλλα και οι βάσεις δεδομένων ήταν οι μόνες πηγές δεδομένων που εξετάζονταν από τις περισσότερες εφαρμογές. Τώρα, τα δεδομένα κατέχουν την μορφή μηνυμάτων ηλεκτρονικού ταχυδρομείου, φωτογραφιών, βίντεο, συσκευών παρακολούθησης, αρχείων PDF, ήχου κ.λπ. και πρέπει να επίσης να εξετάζονται από τις εφαρμογές ανάλυσης δεδομένων. Αυτή η ποικιλία αδόμητων ή/και δομημένων δεδομένων θέτει ορισμένα ζητήματα για την αποθήκευση, την εξόρυξη και την ανάλυση αυτών.



Εικόνα 2 Τα 3V των Μεγάλων Δεδομένων

Πέραν αυτών των χαρακτηριστικών εταιρείες του τομέα, όπως η SAS και η Oracle, έχουν εξετάσει και διατυπώσει επιπλέον χαρακτηριστικά για τα μεγάλα δεδομένα:

- **Μεταβλητότητα(Variability)**

Η μεταβλητότητα αναφέρεται στην ασυνέπεια που μπορεί να αποδειχθεί με τα δεδομένα κατά περιόδους, εμποδίζοντας έτσι τη διαδικασία της αποτελεσματικής διαχείρισης και ανάλυσης των δεδομένων.

Εκτός από τις αυξανόμενες ταχύτητες και ποικιλίες δεδομένων, οι ροές δεδομένων μπορεί να είναι σε μεγάλο βαθμό ασυμβίβαστες με τις περιοδικές κορυφές. Τα ημερήσια, εποχιακά και τα γεγονότα που πυροδότησαν τα κορυφαία φορτία δεδομένων μπορεί να είναι δύσκολο να διαχειριστούν. Ακόμη περισσότερο με μη δομημένα δεδομένα.

- **Αξία(Value)**

Τα δεδομένα έχουν εγγενή αξία, αλλά πρέπει να τα ανακαλύψουμε. Υπάρχει μια σειρά από ποσοτικές και ερευνητικές τεχνικές για την άντληση αξίας από τα δεδομένα, από την ανακάλυψη προτιμήσεων ή τάσεων των καταναλωτών, μέχρι τη σχετική προσφορά ανά τοποθεσία ή την αναγνώριση ενός εξοπλισμού που πρόκειται να αποτύχει. Η τεχνολογική πρόοδος είναι ότι το κόστος της αποθήκευσης δεδομένων και των υπολογισμών έχει μειωθεί εκθετικά, παρέχοντας έτσι μια πληθώρα δεδομένων, από τα οποία οι στατιστικές δειγματοληψίες και άλλες τεχνικές καθίστανται συναφείς και μπορεί να εξαχθεί νόημα από αυτά. Ωστόσο, η εύρεση αξίας απαιτεί επίσης νέες διαδικασίες ανεύρεσης που περιλαμβάνουν έξυπνους και διορατικούς αναλυτές, επιχειρησιακούς χρήστες και στελέχη. Η πραγματική πρόκληση των μεγάλων δεδομένων είναι μια ανθρώπινη, η οποία είναι να μάθει να θέτει τις σωστές ερωτήσεις, αναγνωρίζοντας πρότυπα, ενημερώνοντας τις υποθέσεις και προβλέποντας τη συμπεριφορά.

- **Αξιοπιστία(Veracity)**

Η αξιοπιστία αναφέρεται στις προκαταλήψεις, το θόρυβο και τις ανωμαλίες στα δεδομένα. Είναι τα δεδομένα που αποθηκεύονται και εξάγονται με στόχο την επίλυση του προβλήματος που αναλύεται. Η ποιότητα των δεδομένων, από τα δεδομένα που συλλέγονται, μπορεί να ποικίλει σημαντικά, επηρεάζοντας την ακριβή ανάλυση τους.

Η ποιότητα των δεδομένων αναφέρεται στην κατάσταση ενός συνόλου τιμών ποιοτικών ή ποσοτικών μεταβλητών. Υπάρχουν πολλοί ορισμοί της ποιότητας των δεδομένων, αλλά τα δεδομένα θεωρούνται γενικά υψηλής ποιότητας εάν είναι κατάλληλα για τις χρήσεις που προορίζονται, όπως για τη λειτουργία, τη λήψη αποφάσεων και τον προγραμματισμό. Όσο υψηλότερη η ποιότητα των δεδομένων, τόσο μεγαλύτερη είναι και η αξιοπιστία τους.

5.3.2. Κατηγορίες Μεγάλων Δεδομένων

Τα μεγάλα δεδομένα μπορούν να βρεθούν σε τρεις μορφές:

1. Δομημένα

Οποιαδήποτε δεδομένα που μπορούν να αποθηκευτούν, να προσπελαστούν και να υποστούν επεξεργασία με τη μορφή σταθερού μορφοτύπου ονομάζονται δομημένα δεδομένα. Με το πέρασμα των χρόνων, τα ταλέντα στην επιστήμη των υπολογιστών έχουν επιτύχει μεγαλύτερη επιτυχία στην ανάπτυξη τεχνικών για την εργασία με τέτοια δεδομένα, των οποίων η μορφή

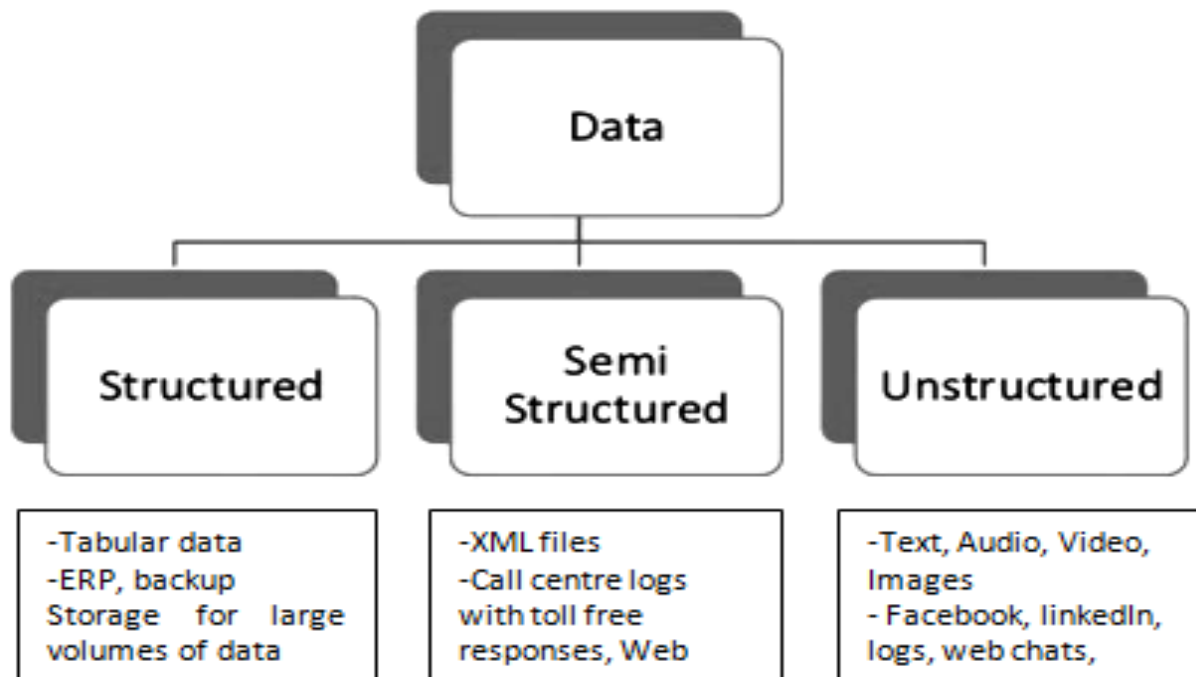
είναι γνωστή εκ των προτέρων, και επίσης στην αποκόμιση αξίας από αυτήν. Ωστόσο, τον τελευταίο καιρό, προβλέπονται ζητήματα όταν το μέγεθος αυτών των δεδομένων αυξάνεται σε τεράστιο βαθμό, τα τυπικά μεγέθη είναι στο εύρος πολλαπλών Exabyte.

2. Μη-Δομημένα

Οποιαδήποτε δεδομένα με άγνωστη μορφή ή δομή ταξινομούνται ως μη δομημένα δεδομένα. Εκτός από το μέγεθος που είναι τεράστιο, τα μη δομημένα δεδομένα δημιουργούν πολλαπλές προκλήσεις όσον αφορά την επεξεργασία τους για να αποκομίσουν αξία από αυτό. Τυπικό παράδειγμα μη δομημένων δεδομένων είναι μια ετερογενής πηγή δεδομένων που περιέχει έναν συνδυασμό απλών αρχείων κειμένου, εικόνων, βίντεο κλπ. Τώρα πια οι οργανισμοί κατέχουν το πλήθος των δεδομένων, αλλά δυστυχώς, συνήθως, δεν ξέρουν πώς να αντλούν αξία από αυτά από τη στιγμή που τα δεδομένα είναι σε ακατέργαστη ή σε αδόμητη μορφή.

3. Ημι-Δομημένα

Τα ημι-δομημένα δεδομένα μπορούν να περιέχουν και τις δύο παραπάνω μορφές. Είναι μια μορφή δομημένων δεδομένων που δεν συμμορφώνονται με την επίσημη δομή των μοντέλων δεδομένων που σχετίζονται, με τις σχεσιακές βάσεις δεδομένων, ή με άλλες μορφές δεδομένων, αλλά περιέχουν ετικέτες ή άλλους δείκτες για να διαχωρίσουν σημασιολογικά στοιχεία και να επιβάλουν ιεραρχίες αρχείων και πεδίων μέσα στα δεδομένα. Ως εκ τούτου, είναι επίσης γνωστή ως αυτό-περιγραφόμενη δομή. Παράδειγμα ημι-δομημένων δεδομένων είναι δεδομένα που περιέχονται σε αρχείο XML.



Εικόνα 3 Τύποι Μεγάλων Δεδομένων

5.4.Εργαλεία και τεχνολογίες μεγάλων δεδομένων

Οι τεχνολογίες ανοιχτού κώδικα σε μεγάλα δεδομένα έχουν κερδίσει έδαφος, λόγω της αποδεδειγμένης δυνατότητας παράλληλης επεξεργασίας και κατανεμημένης αποθήκευσης σε μεγάλο όγκο δεδομένων. Τόσο η παράλληλη επεξεργασία, όσο και οι τεχνικές για την πραγματοποίηση υπολογισμών στα δεδομένα, κατέστησαν δυνατή την επεξεργασία μεγάλων συνόλων δεδομένων σε υψηλή ταχύτητα. Αυτά τα κύρια χαρακτηριστικά αποτέλεσαν ένα σπουδαίο κίνητρο για να εμβαθύνουμε στην ιστορία, την σημασία και το οικοσύστημα του ηγέτη της βιομηχανίας, στον κλάδο των μεγάλων δεδομένων, ApacheHadoop.

Το Hadoop είναι ένα πλαίσιο λογισμικού ανοιχτού κώδικα για την αποθήκευση δεδομένων και την εκτέλεση εργασιών και εφαρμογών σε συστοιχίες υλικού βασικού εξοπλισμού. Με αυτόν τον τρόπο παρέχει τεράστιο αποθηκευτικό χώρο για κάθε είδους δεδομένα, μεγάλη επεξεργαστική ισχύ και την δυνατότητα χειρισμού, θεωρητικά απεριόριστων ταυτόχρονων εργασιών και διεργασιών.

Ιστορία του Hadoop

Με τη ραγδαία ανάπτυξη του Παγκόσμιου Ιστού στα τέλη της δεκαετίας του 1990 και τις αρχές της δεκαετίας του 2000, προέκυψε η δημιουργία μηχανών αναζήτησης και ευρετηρίων που βοηθούσαν στον εντοπισμό σχετικών πληροφοριών, εν μέσω του περιεχομένου που βασιζόταν το κείμενο. Στα πρώτα χρόνια. Τα αποτελέσματα της αναζήτησης επιστρέφονταν από ανθρώπους. Αλλά καθώς ο παγκόσμιος ιστός αυξήθηκε σε εκατομμύρια σελίδες, χρειάστηκε αυτοματοποίηση. Δημιουργήθηκαν προγράμματα ανίχνευσης ιστού, πολλά από τα οποία διεξήχθησαν με πανεπιστημιακά ερευνητικά προγράμματα και ξεκίνησαν νεοσύστατες επιχειρήσεις πάνω στις μηχανές αναζήτησης (π.χ. Yahoo).

Ένα τέτοιο έργο ήταν μια μηχανή αναζήτησης ανοιχτού κώδικα που ονομαζόταν Nutch, δημιούργημα των Doug Cutting και του Mike Cafarella. Ήθελαν να επιστρέφουν τα αποτελέσματα αναζήτησης στον ιστό γρηγορότερα με την κατανομή δεδομένων και υπολογισμών σε διάφορους υπολογιστές, ώστε να μπορούν να πραγματοποιηθούν ταυτόχρονα πολλαπλές διεργασίες. Κατά τη διάρκεια αυτής της περιόδου, η γνωστή μηχανή αναζήτησης Google ήταν σε εξέλιξη. Βασίστηκε στην ίδια ιδέα, δηλαδή την αποθήκευση και την επεξεργασία δεδομένων με κατανεμημένο και αυτοματοποιημένο τρόπο, ούτως ώστε τα σχετικά αποτελέσματα αναζήτησης ιστού να μπορούν να επιστρέφονται ταχύτερα.

Το 2006, ο Cutting προσλήφθηκε από την Yahoo και πήρε μαζί του το έργο Nutch, καθώς και ιδέες που βασίζονταν στην πρόωρη μορφή της Google, σε σχέση με την αυτοματοποίηση της αποθήκευσης και της επεξεργασίας κατανεμημένων δεδομένων. Το έργο Nutch διαιρέθηκε, στο κομμάτι ανίχνευσης ιστού, το οποίο παρέμεινε ως Nutch και στο κομμάτι κατανεμημένης αποθήκευσης και επεξεργασίας, το οποίο ονομάστηκε Hadoop (πήρε την ονομασία από το παιχνίδι του γιου του Cutting). Το 2008, η Yahoo κυκλοφόρησε το Hadoop ως έργο ανοιχτού κώδικα. Σήμερα, η Apache Software Foundation, μια μη-κερδοσκοπική παγκόσμια κοινότητα προγραμματιστών λογισμικού, διαχειρίζονται και συντηρούν το γενικό πλαίσιο και το οικοσύστημα τεχνολογιών του Hadoop.

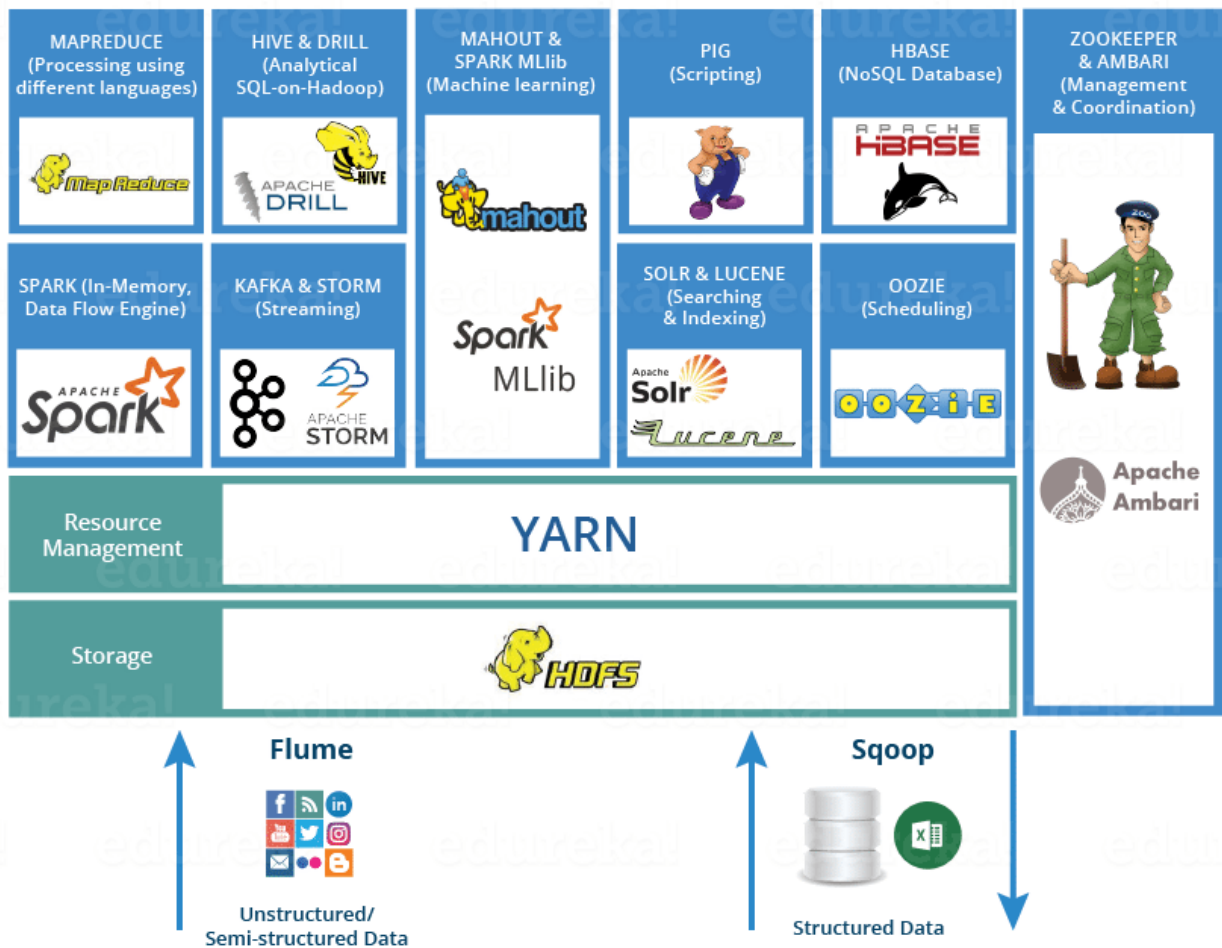
Σημασία του Hadoop

Οι παρακάτω ιδιότητες του Hadoop ενισχύουν την σημασία και την χρησιμότητα του στον τομέα των μεγάλων δεδομένων:

- **Ικανότητα γρήγορης αποθήκευσης και επεξεργασίας τεράστιων ποσοτήτων, οποιουδήποτε είδους δεδομένων.** Με όγκους και ποικιλίες δεδομένων που αυξάνονται διαρκώς, ιδίως από τα κοινωνικά μέσα και το διαδίκτυο των πραγμάτων, αποτελεί βασικό παράγοντα.
- **Υπολογιστική ισχύς.** Το καταναμημένο υπολογιστικό μοντέλο της Hadoop επεξεργάζεται γρήγορα τα μεγάλα δεδομένα. Όσο περισσότεροι υπολογιστικοί κόμβοι, τόσο μεγαλύτερη δύναμη επεξεργασίας.
- **Ανοχή σε σφάλματα.** Η επεξεργασία δεδομένων και εφαρμογών προστατεύεται από τις δυσλειτουργίες του εξοπλισμού και του υλικού. Εάν ένας κόμβος πάψει να λειτουργεί, οι διεργασίες επαναφέρονται αυτόματα σε άλλους κόμβους, βεβαιώνοντας έτσι ότι ο καταναμημένος υπολογισμός δεν θα αποτύχει. Επίσης, πολλαπλά αντίγραφα όλων των δεδομένων αποθηκεύονται αυτόματα σε όλους τους κόμβους.
- **Ευελιξία.** Σε αντίθεση με τις παραδοσιακές σχεσιακές βάσεις δεδομένων, δεν χρειάζεται η προ-επεξεργασία των δεδομένων πριν την αποθήκευσή τους. Η αποθήκευση, καθώς και η εύρεση χρήσης των δεδομένων μπορεί να γίνει σε οποιαδήποτε χρονική στιγμή. Αυτό περιλαμβάνει μη-δομημένα δεδομένα, όπως κείμενο, εικόνες και βίντεο.
- **Χαμηλό κόστος.** Το γενικό πλαίσιο ανοικτού κώδικα του Hadoop, είναι δωρεάν και χρησιμοποιεί υλικό βασικού εξοπλισμού για την αποθήκευση μεγάλων ποσοτήτων δεδομένων.
- **Επεκτασιμότητα.** Η ανάπτυξη και η επέκταση ενός συστήματος είναι αρκετά εύκολη, με στόχο τον χειρισμό περισσότερων δεδομένων, απλά προσθέτοντας κόμβους. Απαιτείται, βέβαια, κάποιο ποσοστό διοίκησης του συστήματος.

Οικοσύστημα του Hadoop

Το οικοσύστημα του Hadoop μπορεί να θεωρηθεί και ως μία εργαλειοθήκη που περιλαμβάνει πολλές υπηρεσίες, όπως η εισαγωγή, η αποθήκευση, η ανάλυση και η συντήρηση δεδομένων.



Εικόνα 4 Το οικοσύστημα του Hadoop

Παρακάτω θα εξεταστούν μερικές από τις υπηρεσίες που υπάρχουν για την κατανόηση της λειτουργίας τους, ξεχωριστά αλλά και σε συνεργασία μεταξύ τους :

- **HDFS**

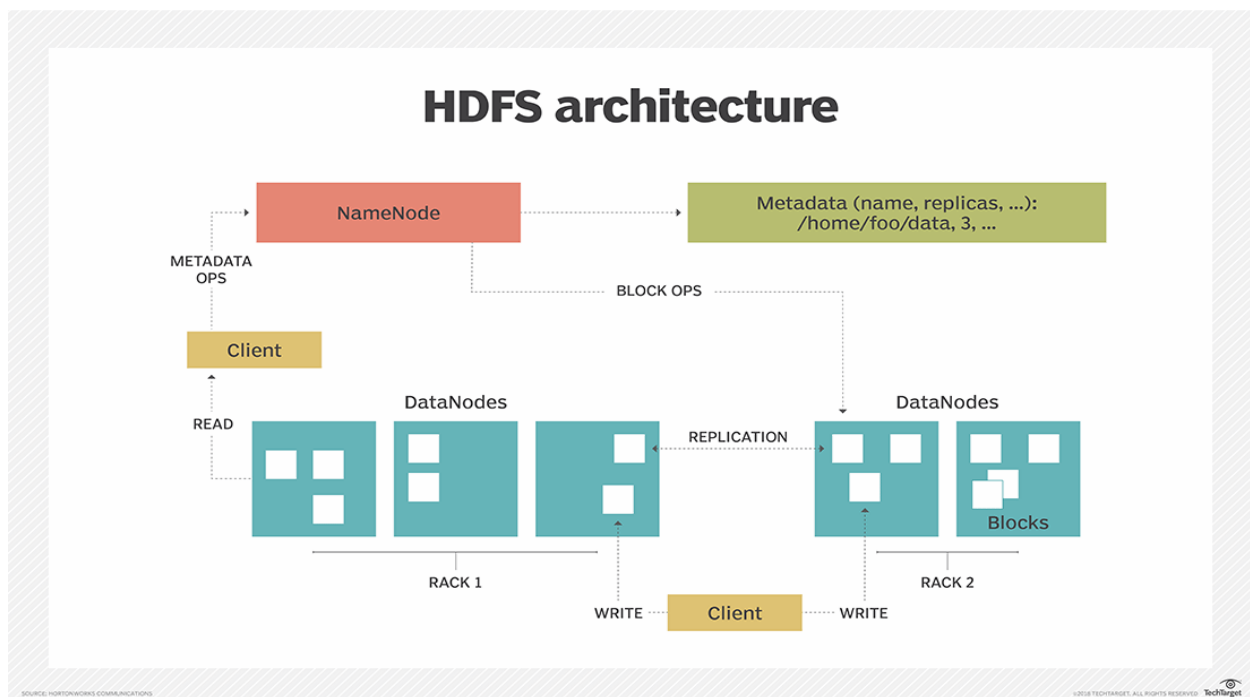
Το σύστημα κατανομής αρχείων του Hadoop(HadoopDistributedFileSystem) προσφέρει έναν τρόπο αποθήκευσης μεγάλων αρχείων σε πολλαπλές μηχανές. Το Hadoop και το HDFS προέρχονται από το GoogleFileSystem(GFS).

Αποτελεί το βασικό στοιχείο ή αλλιώς την «ραχοκοκαλιά» του οικοσυστήματος του Hadoop. Μέσω αυτού επιτρέπεται η αποθήκευση όχι μόνο αρχείων μεγάλου όγκου αλλά και διαφορετικού τύπου, δηλαδή δομημένα, μη-δομημένα και ημι-δομημένα. Δημιουργεί ένα επίπεδο αφαίρεσης από τους πόρους, δίνοντας έτσι την δυνατότητα να δούμε το HDFS ως μία

ενιαία μονάδα. Επίσης, μας βοηθά στην αποθήκευση των δεδομένων μας σε διάφορους κόμβους και στην διατήρηση ενός αρχείου καταγραφής για τα αποθηκευμένα δεδομένα(μεταδεδομένα).

Το HDFS αποτελείται από δύο βασικά στοιχεία, το NameNode και τα DataNodes.

- Το NameNode είναι ο κύριος κόμβος, στον οποίο δεν πραγματοποιείται η αποθήκευση των δεδομένων. Περιέχει μεταδεδομένα, ακριβώς όπως ένα αρχείο καταγραφής(logfile) ή ένας πίνακας περιεχομένων. Επομένως απαιτεί λιγότερο αποθηκευτικό χώρο και περισσότερους υπολογιστικούς πόρους.
- Από την άλλη, όλα τα δεδομένα αποθηκεύονται στα DataNodes και συνεπώς απαιτούν περισσότερους πόρους αποθήκευσης. Τα DataNodes αποτελούνται από υλικό βασικού εξοπλισμού σε ένα κατανεμημένο περιβάλλον, όπως φορητοί, επιτραπέζιοι υπολογιστές, ή εικονικές μηχανές. Εξαιτίας αυτού, οι λύσεις που βασίζονται στο οικοσύστημα του Hadoop, συνήθως, είναι οικονομικά αποδοτικές.
- Η κύρια επικοινωνία στο σύστημα του HDFS γίνεται μέσω του NameNode, κατά την εισαγωγή των δεδομένων. Στη συνέχεια, το ίδιο αναλαμβάνει την εσωτερική αποστολή ενός αιτήματος στα DataNodes να αποθηκεύσουν και να αναπαράγουν τα δεδομένα στους διάφορους υπαρκτούς κόμβους.



Εικόνα 5 Η αρχιτεκτονική του HDFS

Με βάση την λειτουργία του, το HDFS καταφέρνει να είναι αποδοτικό, αξιόπιστο αλλά και εύκολα επεκτάσιμο, δίχως την ανάγκη για αντίγραφα ασφαλείας, καθώς γίνεται η αναπαραγωγή των δεδομένων σε όλους τους κόμβους. Παρόλα αυτά, έχει αδυναμίες, σε θέματα ασφάλειας, περιορισμό της γλώσσας προγραμματισμού μόνο σε Java και την απουσία ευρετηρίων, κάνοντας την πρόσβαση στα δεδομένα πιο αργή.

- **YARN**

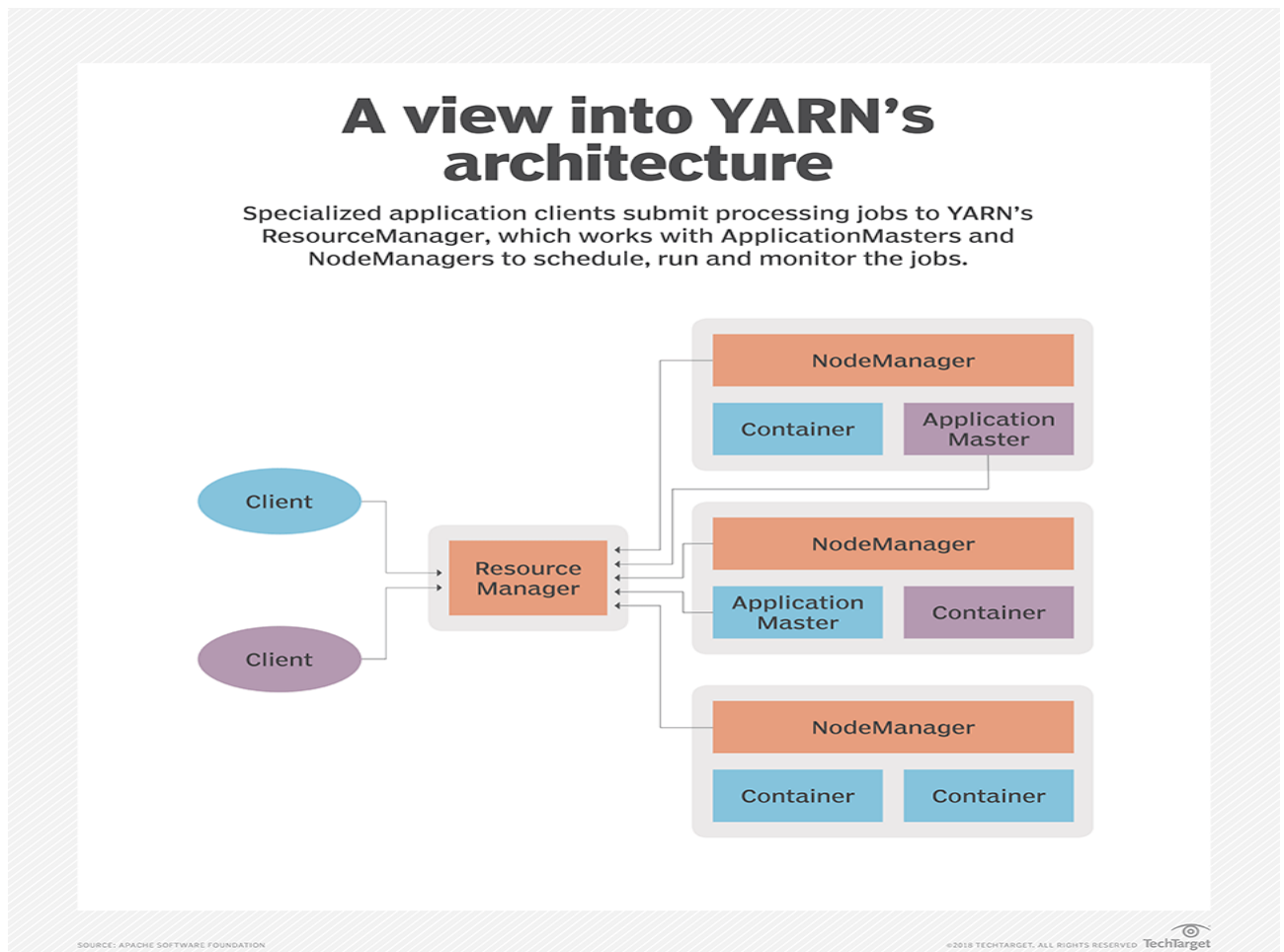
Το Apache Hadoop YARN είναι μία τεχνολογία διαχείρισης πόρων και προγραμματισμού εργασιών μέσα στο κατανεμημένο πλαίσιο επεξεργασίας Hadoop. Αποτελεί ένα από τα βασικά συστατικά του Apache Hadoop, και είναι υπεύθυνο για την κατανομή των πόρων του συστήματος στις διάφορες εφαρμογές που εκτελούνται σε μία συστοιχία υπολογιστών και τον προγραμματισμό εργασιών που πρέπει να εκτελούνται σε διαφορετικούς κόμβους της συστοιχίας.

