

**ΑΕΙ ΠΕΙΡΑΙΑ Τ.Τ.  
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ Τ.Ε.**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Εργαλεία και Τεχνολογίες Επιχειρησιακής Νοημοσύνης**

**Λάμπρος Ι. Βότσης**

**Εισηγητής: Δρ. Νικόλαος Ζάχαρης, Αναπληρωτής Καθηγητής**

**ΑΘΗΝΑ  
ΜΑΡΤΙΟΣ 2017**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Εργαλεία και Τεχνολογίες Επιχειρησιακής Νοημοσύνης**

**Λάμπρος Ι. Βότσης  
Α.Μ. 32293**

**Εισηγητής:**

**Δρ. Νικόλαος Ζάχαρης, Αναπληρωτής Καθηγητής**

**Εξεταστική Επιτροπή:**

**Δρ. Παναγιώτης Γιαννακόπουλος, Καθηγητής  
Δρ. Δημήτριος Νικολόπουλος, Αναπληρωτής Καθηγητής**

**Ημερομηνία εξέτασης 10/03/2017**

(Κενή σελίδα)

## **ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

Ο/Η κάτωθι υπογεγραμμένος, Λάμπρος Βότσης του Ιωάννη, με αριθμό μητρώου 32293 φοιτητής του Τμήματος Μηχανικών Η/Υ Συστημάτων Τ.Ε. του Α.Ε.Ι. Πειραιά Τ.Τ. πριν αναλάβω την εκπόνηση της Πτυχιακής Εργασίας μου, δηλώνω ότι ενημερώθηκα για τα παρακάτω:

«Η Πτυχιακή Εργασία (Π.Ε.) αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο του συγγραφέα, όσο και του Ιδρύματος και θα πρέπει να έχει μοναδικό χαρακτήρα και πρωτότυπο περιεχόμενο.

Απαγορεύεται αυστηρά οποιοδήποτε κομμάτι κειμένου της να εμφανίζεται αυτούσιο ή μεταφρασμένο από κάποια άλλη δημοσιευμένη πηγή. Κάθε τέτοια πράξη αποτελεί προϊόν λογοκλοπής και εγείρει θέμα Ηθικής Τάξης για τα πνευματικά δικαιώματα του άλλου συγγραφέα. Αποκλειστικός υπεύθυνος είναι ο συγγραφέας της Π.Ε., ο οποίος φέρει και την ευθύνη των συνεπειών, ποινικών και άλλων, αυτής της πράξης.

Πέραν των όποιων ποινικών ευθυνών του συγγραφέα σε περίπτωση που το Ίδρυμα του έχει απονείμει Πτυχίο, αυτό ανακαλείται με απόφαση της Συνέλευσης του Τμήματος. Η Συνέλευση του Τμήματος με νέα απόφαση της, μετά από αίτηση του ενδιαφερόμενου, του αναθέτει εκ νέου την εκπόνηση της Π.Ε. με άλλο θέμα και διαφορετικό επιβλέποντα καθηγητή. Η εκπόνηση της εν λόγω Π.Ε. πρέπει να ολοκληρωθεί εντός τουλάχιστον ενός ημερολογιακού θμήνου από την ημερομηνία ανάθεσης της. Κατά τα λοιπά εφαρμόζονται τα προβλεπόμενα στο άρθρο 18, παρ. 5 του ισχύοντος Εσωτερικού Κανονισμού.»

**(Κενή σελίδα)**

## ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή εργασία ασχολείται με τις τεχνολογίες και τα εργαλεία ενός κυρίαρχου τομέα στην διαχείριση και ανάλυση δεδομένων, την Επιχειρησιακή Νοημοσύνη. Ο τομέας της Επιχειρησιακής Νοημοσύνης αποτελεί τομή διαφόρων επιστημών με χρήση ποικίλων τεχνολογιών σε επίπεδο δεδομένων, αρχιτεκτονικής συστημάτων διαχείρισης δεδομένων, προχωρημένης ανάλυσης και οπτικοποίησης πληροφοριών. Η δομή της παρούσας πτυχιακής εργασίας καταστεί δυνατή την κατανόηση του τομέα σε άτομα με βασικές γνώσεις Βάσεων Δεδομένων και προγραμματισμού. Για την επίτευξη αυτού του σκοπού υπάρχουν δυο μέρη, ένα θεωρητικό και μία μελέτη περίπτωσης (case study).

Αρχικά παρουσιάζεται η εξελικτική πορεία του τομέα με αναφορά σε παρακλάδια που ακολουθούν την εξέλιξη όπως τα Big Data. Η ανάλυση των εννοιών και των ορολογιών καταστεί εύκολο στην συνέχεια να αναφερθούν οι συνθήκες για την ανάπτυξη έργων Επιχειρησιακής Νοημοσύνης σε οργανισμούς. Γίνεται αναφορά στο επιχειρηματικό αντίκτυπο, στους παράγοντες επιτυχίας και στους παράγοντες υιοθέτησης καινοτομιών. Αναλυτικά παρουσιάζεται και το μοντέλο ωριμότητας, ένας χάρτης περιγραφής, εξήγησης και αξιολόγησης των σταδίων ανάπτυξης κατά την διάρκεια του κύκλου ζωής ενός έργου. Εκτενής αναφορά στην συνέχεια γίνεται για το πεδίο εφαρμογών, τις τεχνολογίες και τα συστήματα δηλαδή που συνθέτουν ένα περιβάλλον Επιχειρησιακής Νοημοσύνης. Παρουσιάζονται στοιχεία ενδιαφέροντος σχετικά με τα δεδομένα και της πηγές αυτών, γίνεται αναφορά στις υπηρεσίες ενσωμάτωσης και στην διαδικασία ETL ώστε επομένως να αναλυθούν οι υπηρεσίες διαχείρισης δεδομένων με έννοιες όπως Data Warehouse και Data mart. Το τελευταίο κομμάτι του θεωρητικού μέρους ασχολείται συνοπτικά με τα συστήματα OLTP, OLAP, τις τεχνικές Data Mining και την δημιουργία Data Cubes. Όλα τα παραπάνω αποτυπώνονται στην ανάλυση ενός case study όπου η θεωρία έρχεται σε σύγκρουση με την πραγματικότητα για τις απαιτητικές ανάγκες μεγάλου πολυεθνικού οργανισμού.

Το αποτέλεσμα της παρούσας πτυχιακής εργασίας είναι η ουσιαστική κατανόηση των εργαλείων και τεχνολογιών ανάπτυξης έργων Επιχειρησιακής Νοημοσύνης σε έναν οργανισμό.

**(Κενή σελίδα)**

## ABSTRACT

The present thesis concerns the technologies and tools of a dominant sector in data analysis and decision support, Business Intelligence. The field of Business Intelligence is an intersection of various disciplines using a variety of technologies in fields like, data sources management, data warehouse systems and, advanced analysis and visualization. The structure of this thesis enables the comprehension and cognizance of the sector to people with basic understanding of Databases and programming principles. To achieve this goal and foster a clear understanding, there are two parts, a theoretical and a case study.

The first theoretical part presents the evolution of the sector with references to emerging technologies that follows, such as the Big Data. The explanation of concepts and terminologies facilitates the following analysis of the conditions for the development of Business Intelligence projects in organizations. Reference is made to the business impact, the success factors and innovation adoption factors. Further information is presented regarding the maturity model, a descriptive map that assists evaluation of the various stages of development during the lifecycle of a BI project. Extensive reference was then made to the field of applications; the technologies and systems which synthesize a Business Intelligence environment. References are made on the data sources discussion points and therefore to integration services such as ETL process. Following that we get to data management services with analysis of concepts such as Data Warehouse and Data mart. The last part of the theoretical part briefly discusses the OLTP systems, OLAP, Data Mining techniques and the creation of Data Cubes. All these are reflected in the presentation of a case study where the theory comes into contrast with reality for the high demanding needs a large multinational organization.

The conclusive result of the present thesis is the fundamental understanding of the Tools and Technologies required for develop Business Intelligence projects in organizations.



ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ: Επιχειρησιακή Νοημοσύνη, Διαχείριση Δεδομένων, Ανάλυση Δεδομένων, Αποθήκες Δεδομένων, Συστήματα Ανάλυσης

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Επιχειρησιακή νοημοσύνη, ETL διαδικασία, data warehouse, data mart, OLAP, Ανάλυση Δεδομένων

## ΠΕΡΙΕΧΟΜΕΝΑ

<b>ΕΙΣΑΓΩΓΗ</b> .....	<b>1</b>
<b>ΚΕΦΑΛΑΙΟ Α - ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΝΟΗΜΟΣΥΝΗ</b> .....	<b>3</b>
A.1 Εισαγωγή στα Συστήματα Υποστήριξης Αποφάσεων (Decision Support Systems, DSS) .....	3
A.2 Ορισμός Επιχειρησιακής Νοημοσύνης (Business Intelligence, BI).....	4
A.3 Επιστήμη Δεδομένων (Data science).....	5
A.4 Εξελικτική Πορεία της ΕΝ .....	6
<i>Επιχειρησιακή Νοημοσύνη και Ανάλυση 1.0</i> .....	7
<i>Επιχειρησιακή Νοημοσύνη και Ανάλυση 2.0</i> .....	8
<i>Επιχειρησιακή Νοημοσύνη και Ανάλυση 3.0</i> .....	8
Ανάπτυξη του γνωστικού αντικείμενου ΕΝ .....	9
A.5 Περιβάλλον Επιχειρησιακής Νοημοσύνης .....	11
A.6 Τεράστιοι όγκοι δεδομένων (Big Data) .....	12
A.7 Ακαδημαϊκό Υπόβαθρο ΕΝ .....	13
<b>ΚΕΦΑΛΑΙΟ Β – ΥΙΟΘΕΤΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΝΟΗΜΟΣΥΝΗΣ</b> .....	<b>15</b>
B.1 Αντίκτυπο ΕΝ στην ποιότητα λήψης αποφάσεων.....	16
B.2 Μοντέλο Ωριμότητας ΕΝ (BI Maturity Model) .....	17
<i>Πρώτο Στάδιο ωριμότητας - Βρέφος (infant)</i> .....	18
<i>Δεύτερο Στάδιο ωριμότητας - Παιδί (Child)</i> .....	18
<i>Τρίτο Στάδιο ωριμότητας - Έφηβος (Teenager)</i> .....	19
<i>Τέταρτο Στάδιο ωριμότητας - Ενήλικος (Adult)</i> .....	19
<i>Πέμπτο Στάδιο ωριμότητας - Σοφός (Sage)</i> .....	20
<i>Τα χάσματα μεταξύ σταδίων (Gulf &amp; Chasm)</i> .....	20
B.3 Κρίσιμοι Παράγοντες Επιτυχίας.....	22
B.4 Παράγοντες Υιοθέτησης Καινοτομιών .....	25
<b>ΚΕΦΑΛΑΙΟ Γ – ΕΠΙΠΕΔΟ ΕΦΑΡΜΟΓΩΝ</b> .....	<b>27</b>
Γ1. Πηγές Δεδομένων .....	29
Γ1.1 Δεδομένα Πληροφορία και Γνώση .....	29
Γ1.2 Δεδομένα διαχείρισης συναλλαγών ( <i>transactional sources</i> ) .....	30
Γ1.3 Ενδοεταιρική πηγή κύριων δεδομένων ( <i>Enterprise Master Sources</i> ).....	31
Γ1.4 Μη αυτόματες πηγές δεδομένων ( <i>manual/user input</i> ) .....	32
Γ1.5 Εξωτερικά συνεργαζόμενες πηγές δεδομένων ( <i>External sources</i> ).....	33
Γ1.6 Δομημένα, Ημι-δομημένα και Αδόμητα δεδομένα.....	34