Το YARN τυπικά σημαίνει YetAnotherResourceNegotiator, αλλά συνήθως αναφέρεται μόνο από το ακρώνυμο. Η τεχνολογία αποτέλεσε ένα από τα πιο σημαντικά έργα της ApacheSoftwareFoundation το 2012, καθώς ήταν ένα από τα βασικά χαρακτηριστικά που προστέθηκαν στην έκδοση 2.0 του Hadoop, το οποίο κυκλοφόρησε για δοκιμή εκείνο το έτος, και έγινε διαθέσιμο προς το κοινό τον Οκτώβριο του 2013.

Η προσθήκη του YARN επέκτεινε σημαντικά τις δυνατότητες και τις ενδεχόμενες περιπτώσεις χρήσης του Hadoop. Η αρχική μορφή του Hadoop συνέδεε στενά το Hadoop Distributed File System (HDFS) με το πλαίσιο προγραμματισμού MapReduce, του οποίου οι λειτουργίες ήταν: η διαχείριση πόρων της πλατφόρμας μεγάλων δεδομένων και ο χρονοπρογραμματισμός των εργασιών. Ως αποτέλεσμα, τα συστήματα Hadoop της έκδοσης 1.0 δεν μπορούσαν να εκτελέσουν εφαρμογές δίχως το MapReduce, έναν περιορισμό που εξάλειψε το Hadoop YARN.

Το YARN εισήγαγε μια νέα προσέγγιση, η οποία αποσυνέδεε τη διαχείριση πόρων και τον χρονοπρογραμματισμό από το MapReduce, επιτρέποντας έτσι στο Hadoop την δυνατότητα να υποστηρίξει ποικίλους τύπους επεξεργασίας, και ένα ευρύτερο φάσμα εφαρμογών. Κατόρθωσε να αποκεντρώσει την εκτέλεση και την παρακολούθηση των εργασιών επεξεργασίας, διαχωρίζοντας τις διάφορες ευθύνες στα παρακάτω στοιχεία:

- ResourceManager, ο οποίος δέχεται τις υποβολές εργασιών του χρήστη, προγραμματίζει τις εργασίες και κατανέμει τους κατάλληλους πόρους.
- NodeManager, που είναι εγκατεστημένος σε κάθε κόμβο και λειτουργεί ως παράγοντας παρακολούθησης και παρέχει αναφορές στον ResourceManager.
- ApplicationMaster, που δημιουργείται για κάθε εφαρμογή με σκοπό την διαπραγμάτευση των πόρων και την συνεργασία με τον NodeManager για την εκτέλεση και την παρακολούθηση των εργασιών.
- Resourcecontainers, που ελέγχονται από τους NodeManagers και τους ανατίθενται πόροι του συστήματος, τους οποίους διαθέτουν σε μεμονωμένες εφαρμογές.



Εικόνα 6 Η αρχιτεκτονική του YARN

Μέσω της χρήσης του YARN για τον διαχωρισμό των αρχείων του HDFS από το MapReduce, το περιβάλλον του Hadoop έγινε πιο κατάλληλο για εφαρμογές επεξεργασίας σε πραγματικό χρόνο, αλλά και εφαρμογές που δεν μπορούν να περιμένουν την ολοκλήρωση εργασιών παρτίδας(batch). Εκτός από περισσότερες επιλογές τεχνολογιών και εφαρμογών, το YARN προσφέρει επεκτασιμότητα, καλύτερη χρήση πόρων, υψηλή διαθεσιμότητα και βελτιώσεις απόδοσης σε σχέση με το MapReduce.

- **MapReduce**

Το MapReduce αποτελεί βασικό στοιχείο του πλαισίου λογισμικού Apache Hadoop. Βοηθά στη σύνταξη εφαρμογών που επεξεργάζονται μεγάλα σύνολα δεδομένων χρησιμοποιώντας κατανεμημένους και παράλληλους αλγορίθμους μέσα στο περιβάλλον Hadoop. Το MapReduce εξυπηρετεί δύο βασικές λειτουργίες: φιλτράρει και κατανέμει το έργο σε διάφορους κόμβους μέσα σε μία συστάδα ή χάρτη(Map), μια λειτουργία που μερικές φορές αναφέρεται ως χαρτογράφος, και οργανώνει και συγκεντρώνει τα αποτελέσματα από κάθε κόμβο σε μια συνεκτική απάντηση για ένα ερώτημα.

Η αρχική έκδοση του MapReduce περιλάμβανε αρκετές συνιστώσες “δαίμονες”, δηλαδή προγράμματα που τρέχουν στο παρασκήνιο, όπως:

- JobTracker, ο κύριος κόμβος που διαχειρίζεται όλες τις εργασίες και τους πόρους σε,ία συστάδα.
- TaskTrackers, πράκτορες που αναπτύσσονται σε κάθε μηχανή της συστάδας για την εκτέλεση των εργασιών χαρτογράφησης(Map) και συγκέντρωσης(Reduce). και
- JobHistory Server, ένα στοιχείο που παρακολουθεί τις ολοκληρωμένες εργασίες και αναπτύσσεται συνήθως ως ξεχωριστή λειτουργία ή μαζί με το JobTracker.

Με την εισαγωγή της δεύτερης έκδοσης του MapReduce και του Hadoop, οι προηγούμενες συνιστώσες, JobTracker και TaskTracker, αντικαταστάθηκαν από τα αντίστοιχα συστατικά του YARN, που αναφέρθηκαν παραπάνω, τους ResourceManager και NodeManager.

Για την κατανομή των δεδομένων εισόδου και την ανασυγκρότηση των αποτελεσμάτων, το MapReduce λειτουργεί παράλληλα σε μεγάλα μεγέθη συστάδων. Επειδή το μέγεθος της συστάδας δεν επηρεάζει τα τελικά αποτελέσματα μιας εργασίας επεξεργασίας, οι εργασίες μπορούν να χωριστούν σε σχεδόν οποιοδήποτε αριθμό διακομιστών. Επίσης το MapReduce

είναι διαθέσιμο σε πολλές γλώσσες προγραμματισμού, όπως C, C++, Java, Python, Ruby και Perl. Ως εκ τούτου, το MapReduce και το γενικό πλαίσιο του Hadoop απλοποιούν την ανάπτυξη λογισμικού για εφαρμογές μεγάλων δεδομένων.

Επίσης, το MapReduce είναι ανεκτικό σε σφάλματα, με κάθε κόμβο να αναφέρει περιοδικά την κατάστασή του στον κύριο κόμβο(master node). Εάν ένας κόμβος δεν ανταποκρίνεται όπως αναμένεται, ο κύριος κόμβος επαναπροσδιορίζει το κομμάτι της εργασίας του κόμβου σε άλλους διαθέσιμους της συστάδας. Αυτή η ιδιότητα είναι που καθιστά το MapReduce ευέλικτο και πρακτικό, όπως επίσης και ικανό να τρέχει σε φθηνούς διακομιστές βασικού εξοπλισμού.

Ωστόσο, μια πρόκληση που αντιμετωπίζει το MapReduce είναι η υποδομή που χρειάζεται για να τρέξει. Πολλές επιχειρήσεις που θα μπορούσαν να επωφεληθούν από την επεξεργασία μεγάλου όγκου δεδομένων δεν μπορούν να διατηρήσουν το κεφάλαιο και τα γενικά έξοδα που χρειάζονται για μια τέτοια υποδομή. Ως αποτέλεσμα, ορισμένοι οργανισμοί βασίζονται σε δημόσιες υπηρεσίες νέφους για την χρήση Hadoop και MapReduce, οι οποίες προσφέρουν τεράστια επεκτασιμότητα με ελάχιστο κόστος κεφαλαίου ή γενικά έξοδα συντήρησης.

- **Spark**

Το Apache Spark είναι ένα πλαίσιο παράλληλης επεξεργασίας ανοιχτού κώδικα για την διεξαγωγή εφαρμογών ανάλυσης μεγάλου όγκου δεδομένων σε συστοιχίες υπολογιστών. Μπορεί να χειριστεί τόσο την ανάλυση δεσμίδας όσο και την ανάλυση σε πραγματικό χρόνο και το φόρτο εργασίας της επεξεργασίας δεδομένων.

Η πρωταρχική σχεδίαση της τεχνολογίας έγινε το 2009 από ερευνητές του Πανεπιστημίου της Καλιφόρνιας του Berkley, με στόχο την επιτάχυνση των εργασιών επεξεργασίας σε συστήματα Hadoop. Το Φεβρουάριο του 2014, το Spark έγινε ένα από τα κορυφαία έργα της Apache Software Foundation και η έκδοση 1.0 κυκλοφόρησε τον Μάιο του 2014. Η έκδοση 2.0 κυκλοφόρησε τον Ιούλιο του 2016.

Το Spark μπορεί να επεξεργαστεί δεδομένα από μια ποικιλία αποθετηρίων δεδομένων, όπως το Hadoop Distributed File System (HDFS), τις βάσεις δεδομένων NoSQL και διάφορους τύπους σχεσιακών δεδομένων, όπως το Apache Hive. Επιπλέον, υποστηρίζει την επεξεργασία εντός της μνήμης για να ενισχύσει την απόδοση των μεγάλων εφαρμογών ανάλυσης δεδομένων, αλλά

μπορεί επίσης να εκτελέσει μία τυπική επεξεργασία με την χρήση δίσκων, όταν τα σύνολα δεδομένων είναι πολύ μεγάλα για να χωρέσουν στη διαθέσιμη μνήμη του συστήματος.

Το SparkCore, αποτελεί την καρδιά του έργου και παρέχει τη μετάδοση κατανεμημένων εργασιών, τον προγραμματισμό και τη λειτουργικότητα I/O. Επίσης, παρέχει στους προγραμματιστές μια ταχύτερη και πιο ευέλικτη εναλλακτική λύση από το MapReduce, στο οποίο ήταν άμεσα συνδεδεμένες οι πρώιμες εκδόσεις του Hadoop. Με βάση τις εμπειρίες προγραμματιστών του ApacheSpark η εκτέλεση εργασιών γίνεται έως και 100 φορές πιο γρήγορα σε σχέση με το MapReduce κατά την επεξεργασία εργασιών με την χρήση μνήμης και 10 φορές πιο γρήγορα με τη χρήση δίσκων.

Το SparkCore λειτουργεί, εν μέρει, ως ένα επίπεδο διεπαφής προγραμματισμού εφαρμογών (ApplicationProgrammingInterface-API) και υποστηρίζει ένα σύνολο σχετικών εργαλείων για τη διαχείριση και την ανάλυση δεδομένων. Εκτός από την κύρια μηχανή επεξεργασίας που είναι το ίδιο το SparkCore, το περιβάλλον API, του ApacheSpark, συνοδεύεται από ορισμένες βιβλιοθήκες κώδικα για χρήση σε εφαρμογές ανάλυσης δεδομένων. Αυτές οι βιβλιοθήκες περιλαμβάνουν:

- SparkSQL, μια από τις συνηθέστερα χρησιμοποιούμενες βιβλιοθήκες, η οποία επιτρέπει στους χρήστες να αναζητούν δεδομένα αποθηκευμένα σε διαφορετικές εφαρμογές χρησιμοποιώντας την γλώσσα SQL.
- SparkStreaming, η βιβλιοθήκη αυτή επιτρέπει στους χρήστες να δημιουργούν εφαρμογές ανάλυσης και παρουσίασης δεδομένα σε πραγματικό χρόνο.
- MLlib, μια βιβλιοθήκη κώδικα μηχανικής μάθησης, που επιτρέπει στους χρήστες να εφαρμόζουν προηγμένες στατιστικές εργασίες στα δεδομένα της συστάδας και να δημιουργούν εφαρμογές γύρω από αυτές τις αναλύσεις.
- GraphX, Μια ενσωματωμένη βιβλιοθήκη κοινών αλγορίθμων για τον παράλληλο υπολογισμό γραφημάτων.

Η κύρια γλώσσα αλληλεπίδρασης με το ApacheSpark είναι η Scala, η οποία αποτελεί και την γλώσσα δημιουργίας του. Παρόλα αυτά, το Spark διαθέτει κατάλληλα API σύνδεσης για την χρήση πιο γνωστών γλωσσών προγραμματισμού, όπως η Java και η Python. Επίσης, υπάρχει η δυνατότητα λήψης πακέτου προγραμματισμού της R που επιτρέπει στους χρήστες να

χρησιμοποιούν την δημοφιλή, για την επιστήμη των δεδομένων, γλώσσα προγραμματισμού σε μεγαλύτερα κατανεμημένα σύνολα δεδομένων μέσα στο Spark, και να τα χρησιμοποιούν για να χτίσουν εφαρμογές που εκμεταλλεύονται αλγόριθμους μηχανικής μάθησης.

Μία σημαντική ερώτηση που έχει τεθεί για το ApacheSpark σε σχέση με το Hadoop είναι η ακόλουθη:

“Apache Spark: A Killer or Saviour of Apache Hadoop?” – O’Reily

Η απάντηση είναι πως δεν μπορούμε πρακτικά να συγκρίνουμε αυτά τα δύο πλαίσια λογισμικού μεταξύ τους. Το ApacheSpark είναι κατάλληλο για επεξεργασία μεγάλου όγκου δεδομένων σε πραγματικό χρόνο, ενώ το Hadoop σχεδιάστηκε με την προοπτική της αποθήκευσης μη-δομημένων δεδομένων και την εκτέλεση εργασιών παρτίδας. Με τον συνδυασμό των δυνατοτήτων του ApacheSpark, όπως την υψηλή ταχύτητα επεξεργασίας και την προηγμένη ανάλυση δεδομένων, και την υποστήριξη ενσωμάτωσης με τη χαμηλού κόστους λειτουργία του Hadoop σε υλικό βασικού εξοπλισμού, επιτυγχάνεται το καλύτερο αποτέλεσμα.

- **PIG**

Το Apache Pig είναι μια τεχνολογία ανοιχτού κώδικα που προσφέρει έναν μηχανισμό υψηλού επιπέδου για τον παράλληλο προγραμματισμό των εργασιών του MapReduce, που πρέπει να εκτελεστούν σε ομάδες συμπλεγμάτων Hadoop.

Επιτρέπει στους προγραμματιστές να δημιουργούν ρουτίνες εκτέλεσης ερωτημάτων για την ανάλυση μεγάλου όγκου, κατανεμημένων συνόλων δεδομένων, χωρίς να χρειάζεται να κάνουν εργασίες χαμηλού επιπέδου στο MapReduce.

Το ApachePig εξελίχθηκε από ένα έργο τηςYahooResearch και η πρώτη επίσημη περιγραφή του ήταν σε ένα άρθρο που δημοσιεύτηκε το 2008. Ο κύριος σκοπός του Pig είναι η δυνατότητα διαχείρισης όλων των ειδών δεδομένων, συμπεριλαμβανομένων δομημένων και αδόμητων πληροφοριών, και σχεσιακών και ένθετων δεδομένων. Αυτή η δυνατότητα έπαιξε σημαντικό ρόλο στην ονομασία της τεχνολογίας αυτής. Επιπλέον, ενώ η τεχνολογία συνδέεται κατά κύριο λόγο με το Hadoop, αναφέρεται πως είναι δυνατή η χρήση του και με άλλα πλαίσια λογισμικού.

Τα βασικά μέρη του Pig είναι ένας μεταγλωττιστής και μια γλώσσα scripting, γνωστή ως Pig Latin. HPigLatin είναι μια γλώσσα ροής δεδομένων, που προσανατολίζεται στην παράλληλη

επεξεργασία. Οι χρήστες αναφέρουν πως για ορισμένες εργασίες, όπως οι διασυνδέσεις των δεδομένων, είναι πιο εύκολο να δημιουργηθούν με την χρήση PigLatin παρά με την χρήση της Java. Ωστόσο, μέσω λειτουργιών που έχουν οριστεί από το χρήστη (UserDefinedFunctions - UDFs), οι εφαρμογές PigLatin μπορούν να επεκταθούν, ούτως ώστε να συμπεριλάβουν προσαρμοσμένες εργασίες επεξεργασίας γραμμένες σε Java, αλλά και άλλες γλώσσες, όπως JavaScript και Python.

- **Hive**

Το ApacheHive είναι ένα σύστημα αποθήκης δεδομένων ανοιχτού κώδικα για την αναζήτηση και την ανάλυση μεγάλων συνόλων δεδομένων, που αποθηκεύονται κυρίως σε αρχεία Hadoop. Συνήθως, αποτελεί μέλος του συνόλου των συμβατών εργαλείων, που αναπτύσσονται ως μέρος του οικοσυστήματος λογισμικού βασισμένα στο πλαίσιο Hadoop, για τη διαχείριση μεγάλου όγκου δεδομένων σε ένα κατανεμημένο υπολογιστικό περιβάλλον.

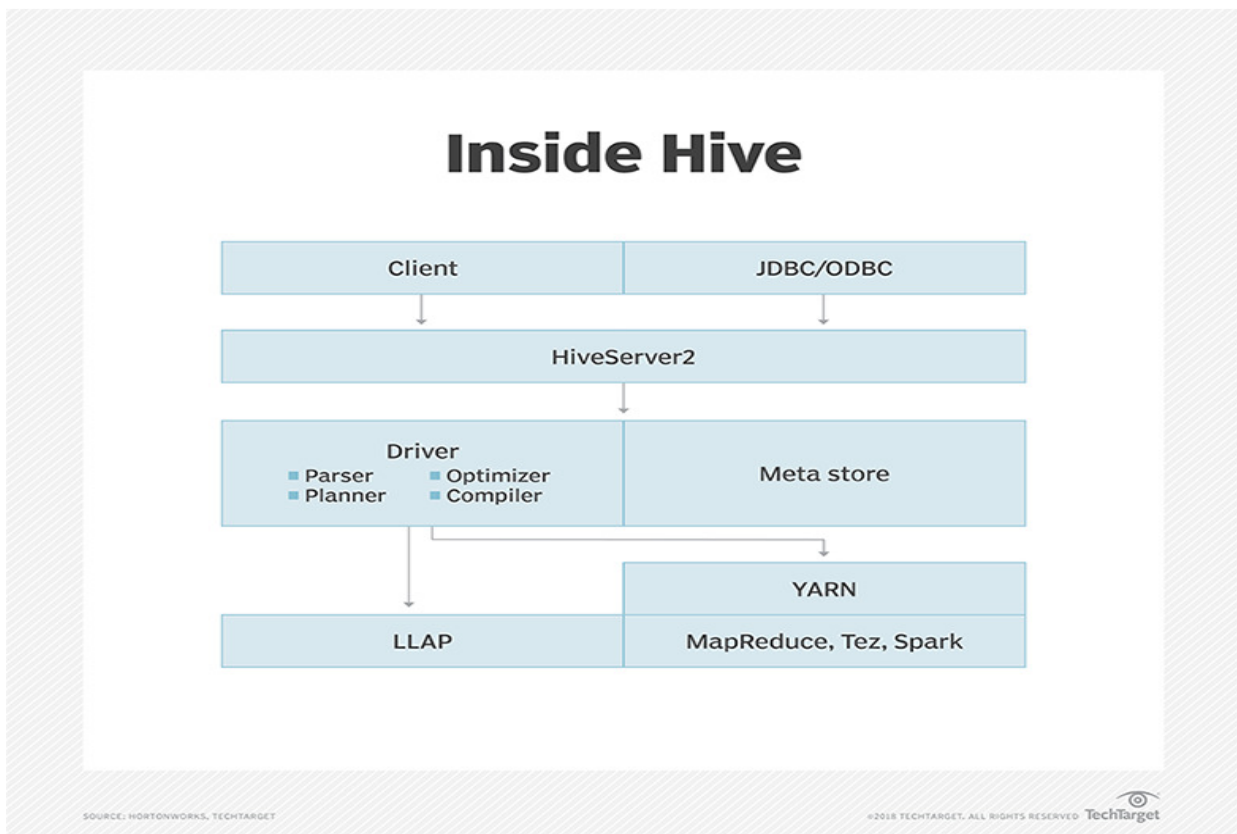
Όπως και το Hadoop, έτσι και το Hive έχει τις ρίζες του σε τεχνικές επεξεργασίας παρτίδας. Ξεκίνησε ως έργο το 2007, από προγραμματιστές του Facebook, οι οποίοι προσπάθησαν να παρέχουν πρόσβαση μέσω της SQL, σε δεδομένα Hadoop, για τους αναλυτές. Επίσης, όπως και το Hadoop, το Hive αναπτύχθηκε για να αντιμετωπίσει την ανάγκη διαχείρισης του τεράστιου όγκου δεδομένων που συσσωρεύονται μέσω της δραστηριότητας του παγκοσμίου ιστού. Η έκδοση 1.0 έγινε διαθέσιμη τον Φεβρουάριο του 2015.

Στις αρχικές εκδόσεις του Hadoop, η επεξεργασία βασιζόταν αποκλειστικά στο πλαίσιο MapReduce και αυτό απαιτούσε από τους χρήστες να κατανοούν τα προηγμένα και πολλές φορές πολύπλοκα στυλ προγραμματισμού της Java, προκειμένου να επιτύχουν τα επιθυμητά ερωτήματα στα δεδομένα. Ως εκ τούτου, το κίνητρο πίσω από το Apache Hive ήταν να απλοποιήσει την ανάπτυξη ερωτημάτων και, με τη σειρά του, να διευρύνει την πρόσβαση των μη-δομημένων δεδομένων, σε μια ευρύτερη ομάδα χρηστών.

Το Hive απαρτίζεται από τρεις κύριες λειτουργίες: σύνοψη δεδομένων, ερωτήματα και ανάλυση. Υποστηρίζει ερωτήματα, τα οποία εκφράζονται σε μια γλώσσα που ονομάζεται HiveQL, ή HQL, μια δηλωτική γλώσσα τύπου SQL που, στην πρώτη της μορφή, μεταφράζει αυτόματα ερωτήματα τύπου SQL σε εργασίες MapReduce, οι οποίες εκτελούνται στην πλατφόρμα

Hadoop. Επιπλέον, η HiveQL υποστηρίζει την σύνδεση προσαρμοσμένων scripts MapReduce σε ερωτήματα.

Όταν τα ερωτήματα SQL υποβάλλονται μέσω του Hive, αρχικά λαμβάνονται από κάποιο πρόγραμμα οδήγησης που δημιουργεί χειριστές συνεδρίας (session handlers), προωθεί το αιτήματα σε έναν μεταγωγιστή μέσω των διεπαφών Java Database Connectivity (JDBC) ή Open Database Connectivity (ODBC), οι οποίες στη συνέχεια προωθούν τις εργασίες για εκτέλεση. Το Hive, επίσης, επιτρέπει την σειριοποίηση ή την παραλληλοποίηση δεδομένων και αυξάνει την ευελιξία στο σχεδιασμό σχήματος με την ενσωμάτωση ενός συστήματος καταλόγου που ονομάζεται Hive-Metastore.



Εικόνα 7 Η εσωτερική λειτουργία του Hive

Το ApacheHive, πλέον, έχει και αυτό εξελιχθεί, ούτως ώστε να είναι συμβατό με περισσότερα εργαλεία του οικοσυστήματος, πέραν του MapReduce. Η ενσωμάτωση του διαχειριστή πόρων YARN στο Hadoop 2.0 βοήθησε τους προγραμματιστές να διευρύνουν τη χρήση του Hive, όπως

έκαναν και με άλλα συστατικά στοιχεία του οικοσυστήματος Hadoop. Με την πάροδο του χρόνου, η HiveQL έχει κερδίσει την υποστήριξη της Apache SparkSQL, και τόσο η HiveQL, όσο και το Hive έχουν αποκτήσει την δυνατότητα εκτέλεσης κατανεμημένων διεργασιών.

- **HBase**

Το ApacheHBase είναι μία τεχνολογία μη-σχεσιακής βάσης δεδομένων, ή αλλιώς μία βάση δεδομένων NoSQL, το οποίο έχει σχεδιαστεί για να εκτελείται πάνω από το DistributedFileSystem (HDFS) του Hadoop. Είναι διαμορφωμένο με βάση το BigTable της Google, το οποίο είναι ένα κατανεμημένο σύστημα αποθήκευσης που έχει σχεδιαστεί για να αντιμετωπίσει μεγάλα σύνολα δεδομένων.

Το HBase έχει σχεδιαστεί ώστε να μπορεί να υποστηρίξει υψηλούς ρυθμούς ενημέρωσης πινάκων και να μπορεί να κλιμακωθεί οριζόντια, δηλαδή αντί να αυξάνεται η χωρητικότητα των ήδη υπάρχον πόρων, να προστίθενται νέοι πόροι. Η εστίαση του Hbase, στην οριζόντια κλιμάκωση, είναι που του δίνει την δυνατότητα υποστήριξης πολύ μεγάλων βάσεων δεδομένων, με πίνακες, που για χάριν παραδείγματος, μπορεί να περιέχουν δισεκατομμύρια σειρές και εκατομμύρια στήλες. Επί του παρόντος, μία από τις πιο σημαντικές χρήσεις, του HBase, είναι η διαχείριση δομημένων δεδομένων για τη βασική υποδομή ανταλλαγής μηνυμάτων του Facebook.

Η δημοφιλία του HBase οφείλεται στην παροχή ισχυρής συνεκτικότητας στα δεδομένα, κατά την ανάγνωση ή την εγγραφή τους, το οποίο διακρίνει, το HBase, από άλλες βάσεις δεδομένων NoSQL. Επίσης, μία σημαντική πτυχή της αρχιτεκτονικής του HBase είναι η χρήση κύριων κόμβων(master nodes) για τη διαχείριση των διακομιστών που κατανέμουν και επεξεργάζονται τμήματα πινάκων δεδομένων.

Όπως και το Hadoop, έτσι και το HBase έχει γραφτεί σε Java, και συνήθως, οι HBase εφαρμογές προγραμματίζονται με τη χρήση Java και διεπαφών προγραμματισμού εφαρμογών πάνω σε αυτή, όπως REST, Avro και Thrift, και όχι με τη χρήση SQL. Ως έργο ανοιχτού κώδικα, η ανάπτυξή του διαχειρίζεται από την Apache Software Foundation και έγινε, το 2010, ένα από τα κορυφαία έργα της Apache.

- **Mahout**

Το ApacheMahout είναι μια βιβλιοθήκη αλγορίθμων μηχανικής μάθησης, που εφαρμόζεται πάνω στο ApacheHadoop και χρησιμοποιεί την ιδεολογική δομή τουMapReduce. Ημηχανική μάθηση είναι υποπεδίο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Εστιάζεται στη δυνατότητα των μηχανών να “μάθουν” χωρίς να έχουν προγραμματιστεί ρητά, και χρησιμοποιείται συνήθως για τη βελτίωση των μελλοντικών επιδόσεων βάσει προηγούμενων αποτελεσμάτων.

Από τη στιγμή που αποθηκεύονται μεγάλα δεδομένα στο σύστημα κατανομής αρχείων Hadoop (HDFS), το Mahout παρέχει τα εργαλεία επιστήμης δεδομένων και τους απαραίτητους αλγόριθμους, για να βρει αυτόματα σημαντικά πρότυπα σε αυτά τα μεγάλα σύνολα δεδομένων. Το έργο ApacheMahout στοχεύει στην ταχύτερη και ευκολότερη μετατροπή μεγάλων δεδομένων σε χρήσιμη πληροφορία.

Το Mahout υποστηρίζει τέσσερις βασικές περιπτώσεις χρήσης της επιστήμης των δεδομένων:

1. Φιλτράρισμα, χρησιμοποιώντας αλγορίθμους εξόρυξης δεδομένων για την αναγνώριση της συμπεριφοράς των χρηστών, κάνει συστάσεις για προϊόντα, όπως για παράδειγμα οι προτάσεις προϊόντων της Amazon.
2. Συσταδοποίηση, λαμβάνει στοιχεία μιας συγκεκριμένης κλάσης (όπως ιστοσελίδες ή άρθρα εφημερίδων) και τα οργανώνει σε ομάδες/συστάδες, έτσι ώστε αντικείμενα που ανήκουν στην ίδια ομάδα να έχουν κοινά χαρακτηριστικά.
3. Κατηγοριοποίηση, μαθαίνοντας από τις ήδη υπάρχουσες κατηγοριοποιήσεις, εκχωρεί μη ταξινομημένα στοιχεία στην καλύτερη κατηγορία.
4. Συχνή εξόρυξη αντικειμένων, αναλύει στοιχεία σε μια ομάδα, όπως αντικείμενα σε καλάθι αγορών, και στη συνέχεια προσδιορίζει ποια στοιχεία εμφανίζονται συνήθως μαζί.

- **Kafka**

Το ApacheKafka είναι ένα κατανεμημένο σύστημα ανταλλαγής μηνυμάτων δημοσίευσης και εγγραφής που έχει σχεδιαστεί για να αντικαταστήσει τους παραδοσιακούς μεσίτες μηνυμάτων. Ένας μεσίτης μηνυμάτων, επίσης γνωστό ως μηχανισμός διασύνδεσης, είναι μια ενδιάμεση

ηλεκτρονική μονάδα που μεταφράζει ένα μήνυμα από το επίσημο πρωτόκολλο ανταλλαγής μηνυμάτων του αποστολέα στο επίσημο πρωτόκολλο μηνυμάτων του δέκτη.

Δημιουργήθηκε και αναπτύχθηκε αρχικά από τη LinkedIn, και στη συνέχεια δόθηκε ως έργο ανοιχτού κώδικα το 2011. Η συνέχεια της ανάπτυξης του Kafka έγινε από την Apache Software Foundation για να εκμεταλλευτεί νέες υποδομές δεδομένων, οι οποίες έγιναν δυνατές λόγω των μαζικά παράλληλων συστοιχιών υπολογιστών βασικού εξοπλισμού.

Επίσης, μία ακόμα χρήση των μεσιτών μηνυμάτων είναι η αποσύνδεση των ροών δεδομένων προς επεξεργασία, και την απομάκρυνση μη απεσταλμένων μηνυμάτων. Το Apache Kafka αποδίδει καλύτερα από τους παραδοσιακούς μεσίτες μηνυμάτων, μέσω βελτιώσεων στην απόδοση, την ενσωματωμένη διαμέλιση, την αναπαραγωγή, την καθυστέρηση και την αξιοπιστία μεταφοράς των πακέτων.

Το Kafka μπορεί να χρησιμοποιηθεί για πολλούς σκοπούς: Ανταλλαγή μηνυμάτων, παρακολούθηση δραστηριότητας ενός ιστοτόπου σε πραγματικό χρόνο, παρακολούθηση λειτουργικών μετρήσεων των κατανεμημένων εφαρμογών, συνάθροιση αρχείων από πολυάριθμους διακομιστές, προμήθεια συμβάντων όπου οι αλλαγές καταστάσεων σε μια βάση δεδομένων καταγράφονται και ταξινομούνται, και συγχρονισμό και επαναφορά δεδομένων σε περίπτωση αποτυχίας του συστήματος.

- **Flume**

Το ApacheFlume είναι μια κατανεμημένη, αξιόπιστη και διαθέσιμη υπηρεσία για την αποτελεσματική συλλογή, συγκέντρωση και μεταφορά μεγάλων ποσοτήτων δεδομένων ροής στο HadoopDistributedFileSystem (HDFS). Έχει μια απλή και ευέλικτη αρχιτεκτονική που βασίζεται σε ροές δεδομένων συνεχούς ροής, και είναι ανθεκτικό σε σφάλματα, μέσω προσαρμόσιμων μηχανισμών αξιοπιστίας για ανακατεύθυνση και ανάκτηση δεδομένων.

Με τη βοήθεια του YARN, τα δεδομένα που λαμβάνονται από το ApacheFlume και άλλες υπηρεσίες που παρέχουν πρωτογενή δεδομένα, συντονίζονται μέσα σε μία συστοιχία Hadoop.

Με άλλα λόγια, το Flume επιτρέπει στους χρήστες του Hadoop να εισάγουν δεδομένα ροής μεγάλου όγκου στο HDFS προς αποθήκευση. Πιο συγκεκριμένα, το Flumeπαρέχει στους χρήστες:

- Την δυνατότητα εισαγωγής δεδομένων ροής από πολλαπλές πηγές στο Hadoop, για αποθήκευση και ανάλυση.
- Ένα πλαίσιο ενδιάμεσης αποθήκευσης από παροδικές αιχμές, όπου ο ρυθμός των εισερχόμενων δεδομένων υπερβαίνει το ρυθμό με τον οποίο μπορούν να εγγραφούν δεδομένα στον προορισμό.
- Συναλλαγές με βάση το κανάλι, ώστε να είναι εγγυημένη η αξιόπιστη παράδοση μηνυμάτων. Όταν ένα μήνυμα μεταφέρεται από έναν πράκτορα σε άλλο, ξεκινούν δύο συναλλαγές, μία για τον πράκτορα που παραδίδει το συμβάν και η άλλη για τον πράκτορα που λαμβάνει το συμβάν. Αυτό εξασφαλίζει την εγγυημένη σημασιολογική παράδοση.
- Οριζόντια κλιμάκωση, για την εισαγωγή νέων ροών δεδομένων και την περαιτέρω αύξηση του όγκου.

Οι επιχειρήσεις χρησιμοποιούν τις ισχυρές δυνατότητες συνεχούς ροής δεδομένων, του Flume, για την εκφόρτωση δεδομένων από ροές υψηλής απόδοσης στο HDFS. Τυπικές πηγές αυτών των ροών είναι τα αρχεία καταγραφής εφαρμογών, τα δεδομένα αισθητήρων και μηχανών, τα δεδομένα γεωγραφικής θέσης και τα κοινωνικά μέσα. Αυτοί οι διαφορετικοί τύποι δεδομένων μπορούν να εισαχθούν στο πλαίσιο του Hadoop για μελλοντική ανάλυση, χρησιμοποιώντας αντίστοιχα εργαλεία, όπως το ApacheHive, ή το ApacheHBase.

• Sqoop

Το ApacheSqoop αναλαμβάνει την αποτελεσματική μαζική μεταφορά δεδομένων μεταξύ του ApacheHadoop και αποθήκες δομημένων δεδομένων, όπως σχεσιακές βάσεις δεδομένων. Βοηθά στην εκφόρτωση ορισμένων εργασιών, όπως επεξεργασίες εξαγωγής, μετατροπής και φόρτωσης(ETL), από την αποθήκη δεδομένων μίας επιχείρησης στο Hadoop, για πιο αποτελεσματική εκτέλεση με πολύ χαμηλότερο κόστος. Μπορεί, επίσης, να χρησιμοποιηθεί για την εξαγωγή δεδομένων από το Hadoop και τη μεταφορά τους σε εξωτερικές αποθήκες δομημένων δεδομένων. Η κύρια διαφορά του με το ApacheFlume είναι, πως το Flume μπορεί να μόνο να εισάγει μη-δομημένα, και ημι-δομημένα δεδομένα στο HDFS. Το Sqoop συνεργάζεται και υποστηρίζεται από σχεσιακές βάσεις δεδομένων, όπως τα Teradata, Netezza, Oracle, MySQL, Postgres και HSQLDB.

Για να ενσωματώσει την μαζική κίνηση των δεδομένων μεταξύ Hadoop και των αποθηκών δεδομένων, το ApacheSqoop αναλαμβάνει την:

- Εισαγωγή διαδοχικών συνόλων δεδομένων από το κύριο υπολογιστικό σύστημα, με αυτό τον τρόπο εξασφαλίζει την αυξανόμενη ανάγκη μεταφοράς δεδομένων από τα υπολογιστικά συστήματα προς το HDFS.
- Απευθείας εισαγωγή των δεδομένων σε αρχεία τύπου ORC,για βελτιωμένη συμπίεση και ελαφριά ευρετηρίαση, με στόχοτην βελτιωμένη απόδοση ερωτημάτων.
- Εισαγωγή δομημένων δεδομένων, για τη μεταφορά ορισμένων δεδομένων από εξωτερικές αποθήκες δεδομένων επιχειρήσεων στο Hadoop, με αποτέλεσμα τη βελτιστοποίηση της οικονομικής απόδοσης για την αποθήκευση, αλλά και την επεξεργασία.
- Παράλληλη μεταφορά δεδομένων, για ταχύτερη απόδοση και βέλτιστη χρήση του υπολογιστικού συστήματος.
- Γρήγορα αντίγραφα δεδομένων, από εξωτερικά συστήματα στο Hadoop.
- Εξισορρόπηση φορτίου, περιορίζοντας την υπερβολική αποθήκευση και επεξεργασία φορτίων σε άλλα συστήματα.

- **Cassandra**

Το ApacheCassandra είναι ένα σύστημα κατανεμημένης βάσης δεδομένων ανοικτού κώδικα που έχει σχεδιαστεί για την αποθήκευση και διαχείριση μεγάλου όγκου δεδομένων σε διακομιστές βασικού εξοπλισμού. Το Cassandra μπορεί να χρησιμεύσει τόσο ως αποθήκη λειτουργικών δεδομένων,πραγματικού χρόνου, για εφαρμογές ηλεκτρονικών συναλλαγών, όσο και ως βάση δεδομένων υψηλής έντασης ανάγνωσης για μεγάλης κλίμακας συστήματα επιχειρηματικής ευφυΐας.

Αρχικά δημιουργήθηκε για το Facebook, και σχεδιάστηκε έτσι ώστε να έχει συμμετρικούς κόμβους διομότιμων(peer-to-peer) κόμβων, αντί για κύριους ή κόμβους σκλάβους, για να εξασφαλίσει ότι δεν μπορεί ποτέ να υπάρχει ένα συγκεκριμένο σημείο αποτυχίας (SinglePointofFailureSPoF). Το Cassandraκατανέμει αυτόματα τα δεδομένα σε όλους τους κόμβους του συμπλέγματος βάσεων δεδομένων, χωρίς να αφαιρεί την δυνατότητα από τον

διαχειριστή να καθορίσει ποια δεδομένα θα αναπαραχθούν, και πόσες αντιγραφές των δεδομένων θα δημιουργηθούν.

Το όνομα Cassandra ήταν εμπνευσμένο από την πανέμορφη μάντισσα της αρχαίας ελληνικής μυθολογίας, η οποία ήταν καταραμένη από τον θεό Απόλλων ώστε να μη πιστεύει κανένας της προφητείες της.

Αφότου το Facebook όρισε το έργο αυτό ως ανοιχτού κώδικα, τράβηξε το ενδιαφέρον της ApacheIncubator και έγινε έργο της το 2008, και ύστερα έγινε ένα από τα κορυφαία έργα της Apache το 2010. Επί του παρόντος, εφαρμογές του Cassandra περιλαμβάνουν το Netflix, Digg, Adobe, Twitter, HP, IBM, Rackspace, Cisco και Reddit.

- **Zookeeper**

Το ApacheZookeeper είναι μια διεπαφή προγραμματισμού εφαρμογών(API)αρχείων ανοιχτού κώδικα, που επιτρέπει στις κατανεμημένες διεργασίες σε μεγάλα υπολογιστικά συστήματα να συγχρονίζονται μεταξύ τους, έτσι ώστε όλοι οι πελάτες που υποβάλλουν αιτήματα να λαμβάνουν δεδομένα με συνέπεια.

Η υπηρεσία Zookeeper, η οποία είναι υποέργο του Apache Hadoop, παρέχεται από ένα σύμπλεγμα διακομιστών, με σκοπό να αποφευχθεί ένα ενιαίο σημείο αποτυχίας(SPoF). Χρησιμοποιεί ένα κατανεμημένο πρωτόκολλο συναίνεσης για να καθορίσει ποιος κόμβος στην υπηρεσία Zookeeper είναι ο ηγέτης/κύριος(master) ανά πάσα στιγμή.

Ο κόμβος ηγέτης αναθέτει μια χρονική σήμανση σε κάθε ενημέρωση έκδοσης για να διατηρήσει την τάξη. Μόλις η πλειοψηφία των κόμβων αναγνωρίσει την παραλαβή μιας ενημέρωσης με χρονική σφραγίδα, ο ηγέτης μπορεί να δηλώσει απαρτία, πράγμα που σημαίνει ότι όλα τα δεδομένα που περιέχονται στην ενημέρωση μπορούν να συντονιστούν με τα στοιχεία του χώρου αποθήκευσης δεδομένων. Η χρήση απαρτίας εξασφαλίζει ότι η υπηρεσία επιστρέφει πάντοτε συνεπείς απαντήσεις.

Το ZooKeeper παρέχει μια πολύ απλή διεπαφή και υπηρεσίες, δια των οποίων τα εξής βασικά πλεονεκτήματα:

- Ταχύτητα. Το ZooKeeper είναι ιδιαίτερα γρήγορο με εργασίες, όπου τα δεδομένα διαβάζονται πιο συχνά απ' ότι εγγράφονται. Πιο συγκεκριμένα, η ιδανική αναλογία ανάγνωσης / εγγραφής είναι περίπου 10: 1.
- Αξιοπιστία. Η υπηρεσία του ZooKeeper αναπαράγεται σε ένα σύνολο εξυπηρετητών, οι διακομιστές των οποίων αντιλαμβάνονται μεταξύ τους. Όσο είναι διαθέσιμος ένας κρίσιμος αριθμός εξυπηρετητών για την λειτουργία της, η υπηρεσία ZooKeeper θα είναι επίσης διαθέσιμη, καθώς δεν υπάρχει κανένα συγκεκριμένο σημείο αποτυχίας.
- Απλότητα. Το ZooKeeper διατηρεί έναν τυπικό ιεραρχικό χώρο ονομάτων, παρόμοιο με τα αρχεία και τους καταλόγους.
- Οργάνωση. Η υπηρεσία διατηρεί αρχείο όλων των συναλλαγών, οι οποίες μπορούν να χρησιμοποιηθούν για αφαίρεση υψηλότερου επιπέδου, όπως τα πρωτόκολλα συγχρονισμού.

- **Lucene & Solr**

Το ApacheLucene είναι μία ανοιχτή βιβλιοθήκη λογισμικού ανάκτησης πληροφοριών, που λειτουργεί με πεδία κειμένου μέσα σε αρχεία εγγράφων. Η ολοκληρωμένη ονομασία του έργου είναι ApacheLuceneProject.

Η διεπαφή προγραμματισμού εφαρμογών (API) του Lucene, παραμένει ίδια ανεξάρτητα από τη μορφή του αρχείου που πρόκειται να εισαχθεί σε ευρετήριο. Υπό την προϋπόθεση ότι οι πληροφορίες του κειμένου μπορούν να ανακτηθούν και να εξαχθούν, το Lucene μπορεί να ευρετηριάσει σχεδόν οποιοδήποτε είδος εγγράφου που περιέχει κείμενο. Το Lucene έχει γίνει δημοφιλές λόγω της χρήσης του σε μηχανές αναζήτησης στο διαδίκτυο, καθώς και για τις λειτουργίες αναζήτησης μεμονωμένων ιστότοπων.

Το ApacheLuceneProject περιλαμβάνει τέσσερα βασικά στοιχεία:

- Το Lucene Core, το οποίο αναλαμβάνει βασικές λειτουργίες, όπως ευρετηρίαση, αναζήτηση, ορθογραφικό έλεγχο, επισημάνσεις και συμβολισμός.
- Το PyLucene, που αποτελεί θύρα υποστήριξης της γλώσσας Python για το LuceneCore.
- Το Solr, μία εφαρμογή δημιουργημένη πάνω στο Lucene, χρησιμοποιεί XML, HypertextTransferProtocol (HTTP), παρέχει διεπαφές προγραμματισμού εφαρμογών για JavaScriptObjectNotation (JSON), Python και Ruby, καθώς και επισημάνσεις,

πολύπλευρη αναζήτηση, προσωρινή αποθήκευση, αναπαραγωγή και ένα περιβάλλον διαχείρισης ιστότοπων.

- Το OpenRelevanceProject, που στοχεύει στη δημιουργία εργαλείων για τη διεξαγωγή δοκιμών σχετικών με την ανάκτηση πληροφοριών, την μηχανική μάθηση και την επεξεργασία φυσικής γλώσσας.

Όπως αναφέρθηκε παραπάνω, το ApacheSolr είναι μία πλατφόρμα αναζήτησης ανοιχτού κώδικα, βασισμένη στη βιβλιοθήκη Lucene. Έγινε γνωστή πλατφόρμα αναζήτησης για ιστότοπους, επειδή μπορεί να ευρετηριάζει και να αναζητά πολλούς ιστότοπους. καθώς και να επιστρέφει προτάσεις για σχετικό περιεχόμενο, βάσει της ταξινόμησης του αιτήματος αναζήτησης. Το Solr είναι επίσης μια δημοφιλής πλατφόρμα αναζήτησης για επιχειρηματική αναζήτηση, επειδή μπορεί να χρησιμοποιηθεί για την ευρετηρίαση και την αναζήτηση εγγράφων και συνημμένων ηλεκτρονικού ταχυδρομείου.

Σύμφωνα με το Apache Lucene Project, το Solr έχει γίνει γνωστό λόγω των δυνατοτήτων που προσφέρει στους διαχειριστές. Μερικές από αυτές τις δυνατότητες είναι: ευρετηρίαση σε σχεδόν πραγματικό χρόνο, αυτοματοποιημένη αναπαραγωγή ευρετηρίων, καταγραφή στατιστικών στοιχείων διακομιστή, αυτόματη ανακατεύθυνση και ανάκτηση, πλούσια ανάλυση και ευρετηρίαση εγγράφων, πολλαπλά ευρετήρια αναζήτησης, εκτεταμένη αποθήκευση σε προσωρινή μνήμη, σχεδιασμένο για κυκλοφορία μεγάλου όγκου, ευελιξία και δυνατότητα επέκτασης και κλιμάκωσης, σύνθετη αναζήτηση πλήρους κειμένου, και ίσο καταμερισμό φορτίου ερωτημάτων.

- **Ambari**

Το ApacheAmbari είναι μια πλήρως ανοιχτή πλατφόρμα διαχείρισης για την παροχή, τη διαχείριση, την παρακολούθηση, και την εξασφάλιση της σωστής λειτουργίας των συμπλεγμάτων του ApacheHadoop.

Το Ambari, επιτρέπει στις επιχειρήσεις να σχεδιάζουν, να εγκαθιστούν και να διαμορφώνουν με ασφάλεια τις υπηρεσίες του ApacheHadoop, διευκολύνοντας έτσι τη συνεχή συντήρηση και διαχείριση συμπλεγμάτων, ανεξάρτητα από το μέγεθος του συμπλέγματος.

Καθιστά απλούστερη τη διαχείριση του Hadoop, παρέχοντας μια σταθερή, και ασφαλή πλατφόρμα λειτουργικού ελέγχου. Το Ambari παρέχει μια διαισθητική διεπαφή χρήστη, Web

UI, καθώς και μία ισχυρή διεπαφή προγραμματισμού εφαρμογών REST API, το οποίο είναι ιδιαίτερα χρήσιμο για την αυτοματοποίηση των λειτουργιών συμπλέγματος.

Με τη χρήση του Ambari, οι διαχειριστές του Hadoop κερδίζουν τα εξής βασικά πλεονεκτήματα:

- Απλοποιημένη εγκατάσταση, διαμόρφωση και διαχείριση. Βοηθάει στη εύκολη και αποτελεσματική δημιουργία, διαχείριση και παρακολούθηση των συμπλεγμάτων του Hadoop. Επιτρέπει την επαναλαμβανόμενη, αυτοματοποιημένη δημιουργία συμπλεγμάτων, με την χρήση της υπηρεσίας AmbariBlueprints.
- Κεντρικές ρυθμίσεις ασφαλείας. Μειώνει έτσι την πολυπλοκότητα της διαχείρισης και της διαμόρφωσης της ασφάλειας του συμπλέγματος για ολόκληρη την πλατφόρμα. Βοηθάει στην αυτοματοποίηση της ρύθμισης, και της διαμόρφωσης σύνθετων ρυθμίσεων ασφαλείας του συμπλέγματος, όπως το Kerberos.
- Πλήρη ορατότητα για την κατάσταση του συμπλέγματος. Επιβεβαιώνει την υγιή κατάσταση και τη διαθεσιμότητα του συμπλέγματος, με μία ολοκληρωμένη προσέγγιση για την παρακολούθησή του. Ρυθμίζει προκαθορισμένες ειδοποιήσεις, βασισμένες σε βέλτιστες πρακτικές, για την παρακολούθηση του συμπλέγματος. Καταγράφει και απεικονίζει κρίσιμες λειτουργικές μετρήσεις, για την ανάλυση και την αντιμετώπιση προβλημάτων.
- Εξαιρετικά επεκτάσιμο και προσαρμόσιμο. Καταφέρει να είναι επεκτάσιμο και προσαρμόσιμο με τη χρήση των εργαλείων AmbariStacks, για τη διαχείριση προσαρμοσμένων υπηρεσιών, και το AmbariViews για την προσαρμογή της διεπαφής του χρήστη, WebUI, του AmbariWeb.

5.5.Χρησιμότητα των μεγάλων δεδομένων

Πλέον τα μεγάλα δεδομένα επηρεάζουν τους οργανισμούς σε όλους σχεδόν τους τομείς. Είναι σημαντικό για κάθε εταιρεία να αντλεί την μέγιστη αξία από αυτά τα δεδομένα, η οποία δεν προέρχεται από την ακατέργαστη μορφή τους, αλλά από την επεξεργασία και την ανάλυσή τους. Η χρήση των μεγάλων δεδομένων γίνεται από ένα ευρύ κοινό εταιρειών και οργανισμών και παρακάτω θα εξετασθούν μερικές από τις βιομηχανικές προκλήσεις που αντιμετωπίζουν και πως αυτοί οι κλάδοι επιλύουν τις προκλήσεις αυτές, μέσω των λύσεων που προσφέρουν τα μεγάλα δεδομένα:

Τραπεζικά συστήματα

Με μεγάλες ποσότητες πληροφοριών που διακινούνται από αμέτρητες πηγές, οι τράπεζες αντιμετωπίζουν την εξεύρεση νέων και καινοτόμων τρόπων διαχείρισης μεγάλων δεδομένων. Παρόλο που είναι σημαντικό να κατανοήσουν τους πελάτες και να αυξήσουν την ικανοποίησή τους, είναι εξίσου σημαντικό να ελαχιστοποιήσουν τον κίνδυνο και την απάτη διατηρώντας παράλληλα τη συμμόρφωση των κανονισμών. Τα μεγάλα δεδομένα μπορούν να δώσουν διορατικές λύσεις, αλλά απαιτείται τα χρηματοπιστωτικά ιδρύματα να παραμένουν ένα βήμα μπροστά μέσω προηγμένων αναλυτικών τεχνικών.

Κυβερνήσεις

Όταν οι κυβερνητικές υπηρεσίες είναι σε θέση να εκμεταλλευτούν και να εφαρμόσουν αναλύσεις στα μεγάλα δεδομένα τους, αποκτούν σημαντικό προβάδισμα όταν πρόκειται για τη διαχείριση επιχειρήσεων κοινής ωφέλειας, τη λειτουργία οργανισμών, την αντιμετώπιση της κυκλοφοριακής συμφόρησης, ή την πρόληψη του εγκλήματος. Αλλά ενώ υπάρχουν πολλά πλεονεκτήματα μέσα από τα μεγάλα δεδομένα, οι κυβερνήσεις πρέπει να αντιμετωπίσουν θέματα διαφάνειας και ιδιωτικότητας.

Βιομηχανίες

Γνωρίζοντας για τις λύσεις που μπορούν να προσφέρουν τα μεγάλα δεδομένα στην βιομηχανία, οι κατασκευαστές μπορούν να αυξήσουν την ποιότητα και την απόδοση, ελαχιστοποιώντας ταυτόχρονα τις διαδικασίες αποβλήτων που είναι καθοριστικές για την ιδιαίτερα ανταγωνιστική αγορά του σήμερα. Όλο και περισσότερες βιομηχανίες εργάζονται σε μια κουλτούρα βασισμένη στην ανάλυση, που σημαίνει ότι μπορούν να επιλύσουν ταχύτερα τα προβλήματα και να κάνουν πιο ευέλικτες επιχειρηματικές αποφάσεις.

Εκπαίδευση

Οι εκπαιδευτικοί που εξοπλίζονται με γνώση που βασίζεται στην πληροφορία μπορούν να επηρεάσουν σημαντικά τα σχολικά συστήματα, τους σπουδαστές και τα προγράμματα σπουδών. Με την ανάλυση μεγάλων δεδομένων, μπορούν να προσδιορίσουν τους σπουδαστές που βρίσκονται σε κίνδυνο, να διασφαλίσουν ότι οι σπουδαστές σημειώνουν επαρκή πρόοδο και μπορούν να εφαρμόσουν ένα καλύτερο σύστημα αξιολόγησης και υποστήριξης των εκπαιδευτικών και των διευθυντών.

Ιατροφαρμακευτική περίθαλψη

Τα αρχεία ασθενών, σχέδια θεραπείας, πληροφορίες συνταγογράφησης, και άλλες πληροφορίες σχετικά με την περίθαλψη του ασθενή αποτελούν μία μεγάλη πρόκληση. Όταν πρόκειται για την ιατροφαρμακευτική περίθαλψη, όλα πρέπει να γίνουν γρήγορα, με ακρίβεια και, σε ορισμένες περιπτώσεις, με ιδιαίτερη διαφάνεια για να πληρούν αυστηρούς κανονισμούς του κλάδου. Όταν τα μεγάλα δεδομένα διαχειρίζονται αποτελεσματικά, οι πάροχοι υγειονομικής περίθαλψης μπορούν να αποκαλύψουν κρυφές ενδείξεις που βελτιώνουν τη φροντίδα και την περίθαλψη των ασθενών.

Λιανεμπόριο

Η δημιουργία καλών σχέσεων με τους πελάτες είναι ζωτικής σημασίας για τη λιανική βιομηχανία και ο καλύτερος τρόπος διαχείρισης είναι η διαχείριση μεγάλων δεδομένων. Οι έμποροι λιανικής πώλησης πρέπει να γνωρίζουν καλούς τρόπους προώθησης προϊόντων στους πελάτες, αποτελεσματικές μεθόδους αντιμετώπισης των συναλλαγών και στρατηγικές για την αποκατάσταση των καθυστερήσεων στις επιχειρήσεις τους. Τα μεγάλα δεδομένα παραμένουν στο επίκεντρο όλων των παραπάνω.

Μέσα Μαζικής Ενημέρωσης

Για να κατανοήσουμε τον τρόπο με τον οποίο τα μέσα μαζικής ενημέρωσης χρησιμοποιούν μεγάλα δεδομένα, είναι πρώτα απαραίτητο να δοθεί κάποιο πλαίσιο στον μηχανισμό που χρησιμοποιείται για τη διαδικασία των μέσων ενημέρωσης. Έχει ειπωθεί από τους Nick Couldry και Joseph Turow ότι οι επαγγελματίες του τομέα της μαζικής ενημέρωσης και της διαφήμισης προσπαθούν να προσεγγίζουν τα μεγάλα δεδομένα ως πολλά σημεία ενεργητικής πληροφόρησης σχετικά με εκατομμύρια άτομα. Η βιομηχανία αυτή, φαίνεται να απομακρύνεται από την παραδοσιακή προσέγγιση της χρήσης συγκεκριμένων μέσων μαζικής ενημέρωσης, όπως εφημερίδες, περιοδικά ή τηλεοπτικές εκπομπές, και αξιοποιεί τον πελάτη, μέσα από τεχνολογίες που φτάνουν στοχευμένα σε αυτόν, σε βέλτιστες χρονικές στιγμές και βέλτιστες τοποθεσίες. Ο απώτερος στόχος είναι να εξυπηρετήσει ή να μεταφέρει ένα μήνυμα ή κάποιο περιεχόμενο που, στατιστικά, είναι σύμφωνο με τη νοοτροπία του καταναλωτή. Για παράδειγμα, οι όμιλοι εκδόσεων προσαρμόζουν όλο και περισσότερο τις διαφημίσεις και το περιεχόμενο για να απευθύνονται σε καταναλωτές που επιλέχθηκαν αποκλειστικά μέσω διαφόρων δραστηριοτήτων εξόρυξης δεδομένων.

Διαδίκτυο των πραγμάτων(IoT)

Τα μεγάλα δεδομένα και το διαδίκτυο των πραγμάτων είναι έννοιες που είναι σχεδόν αλληλένδετες. Τα δεδομένα που εξάγονται από αισθητήρες και άλλες συσκευές από το διαδίκτυο των πραγμάτων παρέχουν μια χαρτογράφηση της δια συνδεσιμότητας των συσκευών. Τέτοιες χαρτογραφήσεις χρησιμοποιούνται από τη βιομηχανία των μέσων ενημέρωσης, τις επιχειρήσεις και τις κυβερνήσεις για να στοχεύσουν με μεγαλύτερη ακρίβεια το κοινό τους και να αυξήσουν την αποτελεσματικότητα των μέσων ενημέρωσης. Εκτός αυτού, καθώς το κόστος των συσκευών και των αισθητήρων του διαδικτύου των πραγμάτων μειώνεται, η χρήση τους από διάφορες παρεμφερείς βιομηχανίες, όπως τα μέσα μαζικής μεταφοράς και η ιατρική περίθαλψη, αυξάνεται. Κατά συνέπεια, η εκθετική αύξηση των συσκευών αυτών και η δραματική αύξηση της παραγωγής και κατανάλωσης των δεδομένων τους αποτελούν ένδειξη της αλληλουχίας των τομέων των μεγάλων δεδομένων και του διαδικτύου των πραγμάτων.

6. Επιχειρηματική ευφυΐα και εξόρυξη δεδομένων

6.1.Τί είναι η εξόρυξη δεδομένων

Ο ορισμός της εξόρυξης δεδομένων, με βάση την Wikipedia, είναι η «εξεύρεση μιας ενδιαφέρουσας, αυτονόητης, ή μη προφανούς και πιθανόν χρήσιμης πληροφορίας ή προτύπων από μεγάλες βάσεις δεδομένων με χρήση αλγορίθμων ομαδοποίησης ή κατηγοριοποίησης και των αρχών της στατιστικής, της τεχνητής νοημοσύνης, της μηχανικής μάθησης και των συστημάτων βάσεων δεδομένων. Στόχος της εξόρυξης δεδομένων είναι η πληροφορία που θα εξαχθεί και τα πρότυπα που θα προκύψουν να έχουν δομή κατανοητή προς τον άνθρωπο έτσι ώστε να τον βοηθήσουν να πάρει τις κατάλληλες αποφάσεις.»

Επίσης αναφέρει ότι «Ο όρος εξόρυξη δεδομένων είναι μία έννοια που συνήθως παραπέμπει σε κάθε είδος φόρμας με μεγάλη ποσότητα δεδομένων ή επεξεργασία δεδομένων, όπως συλλογή,εξαγωγή δεδομένων, τις αποθήκες δεδομένων,ανάλυση δεδομένων και στατιστικής, αλλά επίσης γενικεύεται σε κάθε είδος συστήματος υποστήριξης αποφάσεων συμπεριλαμβανομένου της τεχνητής νοημοσύνης, της εκμάθησης μηχανής και της επιχειρηματικής ευφυΐας. Στην ορθή χρήση του όρου η λέξη κλειδί είναι η ανακάλυψη, που ορίζεται ως η ανίχνευση κάτι καινούριου.»

Αντίστοιχα, η εταιρεία Gartner, η οποία εξειδικεύεται στον τομέα αυτό, ορίζει την εξόρυξη δεδομένων ως «την διαδικασία ανεύρεσης σημαντικών συσχετίσεων, μοτίβων και τάσεων με την κοπή μεγάλων ποσοτήτων δεδομένων που είναι αποθηκευμένα σε αποθετήρια. Η εξόρυξη δεδομένων χρησιμοποιεί τεχνολογίες αναγνώρισης προτύπων, καθώς και στατιστικές και μαθηματικές τεχνικές.»

Εκ των ορισμών διαπιστώνουμε τον πραγματικό στόχο της εξόρυξης δεδομένων, δηλαδή την αυτόματη ή ημι-αυτόματη ανάλυση μεγάλων ποσοτήτων δεδομένα για την εξαγωγή κάποιου προτύπου που ήταν άγνωστο μέχρι εκείνη τη στιγμή.

6.2.Ιστορία και εξέλιξη

Η διαδικασία της εξόρυξης δεδομένων για την εξαγωγή προτύπων, την ανακάλυψη κρυφών συνδέσεων και την πρόβλεψη τάσεων συμβαίνει εδώ και αιώνες. Στη δεκαετία του 60', οι στατιστικολόγοι και οι οικονομολόγοι χρησιμοποίησαν όρους, όπως το «datafishing» ή το «datadredging» για να αναφερθούν σε αυτά που θεωρούσαν ως κακή πρακτική ανάλυσης δεδομένων, χωρίς μια προ υπάρχον υπόθεση. Αργότερα αναφερόταν ως «ανακάλυψη γνώσης από βάσεις δεδομένων», καθώς ο όρος «εξόρυξη δεδομένων» δεν είχε επινοηθεί, μέχρι την δεκαετία του 90'. Πλέον, ο όρος της εξόρυξης δεδομένων περιλαμβάνει τρεις αλληλένδετους επιστημονικούς κλάδους:

- Την στατιστική, δηλαδή την αριθμητική μελέτη των σχέσεων των δεδομένων.
- Την τεχνητή νοημοσύνη, που ορίζεται ως η επιστήμη και μεθοδολογία της δημιουργίας νοούντων μηχανών.
- Και την μηχανική μάθηση, η οποία βασίζεται σε αλγόριθμους που μπορούν να μάθουν από τα δεδομένα και να κάνουν προβλέψεις.

Όσο οι συλλογές δεδομένων αυξάνονταν τόσο σε όγκο, όσο και σε πολυπλοκότητα, η χειρωνακτική ανάλυση των δεδομένων αποτελούσε μη επαρκής λύση και αντικαταστάθηκε από την αυτόματη επεξεργασία δεδομένων. Σε αυτό συνέβαλαν άλλες ανακαλύψεις της επιστήμης των υπολογιστών, όπως τα νευρωνικά δίκτυα, η συσταδοποίηση, οι γενετικοί αλγόριθμοι, τα δέντρα αποφάσεων και η μηχανή υποστήριξης διανυσμάτων.

6.3. Διαδικασία ανακάλυψης γνώσης και τεχνικές εξόρυξης δεδομένων

6.3.1. Διαδικασία ανακάλυψης γνώσης σε βάσεις δεδομένων

Με βάση τα προηγούμενα μπορούμε να διαπιστώσουμε πως οι όροι «ανακάλυψη γνώσης σε βάσεις δεδομένων» και «εξόρυξη δεδομένων» συχνά επικαλύπτουν ο ένας τον άλλον. Στην πραγματικότητα έχουν δοθεί πολλά ονόματα σε αυτή την διαδικασία, όπως εξαγωγή γνώσης, ανακάλυψη πληροφοριών, διερευνητική ανάλυση δεδομένων, συλλογή πληροφοριών και άλλα. Τα τελευταία χρόνια η ανακάλυψη γνώσης σε βάσεις δεδομένων έχει χρησιμοποιηθεί για να αναφερθεί σε μία διαδικασία που αποτελείται από πολλά βήματα, και η εξόρυξη δεδομένων είναι ένα από αυτά. Η διαδικασία της ανακάλυψης γνώσης, συνήθως, αποτελείται από τα ακόλουθα πέντε βήματα :

I. Συλλογή

Τα δεδομένα που απαιτούνται για την μετέπειτα διαδικασία εξόρυξης δεδομένων μπορούν να ληφθούν από πολλές διαφορετικές και ετερογενείς πηγές δεδομένων. Μέσω του πρώτου βήματος αποκτούνται τα δεδομένα από τις διάφορες βάσεις δεδομένων, αρχεία και μη ηλεκτρονικές πηγές.