Γ1.7 Μεταδεδομένα (Metadata) .....	37
Γ1.8 Στοιχεία ενδιαφέροντος σχετικά με τις πηγές δεδομένων: .....	37
Γ2. Υπηρεσίες Ενσωμάτωσης .....	41
Γ2.1 Η Διαδικασία Εξαγωγής-Μετατροπής-Φόρτωσης δεδομένων.....	41
Γ2.2 Σχεδιασμός ETL διαδικασίας. ....	42
Γ2.3 Διαδικασία Εξαγωγής από Πηγές Δεδομένων.....	43
Γ2.4 Στάδιο Μετατροπής και Φόρτωση Δεδομένων .....	43
Γ3. Υπηρεσίες Διαχείρισης Δεδομένων.....	45
Γ3.1 Data Mart.....	45
Γ3.2 Αρχιτεκτονικές Αποθήκης Δεδομένων (Data Warehouse): .....	46
Γ4. Υπηρεσίες Ανάλυσης και Αναφορών .....	52
Γ4.1 OLAP και OLTP .....	52
Γ4.2 Εξόρυξη Δεδομένων (Data Mining).....	52
Γ4.3 Κύβος Δεδομένων (Data Cube) .....	54
Γ4.4 Προχωρημένα Συστήματα Αναλύσεων (Advanced Analytics) .....	54
Γ4.5 Οπτικοποίηση δεδομένων (Data visualization).....	55
<b>ΚΕΦΑΛΑΙΟ Δ – ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ (CASE STUDY).....</b>	<b>57</b>
Δ1. Εισαγωγή .....	57
Δ1.1 Σύγχρονες ανάγκες της αγοράς.....	57
Δ1.1 Η ανάγκη για EN .....	58
Δ1.2 Επιθυμητό Αποτέλεσμα .....	59
Δ2. Ανάλυση δεδομένων συνθηκών και αναγκών.....	60
Δ2.1 Εργαλεία Επιχειρησιακής Νοημοσύνης .....	60
Δ2.2 Σχεδιασμός τοπίου Επιχειρησιακής Νοημοσύνης.....	72
Δ3. Εφαρμογή επιλεχθέντος λύσης .....	76
Δ3.1 Κύκλος Επεξεργασίας Δεδομένων .....	76
Δ3.2 Εισαγωγή Δεδομένων .....	79
Δ3.3 Unix Server (Stored Procedures) .....	87
Δ3.4 Επεξεργασία και αποθήκευση δεδομένων (ETL Process) .....	91
Δ3.5 Ανάλυση και Αναφορές Πωλήσεων στο Business Objects.....	101
<b>ΚΕΦΑΛΑΙΟ Ε – ΣΥΜΠΕΡΑΣΜΑΤΑ .....</b>	<b>105</b>
<b>ΠΑΡΑΠΟΜΠΕΣ .....</b>	<b>107</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ.....</b>	<b>109</b>

## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

<b>Σχήμα 1.1:</b> Τεχνολογικό Περιβάλλον EN .....	<b>5</b>
<b>Σχήμα 1.2:</b> Η εξελικτική πορεία EN και Ανάλυσης .....	<b>10</b>
<b>Σχήμα 1.3:</b> Απλοϊκή απεικόνιση περιβάλλοντος EN .....	<b>11</b>
<b>Σχήμα 1.4:</b> Χαρακτηριστικά Big Data .....	<b>12</b>
<b>Σχήμα 2.1:</b> TDWI Μοντέλο Ωριμότητας Επιχειρησιακής Νοημοσύνης .....	<b>21</b>
<b>Σχήμα 2.2α:</b> Σχέση κόστους αξίας ανά στάδιο του Μοντέλου Ωριμότητας EN και Ανάλυσης .....	<b>21</b>
<b>Σχήμα 2.2β:</b> Σχέση κόστους και αξίας ανά στάδιο του Μοντέλου Ωριμότητας EN και Ανάλυσης .....	<b>21</b>
<b>Σχήμα 2.3:</b> Παράγοντες επιτυχίας Συστημάτων Πληροφοριών κατά τους DeLone και McLean .....	<b>23</b>
<b>Σχήμα 2.4:</b> Συσχετίσεις μεταξύ των παραγόντων υιοθέτησης καινοτομιών .....	<b>26</b>
<b>Σχήμα 3.1:</b> Το σύστημα EN από την εξαγωγή δεδομένων έως την ανάκτηση και ανάλυση .....	<b>27</b>
<b>Σχήμα 3.2:</b> Υπηρεσίες EN στο επίπεδο εφαρμογών.....	<b>28</b>
<b>Σχήμα 3.3:</b> Η πυραμίδα της γνώσης, DIKW Pyramid (Data Information Knowledge Wisdom Pyramid) .....	<b>30</b>
<b>Σχήμα 3.4:</b> Διάγραμμα οντοτήτων-συσχετίσεων (ER-διάγραμμα) πινάκων σε σύστημα σχεσιακής βάσης δεδομένων .....	<b>35</b>
<b>Σχήμα 3.5:</b> πρωτόκολλο SoftX: διεκπεραίωση συναλλαγών μέσω σειράς δομημένων σταδίων .....	<b>39</b>
<b>Σχήμα 3.6.α:</b> Διαφορετικά πακέτα δεδομένων ανά μέθοδο συναλλαγής .....	<b>40</b>
<b>Σχήμα 3.6.β:</b> Ενιαίο πακέτο δεδομένων με χρήση flag ανά μέθοδο συναλλαγής.....	<b>40</b>
<b>Σχήμα 3.7:</b> Τα τρία διακριτά στάδια της ETL διαδικασίας .....	<b>42</b>
<b>Σχήμα 3.8.α:</b> Αρχιτεκτονική ανεξάρτητων data marts .....	<b>46</b>
<b>Σχήμα 3.8.β:</b> Αρχιτεκτονική εξαρτώμενων data marts .....	<b>46</b>

<b>Σχήμα 3.9:</b> Σχεδιασμός αυτόνομων data marts ή DW .....	<b>47</b>
<b>Σχήμα 3.10:</b> Προσέγγιση Kimball: Αρχιτεκτονική διαύλου data mart με συμμορφούμενες διαστάσεις .....	<b>48</b>
<b>Σχήμα 3.11:</b> Προσέγγιση Inmon: Αρχιτεκτονική Hub-and-spoke .....	<b>49</b>
<b>Σχήμα 3.12:</b> Υβριδική προσέγγιση με στοιχεία βασιζόμενα και στις δύο κυρίαρχες αρχιτεκτονικές .....	<b>51</b>
<b>Σχήμα 3.13:</b> Απεικόνιση κύβου δεδομένων 3-διαστάσεων (Πελάτης - Προϊόν - Χρονική Περίοδος) .....	<b>54</b>
<b>Σχήμα 3.14:</b> Παραδείγματα γραφημάτων κατά την απεικόνιση δεδομένων. Η εικόνα των δεδομένων όπως παρουσιάζεται στον τελικό χρήστη.....	<b>55</b>
<b>Σχήμα 4.1:</b> Διαδρομή των δεδομένων για ένα Retail Sales Report.....	<b>73</b>
<b>Σχήμα 4.2:</b> Διαφορετικές αρχιτεκτονικές συστήματος SAP.....	<b>75</b>
<b>Σχήμα 4.3:</b> Οι εφαρμογές του συστήματος SAP ERP (Application Modules).....	<b>76</b>
<b>Σχήμα 4.4:</b> Διαδικασία εκτέλεσης ερωτημάτων στο SAP Business Objects.....	<b>77</b>
<b>Σχήμα 4.5.α:</b> Δυνατότητες ανάλυσης δεδομένων (drilling, filtering, break) .....	<b>79</b>
<b>Σχήμα 4.5.β:</b> Δυνατότητες εξαγωγής αναφορών σε αρχεία pdf, excel και text....	<b>79</b>
<b>Σχήμα 4.6:</b> Περιβάλλον δημιουργίας αναφορών στο εργαλείο MicroStrategy.....	<b>80</b>
<b>Σχήμα 4.7:</b> Παράδειγμα αναφοράς στο εργαλείο MicroStrategy.....	<b>81</b>
<b>Σχήμα 4.8:</b> Απεικόνιση της 4-tier αρχιτεκτονικής του εργαλείου MicroStrategy... <b>Σχήμα 4.9:</b> Τα επιμέρους χαρακτηριστικά του συστήματος του MicroStrategy....	<b>82</b> <b>83</b>
<b>Σχήμα 4.10:</b> Το τοπίο Επιχειρησιακής Νοημοσύνης στο case study.....	<b>84</b>
<b>Σχήμα 4.11:</b> Κύκλος Επεξεργασίας Δεδομένων.....	<b>89</b>
<b>Σχήμα 4.12:</b> Πεδία του εργαλείου Data Upload Tool.....	<b>82</b>
<b>Σχήμα 4.13:</b> Φύλλο List του εργαλείου Data Upload Tool με την αντιστοιχία ημερομηνιών.....	<b>92</b>
<b>Σχήμα 4.14:</b> Φύλλο Correspondence List με τις αντιστοιχίες ημερομηνιών (εβδομάδα – περίοδο – εβδομάδα) .....	<b>93</b>
<b>Σχήμα 4.15.α:</b> Εικόνα εργαλείου για εισαγωγή δεδομένων ανά μήνα.....	<b>94</b>
<b>Σχήμα 4.15.β:</b> Εικόνα εργαλείου για εισαγωγή δεδομένων ανά εβδομάδα.....	<b>94</b>
<b>Σχήμα 4.16:</b> Μηνύματα ενημέρωσης και επιλογής του εργαλείου Data Upload Tool.....	<b>98</b>

<b>Σχήμα 4.17:</b> Διαδικασία μετακίνησης αρχείων στον Unix Server.....	<b>102</b>
<b>Σχήμα 4.18:</b> Το τελικό στάδιο των αναφορών με παραδείγματα τελικών γραφημάτων στο περιβάλλον του SAP Business Objects μετά από ανάλυση των πωλήσεων του case study.....	<b>116</b>

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

<b>Πίνακας 1.1:</b> Δημοσιεύσεις μελετών ανά έτος σχετικά με συστήματα EN και Ανάλυσης .....	<b>9</b>
<b>Πίνακας 3.1:</b> Βασικές διαφορές στις προσεγγίσεις Kimball και Inmon .....	<b>50</b>

**(Κενή σελίδα)**

## ΕΙΣΑΓΩΓΗ

Η τεχνολογία κατακλύζει και αλλάζει όλο και περισσότερο την καθημερινότητα μας επηρεάζοντας τόσο τις προσωπικές ανάγκες όσο και τις επαγγελματικές. Το ίδιο αποτέλεσμα έχει και η τεχνολογία στους οργανισμούς. Όσο περισσότεροι οργανισμοί συνεχίζουν να χρησιμοποιούν την τεχνολογία και τα υπολογιστικά συστήματα για αναπτυξιακούς σκοπούς, όσο και για σκοπούς διευκόλυνσης, τόσο θα αναπτύσσονται νέες τεχνολογίες.

Καθώς οι οργανισμοί συνεχίζουν να χρησιμοποιούν τα συστήματα πληροφορικής για έναν αυξανόμενο αριθμό λειτουργιών σε έναν ανταγωνιστικό, γρήγορα εξελισσόμενο κόσμο, έρχονται αντιμέτωποι με την πρόκληση της επεξεργασίας και ανάλυσης των τεράστιων ποσοτήτων δεδομένων και της μετατροπής αυτών σε όφελος. Οι οργανισμοί διαθέτουν εν δυνάμει τεράστιους όγκους λεπτομερών επιχειρησιακών δεδομένων, ωστόσο οι αναλυτές και οι φορείς λήψης αποφάσεων εξακολουθούν να μην μπορούν να πάρουν τις απαντήσεις που χρειάζονται για να αντιδράσουν αρκετά γρήγορα στις μεταβαλλόμενες συνθήκες της αγοράς. Ο λόγος πολλές φορές είναι ότι τα δεδομένα είναι διάσπαρτα σε διάφορα τμήματα του οργανισμού ή έχουν εμπλακεί σε ένα υποτονικό και όχι σύγχρονο τεχνολογικό περιβάλλον.

Σε αυτές τις περιπτώσεις παρουσιάζεται η Επιχειρησιακή Νοημοσύνη η οποία είναι ένα σύνολο εργαλείων, τεχνολογιών και λύσεων προσφερόμενων σε τελικούς χρήστες με σκοπό την εξαγωγή αποτελεσματικών χρήσιμων πληροφοριών μέσα από ωκεανούς δεδομένων. Η Επιχειρησιακή Νοημοσύνη ποτέ δεν διατίθεται μέσω ενός μοναδικού προϊόντος, μιας τεχνολογίας ή ενός προμηθευτή λογισμικού και την σήμερα ημέρα λογίζεται ως ένα σύνολο εργαλείων που ενισχύουν την αξία των δεδομένων ενός οργανισμού. Όλο και περισσότεροι οργανισμοί τα τελευταία χρόνια αναπτύσσουν τις εν λόγω λύσεις για την ενίσχυση της ανταγωνιστικότητάς τους και την αποτελεσματική και αποδοτική λειτουργία τους.

Η παρούσα πτυχιακή εργασία πραγματεύεται τις τεχνολογίες λύσεων Επιχειρησιακής Νοημοσύνης διαρθρωμένη σε θεωρητικά και πρακτικά μέρη. Αρχικά, στο πρώτο κεφάλαιο «Εισαγωγή στην Επιχειρησιακή Νοημοσύνη» ορίζονται και περιγράφονται βασικές έννοιες με σκοπό την ανάλυση περίπλοκων



εννοιών καθώς και η δημιουργία μιας ξεκάθαρης εικόνας σχετικά με τις εν λόγω τεχνολογίες. Ακολουθώντας, στο κεφάλαιο με τίτλο «Υιοθέτηση Τεχνολογιών Επιχειρησιακής Νοημοσύνης», αποτυπώνονται τα πολλαπλά οφέλη σε έναν οργανισμό συνδέοντας ουσιαστικά τον τεχνολογικό και επιχειρηματικό κόσμο. Εν συνεχεία ακολουθεί το κεφάλαιο «Επίπεδο Εφαρμογών» σκοπός του οποίου είναι η παρουσίαση των διαφόρων τεχνολογικών επιλογών κατά την διαδικασία εφαρμογής, συνδέοντας τις επιχειρησιακές ανάγκες με τις τεχνολογικές δυνατότητες. Προχωρώντας σε περισσότερο πρακτικά θέματα, στο κεφάλαιο «Πρακτική Εφαρμογή» αναλύεται η διαδικασία Επιχειρησιακής Νοημοσύνης όπως αναπτύχθηκε σε μια επιχείρηση. Αναλύονται οι τεχνολογικές αποφάσεις, παρουσιάζονται στιγμιότυπα κώδικα και γραφήματα σχετικά με τις τεχνολογίες και τα λογισμικά τα οποία χρησιμοποιήθηκαν. Η εργασία καταλήγει σε συμπεράσματα συνδυάζοντας γνώσεις από το θεωρητικό και πρακτικό μέρος. Τέλος, συγκεντρωμένα σε παραρτήματα βρίσκονται ακρωνύμια, ορολογίες, ευρετήριο των γραφικών απεικονίσεων καθώς και παραπομπές με σκοπό την ορθή επεξήγηση της εργασίας αλλά κυρίως ως τροφή για μελλοντική έρευνα και μελέτη.

## ΚΕΦΑΛΑΙΟ Α - ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΝΟΗΜΟΣΥΝΗ

### **A.1 Εισαγωγή στα Συστήματα Υποστήριξης Αποφάσεων (Decision Support Systems, DSS)**

Ο όρος Νοημοσύνη άρχισε να χρησιμοποιείται από ερευνητές από την δεκαετία του 1950. Ο όρος Επιχειρησιακή Νοημοσύνη (Business Intelligence, BI) έγινε διάσημος κατά την δεκαετία του 1990 ενώ στα τέλη της δεκαετίας του 2000 ο όρος Επιχειρησιακή Ανάλυση έκανε την εμφάνισή του ως βασικό στοιχείο της Επιχειρησιακής Νοημοσύνης [1]. Η προέλευση του όρου βασίζεται στα Συστήματα Υποστήριξης Αποφάσεων τα οποία τεχνολόγοι και ερευνητές αναπτύσσουν και ερευνούν για περισσότερο από 40 χρόνια [2].

Τα Συστήματα Υποστήριξης Αποφάσεων είναι εργαλεία που διευκολύνουν την λήψη αποφάσεων. Κατά καιρούς έχουν αποδοθεί διάφοροι ορισμοί ένας εκ των οποίων από τους Ralph H. Sprague και Eric D. Carlson στο βιβλίο του 1982, *Building Effective Decision Support Systems*. Σε αυτό το βιβλίο αναφέρονται ως Συστήματα Υποστήριξης Αποφάσεων τα "διαδραστικά υπολογιστικά συστήματα που βοηθούν τους ιθύνοντες να χρησιμοποιούν δεδομένα και μοντέλα για την επίλυση αδόμητων προβλημάτων" [3]. Στη φράση «Συστήματα Υποστήριξης Αποφάσεων» (Decision Support Systems, DSS) έμφαση πρέπει να δίνεται στη λέξη "υποστήριξη" δεδομένου ότι τα DSS είναι εργαλεία που βοηθούν την λήψη αποφάσεων, αλλά δεν αυτοματοποιούν την διαδικασία καθ' αυτή [4].

Σε έναν τομέα της τεχνολογίας τόσο ιδιαίτερο όσο τα Συστήματα Πληροφοριών (Information Systems) και ως εκ τούτου τα DSS και η EN, η καταγραφή της ιστορίας δεν είναι ούτε απλή ούτε γραμμική, η γνώση επεκτείνεται στη χρήση νέων τεχνολογιών και τη δημιουργία νέων εφαρμογών [2]. Επίσης, αν και τα συστήματα EN χρησιμοποιούνται ευρέως στην αγορά, η έρευνα είναι περιορισμένη [5]. Παρακάτω παρουσιάζονται συνοπτικά σημαντικά στοιχεία της EN ωστόσο, λόγω των δυσκολιών που αναφέρθηκαν (έρευνα και οριοθέτηση μεταξύ τεχνολογιών, τεχνικών και επιστημών), μια ξεκάθαρη και ολοκληρωμένη εικόνα για την EN παρουσιάζεται στο Κεφάλαιο Δ – Πρακτική Εφαρμογή (Case Study).

## A.2 Ορισμός Επιχειρησιακής Νοημοσύνης (Business Intelligence, BI)

Ο όρος Επιχειρησιακή Νοημοσύνη (Business Intelligence, BI) χρησιμοποιήθηκε ως κοινή ονομασία για να περιγραφούν «έννοιες και μεθοδολογίες οι οποίες ως στόχο έχουν την βελτίωση λήψεως αποφάσεων βασισμένες σε δεδομένα και πληροφορίες από διάφορα Συστήματα Υποστήριξης Αποφάσεων» στο βιβλίο του Howard Dresner το 1989 [6]. Από εκεί και ύστερα, η EN έχει οριστεί ποικιλοτρόπως με ορισμούς που διαφοροποιούνται σε σχέση με την χρονολογική περίοδο ή την οπτική και τις προτεραιότητες του κάθε ερευνητή ή τεχνολόγου. Για τους σκοπούς αυτής της εργασίας θα αναφέρουμε την EN με τον ορισμό που αποδόθηκε από τους Azvine B., Nauck D. και Majeed B. σε συνέδριο το 2006 (8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing) ο οποίος εστιάζει στην δυνατότητα των οργανισμών να βελτιώσουν την αποδοτικότητά τους και να πετύχουν τους στόχους τους.

Η EN αφορά την συλλογή, την εκτίμηση, την κατανόηση, την ανάλυση και την μετατροπή ενός από τους πιο σημαντικούς και πολύτιμους πόρους που παριστάνονται ως δεδομένα, σε χρηστικές πληροφορίες με σκοπό την βελτίωση του οργανισμού. Τα τυπικά χαρακτηριστικά ενός συστήματος EN περιλαμβάνουν [7]:

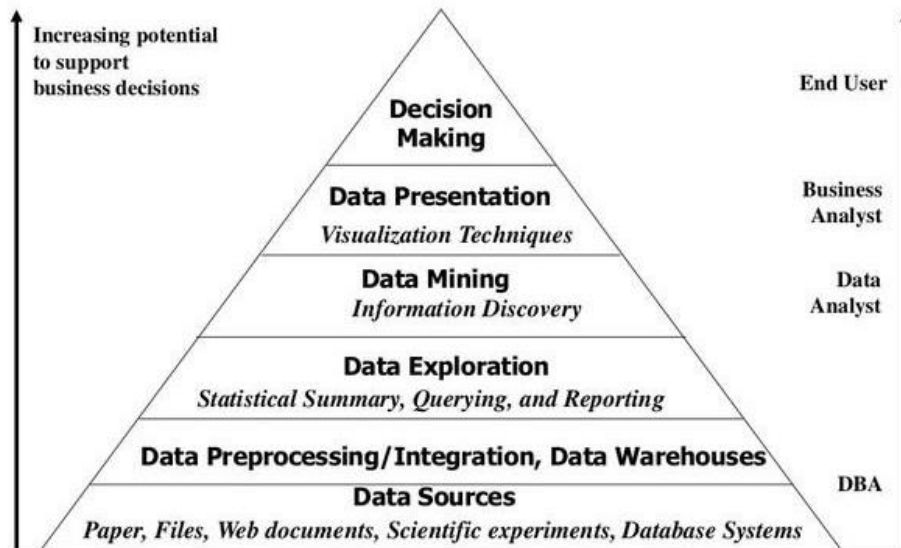
- υποβολή εκθέσεων και απεικόνιση
- ανάλυση τάσεων (ιστορικά και προβλέψεις)
- ανάλυση συμπεριφοράς πελατών
- προγνωστική μοντελοποίηση (ανάλυση πιθανών μελλοντικών σεναρίων)

Αρκετές έννοιες και ορολογίες συνδέονται με την EN. Από τα Συστήματα Υποστήριξης Αποφάσεων, την Επιστήμη Δεδομένων, την Ανάλυση, τα συστήματα Αποθήκης Δεδομένων, τις δυνατότητες του υλικού και λογισμικού στους Η/Υ, καθώς και η ανάδειξη της αρχιτεκτονικής του δικτύου, όλα συνδυάζονται για να δημιουργήσουν ένα πλούσιο τεχνολογικό περιβάλλον.