II. Προεπεξεργασία

Πριν την εφαρμογή αλγορίθμων εξόρυξης δεδομένων, το ερευνώμενο σύνολο δεδομένων πρέπει να συναρμολογείται. Τα δεδομένα που χρησιμοποιούνται από τη διαδικασία ενδέχεται να έχουν λανθασμένα ή ελλείποντα δεδομένα. Μπορεί να υπάρχουν ανωμαλίες δεδομένων από πολλαπλές πηγές που περιλαμβάνουν διαφορετικούς τύπους δεδομένων και μετρήσεις. Μπορεί, επίσης, να υπάρχουν πολλές διαφορετικές δραστηριότητες που εκτελούνται την ίδια στιγμή. Τα λανθασμένα δεδομένα θα πρέπει να διορθώνονται ή να αφαιρούνται, ενώ τα ελλείποντα δεδομένα πρέπει να παρέχονται ή να προβλέπονται, συνήθως χρησιμοποιώντας εργαλεία εξόρυξης δεδομένων.

III. Μετασχηματισμός

Τα δεδομένα από διαφορετικές πηγές πρέπει να μετατραπούν σε κοινή μορφή επεξεργασίας. Ορισμένα δεδομένα μπορούν να κωδικοποιηθούν ή να μετατραπούν σε πιο χρησιμοποιήσιμα σχήματα. Η μείωση των δεδομένων μπορεί να χρησιμοποιηθεί για να μειωθεί ο αριθμός των

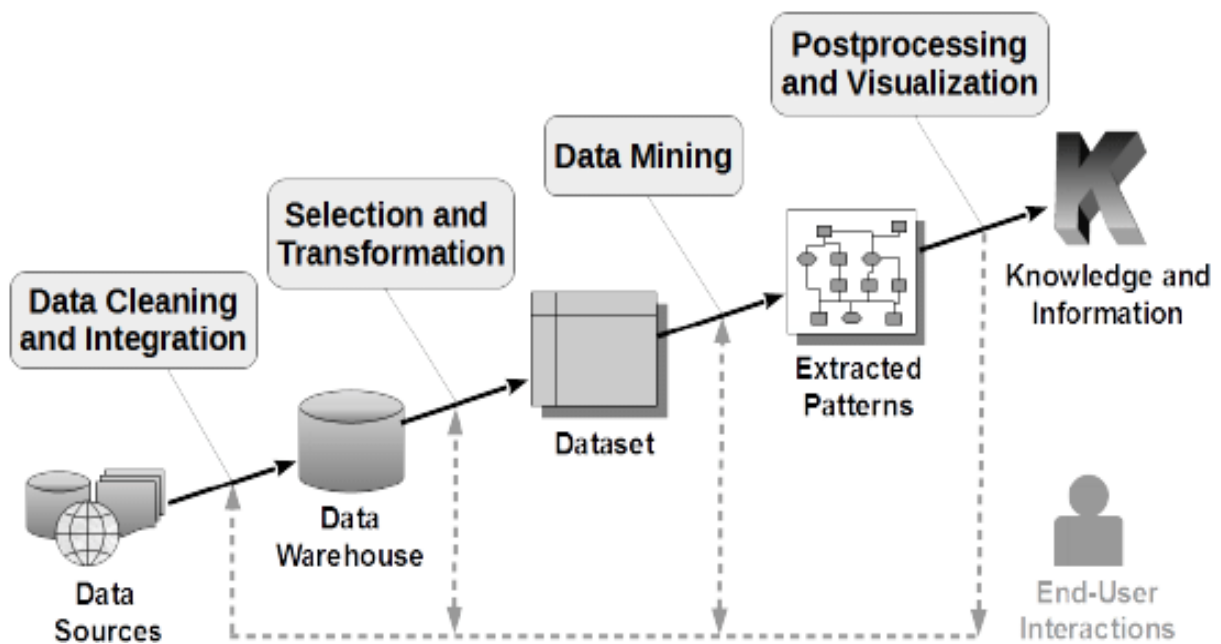
πιθανών μεταβλητών που πρέπει να ληφθούν υπόψη. Για παράδειγμα, η τρέχουσα ημερομηνία και η ημερομηνία γέννησης θα μπορούσαν να αντικατασταθούν από την ηλικία.

IV. Εξόρυξη Δεδομένων

Με βάση την διαδικασία εξόρυξης δεδομένων που ήδη εκτελείται, σε αυτό το βήμα εφαρμόζονται αλγόριθμοι στα μετασχηματισμένα δεδομένα με στόχο την δημιουργία των επιθυμητών αποτελεσμάτων.

V. Ερμηνεία/Αξιολόγηση

Τέλος, η παρουσίαση των αποτελεσμάτων εξόρυξης δεδομένων στους ενδιαφερόμενους χρήστες είναι εξαιρετικά σημαντική, διότι εξαρτάται από τη χρησιμότητα των αποτελεσμάτων. Διάφορες στρατηγικές απεικόνισης και γραφικά περιβάλλοντα χρήστη χρησιμοποιούνται σε αυτό το τελευταίο βήμα.



Εικόνα 8 Οι κύριες φάσεις της ανακάλυψης γνώσης

Υπάρχουν όμως και άλλες παραλλαγές για τον ορισμό των παραπάνω σταδίων, σύμφωνα με το Cross Industry Standard Process for Data Mining (CRISP-DM) όπου τα στάδια πριν την εξόρυξη δεδομένων είναι περισσότερα, ούτως ώστε να διατηρείται η εγκυρότητα και η πληρότητα των

δεδομένων που επεξεργάζονται από τα εργαλεία εξόρυξης. Μία μορφή του μοντέλου αυτού είναι:

- I. Κατανόηση του θέματος
- II. Κατανόηση των δεδομένων
- III. Προετοιμασία των δεδομένων
- IV. Μοντελοποίηση
- V. Αξιολόγηση
- VI. Ανάπτυξη

Επίσης, μία απλοποιημένη μορφή του παραπάνω μοντέλου είναι (1) Προεπεξεργασία, (2) Εξόρυξη δεδομένων, (3) Επικύρωση αποτελέσματος.

6.3.2. Τεχνικές εξόρυξης δεδομένων

Η εξόρυξη δεδομένων περιλαμβάνει επιμέρους διαδικασίες ή τεχνικές. Φυσικά, αυτές οι επιμέρους διεργασίες μπορούν να συνδυαστούν για την απόκτηση πιο εξελιγμένων εφαρμογών εξόρυξης δεδομένων, ανάλογα με το πρόβλημα που πρέπει να επιλυθεί. Οι διαδικασίες είναι οι εξής:

- **Κατηγοριοποίηση(Classification)**

Η διαδικασία της κατηγοριοποίησης τοποθετεί τα δεδομένα σε προκαθορισμένες ομάδες ή κλάσεις. Η διαδικασία αυτή περιέχει κοινά με την εποπτευόμενη μάθηση, κυρίως στο γεγονός ότι, οι ομάδες ή κλάσεις καθορίζονται πριν από την εξέταση των δεδομένων. Μερικά παραδείγματα της εφαρμογής κατηγοριοποίησης είναι, ο καθορισμός λήψης τραπεζικού δανείου και ο εντοπισμός πιστωτικών κινδύνων. Οι αλγόριθμοι κατηγοριοποίησης απαιτούν οι κλάσεις να ορίζονται με βάση τις τιμές των χαρακτηριστικών, των δεδομένων και περιγράφουν συχνά τις κλάσεις εξετάζοντας τα χαρακτηριστικά γνωρίσματα που είναι ήδη γνωστά ότι ανήκουν σε αυτές. Η αναγνώριση μοτίβων είναι ένας τύπος ταξινόμησης, όπου ένα πρότυπο εισόδου κατατάσσεται σε μία από τις διάφορες κλάσεις με βάση την ομοιότητά του με αυτές τις προκαθορισμένες κατηγορίες.

- **Παλινδρόμηση(Regression)**

Η παλινδρόμηση είναι μια ευρέως χρησιμοποιούμενη στατιστική τεχνική μοντελοποίησης για την έρευνα της συσχέτισης μεταξύ μίας εξαρτώμενης μεταβλητής και μιας ή περισσότερων

ανεξάρτητων μεταβλητών. Χρησιμοποιείται με σκοπό την εκχώρηση δεδομένων σε μία πραγματική μεταβλητή πρόβλεψης, όπως ισχύει και στην περίπτωση της κατηγοριοποίησης όταν η μεταβλητή είναι διακριτή, αλλιώς καλείται παλινδρόμηση, αν η μεταβλητή είναι συνεχής.

Στην πραγματικότητα, η παλινδρόμηση προϋποθέτει την εύρεση συναρτήσεων, στις οποίες ταιριάζουν τα σχετικά δεδομένα και ύστερα καθορίζει την καλύτερη συνάρτηση αυτού του είδους και μοντελοποιεί τα δεδομένα που έχουν δοθεί. Το αποτέλεσμα της παλινδρόμησης, δια της χρήσης της ως τεχνική εξόρυξης δεδομένων, είναι ένα μοντέλο, το οποίο χρησιμοποιείται αργότερα για να προβλέψει τις τιμές της κατηγορίας για τα νέα δεδομένα.

- **Πρόβλεψη(Prediction)**

Πολλές εφαρμογές εξόρυξης δεδομένων μπορούν να θεωρηθούν ως προβλέψεις μελλοντικών καταστάσεων των δεδομένων, με βάση ιστορικά και τρέχοντα δεδομένα. Η πρόβλεψη μπορεί να θεωρηθεί και ως ένας τύπος κατηγοριοποίησης. Η διαφορά είναι ότι στην διαδικασία της πρόβλεψης, προβλέπονται οι μελλοντικές καταστάσεις των δεδομένων, και κατά συνέπεια ο καθορισμός τους σε μία κλάση, ενώ στην διαδικασία της κατηγοριοποίησης αυτό γίνεται για τα τρέχοντα δεδομένα. Παρότι οι μελλοντικές τιμές των δεδομένων μπορούν να προβλεφθούν χρησιμοποιώντας και άλλες παρεμφερείς τεχνικές, όπως τεχνικές χρονικής ανάλυσης ή τεχνικές παλινδρόμησης, μπορούν να χρησιμοποιηθούν και άλλες προσεγγίσεις ανάλογα με το πρόβλημα. Μερικές από τις εφαρμογές της πρόβλεψης είναι η αναγνώριση ομιλίας, η μηχανική μάθηση, και η αναγνώριση προτύπων.

- **Συσταδοποίηση(Clustering)**

Η συσταδοποίηση είναι παρόμοια διαδικασία με την της κατηγοριοποίησης, εκτός από το ότι οι ομάδες δεν είναι προκαθορισμένες, αλλά καθορίζονται από τα δεδομένα και μόνο. Η ομαδοποίηση αναφέρεται εναλλακτικά ως μη επιτηρούμενη μάθηση ή κατάτμηση. Μπορεί να θεωρηθεί ως διαχωρισμός ή κατάτμηση των δεδομένων σε ομάδες που μπορεί να μην είναι ήδη διαχωρισμένες. Η συσταδοποίηση επιτυγχάνεται, συνήθως, με τον προσδιορισμό της ομοιότητας μεταξύ των δεδομένων για προκαθορισμένα χαρακτηριστικά. Τα πιο παρόμοια δεδομένα ομαδοποιούνται σε ομάδες. Δεδομένου ότι οι ομάδες δεν είναι προκαθορισμένες, ένας ειδικός τομέας συχνά απαιτείται για να ερμηνεύσει τη σημασία των δημιουργημένων συστάδων.

- **Κανόνες Συσχέτισης(Association Rules)**

Οι κανόνες συσχέτισης, που αναφέρονται εναλλακτικά ως ανάλυση συνδέσεων ή συσχετίσεων, αφορά τη διαδικασία της εξόρυξης δεδομένων που είναι υπεύθυνη για την αποκάλυψη σχέσεων μεταξύ δεδομένων. Η διαδικασία αυτή, αποτελεί μία από τις σημαντικότερες και νεότερες τεχνικές εξόρυξης γνώσης από μεγάλες βάσεις δεδομένων. Οι πληροφορίες και τα δεδομένα που συγκεντρώνονται παράγουν ενδιαφέρουσες συσχετίσεις και πρότυπα, τα οποία βρίσκουν εφαρμογή από τους τομείς της ζωής και της ενασχόλησης του ανθρώπου. Κύριο ρόλο στην ώθηση και μετέπειτα την ευρεία χρήση των κανόνων συσχέτισης ήταν η ανάγκη κατανόησης και ανάλυσης του καλαθιού ηλεκτρονικών αγορών.

- **Ανίχνευση Ανωμαλιών(Anomaly Detection)**

Ανίχνευση ανωμαλιών καλείται η αναγνώριση προτύπων από ένα σύνολο δεδομένων που εμφανίζουν διαφορετική συμπεριφορά από την προσδοκώμενη. Σε πολλές περιπτώσεις, η απλή αναγνώριση του γενικού προτύπου δεν μπορεί να δώσει μια σαφή κατανόηση του συνόλου δεδομένων. Άλλες παρόμοιες ονομασίες που συναντιούνται για τις ανωμαλίες των προτύπων είναι ακραίες τιμές, ασύμφωνες παρατηρήσεις, εξαιρέσεις, παρεκκλίσεις, ιδιαιτερότητες και προσμίξεις, ανάλογα με τους τομείς των εφαρμογών. Στόχος αυτής της διαδικασίας είναι η υψηλού επιπέδου ανίχνευση πιθανών ανωμαλιών, διατηρώντας όμως χαμηλά ποσοστά λανθασμένης προειδοποίησης.

- **Συνοψιση(Summarization)**

Η συνοψιση περιλαμβάνει μεθόδους για την εύρεση μιας συμπαγούς περιγραφής για ένα υποσύνολο δεδομένων, συμπεριλαμβανομένης της απεικόνισης και της δημιουργίας αναφορών. Η διαδικασία της συνοψισης είναι επίσης γνωστή και ως περίληψη ή γενίκευση. Οι πιο εξελιγμένες μέθοδοι της διαδικασίας περιλαμβάνουν, την παραγωγή περιληπτικών κανόνων, τεχνικές πολλαπλών απεικονίσεων, και την ανακάλυψη λειτουργικών σχέσεων μεταξύ των μεταβλητών. Οι τεχνικές περίληψης συχνά εφαρμόζονται για την διερευνητική ανάλυση δεδομένων και την αυτοματοποιημένη παραγωγή αναφορών.

6.4.Εξόρυξη δεδομένων στις επιχειρηματικές διαδικασίες

Με βάση τα προηγούμενα υποκεφάλαια μπορούμε να καταλήξουμε στο συμπέρασμα ότι, η εξόρυξη δεδομένων αποτελεί έναν επιστημονικό κλάδο που έχει παίξει κυρίαρχο ρόλο στη

σύγχρονη επιχείρηση. Η δυνατότητα των τεχνικών της διαδικασίας, της εξόρυξης δεδομένων, στην μετατροπή των δεδομένων σε χρήσιμη πληροφορία, την καθιστά απαραίτητο εργαλείο για κάθε επιχείρηση.

Η διαφοροποίηση της από την επιχειρηματική ευφυΐα φαίνεται εκ του ορισμού. Η εξόρυξη δεδομένων ορίζεται, συνήθως, ως ανάλυση δεδομένων για σχέσεις και πρότυπα που δεν έχουν ανακαλυφθεί μέχρι στιγμής, εφαρμόζοντας στατιστικές και μαθηματικές μεθόδους. Ωστόσο, η επιχειρηματική ευφυΐα αποτελείται από διαδικασίες και μεθόδους για την συστηματική συλλογή, αποθήκευση, ανάλυση και παροχή πρόσβασης σε δεδομένα για να βοηθηθούν οι επιχειρήσεις στη λήψη καλύτερων στρατηγικών και επιχειρηματικών αποφάσεων. Μέσα από τον ορισμό της επιχειρηματικής ευφυΐας μπορούμε να καταλάβουμε πως οι εφαρμογές της περιλαμβάνουν ένα σύνολο από εργαλεία, όπως τα συστήματα υποστήριξης αποφάσεων, τα συστήματα διαχείρισης πληροφοριών, τα συστήματα OnlineAnalyticalProcessing(OLAP) και την εξόρυξη δεδομένων.

Ένα ερώτημα που τίθεται για την εξόρυξη δεδομένων, σε συστήματα επιχειρηματικής ευφυΐας, είναι, ο τρόπος ενσωμάτωσης της διαδικασίας εξόρυξης δεδομένων στις επιχειρηματικές διαδικασίες. Η χρήση της εξόρυξης δεδομένων σε μία επιχείρηση μπορεί να είναι από αναλυτικές διοικητικές διαδικασίες λήψης αποφάσεων, έως και καθημερινές διεργασίες του λειτουργικού επιπέδου της επιχείρησης. Κατά συνέπεια, ο εμπλουτισμός των επιχειρησιακών διαδικασιών με νέες τεχνολογίες προκαλεί τον ανασχεδιασμό τους σε μικρότερο ή μεγαλύτερο επίπεδο.

Για την σωστή ενσωμάτωση της εξόρυξης δεδομένων στις επιχειρησιακές διαδικασίες, έχουν προταθεί σχετικά μοντέλα. Οι Rupnik και Jaklic(2009), έχουν προτείνει μία μεθοδολογία για την ενσωμάτωση αυτή, η οποία αποτελείται από τρία στάδια:

1^ο Στάδιο:

Σκοπός αυτής της φάσης είναι η αξιολόγηση της ετοιμότητας και της αποδοτικότητας της επιχειρησιακής διαδικασίας, και των ατόμων που εμπλέκονται σε αυτήν, για την υλοποίηση της εξόρυξης δεδομένων στην επιχειρησιακή τους διαδικασία.

Εκτός αυτού, η πρώτη φάση ορίζει επίσης την επιχειρηματική αξία της χρήσης της εξόρυξης δεδομένων και αξιολογεί τους κινδύνους και τις ευκαιρίες που μπορεί να υπάρχουν.

2^ο Στάδιο:

Ηδεύτερη φάση είναι υπεύθυνη για την ενσωμάτωση της εξόρυξης δεδομένων στις επιχειρηματικές διαδικασίες. Για την επιτυχή υλοποίηση, υπάρχουν δυο βασικές δραστηριότητες που πρέπει να εκτελεστούν: α) ο σχεδιασμός της μεταβολής των διαδικασιών και β) η ανάλυση και ο σχεδιασμός της εφαρμογής.

Στόχος του πρώτου είναι, να καθορίσει και να σχεδιάσει τις απαραίτητες αλλαγές που πρέπει να γίνουν στις επιχειρηματικές διαδικασίες, ώστε η χρήση της εξόρυξης δεδομένων να έχει προστιθέμενη αξία. Στόχος του δεύτερου είναι, να γίνει η ανάλυση και ο σχεδιασμός της εφαρμογής που θα χρησιμοποιηθεί για την εξόρυξη δεδομένων, και να αναπτυχθεί.

3^ο Στάδιο

Στην τελευταία φάση παίρνει μέρος η υποστήριξη της καθημερινής χρήσης της εξόρυξης δεδομένων στις επιχειρηματικές διαδικασίες. Παρότι αυτή η δραστηριότητα εκτελέστηκε ήδη στο πρώτο στάδιο της μεθοδολογίας, μεταξύ της πρώτης και της τρίτης φάσης μπορεί να υπάρξει μεγάλη χρονική απόκλιση. Ως εκ τούτου, η δραστηριότητα εκτελείται εκ νέου, προκειμένου να προσαρμοστεί στις πιθανές αλλαγές των βάσεων δεδομένων. Μετά από αυτή την δραστηριότητα, η επιχείρηση βρίσκεται σε ετοιμότητα για την καθημερινή χρήση της εξόρυξης δεδομένων, μέσω μίας ή περισσότερων εφαρμογών.

7. Σύγχρονες τάσεις στην επιχειρηματική ευφυΐα

7.1. Προβλέψεις της Gartner για την επιχειρηματική ευφυΐα

ΗGartner, Inc. είναι μία ερευνητική και συμβουλευτική εταιρεία, παγκόσμιας έκτασης, η οποία παρέχει ιδέες, συμβουλές και εργαλεία για τους ηγέτες στον τομέα της πληροφορικής, οικονομικών, εξυπηρέτησης και υποστήριξης πελατών, νομικής συμμόρφωσης, μάρκετινγκ, πωλήσεων και της αλυσίδας εφοδιασμού, σε όλο τον κόσμο.

Το MagicQuadrant αναφέρεται σε μία σειρά εκθέσεων έρευνας αγοράς της Gartner, η οποίες βασίζονται σε ιδιόκτητες μεθόδους ανάλυσης ποιοτικών δεδομένων για να επιδείξουν τάσεις και προβλέψεις της αγοράς, ως προς την κατεύθυνση, την ωριμότητα και τους συμμετέχοντες. Οι αναλύσεις τους διεξάγονται για αρκετές τεχνολογικές βιομηχανίες και ενημερώνονται κάθε 1-2 χρόνια.

Στο Magic Quadrant for Business Intelligence and Analytics του 2018, αναφέρεται εξής προβλέψεις, μέχρι το 2020, :

- Η επαυξημένη ανάλυση δεδομένων, ένα πρότυπο που περιλαμβάνει ερωτήματα και αφήγηση φυσικής γλώσσας, επαυξημένη προετοιμασία δεδομένων, αυτοματοποιημένες διαδικασίες ανάλυσης και οπτική ανακάλυψη, θα αποτελέσει κυρίαρχο παράγοντα για τις νέες αγορές επιχειρηματικής ευφυΐας, μηχανικής μάθησης και ενσωματωμένης ανάλυσης δεδομένων.
- Ο αριθμός των χρηστών, των σύγχρονων διεπαφών επιχειρηματικής ευφυΐας και ανάλυσης, οι οποίες διαφοροποιούνται από τις δυνατότητες ανεύρεσης δεδομένων, θα αυξηθεί με διπλάσιο ρυθμό και θα αποδώσει διπλάσια αξία για τις επιχειρήσεις.
- Η παραγωγή φυσικών γλωσσών και η τεχνητή νοημοσύνη θα αποτελούν κύριο χαρακτηριστικό του 90% των σύγχρονων διεπαφών επιχειρηματικής ευφυΐας.
- Το 50% των ερωτημάτων ανάλυσης θα δημιουργείται μέσω αναζήτησης και επεξεργασίας φυσικής γλώσσας ή φωνής, ή θα δημιουργούνται αυτόματα.
- Οι οργανισμοί που προσφέρουν στους χρήστες τους πρόσβαση σε μία επιμελημένη λίστα εσωτερικών και εξωτερικών δεδομένων θα αποκομίσουν διπλάσια επιχειρηματική αξία από τις επενδύσεις σε αναλύσεις σε αντίθεση με αυτές που δεν θα έχουν δίνουν την δυνατότητα πρόσβασης.
- Ο αριθμός των νέο εισαχθέντων επιστημόνων δεδομένων θα αυξηθεί πέντε φορές πιο γρήγορα σε σχέση με τον αριθμό των ειδικών και έμπειρων επιστημόνων δεδομένων.

7.2. Οι κυριότερες σύγχρονες τάσεις στην επιχειρηματική ευφυΐα

Η Business Application Research Center (BARC) είναι μία από τις κορυφαίες εταιρείες ανάλυσης της βιομηχανίας λογισμικού για επιχειρήσεις, όπως επίσης και μία εταιρεία συμβουλών που παρέχει απαραίτητες πληροφορίες σε περισσότερους από 1000 πελάτες ετησίως.

Οι έρευνες της BARC παρέχουν μία εστιασμένη επισκόπηση στις κορυφαίες λύσεις λογισμικού, τις πρωτοποριακές εξελίξεις και τωρινές ανάγκες, καθώς και τις εξελίξεις στην αγορά σε συγκεκριμένες κατηγορίες εφαρμογών των επιχειρήσεων.

Το BITrend Monitor 2018 είναι μία έρευνα της BARC που αντικατοπτρίζει τις τάσεις που κυριαρχούν στον τομέα της επιχειρηματικής ευφυΐας και της διαχείρισης δεδομένων από την

πλευρά του χρήστη. Η έρευνα αυτή αποκαλύπτει μία ολοκληρωμένη εικόνα σχετικά με τις σημαντικότερες τάσεις της επιχειρηματικής ευφυΐας.

7.2.1. Διαχείριση κύριων δεδομένων & ποιότητας δεδομένων

Η σημασία της ποιότητας των δεδομένων και της διαχείρισης κύριων δεδομένων μπορεί να εξηγηθεί με πολύ απλό τρόπο: οι άνθρωποι μπορούν να λάβουν τις σωστές αποφάσεις μόνο αν τα δεδομένα είναι ακριβή. Μέσω των μηχανισμών συσσωμάτωσής τους, οι αναφορές και αναλύσεις της επιχειρηματικής ευφυΐας μπορούν να βοηθήσουν στην αποκάλυψη προβλημάτων ποιότητας δεδομένων. Οι λειτουργικές διαδικασίες και οι διαδικασίες λήψης αποφάσεων επωφελούνται επίσης από την υψηλή ποιότητα των δεδομένων.

Ο στόχος της διαχείρισης των κύριων δεδομένων είναι η συγκέντρωση και η ανταλλαγή αυτών των δεδομένων, όπως τα κύρια δεδομένα πελατών, προμηθευτών ή προϊόντων, από πολλά συστήματα. Εκτός από ένα "κύριο" σύστημα επιχειρησιακού σχεδιασμού (ERP), πολλές εταιρείες συνεργάζονται επίσης με άλλα συστήματα διαχείρισης πελατειακών σχέσεων (CRM) ή διαχείρισης εφοδιαστικής αλυσίδας (SCM), χρησιμοποιούν υπηρεσίες ιστού, ή χρειάζονται συγχώνευση συστημάτων μετά από εταιρικές συγχωνεύσεις, ή έχουν αποτελεσματικές συνεργασίες. Υπάρχουν αποδεδειγμένες μέθοδοι για την αύξηση της ποιότητας των δεδομένων και την εφαρμογή της διαχείρισης κύριων δεδομένων. Ένα παράδειγμα είναι ο κύκλος ποιότητας δεδομένων, τον οποίο πολλοί πωλητές λογισμικού εφαρμόζουν στα εργαλεία τους.

Στη σημερινή ψηφιακή εποχή, στην οποία τα δεδομένα γίνονται ολοένα και περισσότερο ο πρωτεύων συντελεστής παραγωγής, υπάρχει αυξανόμενη ανάγκη χρήσης ή παράδοσης δεδομένων υψηλής ποιότητας σε νέες υπηρεσίες και προϊόντα. Ο καθοριστικός ρόλος και οι ευθύνες, καθώς και η παρακολούθηση και βελτιστοποίηση των διαδικασιών διασφάλισης ποιότητας, μέσα σε μια καλά σχεδιασμένη λύση λογισμικού για την ποιότητα των δεδομένων και τη διαχείριση βασικών δεδομένων, αποτελούν κρίσιμους παράγοντες επιτυχίας.

7.2.2. Ανακάλυψη & Οπτικοποίηση δεδομένων

Η ανακάλυψη δεδομένων είναι η επιχειρηματική και καθοδηγητική διαδικασία για την ανακάλυψη προτύπων και εξωγενών δεδομένων. Για την αποδοτική και αποτελεσματική αναγνώριση και αξιολόγηση προτύπων και ακραίων τιμών, η ανακάλυψη δεδομένων πρέπει να καλύπτει τρεις λειτουργικούς τομείς με έναν στενά διασυνδεδεμένο τρόπο για την υποστήριξη επαναληπτικών αναλύσεων. Προκειμένου να δοθεί στους χρήστες η δυνατότητα αυτοδύναμης

προσπάθειας ανακάλυψης, πρέπει να υποστηριχθούν πρώτα: ο καθαρισμός δεδομένων, ο εμπλουτισμός και η σύνδεση με διάφορες πηγές δεδομένων για τη δημιουργία συνόλων δεδομένων προς ανάλυση. Αυτά τα σύνολα δεδομένων μπορούν να αναλυθούν με διαδραστικό και εξερευνητικό τρόπο, βασισμένα σε πλούσιες οπτικοποιήσεις ή με την εφαρμογή προηγμένων αναλύσεων με φιλικό προς το χρήστη τρόπο για να βρεθούν σχέδια και πρότυπα που δεν είναι ορατά στο ανθρώπινο μάτι.

Στον τομέα της ανακάλυψης δεδομένων, συμβαίνουν δύο σημαντικές εξελίξεις για την αύξηση της αποτελεσματικότητας και της ποιότητας, αυτή τη στιγμή. Η βελτίωση της καθοδήγησης και εκμάθησης των χρηστών βρίσκεται στην κορυφή των προτεραιοτήτων για τους περισσότερους προμηθευτές. Η μηχανική μάθηση προστίθεται ολοένα και περισσότερο στα εργαλεία εντοπισμού δεδομένων για να καθοδηγήσει τους επιχειρηματικούς αναλυτές μέσω όλων των βημάτων της ανακάλυψης. Για να συνδέσουμε τις προσπάθειες προετοιμασίας και ανακάλυψης διαφορετικών και κατανεμημένων δεδομένων, παρατηρούμε επίσης μια τάση να ενσωματώνουμε την ανακάλυψη δεδομένων με τις διεργασίες επιχειρηματικής ευφυΐας, που είναι συνήθως ισχυρές στη διακυβέρνηση δεδομένων.

7.2.3. Αυτοεξυπηρετούμενη επιχειρηματική ευφυΐα

Η αυτοεξυπηρέτηση επιχειρηματική ευφυΐα βρίσκεται εδώ και πολύ καιρό στις λίστες επιθυμιών των οργανισμών και τα δεδομένα της έρευνας, της BARC, επιβεβαιώνουν ότι εξακολουθεί να είναι υψηλή προτεραιότητα. Η συνεχόμενη υψηλή ζήτηση από τους επιχειρηματικούς χρήστες και οι εξίσου υψηλοί ρυθμοί υλοποίησης ενισχύουν τη σημασία της.

Οι τοπικοί χρήστες της επιχείρησης απαιτούν την πρόσβαση σε δεδομένα οποιαδήποτε στιγμή, οπουδήποτε και από οποιαδήποτε συσκευή, ενώ νέες αναλύσεις και αναφορές πρέπει να παρέχονται σε σύντομο χρονικό διάστημα. Αυτό αυξάνει την πίεση στους οργανισμούς πληροφορικής και επιχειρηματικής ευφυΐας, υποδεικνύοντας ότι οι παραδοσιακές μέθοδοι ανάπτυξης απλά δεν είναι κατάλληλες για τις περισσότερες περιπτώσεις χρήσης. Ως αποτέλεσμα, συνεχίζεται η αύξηση του αριθμού των εφαρμογών που επιτρέπουν στους χρήστες των επιχειρήσεων να δημιουργούν ή να σχεδιάζουν τα δικά τους ερωτήματα, αναφορές, πίνακες ελέγχου ή εξερευνητικά αναλυτικά στοιχεία για ανακάλυψη δεδομένων.

Λόγο του μεγάλου βαθμού αποκεντρωμένης δημιουργίας περιεχομένου, η σύνδεση και η διαχείριση των προσπαθειών των επιχειρηματικών χρηστών καθίσταται όλο και πιο δύσκολη. Η

αυτοεξυπηρετούμενη επιχειρηματική ευφυΐα παρέχει την απαιτούμενη ευελιξία και αυξάνει το χρόνο διάθεσης στην αγορά. Αλλά αυτό δεν πρέπει να γίνεται σε βάρος της ποιότητας των κρίσιμων περιπτώσεων χρήσης ή της αποτελεσματικότητας. Επομένως, είναι σημαντική υποστήριξη της επαναχρησιμοποίησης περιεχομένου και η εύρεση του σωστού επίπεδου αυτοεξυπηρέτησης για όλους τους τύπους περιπτώσεων χρήσης και χρήστες.