Συγκεντρώνοντας όλα τα στοιχεία που προαναφέρθηκαν, με σωστό σχεδιασμό και διαχείριση, αυξάνεται η δυνατότητα για υποστήριξη επιχειρηματικών αποφάσεων και ως εκ τούτου ο αντίκτυπος της διαχείρισης δεδομένων σε έναν

οργανισμό. Ως μια πολυδιάστατη διαδικασία, στην EN περιπλέκονται και συνεργάζονται διάφοροι επαγγελματικοί τομείς όπως,

- Διαχειριστές Βάσεων Δεδομένων (Database Administrators)
- Αναλυτές Δεδομένων (Data Analysts)
- Επιχειρηματικοί Αναλυτές (Business Analysts)
- Τελικοί χρήστες (πχ επαγγελματίες marketing, πωλήσεων, logistics και γενικά διοικητικό προσωπικό)



Σχήμα 1.1: Τεχνολογικό Περιβάλλον EN

### A.3 Επιστήμη Δεδομένων (Data science)

Τα διάφορα χαρακτηριστικά των συστημάτων πληροφορικής είναι σημαντικά και αναλύονται στα επόμενα κεφάλαια, στην ουσία ωστόσο, τα δεδομένα είναι αυτά που ενεργοποιούν την EN. Τα δεδομένα συλλέγονται από ποικίλες πηγές και ενσωματώνονται με διάφορες τεχνολογίες για να ενισχυθεί η χρησιμότητα και η σημασία τους [8]. Τα δεδομένα μετατρέπονται σε πληροφορίες που μεταφέρονται την σωστή στιγμή, στο σωστό μέρος στην καταλληλότερη μορφή με σκοπό την υποβοήθηση λήψης αποφάσεων [5].

Όπως έχει αναφερθεί, ανάλογα με την οπτική, τον σκοπό και την αντίληψη της κάθε εφαρμογής των Συστημάτων Πληροφοριών (Information Systems) μπορεί να γίνει περεταίρω έρευνα, ανάπτυξη αλλά και ορισμός τους. Το ίδιο ισχύει και για

την Επιστήμη των Δεδομένων καθώς μπορούν να προκύψουν θέματα καθαρά μαθηματικής φύσεως, στατιστικής, τεχνολογίας κλπ.

Αρκετές τεχνολογίες σχετίζονται με αυτό που αόριστα ονομάζεται Επιστήμη Δεδομένων με αρκετούς τεχνολόγους να αντιπαρατεθούν γνώμες και θεωρίες περισσότερο σε επίπεδο άρθρων στο διαδίκτυο και παρουσιάσεων σε συνέδρια χωρίς ωστόσο να υπάρχει συνοχή. Όπως αναφέρει ο Gil Press σε άρθρο του στο διαδίκτυο «ο όρος Επιστήμη των Δεδομένων περιγράφει έννοια της αγοράς χωρίς σαφή ορισμό»<sup>1</sup>.

#### **A.4 Εξελικτική Πορεία της EN**

Τις τελευταίες δύο δεκαετίες η EN και η Επιστήμη των Δεδομένων και της Ανάλυσης έχει απασχολήσει όλο και πιο ενεργά την αγορά εργασίας και την ακαδημαϊκή κοινότητα. Σπουδαίες ευκαιρίες αλλά και μεγάλα εμπόδια έχουν ξεπηδήσει από αυτόν τον νεοαναπτυσσόμενο τομέα. Μια έρευνα από το McKinsey Global Institute [9] προέβλεπε ότι κατά το έτος 2018 θα υπάρχουν ανάγκες για περίπου 140,000 με 190,000 άτομα με εις βάθος γνώσεις ανάλυσης δεδομένων καθώς και ανάγκη χιλιάδων διευθυντικών στελεχών με ικανότητες λήψεων σωστών αποφάσεων βάση των αναλύσεων μεγάλου όγκου δεδομένων.

Ο Hal Varian, Chief Economist της Google και επίτιμος καθηγητής στο Πανεπιστήμιο Berkeley της Καλιφόρνια σχολίασε [Hal Varian Answers Your Questions, freakonomics.com, 2008]<sup>2</sup> σχετικά με τις αναφαινόμενες τάσεις αναφερόμενος σε επαγγελματίες τεχνολόγους και μαθητές το εξής: “Τι είναι πανταχού παρόν και φθινό; Τα δεδομένα. Και τι συμπληρώνει την αξία των δεδομένων; Η ανάλυση. Οπότε η συμβουλή μου είναι να παρακολουθήσετε αρκετά μαθήματα σχετικά με το πως να χειραγωγείτε και να αναλύετε δεδομένα. Μαθήματα σχετικά με Βάσεις Δεδομένων, Μηχανική Μάθηση, Οικονομετρία, Στατιστική, Απεικόνιση δεδομένων κ.λπ.”.

<sup>1</sup> <http://www.forbes.com/sites/gilpress/2013/08/19/data-science-whats-the-half-life-of-a-buzzword/#3cab7e24a30>

<sup>2</sup> Όλη η συνέντευξη του Hal Varian στον παρακάτω σύνδεσμο:  
<http://freakonomics.com/2008/02/25/hal-varian-answers-your-questions/>

## Επιχειρησιακή Νοημοσύνη και Ανάλυση 1.0

Ο όρος Ανάλυση απασχολεί τους ερευνητές από το 1950. Στα τέλη του 2000 η Επιχειρησιακή Ανάλυση αντιπροσώπευε το βασικό στοιχείο ενός συστήματος EN. Τα τελευταία χρόνια έχουμε την εμφάνιση όρων όπως Big Data και Internet of Things, όροι και τεχνολογίες που περιπλέκουν όλο και περισσότερο τον τομέα, παρέχοντας ωστόσο και μεγαλύτερες δυνατότητες και ευκαιρίες.

Ως σύστημα με επίκεντρο τα δεδομένα, η EN στηρίζεται στον τομέα της Διαχείρισης Σχεσιακών Βάσεων Δεδομένων (RDBMS) και στις τεχνολογίες συλλογής, εξαγωγής και ανάλυσης. Η Διαχείριση Δεδομένων και η Αποθήκες Δεδομένων αποτελούν τις ρίζες του όρου EN και Ανάλυση 1.0 (BI&A 1.0), του πρώτου και βασικού επιπέδου EN [10]. Ο σχεδιασμός Αποθηκών Δεδομένων, τα στιγμιότυπα αυτών γνωστά και ως Data Marts, οι διαδικασίες Εξαγωγής, Μετατροπής και Φόρτωσης δεδομένων, η Ηλεκτρονική Διαδικασία Ανάλυσης καθώς και η διαδικασία δημιουργίας αναφορών με γραφήματα και πίνακες αποτελούν τον τρόπο ανακάλυψης σημαντικών πληροφοριών μέσα από δομημένα δεδομένα. Οι ανάγκες για αυτού του είδους διαχείρισης δεδομένων έχει γίνει εμπορικός σκοπός αρκετών εταιριών που προσφέρουν λύσεις EN όπως η Microsoft, IBM, Oracle και SAP.

Σύμφωνα με την αναφορά της Gartner [11] από τις 13 άκρως σημαντικές δυνατότητες της EN οι παρακάτω 8 θεωρούνται ως βασικές δυνατότητες των BI&A 1.0:

1. Υποβολή αναφορών (Reporting)
2. Πίνακες ελέγχου (Dashboards)
3. Ad hoc<sup>3</sup> ερωτήματα
4. EN βασισμένη στην αναζήτηση
5. Ηλεκτρονική Διαδικασία Ανάλυσης (On-line Analytical Process, OLAP)
6. Διαδραστική Απεικόνιση (Interactive Visualization)
7. Κάρτες αξιολόγησης (Scorecards)
8. Προγνωστική Μοντελοποίηση (Predictive Modeling)
9. Εξόρυξη Δεδομένων (Data Mining).

---

<sup>3</sup> *Ad hoc* είναι όρος που περιγράφει ότι γίνεται ή συμβαίνει μόνο για ένα συγκεκριμένο σκοπό ή ανάγκη και δεν προβλέπεται πριν αυτό συμβεί

<http://dictionary.cambridge.org/dictionary/english/ad-hoc>

### *Επιχειρησιακή Νοημοσύνη και Ανάλυση 2.0*

Από τις αρχές τους 2000, το Διαδίκτυο προσφέρει ιδιαίτερες ευκαιρίες ανακάλυψης, συλλογής, ανάλυσης και έρευνας δεδομένων μέσα από συστήματα βασισμένα σε ιστοσελίδες, εφαρμογές και άλλες χρήσεις και επιτρέπει στους οργανισμούς να παρουσιάζονται και να διαδρούν με τους πελάτες/χρήστες πιο άμεσα. Πληροφορίες όπως διευθύνσεις IP, δημογραφικά χαρακτηριστικά, καταγραφές διάδρασης κ.α. αποτελούν τον νέο θησαυρό για την ΕΝ. Νέοι τομείς όπως Web Intelligence και Web Analytics αναδύονται μέσω των δυνατοτήτων εργαλείων όπως Google Analytics και διαφόρων εργαλεία ανάλυσης Κοινωνικών Δικτύων (Social and Crowd Sourcing Systems) τα οποία αποτελεσματικά συγκεντρώνουν μεγάλους όγκους αδόμητων και ημιδομημένων δεδομένων, τόσο ασύγχρονα όσο και σύγχρονα (real time) [12]. Τα συστήματα που αναπτύσσονται με χρήση με των παραπάνω τεχνολογιών αναφέρονται ως ΕΝ και Ανάλυση 2.0 (BI&A 2.0).

Μελλοντικά συστήματα BI&A 2.0 θα απαιτούν την ενοποίηση ώριμων και επεκτάσιμων τεχνικών στον τομέα της εξόρυξης κειμένου (π.χ. εξαγωγή πληροφοριών, ταυτοποίηση θεματολογίας, εξόρυξη γνώμης), web mining, ανάλυση κοινωνικών δικτύων, και συνεργασία με αναλύσεις των υπαρχόντων εφαρμογών BI&A 1.0 βασισμένων σε DBMS συστήματα. Στην λίστα δυνατοτήτων του Gartner BI Hype Cycle συμπεριλαμβάνονται πληροφορίες σημασιολογικών υπηρεσιών, ερωταποκρίσεις φυσικής γλώσσας, και ανάλυση περιεχομένου/ κειμένου [13].

### *Επιχειρησιακή Νοημοσύνη και Ανάλυση 3.0*

Σύμφωνα με έρευνα του περιοδικού The Economist<sup>4</sup> το 2011, ο αριθμός των συσκευών κινητής τηλεφωνίας και tablets ξεπέρασε τον αριθμό των laptop και οικιακών Η/Υ για πρώτη φορά το 2011 (480 σε σχέση με 380 εκατομμύρια συσκευές αντίστοιχα). Το ίδιο άρθρο προέβλεπε τον αριθμό των συσκευών κινητής τηλεφωνίας και tablets κατά το 2020 στον αριθμό των 10 δις συσκευών.

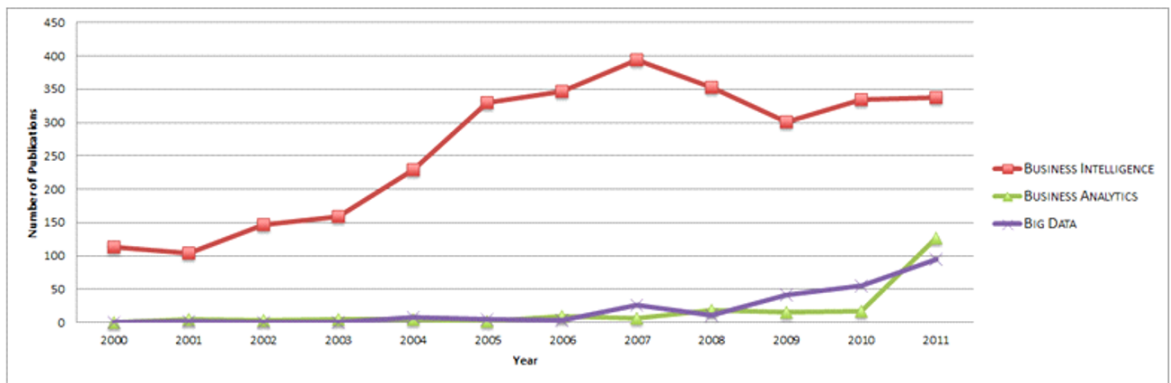
---

<sup>4</sup> Η έρευνα είναι διαθέσιμη στον παρακάτω σύνδεσμο:  
<http://www.economist.com/node/21531109>

Το παραπάνω γεγονός σε συνδυασμό με πληθώρα έξυπνων συσκευών που βασίζονται σε χρήση αισθητήρων και λοιπών οργάνων ανοίγει ένα νέο μεγάλο κεφάλαιο στην τεχνολογία με ονομασία Internet of Things (IoT) το οποίο προκαλεί ενθουσιασμό για τις επερχόμενες δυνατότητες της ΕΝ. Προβλέπονται για το εγγύς μέλλον η ανάπτυξη μη-ολοκληρωμένων εμπορικών συστημάτων [10]. Το μεγαλύτερο μέρος της ακαδημαϊκής έρευνας βρίσκεται σε πρώιμο στάδιο και η κοινότητα αντιμετωπίζει προκλήσεις και ευκαιρίες στις προσπάθειες παραγωγής επιστημονικού και επιχειρηματικού αντίκτυπου συμβατού με τις σύγχρονες ανάγκες [14].

### Ανάπτυξη του γνωστικού αντικειμένου ΕΝ

Στην προσπάθεια μια καλύτερης και βαθύτερης κατανόησης του τομέα αρκετά επιστημονικά περιοδικά, εκπαιδευτικά ιδρύματα και οργανισμοί έχουν δημοσιεύσει μελέτες και ανάλυσης που σκοπό έχουν περισσότερο την εξερεύνηση της ΕΝ και την εδραίωσή της ως θεμελιώδη ανάγκη για την λήψη ορθών αποφάσεων. Στο παρακάτω γράφημα φαίνονται τα στατιστικά δημοσιεύσεων μελετών σχετικά με τομείς αναδυόμενους από τα συστήματα ΕΝ και Ανάλυσης<sup>5</sup> [10].



Keyword	All Years	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Business Intelligence	3,146	113	104	146	159	229	330	346	394	352	201	334	338
Business Analytics	213	0	5	43	4	5	2	9	6	19	16	17	126
Big Data	243	0	1	0	0	7	4	3	26	11	41	44	95
<b>Total</b>	<b>3,602</b>	<b>113</b>	<b>110</b>	<b>149</b>	<b>163</b>	<b>241</b>	<b>336</b>	<b>358</b>	<b>426</b>	<b>382</b>	<b>358</b>	<b>356</b>	<b>560</b>

Πίνακας 1.1: Δημοσιεύσεις μελετών ανά έτος σχετικά με συστήματα ΕΝ και Ανάλυσης

<sup>5</sup> Έρευνα για την περίοδο 2000-2011. Πηγές: Web of Science, Business Source Complete, IEEE Xplore, ScienceDirect, Engineering Village



Παρατηρείτε μια συνεχή σταθερή αύξηση σε μελέτες σχετικά με συμβατικά συστήματα EN ενώ από το 2007 και ύστερα έχουμε την αύξηση του ενδιαφέροντος σχετικά με την Επιχειρησιακή Ανάλυση και τον τομέα των Big Data

Συνοψίζοντας όλα τα αναδυόμενα γνωστικά αντικείμενα που πηγάζουν από τα συστήματα EN και Ανάλυσης χωρίζουμε την εξελικτική πορεία, τις απορρέουσες εφαρμογές και τις αντίστοιχες περιοχές μελέτης όπως φαίνονται στο παρακάτω πίνακα [10].

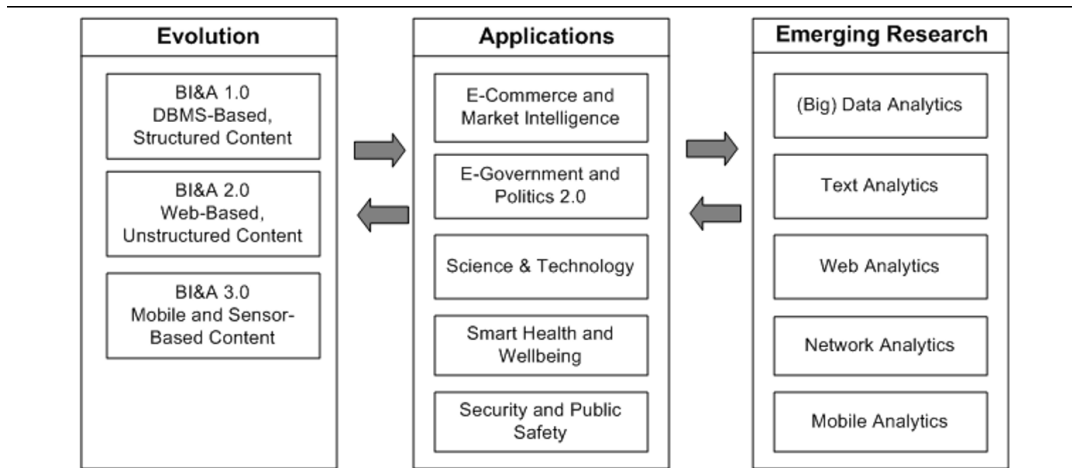


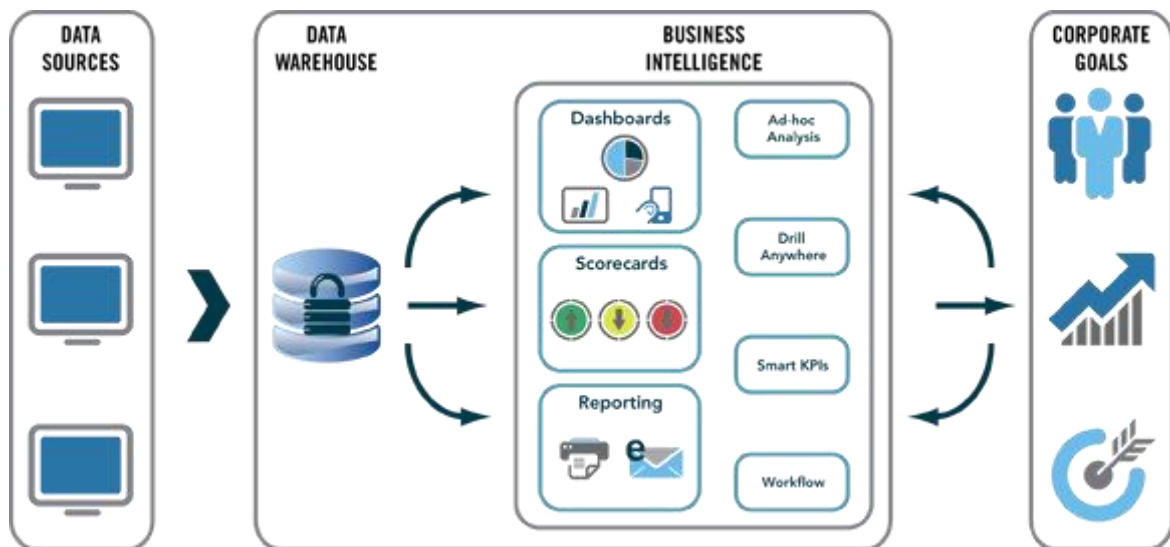
Figure 1. BI&A Overview: Evolution, Applications, and Emerging Research

### Σχήμα 1.2: Η εξελικτική πορεία EN και Ανάλυσης

Η παρούσα εργασία επικεντρώνεται στις τεχνολογίες και τα εργαλεία υλοποίησης συστημάτων EN και Ανάλυσης 1.0, ως εξής αναφερόμενη ως απλά EN για λόγους ευχέρειας και απλούστευσης. Ωστόσο ένας τόσο αναπτυσσόμενος τεχνολογικός τομέας δεν έχει διακριτά σύνορα και όρια από άλλα αντικείμενα μελετών όπως Analytics, Data Science κ.α. Η εις βάθος ανάλυση των μικρών διαφορών δεν αποτελεί αντικειμενικό σκοπό της παρούσας εργασίας γι' αυτό έμφαση θα υπάρχει περισσότερο στην ανάπτυξη της τεχνολογίας και της χρήσης των εργαλείων σε θεωρητικό και πρακτικό επίπεδο με σκοπό την βελτίωση της αποδοτικότητας και την επίτευξη των στόχων των οργανισμών.

## A.5 Περιβάλλον Επιχειρησιακής Νοημοσύνης

Για την δόμηση ενός περιβάλλοντος EN ένας οργανισμός χρειάζεται όχι μόνο τα κατάλληλα εργαλεία και την κατάλληλη Επιχειρησιακή δομή. Χρειάζεται άτομα με τις κατάλληλες ικανότητες που θα μπορέσουν να εκμεταλλευτούν όλες τις δυνατότητες που παρέχονται από αυτή την τεχνολογία. Ιδιαίτερη προσοχή θα πρέπει να δοθεί στην κατανόηση των επιχειρηματικών απαιτήσεων, την στοχοθέτηση και την ανάλυση διαφόρων διαδικασιών που συνδέονται με τον οργανισμό, τον ανταγωνισμό και τα προς παρακολούθηση και μέτρηση στοιχεία. Θα πρέπει να αναγνωριστεί ο στόχος και ο λόγος καθώς ύστερα να αποφασιστεί η πηγή, ο τρόπος λήψης και ο τρόπος μετάφρασης των δεδομένων ώστε να συνθέσουν την λύση που τελικά θα εφαρμοστεί. Στο παρακάτω σχήμα περιγράφεται απλοϊκά ένα περιβάλλον EN που περιέχει όλα τα παραπάνω στοιχεία.



**Σχήμα 1.3:** Απλοϊκή απεικόνιση περιβάλλοντος EN

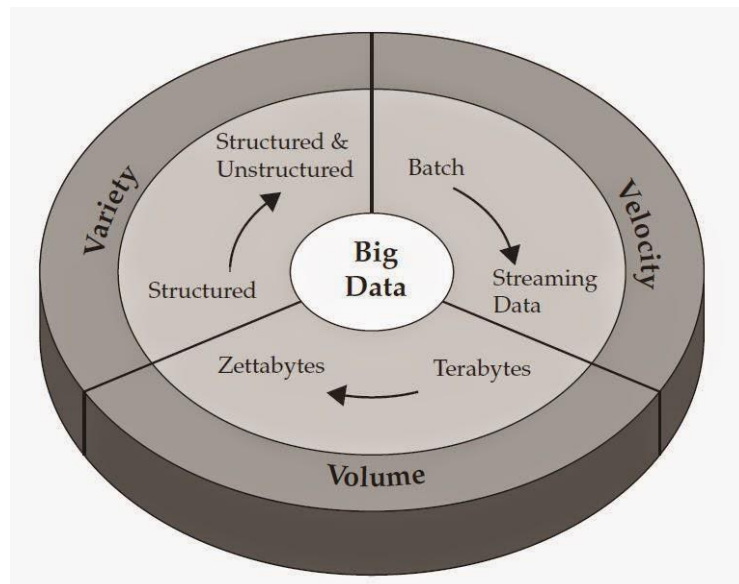
Γίνεται τελικά αντιληπτό το γεγονός πως οι τομείς της πληροφορικής, στατιστικής, μαθηματικών, πωλήσεων, marketing και διοίκησης επιχειρήσεων συνεργάζονται στενά ενώ πλέον συχνή παρουσιάζεται η ανάγκη γνώσεων, τουλάχιστον σε βασικό επίπεδο, όλων των παραπάνω τομέων. Πρακτικά, ένας υπεύθυνος στον τομέα των πωλήσεων χρειάζεται βασική γνώση τεχνολογίας ώστε να μπορεί να κατανοήσει τις δυνατότητες που παρέχονται από την EN. Αντίστοιχα ένας υπεύθυνος στον τομέα της Πληροφορικής χρειάζεται βασικές

γνώσεις πωλήσεων ή οικονομικών ώστε να μπορεί σωστά να κατανοεί και να καλύπτει τις επιχειρηματικές ανάγκες.