7.2.4. Διακυβέρνηση δεδομένων

Σε αντίθεση με τη διακυβέρνηση επιχειρηματικών πληροφοριών, η οποία επικεντρώνεται στην ανάλυση, η διακυβέρνηση δεδομένων επικεντρώνεται στα δεδομένα των λειτουργικών συστημάτων και των συστημάτων ανάλυσης. Η διακυβέρνηση δεδομένων είναι απαραίτητη για εταιρείες με γνώμονα τα δεδομένα, οι οποίες επεκτείνουν τις υπάρχουσες επενδύσεις επιχειρηματικής ευφυΐας με διερευνητικές και επιχειρησιακές αναλύσεις. Η διακυβέρνηση πρέπει να υπερβαίνει τα κλασικά συστήματα επιχειρηματικής ευφυΐας και, από την πλευρά των δεδομένων, να ενσωματώνει λειτουργικά συστήματα. Αυτό εξασφαλίζει ότι οι εταιρείες μπορούν να χρησιμοποιούν νέα ευρήματα και να τροποποιούν άμεσα τις διαδικασίες, αλλά απαιτεί επίσης ευρύτερη σκέψη όσον αφορά τη γνώση, την οργάνωση και την τεχνολογία.

Μια σωστή στρατηγική δεδομένων εννοχιστρώνει τον τρόπο με τον οποίο η επιχειρηματική στρατηγική μεταφράζεται σε δεδομένα και αναλύσεις. Η στρατηγική δεδομένων διαχειρίζεται την εκμετάλλευση των δεδομένων από όλες τις επιχειρηματικές διαδικασίες με στόχο την βελτίωση της αποδοτικότητας και την προώθηση της καινοτομίας των επιχειρήσεων. Η διακυβέρνηση δεδομένων είναι ο κύριος μηχανισμός καθοδήγησης που απαιτείται για την εφαρμογή της στρατηγικής δεδομένων, συμπεριλαμβανομένων πολιτικών και πλαισίων για τη διαχείριση, την παρακολούθηση και την προστασία των δεδομένων, λαμβάνοντας παράλληλα υπόψη τους ανθρώπους, τις διαδικασίες και τις τεχνολογίες. Η καθιέρωση της διακυβέρνησης δεδομένων είναι μια μακροπρόθεσμη προσπάθεια. Πάνω απ' όλα, απαιτεί μια σαφή, συνειδητή διαχειριστική σειρά αποφάσεων σχετικά με τον τρόπο εργασίας και χρήσης των δεδομένων.

7.2.5. Ανάλυση πραγματικού χρόνου

Η ταχύτερη δημιουργία αναφορών και ανάλυση δεδομένων, όχι μόνο όσον αφορά την απόδοση των ερωτημάτων, που εξακολουθεί να είναι ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζουν οι χρήστες με τα εργαλεία επιχειρηματικής ευφυΐας, είναι μια πρόκληση για πολλές εταιρείες. Υπάρχει αυξανόμενη ανάγκη άμεσης διάθεσης των δεδομένων από τα

συστήματα συναλλαγών για την υποστήριξη της ταχύτερης λήψης επιχειρησιακών αποφάσεων βάσει δεδομένων.

Η επιχειρηματική ευφυΐα πραγματικού χρόνου αναφέρεται στην σχεδόν άμεση επεξεργασία και παροχή πληροφοριών σχετικά με επιχειρηματικές λειτουργίες στα συστήματα συναλλαγών, όπως για παράδειγμα το Streaming. Η ανάλυσηπραγματικού χρόνου αφορά την συλλογή γεγονότων ή άλλων νέων δεδομένων, αμέσως μετά την εμφάνισή τους, και την επεξεργασία τους για παρουσίαση ή ανάλυση. Οι συνεχώς αυξανόμενες ποσότητες δεδομένων, ο χρόνος εκτέλεσης των υπολογιστών υψηλών επιδόσεων και η αναγνώριση προτύπων είναι μόνο μερικές από τις προκλήσεις που αντιμετωπίζουν σήμερα οι εταιρείες που εστιάζουν στην επιχειρηματική ευφυΐα με δεδομένα σε πραγματικό χρόνο.

Όπως η οπτικοποίηση ή η προγνωστική ανάλυση, η επιχειρηματική ευφυΐα με δεδομένα σε πραγματικό χρόνο μπορεί να συμπληρώσει την υπάρχουσα στρατηγική επιχειρηματικής ευφυΐας μιας επιχείρησης, με στόχο να αποκτήσει καλύτερη επίγνωση στα δεδομένα της με επιπλέον πολύτιμα ευρήματα. Η κουλτούρα της λήψης αποφάσεων ενός οργανισμού, οι διαθέσιμες δεξιότητες και ο εντοπισμός και η προώθηση κατάλληλων περιπτώσεων χρήσης αποτελούν βασικές πτυχές που πρέπει να εξεταστούν κατά την εξερεύνηση ενός έργου ανάλυσης σε πραγματικό χρόνο.

8. Προμηθευτές και εμπορικά προϊόντα επιχειρηματικής ευφυΐας

8.1. Microsoft

8.1.1. Ιστορία

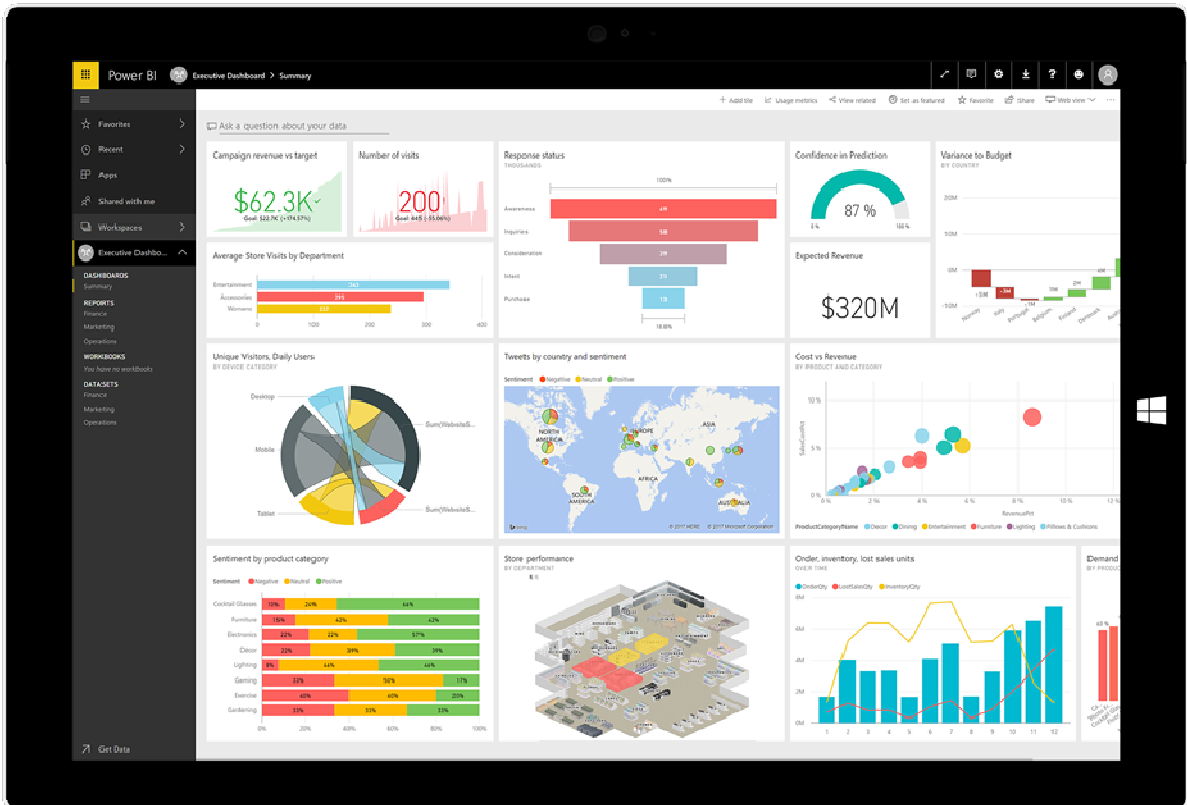
Η Microsoft, η μεγαλύτερη εταιρεία λογισμικού παγκοσμίως ως προς τα έσοδα, είναι ευρύτερα γνωστή κυρίως για το λειτουργικό σύστημα Windows και τη σουίτα εφαρμογών γραφείου MS Office. Επίσης, είναι γνωστή για αρκετά προϊόντα υλικού, όπως τη σειρά MicrosoftSurface που περιέχει tablet και φορητούς υπολογιστές, και τις κονσόλες παιχνιδιών Xbox. Ιδρύθηκε το 1975 από τον Μπιλ Γκέιτς και τον Πολ Άλεν, και τα κεντρικά γραφεία της βρίσκονται στο Redmond, της πολιτείας Ουάσιγκτον, των ΗΠΑ. Στον μακρύ κατάλογο προϊόντων λογισμικού της εταιρείας περιλαμβάνονται και εφαρμογές για επιχειρήσεις, όπως συστήματα ενδοεπιχειρησιακού σχεδιασμού (ERP) και λογισμικό επιχειρηματικής ευφυΐας.

8.1.2. Προϊόντα

Όπως αρκετοί ανταγωνιστές, η Microsoft, εστιάζει σε μεγάλο βαθμό στην παροχή λύσεων βασισμένων στο νέφος (cloud). Το Microsoft Azure είναι η υπηρεσία υπολογιστικού νέφους που δημιουργήθηκε από τη Microsoft για την κατασκευή, τη δοκιμή, την ανάπτυξη και τη διαχείριση εφαρμογών και υπηρεσιών μέσω κέντρων δεδομένων που διαχειρίζεται η ίδια η Microsoft. Μέσω του Azure παρέχει λογισμικό ως υπηρεσία (SaaS), πλατφόρμα ως υπηρεσία (PaaS), και υποδομή ως υπηρεσία (IaaS), και υποστηρίζει πολλές διαφορετικές γλώσσες προγραμματισμού, εργαλεία και πλαίσια λογισμικού, συμπεριλαμβανομένων λογισμικών και συστημάτων, τόσο της ίδιας της εταιρείας, όσο και τρίτων κατασκευαστών. Όχι μόνο η Microsoft αλλά και αρκετοί άλλοι προμηθευτές επιχειρηματικής ευφυΐας και επιχειρηματικού λογισμικού βασίζονται στο Azure για την υποδομή του νέφους τους.

Η Microsoft έχει επεκτείνει τις δυνατότητες επιχειρηματικής ευφυΐας σε όλες τις σειρές προϊόντων του Office, του SharePoint και του SQL Server τα τελευταία χρόνια, παρέχοντας εργαλεία για τη μορφοποίηση των αναφορών, τη δημιουργία εκθέσεων ad-hoc, την ανάλυση και τη δημιουργία ψηφιακών πινάκων (dashboards). Η υπηρεσία Microsoft SQL Server αποτελείται από πολλαπλά προϊόντα που περιλαμβάνουν: ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων, εργαλεία ενσωμάτωσης και διαχείρισης ποιότητας δεδομένων, μία πολυδιάστατη βάση δεδομένων Analysis Services και την υπηρεσία Reporting Services, για τη δημιουργία, τη δημοσίευση και τη διαχείριση αναφορών. Με την ενσωμάτωση της λειτουργικότητας της RevolutionR, που είναι μία βελτιωμένη έκδοση της γλώσσας R από τη Microsoft, στον SQL Server, το προϊόν αποκτά τη δυνατότητα χρήσης σε διάφορα σενάρια προηγμένης ανάλυσης.

Τον Ιούλιο του 2015, η Microsoft ξεκίνησε την νέα γενιά της σειράς προϊόντων Power BI, το οποίο κυκλοφόρησε, αρχικά, το 2013, ως ένα διαφορετικό σύνολο εργαλείων. Σήμερα, είναι ένα προϊόν επιχειρηματικής ευφυΐας βασισμένο στο νέφος και αποτελείται από το PowerBI Desktop, ένα πλήρες σύστημα δημιουργίας ad-hoc εκθέσεων, ψηφιακών πινάκων και ανάλυσης δεδομένων, και PowerBI Service, μία διεπαφή παγκοσμίου ιστού για τη δημοσίευση και την κοινή χρήση περιεχομένου.



Εικόνα 9 Ο κύριος πίνακας ελέγχου του Power BI

Το Power BI διατίθεται στο εμπόριο ως ένα διαδραστικό εργαλείο για την απεικόνιση δεδομένων, προσανατολισμένο στη διευκόλυνση των επιχειρησιακών στελεχών να αναλύουν τα δεδομένα και να μοιράζονται ιδέες κυρίως μέσω των ψηφιακών πινάκων. Η τιμολόγηση και η κυριαρχία των προϊόντων της Microsoft στους προσωπικούς υπολογιστές γραφείου προσφέρουν στην Power BI τεράστια πρόσφυση στην αγορά. Το Power BI περιλαμβάνεται στη σουίτα Azure IoT, και ως μέρος του συνόλου εργαλείων της Microsoft για την επιστήμη των δεδομένων. Μεταξύ άλλων, η σουίτα αυτή προσφέρει το AzureML (Machine Learning) ως πλαίσιο για την ανάλυση δεδομένων. Μέσω ενός ολοκληρωμένου περιβάλλοντος ανάπτυξης που ονομάζεται AzureML Studio, οι χρήστες μπορούν να δημιουργήσουν μοντέλα δεδομένων χρησιμοποιώντας χειρονομίες μεταφοράς και απόθεσης, και απλά διαγράμματα ροής δεδομένων.

8.1.3. Δυνατά σημεία

- **Χαμηλό κόστος:** Η Microsoft αύξησε σημαντικά τη συνδρομητική βάση της μέσα στο 2017. Λόγω αυτού, έχει ασκήσει καθοδικές πιέσεις στις τιμές, της αγοράς της ανάλυσης δεδομένων και επιχειρηματικής ευφυΐας, με τη χαμηλή τιμή συνδρομής για κάθε χρήστη και, πλέον, για κάθε εικονικό διακομιστή. Το κόστος της άδειας χρήσης ήταν ο δεύτερος πιο σημαντικός λόγος για την επιλογή του Microsoft PowerBI, με βάση τις αναφορές των πελατών, με 12% από αυτούς, να το επικαλούνται ως τον κύριο λόγο για την επιλογή τους.
- **Ευκολία χρήσης και οπτική ευχαρίστηση:** Οι βαθμολογίες των πελατών της Microsoft, την τοποθετούν στο κορυφαίο τεταρτημόριο των προμηθευτών, για ευκολία στη χρήση, με το 14% των πελατών να το επικαλούνται ως το κύριο κριτήριο αγοράς. Επίσης, οι βαθμολογίες των πελατών για το Power BI, την τοποθετούν πάλι στους κορυφαίους προμηθευτές, για οπτική εμφάνιση. Στόχος της βασικής στρατηγικής “5X5” της Microsoft είναι να κερδίσει την εμπιστοσύνη και το ενδιαφέρον του πελάτη, όπου η εγγραφή του πελάτη χρειάζεται πέντε δευτερόλεπτα και η κατάπληξή του, με το προϊόν, πέντε λεπτά. Υπάρχουν ορισμένα χαρακτηριστικά μέσα στο προϊόν που συμβάλλουν στη συνολική ευκολία της χρήσης του, συμπεριλαμβανομένου του μοντέλου ανάπτυξης, το οποίο είναι βασισμένο κατά κύριο λόγο στο νέφος. Η Microsoft ήταν από νωρίς στον κλάδο από την άποψη της διεπαφής αναζήτησης φυσικής γλώσσας με το Microsoft QnA (Questions and Answers), το οποίο επιτρέπει στους πελάτες να δημιουργήσουν μια οπτικοποίηση χρησιμοποιώντας όρους αναζήτησης, για τυχόν απορίες σε σχέση με το προϊόν.
- **Όραμα προϊόντων:** Η Microsoft συνεχίζει την πορεία της ως ένας από τους μεγαλύτερους οραματιστές για το μέλλον της επιχειρηματικής ευφυΐας. Το Quick Insights, ένα χαρακτηριστικό του PowerBI, αποτελεί μία βασική μορφή ανακάλυψης δεδομένων. Ενώ αρχικά ήταν διαθέσιμο μόνο στην υπηρεσία νέφους της Microsoft, στα τέλη του 2017 έγινε διαθέσιμο και στην έκδοση Desktop για την ανάλυση διασποράς, όπως, για παράδειγμα, τον εντοπισμό των αλλαγών στις πωλήσεις από το ένα τετράμηνο στο επόμενο. Επίσης, η χρήση διεπαφών συνομιλίας με chatbots, όπως η Cortana, και η πλήρη υποστήριξη της αναπαραγωγής περιεχομένου του PowerBI, σε διαδραστικούς πίνακες, όπως το Surface Hub της Microsoft, αποτελούν ειδοποιό διαφορά

μεταξύ της Microsoft και άλλως προμηθευτών. Άλλα έργα που βρίσκονται σε εξέλιξη είναι η ενσωμάτωση εικονικής πραγματικότητας και Power BI με ταHoloLens της Microsoft.

- **Εμπειρία πελατών:** Η Microsoft μαζί με την Sisense μοιράστηκαν τις υψηλότερες βαθμολογίες πελατών στο MagicQuadrant του 2018 για την επίτευξη επιχειρηματικών οφελών. Συγκεκριμένα, όσον αφορά την παροχή καλύτερης πληροφορίας και γνώσης σε περισσότερους χρήστες. Το υψηλό επίτευγμα αυτού του επιχειρηματικού οφέλους ενισχύεται, τόσο από την χαμηλή τιμή, όσο και από την ανάπτυξη του στο νέφος. Η υιοθεσία των προϊόντων από τον τελικό χρήστη, και η διαθεσιμότητα εξειδικευμένου προσωπικού αποτελούν και αυτά μέρος της συνολικής εμπειρίας των πελατών, για τα οποία η Microsoft αξιολογείται ως άνω του μέσου όρου. Η Microsoft έχει μια ισχυρή κοινότητα συνεργατών, μεταπωλητών και μεμονωμένων χρηστών, και αυτή η κοινότητα διευρύνει τη χρήση του προϊόντος με προεγκατεστημένες εφαρμογές, οπτικοποιήσεις και εκπαιδευτικά βίντεο, πέραν του περιεχομένου που παρέχει απευθείας η Microsoft.

8.1.4. Αδυναμίες

- **Ευέλικτη και αυτοεξυπηρετούμενη ανάλυση μόνο μέσω Azure:** Το Power BI έχει επικεντρωθεί, κυρίως, στις απαιτήσεις για ευέλικτες και αυτοεξυπηρετούμενες αναλύσεις, ενώ οι υπηρεσίες που παρέχει το SQL Server ReportingServices εξυπηρετούν τις ανάγκες του πελάτη για τον προγραμματισμό και την κατανομή αναφορών. Για τους πελάτες της Microsoft, αυτό έχει οδηγήσει στη σύγχυση δύο προϊόντων με διαφορετικές δυνατότητες και διαφορετικές προσεγγίσεις ανάπτυξης. Το 14% των πελατών της Microsoft ανέφεραν την απουσία ή την αδυναμία της λειτουργικότητας ως περιορισμό για της ευρύτερης ανάπτυξης, ενώ το 27% επικαλείται μία σειρά άλλων προβλημάτων. Αυτά τα άλλα προβλήματα περιλαμβάνουν συχνές ενημερώσεις που διαταράσσαν την λειτουργικότητα, και τεκμηριώσεις που δεν είχαν συνοχή σε σχέση με την ημερομηνία κυκλοφορίας. Επίσης, πολλοί χρήστες ανέφεραν προβλήματα με το PowerBIGateway, που χρησιμοποιείται για την σύνδεση εξωτερικών αποθηκών δεδομένων με αυτές της επιχείρησης.
- **Εύρος χρήσης:** Όπως και στο παρελθόν, οι βαθμολογίες της Microsoft από τους πελάτες της, την τοποθετούν στο κατώτατο τεταρτημόριο των προμηθευτών, για το εύρος χρήσης. Το εύρος χρήσης εξετάζει το ποσοστό των χρηστών που χρησιμοποιούν το

προϊόν για μια σειρά λειτουργιών επιχειρηματικής ευφυΐας: από την προβολή αναφορών, τη δημιουργία εξατομικευμένων πινάκων ελέγχου και τη διεξαγωγή απλής ad hoc ανάλυσης, έως την εκτέλεση πολύπλοκων ερωτημάτων, την προετοιμασία δεδομένων και τη χρήση προγνωστικών μοντέλων. Η πλειοψηφία των πελατών, και πιο συγκεκριμένα το 59%, χρησιμοποιούν το PowerBI, κυρίως για την παραμετροποίηση αναφορών και πινάκων ελέγχου, αντί να το χρησιμοποιούν για πιο πολύπλοκες εργασίες. Αυτό το μοτίβο χρήσης υποδηλώνει ότι οι πιο δύσκολες και σύνθετες προετοιμασίες δεδομένων εκτελούνται εκτός του PowerBI ή ότι το προσωπικό των τμημάτων Τεχνολογίας Πληροφορικής (IT) κατασκευάζει κοινούς πίνακες ελέγχου για τους επιχειρηματικούς χρήστες. Μόνο το 20% των οποίων, δημιουργούν δικό τους περιεχόμενο μέσω του PowerBI.

- **Πολλαπλά προϊόντα:** Στον πυρήνα του, το PowerBI είναι ένα αυτόνομο προϊόν. Ωστόσο, ορισμένες λειτουργίες και χαρακτηριστικά που προσφέρει η Microsoft, και αποτελούν μέρος του οράματος των προϊόντων της, καλύπτονται από πολλαπλά προϊόντα και απαιτούν περαιτέρω ενσωμάτωση. Για παράδειγμα, λειτουργίες, όπως το InsighttoAction, απαιτούν το MicrosoftFlow. Το DataCatalog αποτελεί ξεχωριστό προϊόν και τα επιμέρους σύνολα δεδομένων μπορούν να παρέχονται απευθείας από τη Microsoft ή μέσω κάποιου άλλου συνεργάτη. Επιπλέον, ορισμένες δυνατότητες δεν είναι εγγενείς, αλλά είναι δυνατές μέσω του SharePoint ή του Microsoft Teams. Οι επιλογές κλιμάκωσης δεδομένων έχουν βελτιωθεί, με μεγαλύτερα μοντέλα να υποστηρίζονται στην έκδοση Premium του PowerBI, όμως οι επιλογές κλιμάκωσης στο AzureDataLake, AzureSQL ή σε άλλα μέσα αποθήκευσης εσωτερικών εγκαταστάσεων, δεν είναι τόσο απλές.
- **Εμπειρία πωλήσεων:** Η πλειοψηφία των πελατών της Microsoft αξιολόγησε τις πωλήσεις της, ως καλές ή εξαιρετικές. Ωστόσο, η Microsoft είναι ελαφρώς κάτω από το μέσο όρο για αυτό το MagicQuadrant του 2018. Παρόλο που το Power BI είναι μια λύση χαμηλού κόστους, οι πελάτες εξέφρασαν την απογοήτευσή τους για αλλαγές στην τιμολόγηση των πακέτων και την έλλειψη σαφήνειας σχετικά με τη λειτουργικότητα του Power BI που περιλαμβάνεται στο Microsoft Dynamics 365. Πιο συγκεκριμένα, το Microsoft Dynamics 365 χρησιμοποιεί ενσωματωμένες λειτουργίες του PowerBI για τη

δημιουργία αναφορών και πινάκων ελέγχου, όμως όσον αφορά περιεχόμενο εκτός του Dynamics 365, απαιτείται άδεια PowerBIPro ή Premium.

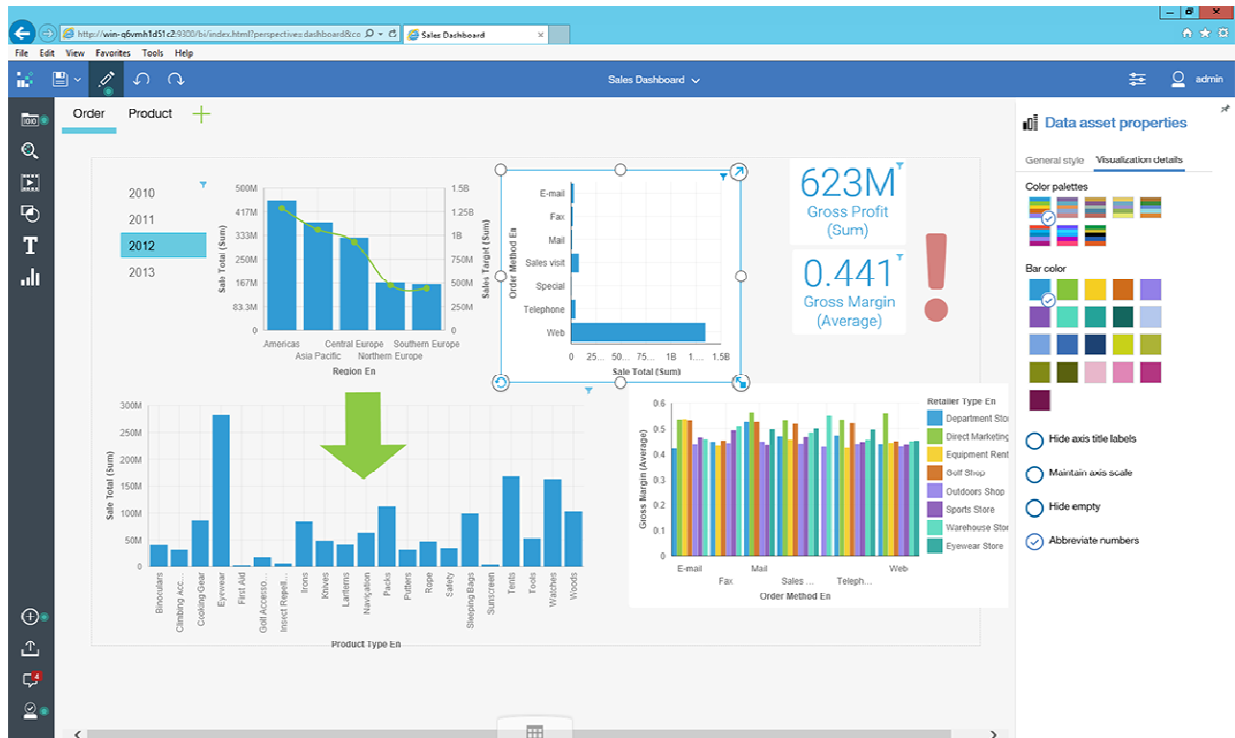
8.2.IBM

8.2.1. Ιστορία

Ως ένας από τους μεγαλύτερους προμηθευτές υλικού, λογισμικού και υπηρεσιών πληροφορικής, η IBM, προσφέρει μία ευρεία εμβέλεια λύσεων, συμπεριλαμβανομένων επιχειρηματικής ευφυΐας και διαχείρισης απόδοσης. Η IBM ιδρύθηκε στην πολιτεία της Νέας Υόρκης το 1911. Αρχικά, η εταιρεία ονομαζόταν ComputingTabulatingRecordingCompany (CTR), αλλά το 1924 μετονομάστηκε σε IBM. Τα προϊόντα της περιλάμβαναν από ζυγαριές και συσκευές χρονομέτρησης για βιομηχανική χρήση μέχρι υπολογιστικές μηχανές και διάτρητες κάρτες. Σήμερα είναι μια πολυεθνική εταιρεία τεχνολογίας υπολογιστών με κεντρικά γραφεία στο Armonk, στην πολιτεία της Νέας Υόρκης. Μέρος των προϊόντων για λύσεις επιχειρηματικής ευφυΐας και ανάλυσης δεδομένων είναι: το IBM Cognos Analytics, το IBM Planning Analytics, το IBM Watson Analytics, το IBM SPSS και την IBM Data Science Experience (DSX).

8.2.2. Προϊόντα

Το IBM Cognos Analytics παρέχει λειτουργίες για διάφορες κατηγορίες εφαρμογών επιχειρηματικής ευφυΐας, όπως πίνακες ελέγχου, μορφοποίηση αναφορών, δημιουργία αναφορών ad hoc και ανάλυση OLAP σε μια διαδικτυακή, ολοκληρωμένη εμπειρία. Το προϊόν χρησιμοποιείται, συνήθως, σε μεγαλύτερα σενάρια που υποστηρίζουν τις ανάγκες πολλών ταυτόχρονων χρηστών καθώς και μεγάλους όγκους δεδομένων. Με έναν κύκλο ανάπτυξης συνεχούς έκδοσης (Continuous Delivery), η IBM συνεχίζει να προσφέρει νέες λειτουργίες μοντελοποίησης δεδομένων του χρήστη, δημιουργίας πινάκων ελέγχου και αναφορών σε τριμηνιαία βάση. Επίσης, οι γεωχωρικές δυνατότητες του Cognos Analytics ενισχύθηκαν με την ενσωμάτωση νέων λειτουργιών χαρτογράφησης, μέσω συνεργασιών με την Mapbox και την Pitney Bowes.



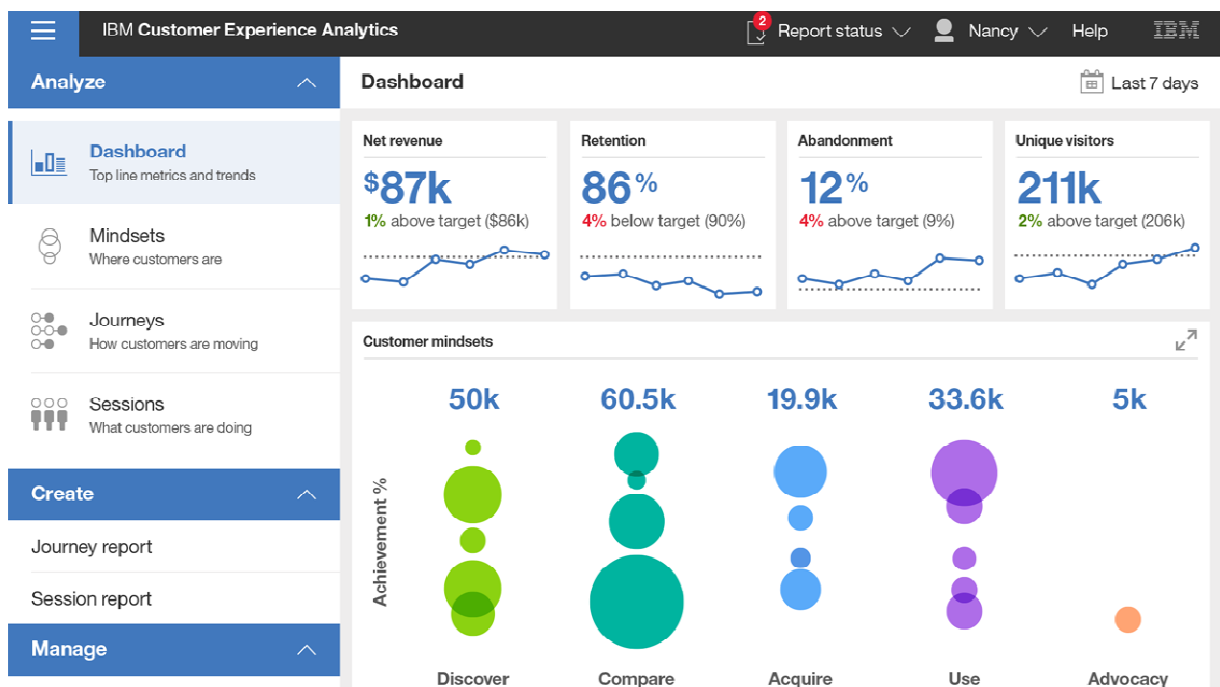
Εικόνα 10 Ο βασικός πίνακας ελέγχου του Cognos Analytics

Για προηγμένες αναλύσεις και εξόρυξη δεδομένων, η IBM προσφέρει τα προϊόντα της σειράς SPSS, συμπεριλαμβανομένων των SPSS Modeler και SPSS Statistics. Η SPSS προσφέρει καλή υποστήριξη για τη δημιουργία, τη δοκιμή και την ανάπτυξη μεμονωμένων μοντέλων εξόρυξης δεδομένων. Ωστόσο, η λύση εξόρυξης δεδομένων που προσφέρει η SPSS δεν είναι ακόμα, πλήρως ενσωματωμένη με το Cognos Analytics και το Watson Analytics. Επιπλέον, η IBM έχει κυκλοφορήσει νέα εργαλεία για την επιστήμη των δεδομένων, που βασίζονται σε φορητούς υπολογιστές, για την ανάπτυξη με τη χρήση των γλωσσών R και Python: το Watson Studio για εφαρμογές υπολογιστικού νέφους και το Data Science Experience Local για την εσωτερική ανάπτυξη. Η διασύνδεση του SPSS Modeler έχει ενσωματωθεί και στα δύο νέα εργαλεία.