Εκ κατακλείδι, επιτακτική είναι η ανάγκη σε μια επιχείρηση να διοικούν και να εργάζονται άτομα που μπορούν να συνδέσουν τους παραπάνω τομείς, να αναγνωρίζουν τυχών περιορισμούς και δυνατότητες και να αυξήσουν την παραγωγικότητα μέσα από την συνεργασία.

## A.6 Τεράστιοι όγκοι δεδομένων (Big Data)

Το 2011 ένας νέο όρος έκανε την εμφάνισή του σε συνέδρια EN, στους κύκλους των επαγγελματιών του χώρου και ακαδημαϊκά, ο όρος Big Data ο οποίος είναι παραπλανητικός καθώς ο όγκος δεν είναι η μοναδική πρόκληση. Με λίγα λόγια, ο όρος Big Data ισχύει για πληροφορίες που δεν μπορούν να υποβληθούν σε επεξεργασία ή να αναλυθούν χρησιμοποιώντας παραδοσιακές διαδικασίες ή εργαλεία. Τρία χαρακτηριστικά καθορίζουν τα Big Data: όγκος, ποικιλομορφία, και ταχύτητα δημιουργίας δεδομένων (όπως φαίνεται στο παρακάτω Σχήμα ) [15].



Σχήμα 1.4: Χαρακτηριστικά Big Data

Μια άλλη αναφορά στα Big Data προσδιορίζει τον όρο ως σειρές δεδομένων των οποίων το μέγεθος ξεπερνά τις δυνατότητες συλλογής, αποθήκευσης, διαχείρισης και ανάλυσης ενός τυπικού λογισμικού Βάσεων Δεδομένων [9]. Ως το τελευταίο κεφάλαιο στην EN, τα Big Data ήρθαν στο προσκήνιο ως μια ανάγκη για την παρουσίαση ενός νέου τρόπου διαχείρισης μη ορθόδοξων πηγών δεδομένων

(π.χ. κοινωνικά δίκτυα, αισθητήρες) προχωρημένων τεχνολογιών (π.χ. Hadoop αρχιτεκτονικές, οπτικοποίηση, προγνωστική ανάλυση) και μη συμβατικών συνδυασμών χρηστών [16]. Εξ αρχής διαφαινόταν η δυσκολία στην διαχείριση τέτοιων δεδομένων. Η πολυπλοκότητα στην ανακάλυψη, διαχείριση και αποθήκευση οδήγησε αρκετές μελέτες στο συμπέρασμα ότι η δημοτικότητα των Big Data θα κατέρρευε μελλοντικά [17].

Ωστόσο μια τέτοια πρόβλεψη αποδείχτηκε λανθασμένη καθώς από τότε όλο και περισσότεροι οργανισμοί και ακαδημαϊκοί απασχολούνται στην ανάπτυξη τεχνολογιών και γνώσης πάνω στην διαχείριση των Big Data. Το 2012 στο διεθνές συνέδριο Business Intelligence Conference που φιλοξενήθηκε από το Special Interest Group on Decision Support Knowledge and Data Management Systems (SIGDSS) και το Teradata University Network (TUN) τέθηκαν στόχοι εκ των οποίων “ο εντοπισμός βέλτιστων πρακτικών καθώς και των κενών στην έρευνα και τη διδασκαλία της EN, με έμφαση στα Big Data” γεγονός που υποδεικνύει την ανάγκη για εξερεύνηση του τομέα από το ακαδημαϊκό επίπεδο με γνώμονα το αντίκτυπο στους οργανισμούς.

## **A.7 Ακαδημαϊκό Υπόβαθρο EN**

Σε έρευνα που διεξαχθεί στα πλαίσια του διεθνούς συνεδρίου BIC3<sup>6</sup> (Business Intelligence Conference), αναπτύχθηκαν σημαντικά και ενδιαφέροντα συμπεράσματα σχετικά με την ακαδημαϊκή έρευνα. Χαρακτηριστικά, μια απάντηση επαγγελματία στον χώρο αναφέρει “υπάρχει μεγάλη ανάγκη για άτομα με ικανότητες για EN και Analytics. Χρειαζόμαστε (σ.σ. η αγορά εργασίας) περισσότερους επιστήμονες δεδομένων με ικανότητες στην στατιστική, μοντελοποίηση και κατανόηση των δομημένων και αδόμητων δεδομένων.” Παράγοντας που περιπλέκει μια ήδη περίπλοκη δομή για την ακαδημαϊκή διδασκαλία είναι ο ρυθμός ανάπτυξης του τομέα. Στην ίδια έρευνα η απάντηση ενός ακαδημαϊκού περιγράφει γλαφυρά την ακαδημαϊκή κατάσταση. “Έχω δημιουργήσει μια διδακτική εμπειρία για τους μαθητές μου προσαρμοσμένη στην δουλειά μου στην αγορά εργασίας. Δεν χρησιμοποιώ σημειώσεις και βιβλία. Το μάθημα μπορεί να διδαχθεί ως σύγχρονη ή ασύγχρονη εμπειρία γνώσης”.

---

<sup>66</sup> Τα αποτελέσματα της έρευνας είναι διαθέσιμα στον παρακάτω σύνδεσμο:  
<http://www2.commerce.virginia.edu/bic3/content/2012%20State%20of%20BI%20FINAL.pdf>

Είναι σαφές πως ακαδημαϊκά υπάρχει ανάγκη για ταχύ εναρμονισμό με τις ανάγκες τις αγορές και την ανάπτυξη της τεχνολογίας. Λογικό φαίνεται το αποτέλεσμα της έρευνας σχετικά με τα διαθέσιμα προγράμματα πανεπιστημίων στον τομέα EN καθώς το 2012 (οπότε και διεξήχθη η έρευνα) αριθμούσε ο ακαδημαϊκός χώρος 131 προγράμματα πλήρους φοίτησης σε αντιπαράθεση με τα 15 προγράμματα το 2010. Πλέον ο αριθμός αυτών των προγραμμάτων έχει αυξηθεί, καθώς όλα τα ιδρύματα απασχόλησης στον τομέα της τεχνολογίας και των θετικών επιστημών προσφέρουν μόρφωση σε γενικό ή ειδικό επίπεδο (π.χ. Big Data στις Επιστήμες Υγείας) ενώ ενδιαφέρουσα τάση αφορά τα προγράμματα που προσφέρονται ως διεπιστημονική και διατομεακή εκπαίδευση (π.χ. επαγγελματίες του Marketing να επιλέγουν προγράμματα EN και Big Data).

Τα συμπεράσματα συνεδρίου BIC3 σχετικά με το αναπτυσσόμενο ενδιαφέρον των φοιτητών στον τομέα καθώς και την αυξανόμενη δράση και έρευνα από τα πανεπιστημιακά ιδρύματα σε σχέση με τις τάσεις της αγοράς μπορεί να συμπυκνωθεί στα παρακάτω σημεία ενδιαφέροντος:

- Ο αριθμός των προγραμμάτων που προσφέρουν εκ βάθους κατανόηση του τομέα έχει αυξηθεί δραματικά από το 2010 και ύστερα
- Η πρόσβαση και χρήση εργαλείων EN για εκπαιδευτικούς σκοπούς έχει αυξηθεί
- Η προσφορά συνεχίζει να υπολείπεται της προσφοράς για ακαδημαϊκούς απόφοιτους με ειδίκευση στον τομέα.
- Σε έναν κόσμο δεδομένων, οι θεμελιώδεις γνώσεις παραμένουν κρίσιμης σημασίας.
- Οι εργοδότες δεν είναι ικανοποιημένοι με την πρακτική εμπειρία των αποφοίτων.

Ένας τομέας τόσο ταχύτατα αναπτυσσόμενος όσο η EN προκαλεί ενθουσιασμό τόσο σε ακαδημαϊκό τόσο και σε επιχειρηματικό επίπεδο καθώς οι μελλοντικές ευκαιρίες δύσκολα γίνονται αντιληπτές σε αυτό το πρώιμο στάδιο. Το σίγουρο είναι πως μεγάλοι και μικρότεροι οργανισμοί συνεχίζουν να αναβαθμίζουν τον ρόλο της τεχνολογίας και της επιστήμης των δεδομένων ενώ παράλληλα τα Πανεπιστημιακά Ιδρύματα και ακαδημαϊκοί συνεχίζουν να επικεντρώνονται στην έρευνα και εκπαίδευση.

## ΚΕΦΑΛΑΙΟ Β – ΥΙΟΘΕΤΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΝΟΗΜΟΣΥΝΗΣ

Τα συστήματα EN αποτελούν προτεραιότητα για πολλούς CIO για πάνω από μια δεκαετία ωστόσο, λίγα πράγματα είναι γνωστά σχετικά με την αποτελεσματική διαχείριση αυτών των συστημάτων σε σχέση με τεχνικές γνώσεις στην φάση της υλοποίησης [18]. Πολλές και ενδιαφέρουσες τεχνολογίες έχουν αναπτυχθεί μέσα από τις ανάγκες της αγοράς όπως Real Time Analysis, Big Data, Analytics βασισμένα σε cloud τεχνολογίες καθώς επίσης ολοένα και πιο εύχρηστα, ευέλικτα και σύγχρονα λογισμικά EN διατίθενται παρά το απαιτητικό μακροοικονομικό περιβάλλον [19].

Λαμβάνοντας υπόψη όλες τις εξελίξεις στον τομέα, συχνά ισχυρίζεται το γεγονός ότι τα συστήματα έχουν ξεπεράσει τους επιχειρηματικούς περιορισμούς των περασμένων ετών ωστόσο η αφθονία των διαθέσιμων δεδομένων την σήμερα ημέρα επαναφέρει το ερώτημα για το αν πραγματικά έχει ξεπεραστεί το πρόβλημα του «πνιγμού στα δεδομένα, λόγω δίψας για πληροφορίες» [20]. Πολλές μελέτες έχουν σταθεί σε εμπειρικά αποτελέσματα όσον αφορά την επιτυχία αρχιτεκτονικών Αποθηκών Δεδομένων, στην επιλογή συγκεκριμένου προμηθευτή λογισμικού για λειτουργίες ανάλυσης, στην επίδραση της EN στην υποστήριξη λήψεων αποφάσεων ωστόσο ενδιαφέροντα ερωτήματα παραμένουν ανοιχτά ιδίως σε σχέση με τους μηχανισμούς που οδηγούν στην επιτυχία των συστημάτων [21]. Υπάρχουν αυξημένες ανάγκες για τεχνικές, τεχνολογικές, μαθηματικές οικονομικές και επιχειρηματικές γνώσεις. Όπως γίνεται αντιληπτό, η διαχείριση της πολυπλοκότητας της EN δεν μπορεί να είναι μια αυτόματη διαδικασία που επαφίεται σε μερικούς εξειδικευμένους τεχνολόγους που γνωρίζουν όλα τα μέρη και τις συνδέσεις των συστημάτων. Η EN απαιτεί αλλαγές στον τρόπο που οι άνθρωποι και οι οργανισμοί δουλεύουν και συνεργάζονται [22].