Το IBM Planning Analytics, πρώην γνωστό ως TM1, το οποίο βρισκόταν στην αγορά από τις αρχές της δεκαετίας του 1980, είναι ουσιαστικά μια υψηλής απόδοσης, πολυδιάστατη βάση δεδομένων μνήμης για τον ενδοεπιχειρησιακό σχεδιασμό και την ανάλυση OLAP. Το Planning Analytics απευθύνεται σε έμπειρους χρήστες που αναπτύσσουν εφαρμογές ενδοεπιχειρησιακού σχεδιασμού και ανάλυσης. Η λύση είναι συμβατή τόσο στο Excel, όσο και

στο front-endeφαρμογών διαδικτύου και προσφέρει δυνατότητες για τη δημοσίευση περιεχομένου στον παγκόσμιο ιστό.

Το IBM WatsonAnalytics είναι ένα προϊόν που βασίζεται στο νέφος και επικεντρώνεται στην ανάλυση με βάση την αναζήτηση, την οπτική ανάλυση, την προγνωστική ανάλυση, καθώς και την ανακάλυψη δεδομένων. Αυτή η λύση απευθύνεται σε επιχειρηματικούς χρήστες και προσφέρει καλή υποστήριξη για την ανακάλυψη δεδομένων με γνωστικές δυνατότητες πρόβλεψης και μάθησης.



Εικόνα 110 κύριος πίνακας ελέγχου του Watson Analytics

8.2.3. Δυνατά σημεία

- **Επαυξημένη ανάλυση και ευκολία χρήσης:** Η IBM κατέχει ηγετική θέση στην ενσωμάτωση επαυξημένης ανάλυσης ή “έξυπνων” δυνατοτήτων στα προϊόντα της, αρχικά στο WatsonAnalytics και όλο και περισσότερο στο CognosAnalytics, με στόχο την απλοποίηση της προετοιμασίας των δεδομένων, την δημιουργία βασικών απεικονίσεων και την προηγμένη προγνωστική ανάλυση. Αυτές οι δυνατότητες διευκολύνουν τη χρήση των προϊόντων και επιτρέπουν στους επιχειρηματικούς χρήστες

και στους επιστήμονες δεδομένων να αξιοποιήσουν στο έπακρο τις αναλυτικές τους ικανότητες.

- **Πλατφόρμα "Όλα σε ένα"(All-in-one):** Η δυνατότητα δημιουργίας περιεχομένου ανάλυσης που είτε επικεντρώνεται στην πρόγνωση, είτε είναι διερευνητικό, δηλαδή περιλαμβάνει πειραματισμούς για την επίλυση νέων προβλημάτων και βελτιστοποιείται για περιοχές αβεβαιότητας, είναι η ειδοποιός διαφορά της IBM σε μια αγορά που συχνά υπαγορεύει την απαίτηση ξεχωριστών εργαλείων για την δημιουργία βασικών αναφορών και την οπτική ή την διερευνητική ανάλυση. Οι πελάτες της IBM επιλέγουν το Cognos Analytics, κυρίως για την ανώτερη λειτουργικότητά του. Η σύγχρονη, αυτή, πλατφόρμα περιλαμβάνει μέχρι και λειτουργίες χρονοπρογραμματισμού και ειδοποιήσεων, χαρακτηριστικά, τα οποία ιστορικά έχουν στερηθεί από τις σύγχρονες πλατφόρμες.
- **Παγκοσμίως γνωστός, υπεύθυνος προμηθευτής:** Η IBM έχει μια ευρεία παγκόσμια παρουσία και την ικανότητα να υποστηρίζει πελάτες σε όλες τις γεωγραφικές περιοχές, κάτι που αποτελεί σταθερό μοτίβο με τους μεγάλους προμηθευτές και σε αντίθεση με σχεδόν τους μισούς προμηθευτές του MagicQuadrant του 2018. Δείχνει επίσης την προσοχή και έχει μία ενεργή συμμετοχή σε κοινωνικές πρωτοβουλίες. Για παράδειγμα, την παροχή λογισμικού δωρεάν, ή με μειωμένο κόστος στον ακαδημαϊκό χώρο και την θέσπιση μιας εταιρικής πολιτικής για τις περιβαλλοντικές υποθέσεις.
- **Μεγάλη βάση έμπειρων χρηστών:** Η IBM έχει μια ισχυρή βάση χρηστών επιχειρηματικής ευφυΐας, πολλοί από τους οποίους μετά από διερεύνηση άλλων, πιο σύγχρονων προϊόντων, ήταν πρόθυμοι να συνεχίσουν στην πλατφόρμας της IBM, εάν η προσφορά των προϊόντων ήταν καλή. Οι υπάρχοντες πελάτες του Cognos BI μπορούν να αναβαθμίσουν στο Cognos Analytics, ως μέρος της συντήρησης, συνεπώς το κόστος των αδειών δεν αποτελεί εμπόδιο στην μετάβαση. Το WatsonAnalytics, από την άλλη πλευρά, διαθέτει άδεια ανά χρήστη, αλλά διατίθεται μόνο μέσω ψηφιακής εγγραφής. Η τιμολόγηση των προσφορών Plus, 30\$ ανά χρήστη/ανά μήνα, ή Professional, 80\$ ανά χρήστη/ ανά μήνα, είναι χαμηλότερες απ' ότι άλλων προϊόντων, με ισχυρές δυνατότητες ανάλυσης και έναν εύκολο τρόπο διερεύνησης και δοκιμής. Μέσα από τις βαθμολογίες των πελατών της IBM για το WatsonAnalytics, μπορούμε να διαπιστώσουμε πως, το χαμηλό κόστος άδειας χρήσης αποτελεί τον κύριο λόγο για την αγορά του προϊόντος.

8.2.4. Αδυναμίες

- **Αργή αναβάθμιση προϊόντων:** Το Cognos Analytics αποτελεί μια ενδιαφέρουσα και καινοτόμο πρόταση, που όμως, βάσει των αναφορών, των πελατών της Gartner, δεν έχει κερδίσει έδαφος, τόσο από νέους πελάτες, όσο και από αναβαθμίσεις από ήδη υπάρχοντες πελάτες. Πολλοί χρήστες του Cognos Analytics δεν έχουν αναβαθμίσει στην ανασχεδιασμένη έκδοση και συνεχίζουν να διερευνούν πιο σύγχρονες επιλογές από άλλους προμηθευτές. Η διακλαδισμένη στρατηγική προϊόντων της IBM συνεχίζει να προκαλεί σύγχυση και ανησυχία, όσον αφορά τη δέσμευσή της, να παρέχει μια ολοκληρωμένη και συνεκτική υπηρεσία ανάλυσης και μία πλατφόρμα επιχειρηματικής ευφυΐας.
- **Περιορισμένη πολυπλοκότητα της ανάλυσης:** Με βάση την κατανόηση της αγοράς, οι βαθμολογίες ευκολίας της χρήσης που ανέφεραν οι πελάτες της IBM περιθωριοποιούνται από τις χαμηλότερες βαθμολογίες για την πολυπλοκότητα της ανάλυσης, τοποθετώντας την IBM στο χαμηλότερο τεταρτημόριο για αυτό το χαρακτηριστικό. Το 45% των πελατών ανέφεραν ότι χρησιμοποιούν την πλατφόρμα, κυρίως για παραμετροποίηση των πινάκων ελέγχου, και εκτελούν λιγότερο συχνά πιο εξελιγμένες εργασίες, όπως η προετοιμασία δεδομένων αυτοεξυπηρέτησης, ή η διαδραστική οπτική εξερεύνηση.
- **Κενά στις δυνατότητες προϊόντων:** Παρόλο που οι πελάτες της IBM ανέφεραν την ανώτερη λειτουργικότητα ως ένα από τους λόγους για την αγορά του Cognos Analytics, οι συνολικές βαθμολογίες των προϊόντων της IBM, και για τα δύο προϊόντα, ήταν οι χαμηλότερες από οποιονδήποτε άλλο προμηθευτή στο MagicQuadrant του 2018. Πιο συγκεκριμένα, το 25% των πελατών ανέφερε ότι η αδυναμία, ή η έλλειψη λειτουργικότητας του προϊόντος αποτελεί πρόβλημα της πλατφόρμας, σε σύγκριση με τον μέσο όρο 10% για τους άλλους πωλητές. Και τα δύο προϊόντα έχουν περιορισμούς, όσον αφορά τις δυνατότητες λειτουργίας σε smartphones, καθώς και την επεκτασιμότητα και την πολυπλοκότητα των μοντέλων. Κάθε προϊόν έχει τα δικά του κενά. Για παράδειγμα, το WatsonAnalytics είναι λιγότερο ώριμο σε σχέση με το Cognos Analytics, όσον αφορά τη διαχείριση και την ασφάλεια, αλλά υποστηρίζει πιο πολύπλοκη ανάλυση με προηγμένες μεθόδους εντοπισμού προτύπων. Το Cognos Analytics είναι καλύτερο για τη διαχείριση μεταδεδωμένων σε σχέση με το WatsonAnalytics.

- **Αρνητική εμπειρία πελατών και εμπειρία πωλήσεων:** Μετά από μια βελτίωση το προηγούμενο έτος, οι βαθμολογίες για την εμπειρία των πελατών και την εμπειρία πωλήσεων ήταν χαμηλότερες φέτος. Η εμπειρία των πελατών της IBM βρισκόταν στο κατώτατο τεταρτημόριο των προμηθευτών για το MagicQuadrant του 2018, λόγω του συνολικούφέλους των επιχειρήσεων και της διαθεσιμότητας των εξειδικευμένων πόρων, που ήταν κάτω του μέσου όρου. Η εμπειρία πωλήσεων μειώθηκε ιδιαίτερα για το Cognos Analytics, τοποθετώντας την IBM στο κατώτατο τεταρτημόριο των πωλητών, παρά το γεγονός ότι το 58% των πελατών αξιολογούν την εμπειρία τους ως "εξαιρετική".

8.3.Qlik

8.3.1. Ιστορία

Η Qlik, η οποία ιδρύθηκε αρχικά το 1993 στο Lund της Σουηδίας, μετακόμισε τις έδρες της στις Ηνωμένες Πολιτείες το 2005, αφού συγκέντρωσε κεφάλαια από διάφορες εταιρείες επιχειρηματικών κεφαλαίων. Το QlikView, το ουσιαστικά άγνωστο προϊόν της εταιρείας εκείνη την εποχή, κυκλοφόρησε πολύ επιθετικά μετά την επένδυση της Σουηδικής εταιρείας VC Industrifonden. Λόγω αυτού, η Qlik κέρδισε την προσοχή επενδυτών και το 2010 μπήκε δημόσια στο χρηματιστήριο NASDAQ. Το 2016, η Qlik εξαγοράστηκε από την εταιρία ιδιωτικών μετοχών ThomaBravo και αποσύρθηκε από το χρηματιστήριο. Μέχρι την διαθεσιμότητα του Qlik Sense το 2014, η Qlik ήταν μια εταιρεία “ενός προϊόντος”. Σήμερα, ο συγκεκριμένος πωλητής παρέχει μία γκάμα λύσεων για οπτική ανάλυση και επιχειρηματική ευφυΐα.

8.3.2. Προϊόντα

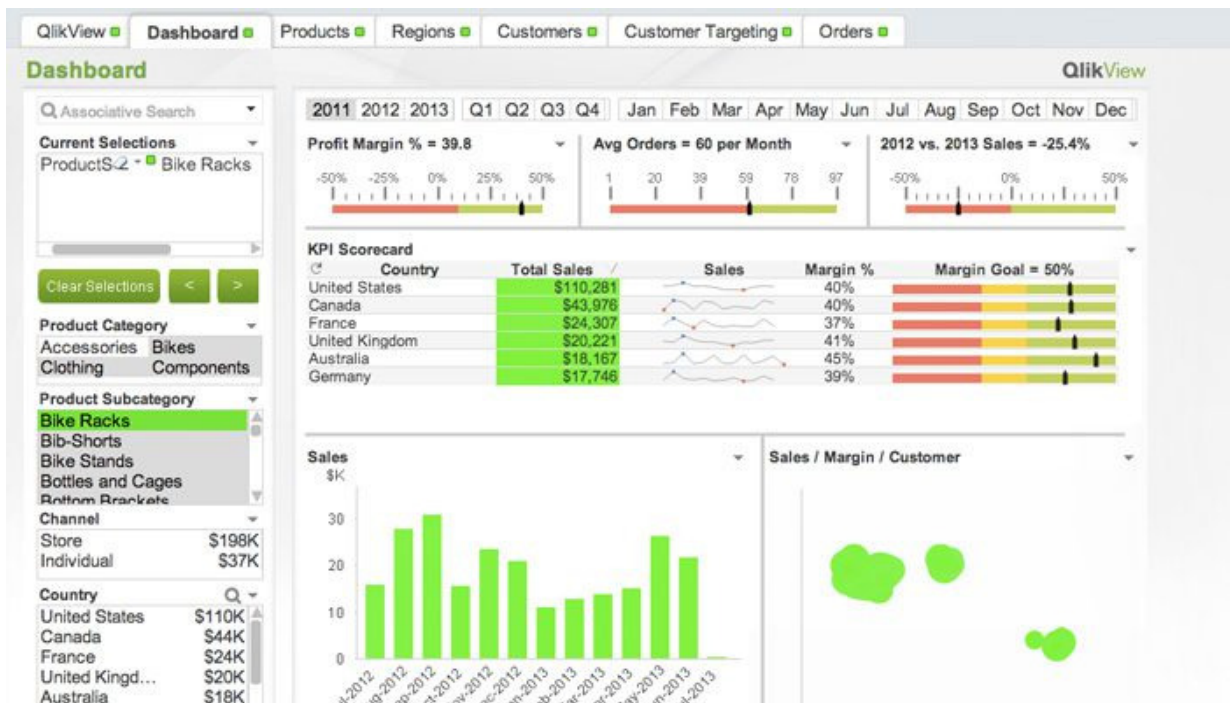
Η πλατφόρμα της Qlik αποτελείται από διάφορα προϊόντα, όπως το Qlik SenseEnterprise και το Qlik SenseCloud, καθώς και την πλατφόρμα Qlik Analytics για προγραμματιστές, το QlikView και το Qlik NPrinting, που αποκτήθηκαν μαζί με την εταιρεία Vizubi το 2015, για βελτιώσεις στην εκτύπωση και την διάταξη σελίδας.

Το Qlik Sense αποτελεί μία λύση αυτοεξυπηρετούμενης οπτικοποίησης δεδομένων και ανακάλυψης δεδομένων, που παρέχει άμεσα αποτελέσματα ανάλυσης. Λειτουργεί με τη βοήθεια του Qlik IndexingEngine(QIX), και παρέχει ευέλικτη πρόσβαση στα σύνολα δεδομένων που είναι αποθηκευμένα στην μνήμη. Η Qlik βελτίωσε σημαντικά την φιλικότητα των πρόσθετων προετοιμασίας δεδομένων, προς το χρήστη, προσθέτοντας την οπτική προετοιμασία και

παρουσίαση δεδομένων, αντικαθιστώντας τον προσαρμοσμένο κώδικα που ήταν προηγουμένως απαραίτητος. Αυτές οι βελτιώσεις αυξάνουν την παραγωγικότητα της προετοιμασίας δεδομένων για τους προγραμματιστές και επίσης καθιστούν την προετοιμασία των δεδομένων προσιτή για χρήστες με λιγότερες τεχνικές ικανότητες.

Το Qlik NPrinting είναι μια εφαρμογή δημιουργίας, διανομής και χρονοπρογραμματισμού αναφορών, η οποία μπορεί να χρησιμοποιηθεί για τη δημιουργία αναφορών που βασίζονται σε περιεχόμενο του Qlik Sense ή του QlikView. Επιτρέπει στους οργανισμούς να δημιουργούν αναφορές σε ποικίλες δημοφιλείς μορφές, συμπεριλαμβανομένων αρχείων PDF και Office.

Το QlikView είναι ένα εργαλείο δημιουργίας πινάκων ελέγχου και ανάλυσης, βασισμένο στην τεχνολογία μνήμης, που απευθύνεται σε επιχειρηματικούς χρήστες. Ωστόσο, ενώ το QlikView δίνει την δυνατότητα στους επιχειρηματικούς χρήστες να αντιμετωπίσουν τις κοινές προκλήσεις που παρατηρούνται στα συστήματα επιχειρηματικής ευφυΐας των τμημάτων Τεχνολογίας της Πληροφορικής (IT), η μεγάλη ποσότητα προγραμματισμού που χρειάζεται για τις πολύπλοκες περιπτώσεις εγκατάστασης του QlikView, απαιτεί εκπαιδευμένο προσωπικό για την παραγωγή λύσεων κλιμακούμενης και υψηλής απόδοσης.



Εικόνα 12 Ο πίνακας ελέγχου του QlikView

8.3.3. Δυνατά σημεία

- **Προσαρμόσιμα προϊόντα για ισχυρές εφαρμογές:** Οι πελάτες της Qlik χρησιμοποιούν συχνά το QlikView και το Qlik Sense ως ένα είδος αποθήκης δεδομένων, λόγω της δυνατότητας του Qlik AssociativeEngine να υποστηρίζει πολλαπλές πηγές δεδομένων, σύνθετα μοντέλα δεδομένων και περίπλοκους υπολογισμούς. Αν και η αυτοεξυπηρετούμενη ανάλυση και η επιχειρηματική ευφυΐα αντικατοπτρίζουν ένα μεγάλο μέρος της τρέχουσας αγοράς. Μέσω του scriptengine της Qlik υποστηρίζονται πολύπλοκοι μετασχηματισμοί δεδομένων από πολλαπλές πηγές δεδομένων, για κεντρική και ευέλικτη παροχή επιχειρηματικών πληροφοριών, για διαδραστικούς πίνακες ελέγχου, προς μαζική κατανάλωση.
- **Διαφοροποίηση του μάρκετινγκ:** Κατά τη διάρκεια του περασμένου έτους, η Qlik σαν εταιρεία, έχει κάνει περισσότερα για να παρέχει σαφή παραδείγματα στις επιχειρήσεις. Η βελτίωση της κατανόησης των δεδομένων, ως μέρος ενός γενικού προγράμματος για την ανάλυση και την επιχειρηματική ευφυΐα, ήταν βασικός στόχος κατά τη διάρκεια αυτής της περιόδου. Αυτός ο στόχος υπερβαίνει τις δυνατότητες των προϊόντων της, με τον προμηθευτή να χορηγεί εικονικές συνεδρίες, σεμινάρια, συναντήσεις και blogs και να παρέχει εκπαιδευτικές υπηρεσίες περί του θέματος. Επίσης, με την αύξηση του ενδιαφέροντος του “δεδομένα για το καλό”, αξίζει να σημειωθεί ότι το περιοδικό FastCompany αναγνώρισε την Qlik ως μία από τις δέκα πιο καινοτόμες εταιρείες για το κοινωνικό αγαθό. Το πρόγραμμα της εταιρείας, ChangeOurWorldCorporateSocialResponsibility (CSR), το οποίο ξεκίνησε το 2010, πλέον παρέχει δωρεάν λογισμικό και υπηρεσίες για περισσότερους από 300 μη κερδοσκοπικούς οργανισμούς.
- **Όραμα προϊόντων:** Η Qlik υπήρξε από νωρίς στην αγορά με ορισμένα στοιχεία του οράματος της για τα προϊόντα της. Η θέση της στην αγορά επιτρέπει σε συνέταιρους και συνεργάτες να αναπτύξουν περιεχόμενο για να επεκτείνουν περαιτέρω την πλατφόρμα, ή να αποφέρουν οικονομικό όφελος από ενδοεπιχειρησιακές εφαρμογές. Η επιτυχία της στην αγορά αποτέλεσε τον κύριο λόγο για την εξαγορά συγκεκριμένων εταιρειών που θα την βοηθούσαν στην πορεία του οράματος της, όπως η NPrinting και το Idevio. Πιο συγκεκριμένα, η Qlik έχει εδραιωθεί σε τομείς όπως την επαυξημένη ανάλυση, την

ικανότητα ανάλυσης τόσο τοπικών και σταθερών δεδομένων, όσο και streaming ή δεδομένων ροής, την στρατηγική πολλαπλών νεφών και τα μεγάλα δεδομένα.

- **Δίκτυο συνεργατών:** Με ένα δίκτυο, με περισσότερους από 500 συνέταιρους ενοποίησης συστημάτων πληροφορικής, και 1700 συνεργάτες σε όλο τον κόσμο, εκτιμάται πως, περίπου το 70% των υλοποιήσεων της Qlik είναι υπό την καθοδήγηση των συνεργατών της. Αυτοί οι συνεργάτες έχουν συχνά μακροχρόνιες σχέσεις με τους πελάτες τους και κατανοούν τις ιδιαίτερες απαιτήσεις τους. Επίσης, συνεισφέρουν επεκτάσεις προϊόντων, προεγκατάσταση εφαρμογών και εκπαίδευση, μέσω της αγοράς ή της κοινότητας. Οι πελάτες της Qlik σημείωσαν τη διαθεσιμότητα εμπειρών χρηστών, των προϊόντων της στην αγορά ως άνω του μέσου όρου.

8.3.4. Αδυναμίες

- **Αυτοεξυπηρετούμενη επιχειρηματική ευφυΐα:** Ένα από τα βασικά χαρακτηριστικά του QlikView ήταν η ταχεία υλοποίηση της προσέγγισης διερευνητικών εφαρμογών πινάκων ελέγχου, αντί της αυτοεξυπηρετούμενης προετοιμασίας και ανάλυσης δεδομένων. Αντίστοιχα, το Qlik Sense, στόχευε στον εκσυγχρονισμό της διεπαφής και καθιστώντας την πιο ανοιχτή και επεκτάσιμη, αλλά αποσκοπούσε επίσης να φέρει περισσότερες λύσεις στον τομέα της αυτοεξυπηρετούμενης ανάλυσης. Μέχρι σήμερα, το πρόβλημα αυτό δεν έχει αντιμετωπιστεί πλήρως. Η ευκολία με την οποία ο χρήστης μπορεί να ολοκληρώσει μια πλήρη αναλυτική ροή εργασίας, όπως πρόσβαση σε μια νέα πηγή δεδομένων, συνδυασμό πολλαπλών πηγών δεδομένων και στη συνέχεια διερεύνηση των δεδομένων αυτών με τη χρήση του Qlik Sense, είναι ασθενέστερη από αυτή των αρχικών ανταγωνιστών.
- **Κόστος λογισμικού:** Από τους πελάτες της Qlik, το 27% ανέφεραν ότι το κόστος του λογισμικού αποτελεί περιορισμό για την ευρύτερη χρήση του, και ο προμηθευτής βρίσκεται στο ανώτατο τεταρτημόριο για αυτόν τον περιορισμό, εδώ και πολλά χρόνια. Σε μια αγορά με καθοδική πίεση στην τιμολόγηση, το κόστος θα συνεχίσει να αποτελεί πρόκληση, εκτός και αν η Qlik μπορεί να διατυπώσει σαφέστερα τις ειδοποιούς διαφορές για την προστιθέμενη αξία. Επίσης, η τιμολόγηση των κύριων προϊόντων της είναι δραστικά διαφορετική. Το Qlik Sense χρησιμοποιεί ένα μοντέλο βασισμένο στον χρήστη, ενώ το QlikView έχει πιο ευέλικτες επιλογές τιμολόγησης, αλλά συχνά βασίζεται σε

διακομιστές. Όπως συμβαίνει και με αρκετούς άλλους προμηθευτές στο MagicQuadrant του 2018, η Qlik μεταβαίνει σε μοντέλο συνδρομητικής τιμολόγησης.

- **Χαμηλό μερίδιο αγοράς:** Στα πλαίσια της αγοράς, η Qlik κατείχε την πρώτη και την δεύτερη θέση σε μετρήσεις ενδιαφέροντος των πελατών της Gartner, βάσει ερωτήσεων και αναζητήσεων στην ιστοσελίδα της Gartner. Μετά το 2017, έχει πέσει στην τρίτη θέση με διαφορά. Επειδή η Qlik είναι πλέον ιδιωτική, δεν αναφέρει την αύξηση των συνολικών εσόδων της, παρόλο που ο αριθμός των απασχολούμενων παγκοσμίως μειωνόταν κατά 2,8% σε ετήσια βάση, μέχρι και το τρίτο τρίμηνο του 2017. Επιπλέον, η Qlik μείωσε κατά 10% το εργατικό δυναμικό της, τον Ιανουάριο του 2018, κυρίως στις πωλήσεις, καθώς μετατοπίζει την προσοχή της στις πωλήσεις επιχειρήσεων με υποψήφιους πελάτες πώλησης σε μικρές ή μεσαίες επιχειρήσεις. Σημεία ανησυχίας αποτελούν και οι αλλαγές στην εκτελεστική ηγεσία, και ιδιαίτερα στις θέσεις του Διευθύνων Σύμβουλου, του Διευθυντή Τεχνολογιών Πληροφορικής και Επικοινωνιών, και του Αντιπρόεδρου των Πωλήσεων.
- **Προκλήσεις μετεγκατάστασης:** Οι βαθμολογίες των πελατών της Qlik για τη συνολική υποστήριξη είναι πάνω από το μέσο όρο, με ελαφρά βελτίωση, χρόνο με το χρόνο. Ωστόσο, τα αποτελέσματά της για την ικανοποίηση με την εμπειρία μετεγκατάστασης, την τοποθετούν στο κατώτατο τεταρτημόριο των προμηθευτών, μια πτώση που συμβαίνει τα τελευταία χρόνια. Η μετεγκατάσταση μπορεί να περιλαμβάνει αναβάθμιση σε σειρές προϊόντων, καθώς και μεταφορά περιβάλλοντος από το QlikView στο Qlik Sense. Επειδή η κύρια στρατηγική της Qlik στην αγορά ήταν για τη συνύπαρξη του QlikView με το QlikSense, τα εργαλεία μετεγκατάστασης έλειπαν σε μεγάλο βαθμό, μέχρι τον Ιούνιο του 2017. Περίπου το 18% των πελατών της Qlik σχεδιάζουν να διακόψουν ή να μειώσουν τη χρήση του QlikView και αναφέρουν πως το QlikSense δεν θα είναι απαραίτητα η αντικατάσταση. Η έλλειψη απλών εργαλείων μετεγκατάστασης σημαίνει ότι οι πελάτες ενδέχεται να αλλάξουν σε άλλο προμηθευτή.

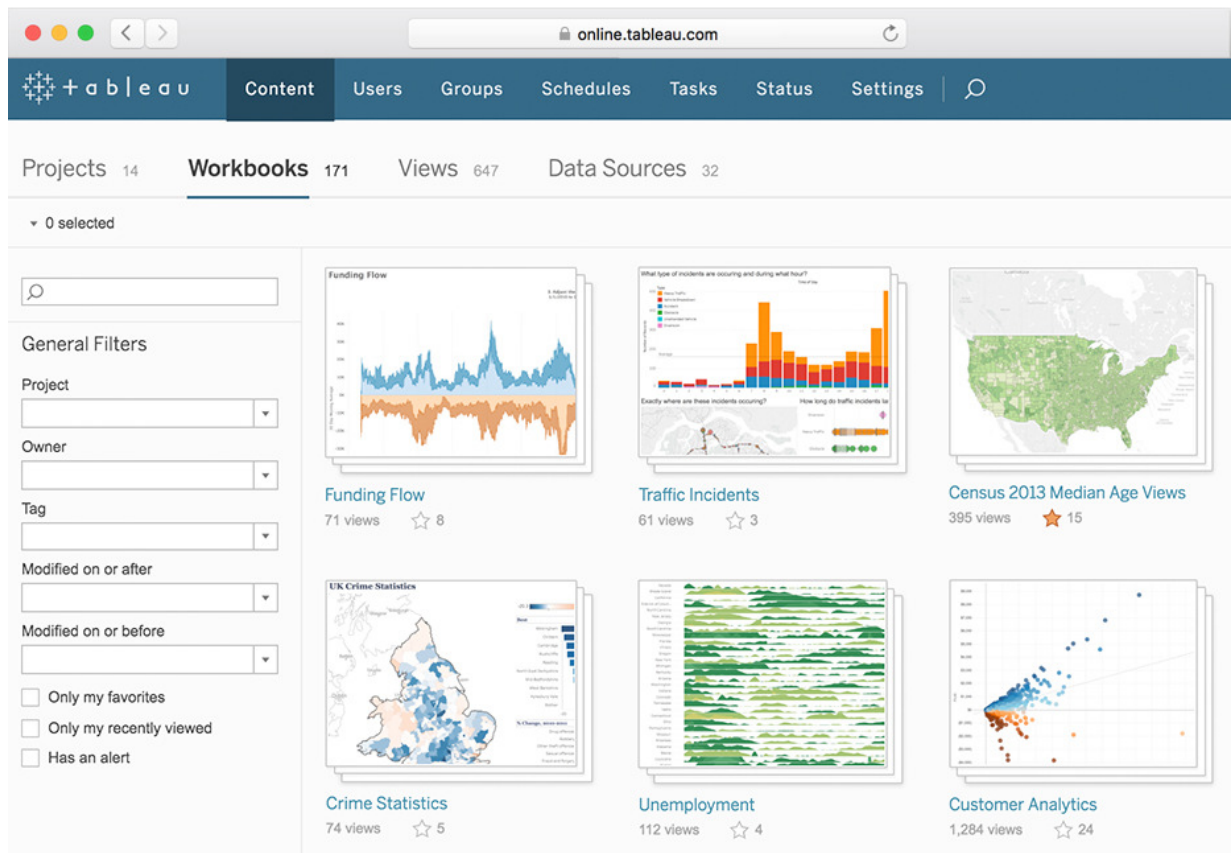
8.4. Tableau

8.4.1. Ιστορία

Η TableauSoftware προέκυψε από μία επιστημονική έρευνα στο Πανεπιστήμιο του Στάνφορντ, μεταξύ 1999 και 2002. Ιδρύθηκε στο MountainView της Καλιφόρνια. τον Ιανουάριο του 2003 από τους ChristianChabot, PatHanrahan και ChrisStolte, οι οποίοι εξειδικεύονταν σε τεχνικές οπτικοποίησης δεδομένων, για την εξερεύνηση και την ανάλυση σχεσιακών βάσεων δεδομένων και κύβων δεδομένων. Η Tableauακολουθεί μια στρατηγική παράδοσης λογισμικού που απαιτεί ελάχιστη εκπαίδευση και επιτρέπει στους επιχειρηματικούς χρήστες να ερμηνεύουν τα δικά τους δεδομένα, κυρίως μέσω διαδραστικής απεικόνισης. Το 2018, η Tableau εισήγαγε την ιδέα μιας πλατφόρμας ανάλυσης για να δώσει έμφασηστις δυνατότητές της, οι οποίες είναι συγκεντρωμένες στα προϊόντα της: TableauDesktop, TableauServer, TableauOnline και TableauPrep, τα οποία πωλούνται ξεχωριστά, ή ως πακέτα σε νέες συνδρομητικές προσφορές.

8.4.2. Προϊόντα

Το Tableau προσφέρει μία φιλική προς το χρήστη πλατφόρμα οπτικής ανάλυσης και ανακάλυψης δεδομένων που αποτελείται από: ένα desktop και ένα webclient για ανάπτυξη και συγγραφή και ένα server για κεντρική ανάπτυξη, κοινή χρήση και συνεργασία. Το διαισθητικό περιβάλλον εργασίας, η ενσωματωμένη νοημοσύνη και η χρήση της μνήμης για τη βελτιστοποίηση της απόδοσης, συμβάλλουν στη δημοτικότητα αυτής της λύσης για σενάρια οπτικής ανάλυσης, δημιουργία πινάκων ελέγχου και ανακάλυψη δεδομένων.



Εικόνα 13 Η αρχική οθόνη του Tableau

Η ανοιχτή προσέγγιση της Tableau σε διάφορες πηγές δεδομένων αποτελεί ένα από τα δυνατά του χαρακτηριστικά. Οι λύσεις που προσφέρει στους χρήστες, τους επιτρέπουν να συνδυάζουν και να αναλύουν δεδομένα από περισσότερες από 65 διαφορετικές πηγές δεδομένων. Η προετοιμασία των δεδομένων στο TableauDesktop γίνετα πιο γρήγορα, καθώς πολλές μετατροπές μπορούν να γίνουν άμεσα κατά την ανάλυση δεδομένων, επιτρέποντας μια πραγματικά επαναληπτική προσέγγιση στην ανακάλυψη δεδομένων. Με την έκδοση του 2018.1 TableauPrep, η προετοιμασία των δεδομένων έχει ενισχυθεί με βαθύτερη λειτουργικότητα και μία οπτική προσέγγιση, με προτάσεις διαμόρφωσης δεδομένων, προκαθορισμό χαρακτηριστικών και ενισχυμένη ανιχνευσιμότητα.

Το TableauServer αποτελεί ένα κεντρικό αποθετήριο για χρήστες, πηγές δεδομένων και οπτικοποιήσεις δεδομένων. Οι χρήστες μπορούν να αλληλεπιδρούν με τους πίνακες ελέγχου του διακομιστή, χωρίς καμία εγκατάσταση στην συσκευή που χρησιμοποιούν. Επίσης, το

TableauServer παρέχει αξιόπιστη και ισχυρή ασφάλεια για τους πίνακες ελέγχου, όπως και επιτρέπει στους εξουσιοδοτημένους χρήστες να κατεβάζουν και να επεξεργάζονται τους πίνακες αυτούς.

Τον τελευταίο καιρό, η Tableau έχει επικεντρωθεί έντονα στη βελτίωση και την εξέλιξη της πλατφόρμας της, για αυτοεξυπηρετούμενη ανάλυση δεδομένων, σε μια σύγχρονη πλατφόρμα επιχειρηματικής ευφυΐας. Λειτουργίες, όπως η πιστοποίηση πηγής δεδομένων, οι προτάσεις πηγών δεδομένων και μια λειτουργία που ονομάζεται TableauBridge έχουν προστεθεί στις πιο πρόσφατες κυκλοφορίες. Το TableauBridge στοχεύει στη διασύνδεση του TableauOnline με πηγές δεδομένων εσωτερικών εγκαταστάσεων για να υποστηρίξει σενάρια υβριδικών υπολογιστικών νεφών. Το 2018, η Tableau εξαγόρασε την EmpiricalSystems, μία νεοσύστατη εταιρεία που ειδικεύεται στην αυτοματοποίηση στατιστικών, για να εμπλουτίσει τα προϊόντα της με περαιτέρω χαρακτηριστικά ανάλυσης.

8.4.3. Δυνατά σημεία

- **"Χρυσό πρότυπο" για διαδραστική οπτική εξερεύνηση:** Τα βασικά πλεονεκτήματα του Tableau συνεχίζουν να είναι η διαισθητική και διαδραστική οπτικοποίηση και εξερεύνηση δεδομένων, και οι αναλυτικές ικανότητες των πινάκων ελέγχου για σχεδόν οποιαδήποτε πηγή δεδομένων, αξιοποιώντας έτσι ένα εκτεταμένο σύνολο συνδέσμων δεδομένων, τόσο με την χρήση τεχνολογιών μνήμης, όσο και με την άμεση πρόσβαση ερωτημάτων σε μεγαλύτερα σύνολα δεδομένων. Αυτός ο συνδυασμός, ο οποίος περιλαμβάνει προηγμένες λειτουργίες, όπως η πρόγνωση, η ομαδοποίηση, η αυτοματοποιημένη γεωκωδικοποίηση και η υποβοηθούμενη επεξεργασία μοτίβων, επιτρέπει στους χρήστες να πραγματοποιούν βαθύτερη εξερεύνηση και χειρισμό των δεδομένων τους, πιο εύκολα και γρήγορα από ό, τι με τις περισσότερες ανταγωνιστικές πλατφόρμες. Πιο συγκεκριμένα, οι πελάτες της Tableau συνεχίζουν να επενδύουν στα προϊόντα της, για την καλή εμπειρία του χρήστη, την ευκολία χρήσης και τη λειτουργικότητα τους, για τα οποία βαθμολογείται μεταξύ των καλύτερων προμηθευτών.
- **Εστίαση στην εμπειρία και την επιτυχία των πελατών:** Οι βαθμολογίες των πελατών τοποθετούν το Tableau για όλες τις πτυχές της εμπειρίας και των δραστηριοτήτων του πελάτη, στο κορυφαίο τεταρτημόριο των πωλητών για το MagicQuadrant του 2018. Αυτές περιλαμβάνουν, μια κορυφαία βαθμολογία για την επίτευξη επιχειρηματικών

οφελών και το υψηλότερο αποτέλεσμα για την ενεργοποίηση των τελικών χρηστών, δύο πολύ βασικοί παράγοντες της επιτυχίας. Επίσης, οι βαθμολογίες των πελατών του Tableau, το τοποθετούν στο κορυφαίο τεταρτημόριο των προμηθευτών για την ηθική και τον πολιτισμό, με το 94% των πελατών να το βαθμολογούν ως "εξαιρετικό".

- **Διεύρυνση της εφαρμογής και της ανάπτυξης:** Ένας αυξανόμενος αριθμός πελατών της Tableau, σε ποσοστό μεγαλύτερο του 55%, το οποίο είναι πάνω από του μέσου όρου, το χρησιμοποιεί για να δώσει τη δυνατότητα στους χρήστες να παρέχουν περιεχόμενο στους καταναλωτές με ευέλικτο και επαναληπτικό τρόπο. Άλλοι πελάτες αναφέρουν πως το χρησιμοποιούν για να επιτρέψουν την πλήρη αποκεντρωμένη ανάλυση από τους επιχειρηματικούς χρήστες. Ο μέσος όρος χρήσης συνεχίζει να αυξάνεται κάθε χρόνο, καθώς οι οργανισμοί καθιερώνουν το Tableau σε υψηλότερα ποσοστά και έχουν διευρύνει τη χρήση του σε όλη την επιχείρηση. Επιπλέον, οι περισσότεροι πελάτες θεωρούν ότι η πλατφόρμα της Tableau είναι είτε "ένα από τα καλύτερα" είτε "το καλύτερο" επιχειρηματικό πρότυπο.
- **Ευέλικτες επιλογές ανάπτυξης:** Το Tableau προσφέρει αρκετές επιλογές ανάπτυξης και χρήσης, είτε στο νέφος, με το TableauOnline, είτε σε τοπικά συστήματα. Αρχικά, το Tableau ήταν διαθέσιμο μόνο για τοπική εγκατάσταση, όμως σε σύντομο χρονικό διάστημα έγινε διαθέσιμο και μέσω του νέφους. Επίσης, οι επιλογές της χρήσης μέσω νέφους έχουν εξελιχθεί ώστε να παρέχουν, στο ίδιο πακέτο, εικονικές μηχανές για γνωστές πλατφόρμες υπολογιστικού νέφους, όπως την AWS και το Microsoft Azure, με στόχο την απλοποίηση της εγκατάστασης. Κατά τη διάρκεια του προηγούμενου έτους, έχει προσθέσει υποστήριξη για την πλατφόρμα Google Cloud, καθώς και υβριδική υποστήριξη δεδομένων σε τοπικές πηγές δεδομένων, μέσω του νέφους.

8.4.4. Αδυναμίες

- **Η ενσωμάτωση της επικρατούσας αγοράς:** Η οπτική εξερεύνηση δεδομένων, ενώ εξακολουθεί να αποτελεί την ειδοποιό διαφορά της Tableau, πλέον προσφέρεται σε κάποια μορφή από τους περισσότερους προμηθευτές της αγοράς. Η καθοδική πίεση των τιμών από τις άδειες χαμηλού κόστους και την πληθώρα συνδρομητικών επιλογών, έχουν επηρεάσει το ανταγωνιστικό περιβάλλον. Αυτό έχει προκαλέσει ένα, όλο και πιο πολύ, ανταγωνιστικό επιχειρηματικό περιβάλλον, και καθώς η διαφοροποίηση των χαρακτηριστικών ελαχιστοποιείται, οι ανταγωνιστικές επιλογές αυξάνονται, και η τιμή

και το πλήθος των χαρακτηριστικών που προσφέρει ένα προϊόν αποτελούν τον κύριο παράγοντα στην απόφαση αγοράς, περισσότερο απ' ό τι σε προηγούμενα έτη.

- **Τιμολόγηση και πακέτα:** Το κόστος της άδειας χρήσης λογισμικού, ιδίως επειδή οι επιλογές χαμηλού κόστους αυξάνονται και βελτιώνονται, εξακολουθεί να αποτελεί πρόκληση για την Tableau. Ένα από τα λίγα αδύνατα σημεία της Tableau, που φαίνεται και στις βαθμολογίες πελατών της, είναι το κόστος των αδειών της, με το υψηλότερο ποσοστό χρηστών να αναφέρει πως το κόστος αποτελεί περιορισμό για την ευρύτερη ανάπτυξη. Λόγω της όλο και πιο πολύ αυξανόμενης πίεσης, της αγοράς, που δεχόταν η Tableau, ανταποκρίθηκε σε αυτήν αλλάζοντας την τιμολόγηση σε συνδρομητική, παρέχοντας ξεχωριστές επιλογές για την τιμολόγηση των επιχειρήσεων και καθιερώνοντας μία πιο ευέλικτη νοοτροπία, όσον αφορά την προεξόφληση σε μεγάλες συμφωνίες.
- **Έλλειψη υποστήριξης σύνθετων μοντέλων δεδομένων:** Οι οργανισμοί αναζητούν συνεχώς, χρήσιμες επιχειρηματικές πληροφορίες, από μεγαλύτερους και πιο σύνθετους συνδυασμούς δεδομένων, που απαιτούν πιο σύνθετα μοντέλα δεδομένων. Παρόλο που αντίστοιχες λειτουργίες υποστηρίζονται από το Tableau, ένα ευρύ φάσμα επιλογών συνδεσιμότητας πηγών δεδομένων, όπως μοντέλα πολλαπλών επιπέδων, πρέπει να δημιουργηθούν είτε εκτός του Tableau, σε κάποια αποθήκη δεδομένων, είτε μέσω συνεργατών αυτοεξυπηρετούμενης προετοιμασίας δεδομένων. Ωστόσο, η Tableau ετοιμάζει ένα αυτόνομο εργαλείο προετοιμασίας δεδομένων, με το όνομα ProjectMaestro, και βρίσκεται, επί του παρόντος, σε έκδοση beta, το οποίο έχει σκοπό να διευκολύνει τους χρήστες να διαμορφώνουν και να εναρμονίζουν μεγάλα και περίπλοκα σύνολα δεδομένων χωρίς να καταφεύγουν σε άλλο εργαλείο.
- **Όραμα προϊόντων:** Η Tableau επενδύει, αλλά δεν είναι ο οδηγός, στο επόμενο κύμα ανατρεπτικών καινοτομιών. Η συντριπτική πλειοψηφία των επενδύσεων στα προϊόντα της επικεντρώνεται στην κάλυψη χασμάτων σε διάφορους τομείς, όπως την επεκτασιμότητα, την ευέλικτη ανάπτυξη συμπεριλαμβανομένου του υπολογιστικού νέφους, την υποστήριξη μεγαλύτερων και πιο σύνθετων διακυβερνωμένων συνόλων δεδομένων, και τη διευκόλυνση της οπτικής εξερεύνησης. Παρόλα αυτά, η Tableau έχει καθυστερήσει αρκετά, σε σχέση με άλλες νεοσύστατες εταιρείες και μεγαλύτερους

προμηθευτές, στις επενδύσεις, οι οποίες θα παίξουν σημαντικό ρόλο στο μέλλον της ανάλυσης δεδομένων και της επιχειρηματικής ευφυΐας.

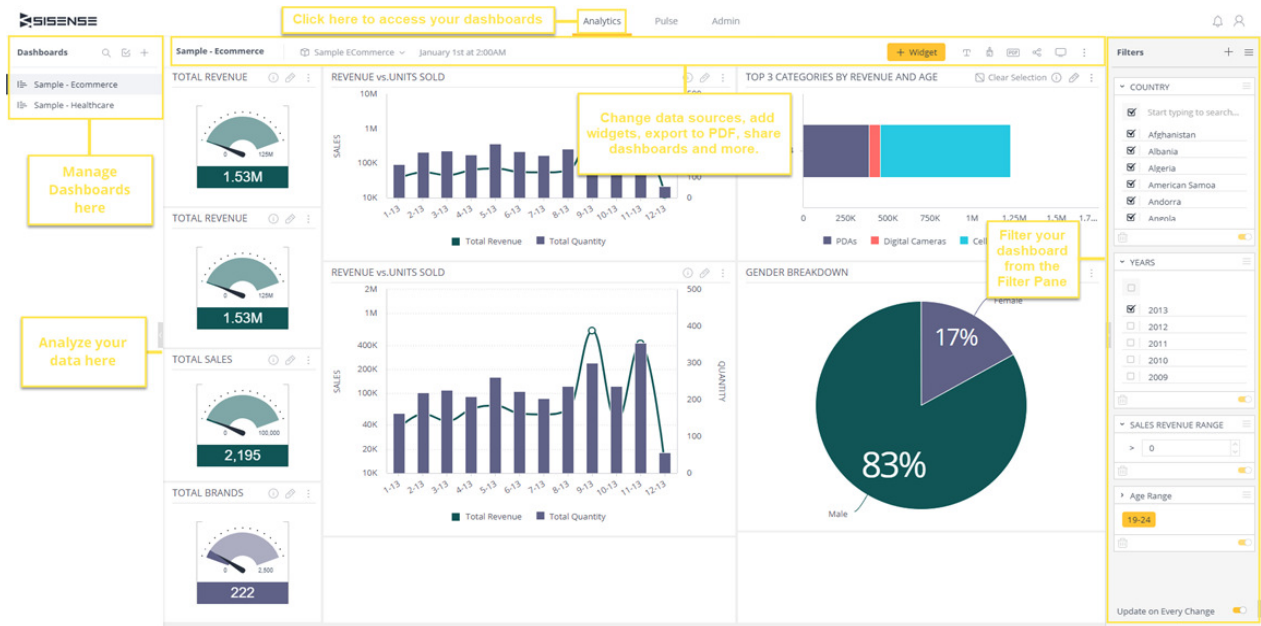
8.5.Sisense

8.5.1. Ιστορία

Με έδρα τη Νέα Υόρκη, η Sisense είναι ένας αναπτυσσόμενος πωλητής επιχειρηματικής ευφυΐας και ανάλυσης δεδομένων, που προσφέρει ένα σύγχρονο προϊόν κατάλληλο για δημιουργία πινάκων ελέγχου και ανάλυσης, καθώς και πιο εξερευνητικές περιπτώσεις χρήσης σε μια ενιαία πλατφόρμα. Η Sisense ιδρύθηκε το 2004 στο Τελ Αβίβ από τους Elad Israeli, Eldad Farkash, Aviad Harell, Guy Boyangu και Adi Azaria, και σήμερα εξυπηρετεί μεσαίες και μεγάλες εταιρείες ανά τον κόσμο, με τα προϊόντα της.

8.5.2. Προϊόντα

Το Sisense διατίθεται στο εμπόριο ως προϊόν "ενιαίας στοίβας"(single-stack), δηλαδή μια πλατφόρμα που περιλαμβάνει όλα τα χαρακτηριστικά και τη λειτουργικότητα ενός συστήματος, και προσπαθεί να απλοποιήσει την ανάλυση για σύνθετα δεδομένα και να παρέχει ένα εύχρηστο περιβάλλον πίνακα ελέγχου, όπου οι χρήστες έχουν την επιλογή να ξεκινούν με προκαθορισμένους πίνακες ελέγχου για επιλεγμένες πηγές δεδομένων ή εργαλεία μοντελοποίησης και ενσωμάτωσης δεδομένων από διαφορετικές πηγές δεδομένων. Το Sisense μπορεί να καταναλώσει δεδομένα από υπολογιστικά φύλλα, το οικοσύστημα του Hadoop, εφαρμογές ιστού και σχεσιακές βάσεις δεδομένων και να το φορτώνει στη ενσωματωμένη αποθήκη δεδομένων του, τοElastiCube, χωρίς την ανάγκη προ-συσσωμάτωσης ή προ-υπολογισμού. Το ElastiCube, παρά το όνομά του, δεν βασίζεται σε κάποια τεχνολογίαβάσης κύβου δεδομένων, αλλά στοMonetDB. Η απόδοση των ερωτημάτων βελτιώνεται στο Sisense, αυξάνοντας όλο και περισσότερο τις κρυφές μνήμες CPU και τις οδηγίες διάνυσμα για υπολογισμούς.



Εικόνα 14 Η αρχική οθόνη του Sisense

Αν και υποχρεωτική, η εκ των προτέρων μοντελοποίηση για την δημιουργία των Elasticubes, είναι απλή και πολύ ευέλικτη, καθώς έχει μειωθεί σκόπιμα σε ένα ελάχιστο απαιτούμενο. Στους επιχειρηματικούς χρήστες προσφέρονται καλές επιλογές συνδεσιμότητας και η δυνατότητα των περισσότερων μετασχηματισμών δεδομένων να γίνουν με την οπτική ανάλυση των δεδομένων. Το Elasticube Manager, το περιβάλλον μοντελοποίησης της Sisense, μεταφέρθηκε πρόσφατα σε έναν σύγχρονο web-client, ο οποίος επί του παρόντος στερείται της πλήρους συνδεσιμότητας του ισχυρού client που προορίζεται να αντικαταστήσει.

8.5.3. Δυνατά σημεία

- Βελτιώσεις προϊόντων και ισχυρό όραμα:** Το Sisense 7 προσθέτει σχετικές βελτιώσεις στο προϊόν και κλείνει αρκετά κενά. Προσφέρει καλύτερη υποστήριξη υπολογιστικού νέφους και “έξυπνων” συσκευών, βελτιωμένη αυτοεξυπηρετούμενη προετοιμασία δεδομένων και διαχειριστή Elasticube, και ενισχυμένη ενσωμάτωση ροής εργασιών, σε μια λύση που βασίζεται σε φυλλομετρητή. Εκτός από αυτές τις βελτιώσεις του προϊόντος, οι βαθμολογίες των πελατών, τοποθετούν το Sisense στο κορυφαίο τεταρτημόριο των προμηθευτών, για εύκολη χρήση στην ανάπτυξη και την κατανάλωση περιεχομένου, όπως και για την οπτική ελκυστικότητα. Από την πλευρά του προσωπικού

Τεχνολογίας της Πληροφορικής, οι αναφορές τους δείχνουν, επίσης, μια καλή προοπτική για το προϊόν, με κορυφαίες βαθμολογίες στην ευκολία χρήσης για τη διοίκηση και την εμπειρία μετεγκατάστασης. Επιπλέον, η Sisense βρίσκεται στο κορυφαίο τεταρτημόριο για το όραμά της, σχετικά με τη στρατηγική των προϊόντων της, καθώς στοχεύει σε ένα μέλλον με περισσότερες καινοτόμες δυνατότητες.

- **Ισχυρός Κατασκευαστής Ανάλογου Εξοπλισμού(OEM):** Περίπου το 53% των πελατών της Sisense αξιοποιούν το προϊόν λόγω του κατασκευαστή και λόγω της ενσωματωμένης ανάλυσης δεδομένων, και αποτελεί το υψηλότερο ποσοστό από όλους τους προμηθευτές που αξιολογήθηκαν για το MagicQuadrant του 2018. Αυτή η θέση προέρχεται από ένα ισχυρό και εξειδικευμένο δίκτυο συνεργατών, που αξιοποιεί το προϊόν για την παροχή λύσεων ανάλυσης δεδομένων και τη δημιουργία εσόδων μέσω δεδομένων των πελατών. Ο συγκεκριμένος προμηθευτής συνεργάζεται επίσης με άλλους κατασκευαστές, για την κατάλληλη τιμολόγηση και παροχή πακέτων ανάλογες με τις περιπτώσεις χρήσης, κάτι που αποτελεί ειδοποιό διαφορά για την Sisense, στη σύγχρονη αγορά. Επίσης, το 2017, η Sisense ξεκίνησε το πρόγραμμα επιτάχυνσης νεοσύστατων επιχειρήσεων Sisense Startup, για να παραχωρήσει το λογισμικό της δωρεάν, για ένα χρόνο, σε νεοφερμένους χρήστες, που επιθυμούν να αναπτύξουν λύσεις με το προϊόν της. Μέχρι το τέλος του 2017, περισσότεροι από 700 οργανισμοί είχαν εγγραφεί σε αυτό το πρόγραμμα.
- **Ισχυρή δυναμική στην αγορά, που υποστηρίζεται από σταθερή εμπειρία και λειτουργίες των πελατών:** Οι βαθμολογίες των πελατών τοποθετούν την Sisense στο κορυφαίο τεταρτημόριο των προμηθευτών, για την εμπειρία και τις λειτουργίες των πελατών, συμπεριλαμβανομένης της υποστήριξης, της ποιότητας των προϊόντων και της κορυφαίας εμπειρίας μετεγκατάστασης. Αυτή η ικανότητα βοηθά την εταιρεία να διατηρήσει μια ισχυρή δυναμική της αγοράς. Τα συνολικά έσοδα του συγκεκριμένου προμηθευτή αυξήθηκαν κατά 70% για το 2017, με 1.000 νέους χρήστες. Επιπλέον, οι βαθμολογίες των πελατών έθεσαν τη Sisense στο κορυφαίο τεταρτημόριο για την ηθική, τον πολιτισμό και την ποικιλομορφία του, με το 100% των αναφορών να αξιολογούν τον πωλητή ως "εξαιρετικό".
- **Επιχειρηματικό πρότυπο σε μεσαίες επιχειρήσεις:** Η Sisense θεωρείται το επιχειρηματικό πρότυπο, με βάση το 77% των πελατών της, που αποτελεί, επίσης, το

κορυφαίο αποτέλεσμα για τους προμηθευτές σε αυτό το MagicQuadrant. Αντί το προϊόν της να είναι μια λύση για πολύ συγκεκριμένες περιπτώσεις χρήσης που πωλούνται ως μια υπηρεσία, οι πελάτες υιοθετούν τη λύση ως το κύριο, και σε πολλές περιπτώσεις, αποκλειστικό εργαλείο ανάλυσης που χρησιμοποιείται για ένα ευρύ φάσμα χρήσεων. Αυτό το φάσμα των δυνατοτήτων που προσφέρονται, μέσω μίας ενοποιημένης ροής εργασιών βασισμένες στο διαδίκτυο, συμβάλει στη στήριξη της διεξόδου της Sisense, στην αγορά.

8.5.4. Αδυναμίες

- **Μέγεθος ανάπτυξης:** Αν και αποτελεί επιχειρηματικό πρότυπο, το μέσο μέγεθος των αναπτύξεων της Sisense είναι λίγο πάνω από 300 χρήστες. Αυτό την τοποθετεί στο κάτω τεταρτημόριο των προμηθευτών του MagicQuadrant και πιθανότατα οφείλεται από το ποσοστό χρήσης της σε μικρότερες οργανώσεις. Η Sisense έχει σχετικές αναφορές από μεγάλες εγκαταστάσεις, αλλά οι οργανισμοί που σχεδιάζουν να επεκταθούν πέρα από μερικούς εκατοντάδες χρήστες, συνήθως συζητούν με προμηθευτές που να ανταποκρίνονται στις απαιτήσεις κλιμάκωσης τους.
- **Μικρός όγκος δεδομένων:** Κατά την αξιολόγηση του όγκου των αναλυτικών δεδομένων, οι πελάτες της Sisense αναφέρουν όγκους δεδομένων μεταξύ των μικρότερων σε αυτό το MagicQuadrant. Επιπλέον, μετά την φόρτωση των δεδομένων στον engine του Sisense, οι περισσότεροι πελάτες αναλύουν λιγότερο από 50GB σε μια ενιαία εφαρμογή. Αυτά τα αποτελέσματα δεν ευθυγραμμίζονται με την ειδοποιό διαφορά του προϊόντος, το οποίο είναι ικανό να χειρίζεται πολύ μεγάλα, σύνθετα σύνολα δεδομένων και να παρέχει υψηλή απόδοση. Οι περισσότερες εφαρμογές φαίνεται να βασίζονται σε μπλοκ μηνυμάτων διακομιστή (ServerMessageBlocks - SMB), εγγενώς λιγότερο απαιτητικές, οι οποίες συνήθως δεν χρειάζεται να αντέξουν την ανάλυση μεγάλου όγκου δεδομένων. Παρόλο που δεν έχουν εντοπιστεί προβλήματα απόδοσης από τους πελάτες, οι οργανισμοί που απαιτούν ανάλυση σε σημαντικά μεγαλύτερους όγκους δεδομένων θα πρέπει να δοκιμάσουν τις δυνατότητες κλιμάκωσης, πριν δεσμευτούν στο προϊόν.
- **Περιορισμοί προϊόντος:** Το Sisense 7.0 (Οκτώβριος 2017) αντιμετώπισε αρκετούς περιορισμούς που εντοπίστηκαν από τους πελάτες, παρόλα αυτά παραμένουν αρκετοί περιορισμοί στο προϊόν. Οι δυνατότητες διαδραστικής οπτικοποίησης, που αποτελούν

ένα βασικό παράγοντα αγοράς, βαθμολογήθηκαν ως μόνο "καλές", και η βαθμολογία αυτή, είναι χαμηλότερη από πολλούς από τους άμεσους ανταγωνιστές της Sisense. Επίσης, η επαυξημένη ανακάλυψη δεδομένων, μια βασική δυνατότητα για πλατφόρμες της επόμενης γενιάς, αντιπροσωπεύεται ελάχιστα στην πλατφόρμα της Sisense, ιδίως σε τομείς όπως η προετοιμασία δεδομένων.

- **Περιορισμένη γεωγραφική κάλυψη:** Η γεωγραφική κάλυψη της Sisense είναι περιορισμένη, με κέντρα υποστήριξης και επιχειρήσεις στις ΗΠΑ, το Ισραήλ και το Κίεβο της Ουκρανίας. Στη συνέχεια, η Sisense αξιοποιεί το δίκτυο συνεργατών της, το οποίο είναι άνω των 400, συμπεριλαμβανομένων των μεταπωλητών και των κατασκευαστών ανάλογου εξοπλισμού(OEM), για άλλες περιοχές. Αυτή η έλλειψη άμεσης τοπικής παρουσίας μπορεί να αποτελέσει πρόκληση για ορισμένους πιθανούς πελάτες.

9. Επιχειρηματική Ευφυία & Λογισμικό Ανοιχτού Κώδικα

9.1.Τί είναι το λογισμικό ανοιχτού κώδικα

Με τον όρο “λογισμικό ανοιχτού κώδικα”, εννοείται λογισμικό του οποίου ο πηγαίος κώδικας είναι ελεύθερος για να επιθεωρηθεί, να τροποποιηθεί και να βελτιωθεί από τρίτους.

Το λογισμικό ανοιχτού κώδικα δεν σημαίνει απαραίτητως δωρεάν ή ελεύθερο λογισμικό, αλλά αναφέρεται στο γεγονός πως επιτρέπεται σε κάθε χρήστη να εξετάσει και να χρησιμοποιήσει τη γνώση και τις δυνατότητες που προσφέρει ο παρεχόμενος πηγαίος κώδικας. Στην πράξη, τα περισσότερα προγράμματα ανοιχτού κώδικα παρέχονται δωρεάν και μπορούν να χαρακτηριστούν ελεύθερα.

9.2.Πλεονεκτήματα

- **Κοινότητα:** Οι λύσεις ανοιχτού κώδικα που προσανατολίζονται προς τις επιχειρήσεις έχουν συχνά αναπτυσσόμενες κοινότητες γύρω τους, οι οποίες δεσμεύονται από μια κοινή προσπάθεια να υποστηρίξουν και να βελτιώσουν μια λύση που ωφελεί τόσο την ίδια την κοινότητα, όσο και την επιχείρηση.
- **Διαφάνεια:** Ο πηγαίος κώδικας ανοιχτού λογισμικού είναι ανά πάσα στιγμή διαθέσιμος για πλήρη προβολή και λήψη από οποιονδήποτε, καθώς και οι συζητήσεις για το πως η κοινότητα αναπτύσσει λειτουργίες και αντιμετωπίζει σφάλματα. Αντίθετα, ο κλειστός

κώδικας, παράγεται με μυστικότητα και μπορεί να έρθει με απρόβλεπτους περιορισμούς και άλλες ανεπιθύμητες εκπλήξεις.

- **Αξιοπιστία:** Λόγω των περισσότερων ατόμων που δουλεύουν πάνω σε έργα ανοιχτού κώδικα, συνήθως η αξιοπιστία του είναι ανώτερη. Με μια παγκόσμια κοινότητα που υποστηρίζει μια βάση κώδικα, αντί για μια ομάδα εντός μίας εταιρείας, ο ανοιχτός κώδικας αναπτύσσεται σε διαδικτυακά φόρουμ, και καθοδηγείται από εμπειρογνώμονες και ειδικούς.
- **Ασφάλεια:** Όπως συμβαίνει και με την αξιοπιστία, ο πηγαίος κώδικας των λογισμικών ανοικτού κώδικα είναι συχνά πιο ασφαλής, διότι εξετάζεται πολύ πιο λεπτομερώς και ελέγχεται από την κοινότητα και τα τυχόν προβλήματα που προκύπτουν τείνουν να διορθώνονται με μεγαλύτερη επιμέλεια.
- **Χαμηλό κόστος:** Αν και οι λύσεις ανοιχτού κώδικα θα πρέπει να θεωρηθούν ως κάτι περισσότερο από ελεύθερο λογισμικό, το γεγονός ότι δεν απαιτούνται αμοιβές αδειοδότησης, παραμένει ένα σημαντικό πλεονέκτημα όταν εξετάζουμε το συνολικό κόστος ανάπτυξης μιας λύσης.