Κατά συνέπεια, η ερευνητική ερώτηση που τίθεται σε αυτό το κεφάλαιο έχει ως εξής: Ποιο είναι το αντίκτυπο της εφαρμογής EN και πως μπορεί να επιτευχθεί ο στόχος για βελτίωση της διαδικασίας λήψεως αποφάσεων;

## **B.1 Αντίκτυπο EN στην ποιότητα λήψης αποφάσεων**

Από θεωρητικής πλευράς, οι δυνατότητες της αποτελεσματικής διαχείρισης συστημάτων EN αποτελούν αντανάκλαση της αξιοποίησης των διαθέσιμων πόρων [23]. Για την επίτευξη αυτής της αξιοποίησης απαιτείτε στενή συνεργασία του τεχνολογικού και του διοικητικού δυναμικού των οργανισμών καθ' όλη τη διάρκεια του κύκλου ζωής της λύσης EN. Τα καίρια σημεία στα οποία οι δύο “κόσμοι” πρέπει να επιδείξουν αगाστή συνεργασία δεν είναι άλλα από το να ταιριάζουν οι επιχειρηματικές αποφάσεις με τις ανάγκες πληροφόρησης - να τεθούν οι σωστές ερωτήσεις, να καθιερωθεί η συνεχής υποστήριξη της διοίκησης - και να δεσμευτούν οι τελικοί χρήστες για την εκπαίδευση και την χρήση των νέων εργαλείων.

Στη διαχείριση της EN επαφίεται η ευθύνη για το σχεδιασμό, την υλοποίηση και τη λειτουργία της EN τόσο κατά το στάδιο των δεδομένων όσο και στο στάδιο της πληροφορίας. Ο σχεδιασμός ξεκινάει, ιδανικά, με ανάλυση απαιτήσεων, δηλαδή την πρόβλεψη των μελλοντικών αποφάσεων (θέτοντας τις σωστές ερωτήσεις), οι οποίες στη συνέχεια μεταφράζονται σε ανάγκες πληροφόρησης, και εν συνεχεία καθορίζει τις απαιτήσεις δεδομένων. Η διαδικασία υλοποίησης ακολουθεί την αντίθετη κατεύθυνση, αλλά και οι δύο διαδικασίες απαιτούν διατμηματική διαχείριση για τη διασφάλιση της συνεργασίας μεταξύ των χρηστών των πληροφοριών (υπεύθυνοι λήψης αποφάσεων) και των χειριστών δεδομένων (τμήμα IT) [24]. Βάση των παραπάνω καταλήγουμε στα εξής συμπεράσματα:

- Η ποιότητα διαχείρισης της EN σχετίζεται θετικά με την ποιότητα
  - των ληφθέντων επιχειρηματικών αποφάσεων
  - των δεδομένων.
  - των πληροφοριών
  - της λύσης EN
- Η ποιότητα των ληφθέντων επιχειρηματικών αποφάσεων σχετίζεται θετικά με την ποιότητα των πληροφοριών
- Η ποιότητα των πληροφοριών σχετίζεται θετικά με την ποιότητα των δεδομένων.
- Η ποιότητα ληφθέντων επιχειρηματικών αποφάσεων σχετίζεται θετικά με την ποιότητα την λύσης EN.

Εν κατακλείδι, και λαμβάνοντας υπόψιν το γεγονός πως η απόδοση επένδυσης όσον αφορά έργα σχετικά με την ΕΝ είναι δύσκολο να υπολογιστεί, φτάνουμε στο συμπέρασμα πως μόνο και μόνο όταν οι τεχνολογικοί πόροι και οι επιχειρηματικές ανάγκες εναρμονιστούν μέσω σωστής διαχείρισης οι οργανισμοί μπορούν να λάβουν τα μέγιστα οφέλη από την ΕΝ. Η διαδικασία αυτή, όπως και κάθε διαδικασία ανάπτυξης νέων τεχνικών και τεχνολογιών στους οργανισμούς, παρακολουθείται μέσω ενός μοντέλου ωριμότητας σκοπός του οποίου είναι να περιγράψουν, εξηγήσουν και αξιολογήσουν τα στάδια ανάπτυξης του έργου [25].

## **B.2 Μοντέλο Ωριμότητας ΕΝ (BI Maturity Model)**

Τα μοντέλα ωριμότητας όπως αναφέρθηκε παραπάνω, σκοπό έχουν να περιγράψουν, εξηγήσουν και αξιολογήσουν τα στάδια ανάπτυξης κατά την διάρκεια του κύκλου ζωής ενός έργου ανάπτυξης μιας νέας τεχνολογίας ή διαδικασίας. Ο τρόπος σκέψης πίσω από τα μοντέλα ωριμότητας είναι αρκετά απλός και ευκατανόητος καθώς βασίζεται στο γεγονός ότι ένα έργο αναπτύσσεται σε χρονικές περιόδους η κάθε μία των οποίων έχει ιδιαίτερα χαρακτηριστικά, διέπτετε από ξεχωριστούς κανόνες, ενέχει κινδύνους και παρουσιάζει ευκαιρίες ξεχωριστούς ανά επίπεδο. Κατά μια έννοια θα μπορούσαμε να πούμε ότι παρομοιάζουμε τον κύκλο ανάπτυξης με τον κύκλο ζωής των ανθρώπων. Αρκετά μοντέλα ωριμότητας χωρίζουν τα στάδια σε αντίστοιχα στάδια της ανάπτυξης ενός ανθρώπου (στάδιο νεογνού, εφήβου, ενήλικα κοκ) και κάπως έτσι είναι εύκολο ακόμα και για άτομα άλλων κλάδων ή τμημάτων να παρακολουθήσουν και να κατανοήσουν την διαδικασία ανάπτυξης καθώς επίσης και να συγκρίνουν οργανισμούς και έργα μεταξύ τους μελετώντας το στάδιο στο οποίο βρίσκονται [25].

Τα μοντέλα ωριμότητας ΕΝ είναι σημαντικά στην διαδικασία ανάπτυξης καθώς περιγράφουν την διαδρομή και βοηθούν τον οργανισμό να προβεί στις κατάλληλες ενέργειες που θα εναρμονίσουν τις ανάγκες του οργανισμού με τις απαραίτητες διαθέσιμες τεχνολογίες. Τα μοντέλα ωριμότητας προέρχονται από το Μοντέλο Ωριμότητας Δυνατοτήτων (Capability Maturity Model, CMM) το οποίο αναπτύχθηκε για να περιγράψει τα στάδια ανάπτυξης λογισμικού βασισμένο στην εργασία με τίτλο «Διατριβή Ωριμότητας» (Maturity Thesis) [26].



Εν μέσω πληθώρας εργασιών, δημοσιεύσεων και διατριβών οι περισσότεροι ακαδημαϊκοί χρησιμοποιούν το μοντέλο του W. Eckerson καθώς είναι αρκετά χρήσιμο για να εκθέσουν και εξακριβώσουν τις ανάγκες και τα χαρακτηριστικά της τεχνικής πλευράς των έργων ανάπτυξης EN [27]. Ο W. Eckerson ανέπτυξε το TDWI<sup>7</sup>'s Business Intelligence Maturity Model το 2004<sup>8</sup>, το μοντέλο αυτό επικεντρώνεται περισσότερο στα τεχνικά σημεία της ανάπτυξης της EN σε έναν οργανισμό. Η ωριμότητα αξιολογείται σε οκτώ κομβικά σημεία

1. πεδίο εφαρμογής
2. χορηγία
3. χρηματοδότηση
4. αξία
5. αρχιτεκτονική
6. δεδομένα
7. ανάπτυξη
8. παράδοση

και χωρίζεται σε πέντε διαφορετικά στάδια ωριμότητας

#### *Πρώτο Στάδιο ωριμότητας - Βρέφος (infant)*

Αυτό το στάδιο περιέχει ένα μικρότερο στάδιο, το προγενέθλιο (prenatal). Σε αυτό το στάδιο χρησιμοποιούνται ως πηγές δεδομένων τοπικά αρχεία που περιέχουν δεδομένα χωρίς συγκεκριμένη δομή. Είναι αρχεία υπολογιστικών φύλλων ονομαζόμενα ως Spreadmarts τα οποία οι εργαζόμενοι αναπτύσσουν σύμφωνα με μερικούς κανόνες αόριστα δομημένους οι οποίοι προκαλούν σύγχυση όταν διαφορετικοί χρήστες επεμβαίνουν στα δεδομένα. Ωστόσο αποτελεί μια πρώτη προσπάθεια για διαχείριση δεδομένων.

#### *Δεύτερο Στάδιο ωριμότητας - Παιδί (Child)*

Σε αυτό το στάδιο τα δεδομένα συγκεντρώνονται και διαχειρίζονται για τους σκοπούς ενός τμήματος του οργανισμού και καλύπτουν τις ανάγκες των

---

<sup>7</sup> TDWI είναι ακρωνύμιο του οργανισμού τεχνολογιών δεδομένων The Data Warehouse Institute

<sup>8</sup> Περισσότερες πληροφορίες σχετικά με το Business Intelligence Maturity Model από τον οργανισμό TDWI στον παρακάτω σύνδεσμο:

[http://www.tdwi.eu/wissen/whitepaper/?no\\_cache=1&tx\\_mwknowledgebase\\_pi1%5BshowUid%5D=46](http://www.tdwi.eu/wissen/whitepaper/?no_cache=1&tx_mwknowledgebase_pi1%5BshowUid%5D=46)

εργαζομένων σε αυτό το τμήμα μόνο. Δεν παρέχονται υπηρεσίες ενσωμάτωσης και διαμοιρασμού με άλλα τμήματα εντός του οργανισμού καθώς επίσης και δεν μοιράζονται εκτός. Ωστόσο αναπτύσσονται τα πρώτα εργαλεία ανάλυσης και πιθανά και η απόκτηση κάποιου λογισμικού ανάλυσης και απεικόνισης των δεδομένων με σκοπό την ανάπτυξη του τμήματος. Τα δεδομένα προέρχονται συνήθως παρέχονται από ήδη καθιερωμένες τεχνολογίες.

### *Τρίτο Στάδιο ωριμότητας - Έφηβος (Teenager)*

Ο οργανισμός αναγνωρίζει σιγά σιγά την ανάγκη για καθολική εφαρμογή EN και διαμορφώνει τους κανόνες και τις αρχές του έργου. Μελετάει τις διαθέσιμες τεχνολογίες, μεθοδολογίες, ανοίγει δίαυλους επικοινωνίας με γνώστες του αντικειμένου στην αγορά και συμβουλευεται εξωτερικούς συνεργάτες. Μια πλατφόρμα διαλέγεται με σκοπό την ανάπτυξη μιας ολοκληρωμένης λύσης από μία ομάδα εργαζομένων οι οποίοι πλέον αποτελούν ένα ξεχωριστό τμήμα και κομμάτι του οργανισμού. Αναγνωρίζεται η ανάγκη για την ανάπτυξη αρχιτεκτονικής Αποθήκης Δεδομένων δομημένη κεντρικά ώστε να εξυπηρετεί βασικές ανάγκες των διαφόρων τμημάτων. Αρχίζουν να γνωρίζουν σταδιακά όλοι οι εργαζόμενοι τις πληροφορίες που απορρέουν από τα δεδομένα του οργανισμού έχοντας πρόσβαση σε αναφορές, γραφήματα και πίνακες γενικής χρήσης με απλές δυνατότητες ατομικής διαχείρισης.