9.3.Μειονεκτήματα

- **Δυσκολία χρήσης:** Ορισμένες εφαρμογές ανοιχτού κώδικα ενδέχεται να είναι πιο δύσκολες ως προς την εγκατάσταση και τη χρήση. Άλλες ενδέχεται να μην έχουν φιλικές διεπαφές, ή χαρακτηριστικά που να είναι γνώριμα προς τον χρήστη. Αυτό μπορεί να επηρεάσει την παραγωγικότητα και να αποτρέψει τον τελικό χρήστη να υιοθετήσει ή να χρησιμοποιήσει τα προγράμματα με ευκολία.
- **Προβλήματα συμβατότητας:** Πολλοί τύποι εμπορικού υλικού χρειάζονται εξειδικευμένους οδηγούς για την εκτέλεση προγραμμάτων ανοιχτού κώδικα, τους οποίους, συχνά, διατίθενται μόνο από τον κατασκευαστή του εξοπλισμού. Αυτό μπορεί ενδεχομένως να αυξήσει το κόστος ενός έργου. Ακόμη και αν υπάρχει το απαραίτητο πρόγραμμα οδήγησης ανοιχτού κώδικα, ενδέχεται να μην λειτουργεί τόσο σωστά με το λογισμικό, όσο με ένα εμπορικό πρόγραμμα οδήγησης.
- **Ευθύνη και εγγυήσεις:** Με το εμπορικό λογισμικό, ο προμηθευτής, συνήθως, παρέχει αποζημίωση και εγγύηση στο πλαίσιο προδιαγραφών μίας τυπικής άδειας χρήσης. Αυτό οφείλεται στο γεγονός ότι έχουν πλήρη έλεγχο και δικαιώματα πνευματικής ιδιοκτησίας

πάνω στο προϊόν και τον υποκείμενο κωδικό του. Οι συμβατικές άδειες λογισμικού ανοιχτού κώδικα περιέχουν τυπικά μόνο περιορισμένη εγγύηση και καμία ευθύνη ή προστασία αποζημίωσης λόγω παραβίασης.

- **Κρυφές δαπάνες:** Λογισμικό που είναι ελεύθερο, εκ των προτέρων, αλλά αργότερα κοστίζει χρήματα για να συνεχίσει η λειτουργία του, μπορεί να είναι ένα μεγάλο βάρος, ειδικά αν δεν έχει εξετασθεί αυτό το κρυφό κόστος από την αρχή.

9.4. Προμηθευτές επιχειρηματικής ευφυΐας ανοιχτού κώδικα

9.4.1. Jaspersoft Community

9.4.1.1. Ιστορία

Η Jaspersoft αρχικά ονομαζόταν Panscorpic και ιδρύθηκε από τον Al Campa, Διευθύνοντα Σύμβουλο, και τον Raj Bhargava, Αντιπρόεδρο Προϊόντων το 2001. Η Panscorpic συγκέντρωσε 23 εκατομμύρια δολάρια από τις εταιρείες DollCapital, DiscoveryVentures, MorgenthalerVentures και Partech, και το 2004 η Panscorpic συνεργάστηκε με τον Teodor Danciu, ο οποίος απέκτησε την πνευματική ιδιοκτησία της και άλλαξε το όνομα της εταιρείας σε Jaspersoft. Στις 28 Απριλίου 2014, η TIBCO ανακοίνωσε ότι απέκτησε την Jaspersoft για περίπου 185 εκατομμύρια δολάρια.

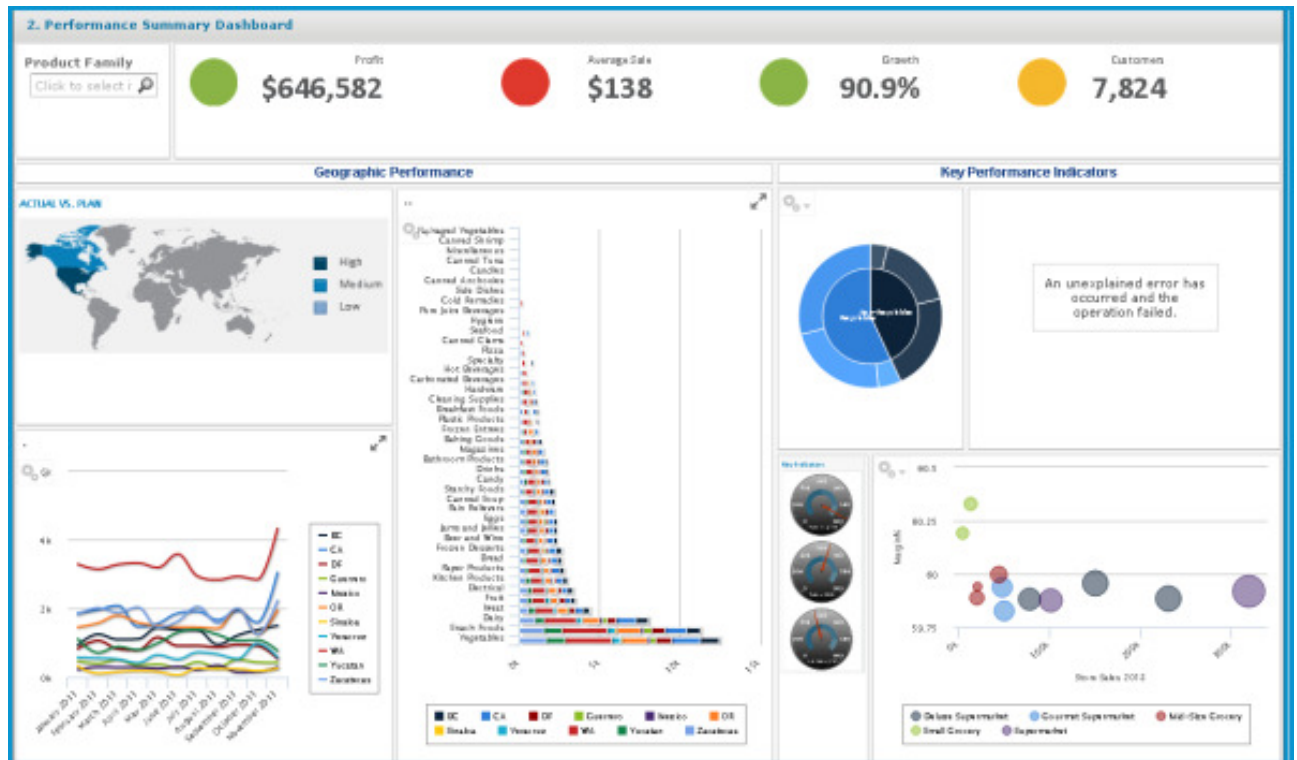
9.4.1.2. Προϊόντα

Το JasperReports είναι ένα από τα πιο δημοφιλή και ευρέως χρησιμοποιούμενα εργαλεία αναφορών ανοιχτού κώδικα. Χρησιμοποιείται σε εκατοντάδες χιλιάδες περιβάλλοντα παραγωγής και διαθέτει εκδόσεις που υποστηρίζονται από την κοινότητα και για εμπορική χρήση.

Το JasperReports αποτελείται από διάφορα στοιχεία, όπως:

- **JasperReports Server:** Ένα πρόγραμμα για το σχεδιασμό αναφορών που μπορεί να ενσωματωθεί ή να χρησιμοποιηθεί ανεξάρτητα.
- **JasperReports Library:** Μια βιβλιοθήκη γραμμένη αποκλειστικά σε Java, που επιτρέπει στο χρήστη να χρησιμοποιεί δεδομένα από οποιαδήποτε πηγή και να τα εξάγει σε αναφορές, σε διάφορες μορφές, όπως HTML, PDF, Excel και άλλες.
- **Jaspersoft ETL:** Μία μηχανή ενσωμάτωσης δεδομένων που μετατρέπει τα πρωτογενή δεδομένα από διάφορες πηγές, σε επιχειρηματικές πληροφορίες προς κατανάλωση.

- **Jaspersoft Studio:** Ένα πρόγραμμα, βασισμένο στο Eclipse, που επιτρέπει τον σχεδιασμό και την προσαρμογή αναφορών που θα ενσωματωθούν στο JasperReportsServer.



Εικόνα 15 Η κύρια οθόνη του JasperSoft

Το JasperReports υποστηρίζεται από άριστη τεκμηρίωση, μία σελίδα wiki και πρόσθετους πόρους. Λόγω του ότι είναι γραμμένο σε Java, μπορεί να εκτελείται σε Windows, Linux και Mac και είναι διαθέσιμο για δωρεάν λήψη. Η τελευταία έκδοση κυκλοφορεί από τον Δεκέμβριο του 2015 και είναι αδειοδοτημένη από την AfferoGeneralPublicLicense(AGPL).

9.4.2. Birt

9.4.2.1. Ιστορία

Το BIRT είναι μια πλατφόρμα τεχνολογίας ανοιχτού κώδικα που χρησιμοποιείται για τη δημιουργία οπτικοποιήσεων δεδομένων και αναφορών. Το έργο BIRT προτάθηκε για πρώτη φορά και χρηματοδοτήθηκε από την Actuate Corporation, όταν η Actuate προσχώρησε στο Eclipse Foundation, στις 24 Αυγούστου 2004. Στη συνέχεια, εγκρίθηκε και έγινε ένα από τα

κορυφαία έργα στην κοινότητα της Eclipse, στις 6 Οκτωβρίου 2004. Η κοινότητα που συμβάλλει στο έργο περιλαμβάνει κολοσσούς, όπως την IBM και την InnoventSolutions.

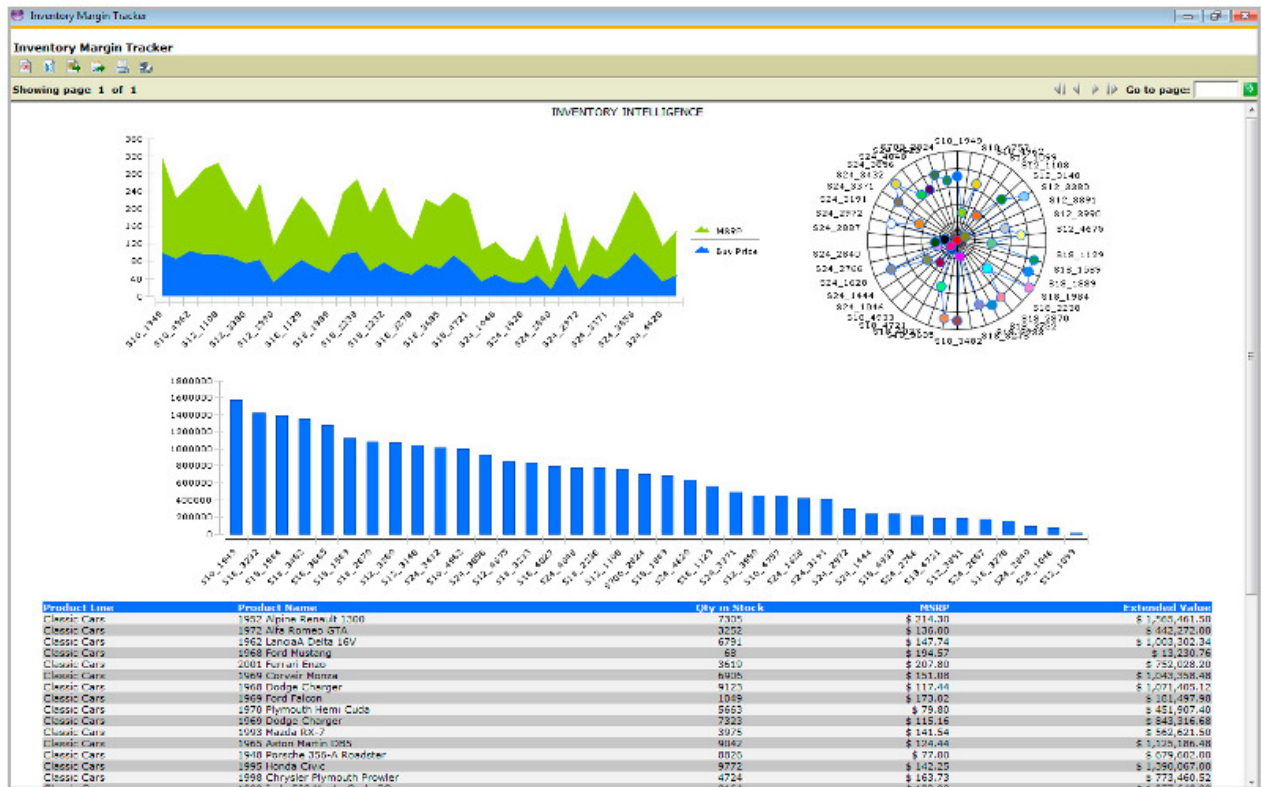
9.4.2.2. Προϊόντα

Το BIRT αποτελείται από πολλά στοιχεία. Τα βασικά συστατικά περιλαμβάνουν:

- **BIRT ReportDesigner:** Ένα πρόγραμμα για τη δημιουργία και το σχεδιασμό αναφορών. Τα σχέδια αυτά αποθηκεύονται σε ανοιχτή μορφή τύπου XML.
- **BIRT Runtime:** Ένα περιβάλλον εκτέλεσης για τη δημιουργία των σχεδίων αναφοράς που μπορεί να αναπτυχθεί σε οποιοδήποτε περιβάλλον της Java.

Επίσης, το BIRT περιέχει τρία επιπλέον στοιχεία:

- **ReportEngine:** Μία μηχανή που χρησιμοποιεί τα αρχεία σχεδιασμού αναφορών για την παραγωγή και τη δημιουργία αναφορών. Χρησιμοποιώντας το ReportEngine API μπορεί να ενσωματωθεί σε οποιαδήποτε εφαρμογή Java / Java EE.
- **ChartingEngine:** Μία μηχανή που χρησιμοποιείται για τον σχεδιασμό και τη δημιουργία διαγραμμάτων, είτε αυτόνομα, είτε ενσωματωμένα στο BIRT Reports. Το ChartingEngine API επιτρέπει στους προγραμματιστές Java/Java EE να προσθέσουν δυνατότητες δημιουργίας διαγραμμάτων στις εφαρμογές τους.
- **BIRT Viewer:** Το BIRT παρέχει ένα δείγμα προγράμματος επισκόπησης, που χρησιμοποιείται για την προεπισκόπηση αναφορών μέσα στο Eclipse.



Εικόνα 16 Η κύρια οθόνη του BIRT

Το BIRT είναι γραμμένο σε Java και είναι αδειοδοτημένο βάσει της δημόσιας άδειας της Eclipse και η πιο πρόσφατη έκδοση του είναι διαθέσιμη σε Windows, Linux και Mac.

9.4.3. Pentaho

9.4.3.1. Ιστορία

Το Pentaho είναι ένα λογισμικό επιχειρηματικής ευφυΐας που παρέχει λειτουργίες, όπως ενσωμάτωση δεδομένων, υπηρεσίες OLAP, δημιουργία αναφορών, πίνακες ελέγχου πληροφοριών, εξόρυξη δεδομένων και δυνατότητες εξαγωγής, μετασχηματισμού και φόρτωσης (ETL). Η Pentaho έχει την έδρα της στο Ορλάντο, Φλόριντα. Αποκτήθηκε από τη HitachiData Systems το 2015, και τον Σεπτέμβριο του 2017, έγινε μέρος της HitachiVantara, μιας νέας εταιρείας που ενοποιεί τις λειτουργίες των Pentaho, HitachiDataSystems και HitachiInsightGroup.

9.4.3.2. Προϊόντα

Η Pentaho προσφέρει μια ολοκληρωμένη σουίτα επιχειρηματικής ευφυίας, η οποία καλύπτει μια σειρά από περιπτώσεις χρήσης, από την δημιουργία αναφορών έως και την εξόρυξη δεδομένων. Η σουίτα Pentaho BI περιλαμβάνει πολλά έργα ανοικτού κώδικα, από τα οποία τα πιο βασικά είναι:

- **PentahoReporting:** Αποτελεί, το ίδιο, μία σουίτα εργαλείων ανοιχτού κώδικα που επιτρέπει στους χρήστες να δημιουργήσουν αναφορές των δεδομένων τους σε μορφή PDF, Excel, HTML, απλού κειμένου, αρχείου RichText, XML και CSV.
- **PentahoServer:** Πρόκειται για μια εφαρμογή J2EE που παρέχει υποδομή για την εκτέλεση και την προβολή αναφορών μέσω διεπαφής χρήστη που βασίζεται στον ιστό.



Εικόνα 17 Ο βασικός πίνακας ελέγχου του Pentaho

Το Pentaho υποστηρίζεται από πολλούς πόρους της κοινότητας, όπως τεκμηριώσεις, σελίδα στο wiki και άλλες. Το εργαλείο λειτουργεί με τη χρήση Java EnterpriseEdition και μπορεί να χρησιμοποιηθεί σε Windows, Linux και Mac. Η πιο πρόσφατη σταθερή έκδοση της είναι η

έκδοση 8.1 από τον Μάιο του 2018 και είναι αδειοδοτημένο από την GeneralPublicLicense(GPL).

10. Σύνοψη

Με την αυξανόμενη σημασία των δεδομένων, όχι μόνο να υποστηρίζουν τις επιχειρηματικές αποφάσεις, αλλά και να αυξάνουν την αποδοτικότητα και την αποτελεσματικότητα των επιχειρησιακών διαδικασιών, καθώς και να υποστηρίζουν τον αυξανόμενο αριθμό προϊόντων, υπηρεσιών και επιχειρηματικών μοντέλων που βασίζονται σε δεδομένα, πλέον, μία σύγχρονη πλατφόρμα επιχειρηματικής ευφυΐας είναι απαραίτητη για κάθε επιχείρηση που επιθυμεί να επιτύχει την προσαρμογή της, στην ψηφιοποίηση των αγορών.

Η αγορά επιχειρηματικής ευφυΐας είναι μία διαρκώς αυξανόμενη και ιδιαίτερα ανταγωνιστική αγορά. Απαρτίζεται τόσο από μεγάλους οργανισμούς και εταιρείες, με χρόνια εμπειρίας και μακρύ ιστορικό επιτυχιών, όσο και από αμέτρητους μικρότερους προμηθευτές, με εναλλακτικές και μερικές φορές καινοτόμες προσεγγίσεις στις λύσεις επιχειρηματικής ευφυΐας. Συνολικά, η αγορά ανάλυσης δεδομένων και επιχειρηματικής ευφυΐας, εκτιμάται ότι θα αυξηθεί κατά 10% το 2018, με αναμενόμενη αύξηση 8% έως το 2021, όπως αντικατοπτρίζεται στην τρέχουσα εκτίμηση της Gartner.

Κατά τη διάρκεια του έτους οι κύριες τάσεις που κυριάρχησαν στην επιχειρηματική ευφυΐα, μεταξύ άλλων, ήταν η διαχείριση κύριων δεδομένων και ποιότητας δεδομένων, η ανακάλυψη και οπτικοποίηση δεδομένων, η αυτοεξυπηρετούμενη επιχειρηματική ευφυΐα, η διακυβέρνηση δεδομένων και η ανάλυση σε πραγματικό χρόνο. Με την αύξηση της πολυπλοκότητας στο περιβάλλον της επιχειρηματικής ευφυΐας, ο προσδιορισμός των τάσεων και των εξελίξεων στην αγορά αποτελεί βασικό παράγοντα για την αποτελεσματική λήψη αποφάσεων.

Στα πλαίσια της πτυχιακής εργασίας εξετάστηκαν αναλυτικά μερικοί από τους προμηθευτές εμπορικών προϊόντων επιχειρηματικής ευφυΐας, όπως και προμηθευτές επιχειρηματικής ευφυΐας ανοιχτού κώδικα. Καθώς, τα πλεονεκτήματα και τα μειονεκτήματα των λογισμικών ανοιχτού κώδικα έχουν αναλυθεί συνοπτικά σε παραπάνω κεφάλαιο, παρακάτω παρατίθεται συγκριτικός πίνακας των προμηθευτών εμπορικών προϊόντων:

Συγκριτικός Πίνακας Προμηθευτών Επιχειρηματικής Ευφυίας

Microsoft

- + Δυνατή σειρά προϊόντων για δημιουργία και μορφοποίηση αναφορών, δημιουργία adhoc αναφορών και πινάκων ελέγχου, και ανάλυση δεδομένων
 - + Το SQLServer είναι ευρέως γνωστό σύστημα διαχείρισης βάσεων δεδομένων που αποτελείται από σχεσιακή αποθήκευση δεδομένων, μοντελοποίηση OLAP, υποστήριξη χωρικών δεδομένων και ενσωματωμένη εξόρυξη δεδομένων
 - + Το Excel είναι ένα πολύ γνωστό και ευρέως χρησιμοποιούμενο εργαλείο ανάλυσης και αυτοεξυπηρετούμενης επιχειρηματικής ευφυίας
 - + Το PowerBI, το οποίο αποτελεί τη λύση οπτικής ανακάλυψης και επιχειρηματικής ευφυίας της Microsoft, προσφέρεται σε ελκυστική τιμή, γεγονός που καθιστά εύκολο για οργανισμούς και μεμονωμένους χρήστες να ξεκινήσουν με αυτό.
 - + Εκτεταμένο δίκτυο επιχειρηματικών συνεργατών, που παρέχουν τις δυνατότητες που προσφέρει η Microsoft σε όλο τον κόσμο.
-
- Τα προϊόντα της σειράς Power BI είναι ενσωματωμένα μόνο σε επίπεδο δεδομένων και δεν διαθέτουν κεντρικό αποθετήριο μεταδεδομένων.
 - Ορισμένα εργαλεία έχουν επικαλυπτόμενες λειτουργίες. Αυτό θα μπορούσε να προκαλέσει σύγχυση σε όσους θέλουν να αξιολογήσουν τα προϊόντα με βάση τις περιπτώσεις χρήσης τους.
 - Τα θέματα χρονοπρογραμματισμού και διαχείρισης εταιρικής απόδοσης καλύπτονται από λύσεις συνεργατών
-

IBM

- + Το εκτεταμένο χαρτοφυλάκιο προϊόντων της, περιλαμβάνει το CognosAnalytics για δημιουργία και μορφοποίηση adhoc αναφορών, ανάλυση OLAP, οπτική και προηγμένη ανάλυση και δημιουργία πίνακα ελέγχου, και το PlanningAnalytics για το προγραμματισμό και τη διαχείριση της απόδοσης.
 - + Ευρείες δυνατότητες για εξόρυξη δεδομένων και προηγμένη ανάλυση με SPSS και DSX, καθώς και οπτική ανακάλυψη και επιχειρηματική ευφυία με το IBM WatsonAnalytics.
 - + Πολλαπλές επιλογές ανάπτυξης, όπως στο νέφος, τοπικές εγκαταστάσεις ή υβριδικά μοντέλα.
 - + Συνεχόμενες επενδύσεις και καινοτομίες στην επιχειρηματική ευφυία, προηγμένη ανάλυση και διαχείριση δεδομένων, για χρήστες πληροφορικής και επιχειρήσεις
 - + Εδραιωμένη και αυξανόμενη κοινότητα συνεργατών, με παγκόσμια υποστήριξη προϊόντων.
-
- Περιορισμένη ενσωμάτωση μεταξύ των προϊόντων Cognos Analytics, Planning Analytics, SPSS και Watson Analytics.
 - Πιθανή σύγχυση πελατών σχετικά με την επωνυμία των προϊόντων, π.χ. IBM Cognos BI και Cognos Analytics, IBM Planning Analytics και Cognos TM1
 - Το IBM Watson Analytics είναι διαθέσιμο μόνο ως λύση σύννεφο. Εξάγει τα δεδομένα και τα φορτώνει στην ενσωματωμένη υπηρεσία αποθήκευσης στο νέφος, που μπορεί να προκαλέσει πιθανές καθυστερήσεις.
-

Qlik

- + Μεγάλη γκάμα εργαλείων για τη δημιουργία αναφορών και πινάκων ελέγχου, ανάλυση και δημιουργία μεμονωμένων εφαρμογών.

- + Καλή συσχετιστική και καθοριζόμενη πλοήγηση σε δεδομένα με αναζήτηση στο διαθέσιμο περιεχόμενο.
 - + Καλές δυνατότητες μορφοποίησης και δημιουργίας διαγραμμάτων.
 - + Γρήγορη ανταπόκριση επεξεργασίας με τη χρήση τεχνολογιών μνήμης.
 - + Πηγές δεδομένων για επαυξημένη ανάλυση, με επιμελημένα και έτοιμα προς χρήση εξωτερικά δεδομένα.
-
- Δεν υπάρχουν δυνατότητες χρονοπρογραμματισμού από τις λύσεις που προσφέρουν τα προϊόντα της.
 - Περιορισμένη ενσωματωμένη λειτουργικότητα για τη διακυβέρνηση δεδομένων και τους κύριους δείκτες απόδοσης, για τον καθορισμό της απόδοσης σε διαφορετικές εφαρμογές
 - Οι σύνθετες υλοποιήσεις του QlikView και του Qlik Sense και η ενσωμάτωση δεδομένων απαιτούν προσαρμοσμένο κώδικα, παρόλο που οι πρόσφατες εκδόσεις του Qlik Sense έχουν μειώσει αυτή την ανάγκη.
-

Tableau

- + Εύκολο στη χρήση περιβάλλον και κατ' επέκταση δυνητικά υψηλή αποδοχή από τον χρήστη.
 - + Οπτική ανάλυση με ενσωματωμένη νοημοσύνη και υψηλή διαδραστικότητα κατάλληλη για επιχειρηματικούς χρήστες.
 - + Δυνατότητες προετοιμασίας και ανάλυσης δεδομένων για χρήστες που έχουν πρόσβαση σε διαφορετικές πηγές δεδομένων, συμπεριλαμβανομένων των δυνατοτήτων για συνδέσεις μεταξύ βάσεων δεδομένων και έξυπνες προτάσεις σύνδεση με τη χρήση μηχανικής μάθησης.
 - + Ποικίλες επιλογές διαδραστικής απεικόνισης με συστάσεις οπτικοποίησης.
 - + Δυνατότητες ανάλυσης δεδομένων και δημιουργίας αναφορών και εκτός σύνδεσης.
-
- Μη διαθέσιμες προκαθορισμένες λειτουργίες για χρονοπρογραμματισμό και προϋπολογισμό.

- Κίνδυνος δημιουργίας υπερβολικών αναφορών, χωρίς την κατάλληλη διαχείριση.
 - Το Tableau Prep ως ξεχωριστό εργαλείο για την προετοιμασία δεδομένων απαιτεί από τους χρήστες να αλλάζουν εργαλεία κατά την επαναληπτική ανακάλυψη.
-

Sisense

- + Ενσωματωμένο προϊόν για την ανακάλυψη δεδομένων και την χρήση πινάκων ελέγχου σε ένα πλήρες πακέτο, το οποίο είναι φιλικό προς τους επιχειρηματικούς χρήστες.
 - + Προσφέρει γρήγορη εσωτερική αποθήκευση δεδομένων σε στήλες, σε συνδυασμό με την ιδιόκτητη τεχνολογία in-chip, για επιτάχυνση της απόδοσης.
 - + Περιέχει API και βιβλιοθήκη JavaScript για σκοπούς ενσωμάτωσης
 - + Ολοκληρωμένες δυνατότητες ενσωμάτωσης και μοντελοποίησης δεδομένων από επιχειρηματικούς χρήστες.
 - + Πίνακες ελέγχου με βάση την HTML5 και διαδραστικές οπτικοποιήσεις που απευθύνονται σε επιχειρηματικούς χρήστες.
-
- Δεν υπάρχει προκαθορισμένη λειτουργικότητα για χρονοπρογραμματισμό.
 - Περιορισμένες δυνατότητες στον τομέα της εκτύπωσης και της κατανομής των αναφορών, σε σύγκριση με ορισμένους ανταγωνιστές σε αυτόν τον τομέα.
 - Η τοπική παρουσία περιορίζεται μόνο στις Ηνωμένες Πολιτείες και το Ισραήλ, αν και ο προμηθευτής επεκτείνεται προς το παρόν και σε άλλες περιοχές.

Βιβλιογραφία

- [1] <https://www.gartnerinfo.com/cios9/CIOLeadershipForum2015Profile.pdf>
- [2] <https://www.gartnerinfo.com/cios10/CIOLeadershipForum2016Profile.pdf>
- [3] https://www.gartner.com/imagesrv/cio/pdf/Gartner_CIO_Agenda_2017.pdf
- [4] <https://www.gartner.com/it-glossary/business-intelligence-bi/>
- [5] https://en.wikipedia.org/wiki/Business_intelligence
- [6] https://en.wikipedia.org/wiki/Decision_support_system
- [7] <https://searchbusinessanalytics.techtarget.com/definition/business-intelligence-BI>
- [8] <https://web.archive.org/web/20080913121526/http://www.research.ibm.com:80/journal/rd/024/ibmrd0204H.pdf>
- [9] <https://archive.org/details/cyclopaediacomm00devegoog>
- [10] <https://archive.org/stream/decisionsupports1980keen?ref=ol#page/8/mode/2up>
- [11] https://repository.kallipos.gr/bitstream/11419/1227/2/Kef._1.pdf
- [12] <https://www.linkedin.com/pulse/why-business-intelligence-bi-important-your-company-yass>
- [13] <https://www.gartner.com/newsroom/id/3612617>
- [14] <http://im1.im.tku.edu.tw/~cjou/bi2009/4.pdf>
- [15] https://link.springer.com/content/pdf/10.1007%2F978-3-642-03422-0_6.pdf
- [16] https://www.sas.com/el_gr/insights/big-data/what-is-big-data.html
- [17] Fox, Charles (2018-03-25). Data Science for Transport. Springer.
- [18] https://www.sas.com/el_gr/insights/big-data/what-is-big-data.html#dmhistory
- [19] <https://www.guru99.com/what-is-big-data.html#3>
- [20] <https://datafloq.com/read/big-data-history/239>
- [21] <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

- [22] <https://opensource.com/life/14/8/intro-apache-hadoop-big-data>
- [23] Report-VII-BigDataChallengesandSolutions.pdf
- [24] oea-big-data-guide-1522052
- [25] https://en.wikipedia.org/wiki/Data_quality
- [26] <https://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
- [27] https://en.wikipedia.org/wiki/Semi-structured_data
- [28] https://en.wikipedia.org/wiki/Big_data
- [29] <https://www.gartner.com/it-glossary/data-mining>
- [30] https://el.wikipedia.org/wiki/%CE%95%CE%BE%CF%8C%CF%81%CF%85%CE%BE%CE%B7_%CE%B4%CE%B5%CE%B4%CE%BF%CE%BC%CE%AD%CE%BD%CF%89%CE%BD
- [31] https://www.sas.com/en_us/insights/analytics/data-mining.html
- [32] Dunham - Data Mining
- [33] https://repository.kallipos.gr/bitstream/11419/1226/3/Epixeirimatiki_Efyia_kai_exorixi_Dedomeno_n_pdf.pdf
- [34] https://www.intechopen.com/books/data_mining_and_knowledge_discovery_in_real_life_applications/the_deployment_of_data_mining_into_operational_business_processes
- [35] <http://barc-research.com/about-us/>
- [36] BARC_BI_Trend_Monitor_2018_Online.pdf
- [37] https://www.edureka.co/blog/hadoop-ecosystem?utm_source=youtube&utm_campaign=hadoop-ecosystem-091216-wr&utm_medium=description
- [38] <https://hadoopecosystemtable.github.io/>
- [39] https://www.sas.com/el_gr/insights/big-data/hadoop.html
- [40] <https://searchdatamanagement.techtarget.com/definition>
- [41] <https://whatis.techtarget.com/definition>

[42] <https://hortonworks.com/apache/>

[43] <https://www.gartner.com/doc/reprints?id=1-4PUK79W&ct=180130&st=sb>

[44] <http://www.oracle.com/us/corporate/analystreports/barc-score-bi-platforms-4071728.pdf>

[45] <https://opensource.com/article/17/8/enterprise-open-source-advantages>

[46] <https://www.nibusinessinfo.co.uk/content/disadvantages-open-source-software>