### *Τέταρτο Στάδιο ωριμότητας - Ενήλικος (Adult)*

Πλέον ο οργανισμός λειτουργεί και παίρνει αποφάσεις σύμφωνα με τις πληροφορίες που αντλούνται από την EN. Ο τομέας αναλαμβάνει στρατηγική σημασία πέραν της απλής πληροφόρησης και οι καθημερινές διαδικασίες παρακολουθούνται και ελέγχονται μέσω της EN. Το τμήμα ανάπτυξης και διαχείρισης EN αποτελεί αυτόνομο κομμάτι του οργανισμού με δικό του καταρτισμένο προσωπικό και ανεξάρτητο από άλλα τεχνολογικά τμήματα. Τα κύρια χαρακτηριστικά ενός οργανισμού σε αυτό το στάδιο είναι κυρίως η συγκεκριμένη δομή των δεδομένων και της Αποθήκης Δεδομένων. Τα αποτελέσματα ανάλυσης παραδίδονται σε μικρό χρονικό διάστημα μετά το συμβάν που αφορούν (αν όχι σε πραγματικό χρόνο). Τέλος, πιθανόν να

υλοποιείται και η προγνωστική ανάλυση. Σε αυτό το στάδιο πλέον, η χρήση πολυσύνθετων τεχνολογιών και εργαλείων παρέχει ακρίβεια και αξιοπιστία σε όλα τα επίπεδα του οργανισμού.

#### *Πέμπτο Στάδιο ωριμότητας - Σοφός (Sage)*

Σε αυτό το επίπεδο οι βασικοί πυλώνες του οργανισμού συνεργάζονται δημιουργώντας ένα κέντρο αριστείας (Center of Excellence) [28] ενώ χρησιμοποιούνται αρχιτεκτονικές προσανατολισμένες στις υπηρεσίες (Service Oriented Architecture, SOA) για την περαιτέρω ανάπτυξη της EN [29]. Οι δυνατότητες του συστήματος EN συμπεριλαμβάνουν την δυνατότητα δημιουργίας αναφορών από τους χρήστες εξυπηρετώντας ιδιαίτερες ανάγκες. Ξεχωριστές ομάδες εντός των διαφόρων τμημάτων επωμίζονται την δημιουργία εξειδικευμένων λύσεων ανάλυσης και παρουσίασης δεδομένων. Ο αριθμός των χρηστών έχει αυξηθεί δραματικά όπως επίσης και των ατόμων με ειδικά πιστοποιητικά χρήσης και ανάπτυξης τεχνολογιών.

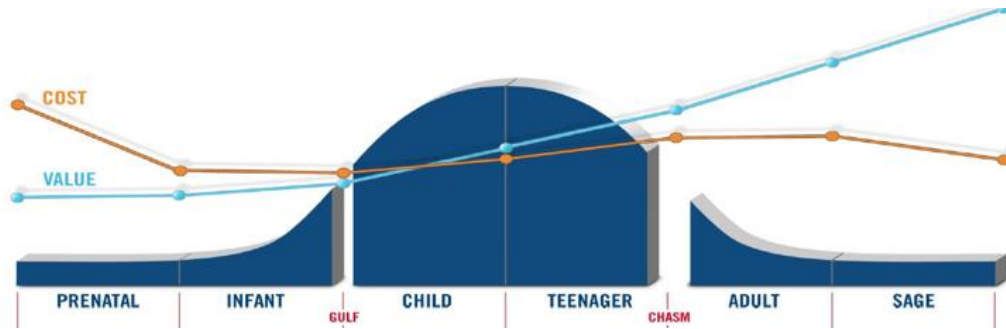
#### *Τα χάσματα μεταξύ σταδίων (Gulf & Chasm)*

Κατά την μετάβαση από το στάδιο βρέφους στο στάδιο του σοφού, οι οργανισμοί έρχονται αντιμέτωποι με δύο μεγάλα χάσματα στα οποία έχουν αποδοθεί οι ορισμοί Gulf και Chasm.

Κατά το χάσμα Gulf, το οποίο παρατηρείτε μεταξύ των δύο πρώτων σταδίων (βρέφος και παιδί), υπάρχει ο κίνδυνος ο οργανισμός να μην προχωρήσει από το στάδιο βρέφος στο στάδιο παιδί για διάφορους λόγους. Τα εμπόδια δεν είναι τεχνολογικά (πχ δυνατότητα ανάπτυξης ολοκληρωμένης λύσης Αποθήκης Δεδομένων) αλλά αφορούν κυρίως τον λανθασμένο επιχειρηματικό σχεδιασμό, την κακή ποιότητα δεδομένων, την αρνητική κουλτούρα του οργανισμού καθώς και την έντονη χρήση spreadsheets τα οποία οι χρήστες δεν δέχονται να καταργήσουν καθώς μοιάζουν καλύτερη ιδέα από μια κεντροποιημένη λύση.

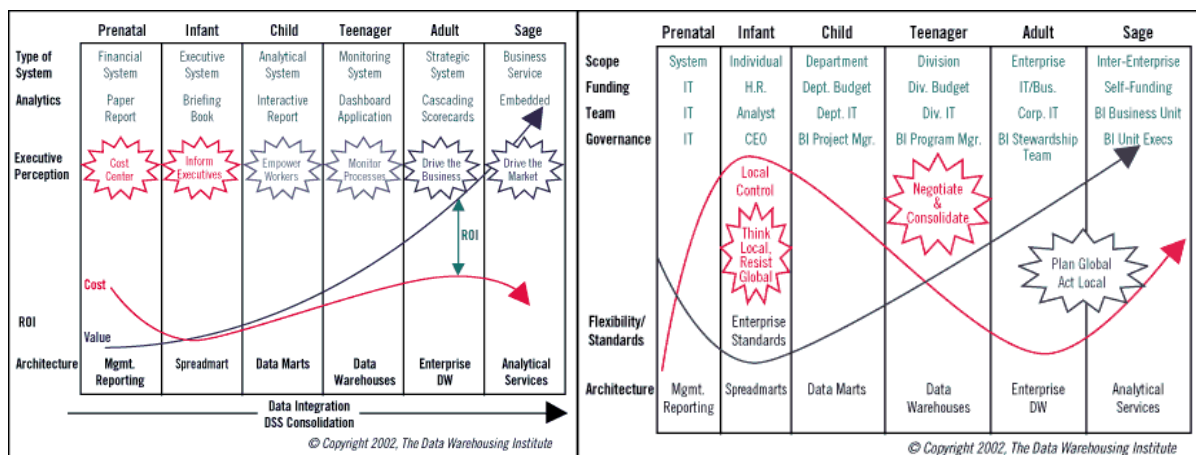
Κατά το χάσμα Chasm εμφανίζονται εμπόδια που δυσκολεύουν την μετάβαση από το στάδιο εφήβου στο στάδιο ενηλίκου τα οποία προέρχονται κυρίως από το διοικητικό κομμάτι του οργανισμού. Τέτοια εμπόδια αφορούν την αδιαφορία της διοίκησης για περαιτέρω διάθεση πόρων προς τις προχωρημένες τεχνολογίες και

την εκπαίδευση των χρηστών, την εισαγωγή γραφειοκρατίας και κανόνων με αποτέλεσμα την ακινητοποίηση των τεχνολογιών καθώς οι διαδικασίες κολλάνε σε διαδικαστικά θέματα. Από την πλευρά της ανάπτυξης μια ενδεχόμενη στυφνή αρχιτεκτονική ενδέχεται να προκαλέσει δυσκολίες στην μεταφορά σε νέες δομές.



Σχήμα 2.1: TDWI Μοντέλο Ωριμότητας Επιχειρησιακής Νοημοσύνης

Το μοντέλο ωριμότητας αναλύεται περαιτέρω παραθέτοντας τα παρακάτω γραφήματα όπου και διαφαίνονται όλα τα στάδια σε συνδυασμό με τα κομβικά σημεία κατά τα οποία αξιολογείται το μοντέλο ωριμότητας. Σε πιο λεπτομερή ανάλυση, στο αριστερό γράφημα έχουμε τις σχέση κόστους και αξίας υπολογίζοντας την διαφορά τους ως την απόδοση επένδυσης (σε θεωρητικό επίπεδο) ενώ στο δεξιό γράφημα φαίνεται η διαφορά μεταξύ τοπικού ελέγχου της τεχνολογίας και αξίας για τον οργανισμό εξάγοντας συμπεράσματα για την κουλτούρα που απαιτείται για κάθε μετάβαση.



Σχήμα 2.2.α: Σχέση κόστους αξίας ανά στάδιο του Μοντέλου Ωριμότητας EN και Ανάλυσης.

Σχήμα 2.2.β: Σχέση κόστους και αξίας ανά στάδιο του Μοντέλου Ωριμότητας EN και Ανάλυσης

Έχοντας ως βάση τα παραπάνω ο Eckerson εκτιμάει πως οι οργανισμοί μπορούν εύκολα να καταλάβουν σε ποιο στάδιο βρίσκονται και να λάβουν τις σωστές αποφάσεις για την αποφυγή των χασμάτων και το πέρασμα από το ένα στάδιο στο άλλο έτσι ώστε να μπορέσουν να αξιοποιήσουν στο μέγιστο της δυνατότητας της EN.<sup>9</sup>

### **B.3 Κρίσιμοι Παράγοντες Επιτυχίας**

Η επιτυχημένη εφαρμογή EN παραμένει μία τεχνολογική και επιχειρηματική πρόκληση. Πολλά έργα EN ενέχουν υψηλό ρίσκο παρόλα αυτά όλο και περισσότεροι οργανισμοί βλέπουν με ενθουσιασμό το ρίσκο αυτό [30]. Ο ενθουσιασμός προέρχεται από τις ευκαιρίες και την ανάπτυξη που προσφέρει η EN καθώς συχνά περιγράφεται ως “μια ευρεία κατηγορία τεχνολογιών, εφαρμογών και διαδικασιών για τη συλλογή, αποθήκευση, πρόσβαση, και ανάλυση των δεδομένων βοηθώντας τους χρήστες να λάβουν καλύτερες αποφάσεις” [31]. Ωστόσο η διαδικασία της εφαρμογής δεν επαφίεται μονάχα σε αποφάσεις σχετικά με την επιλογή του κατάλληλου λογισμικού και υλικού αλλά πολυσύνθετες απαιτήσεις υποδομών και πόρων. Παρά την έντονη ανάπτυξη της EN και την πολυπλοκότητα των εφαρμογών συστημάτων EN, οι Κρίσιμοι Παράγοντες Επιτυχίας (Critical Success Factors, CSFs) των πρωτοβουλιών εφαρμογής τέτοιων συστημάτων παραμένουν ελάχιστα κατανοητοί [30].

Οι Κρίσιμοι Παράγοντες Επιτυχίας, CSFs σαν όρος χρησιμοποιήθηκε για πρώτη φορά από τον D. Ronald Daniel στο άρθρο του 1961 στο Harvard Business Review με τίτλο “Management Information Crisis”. Ο D. Ronald Daniel τόνισε τα είδη των πληροφοριών που απαιτούνται για τη στήριξη δραστηριοτήτων διοίκησης. Ανέφερε, μεταξύ άλλων, πως ένα πληροφοριακό σύστημα θα πρέπει να είναι επιλεκτικό και επικεντρωμένο στην παροχή λεπτομερειών γύρω από τρεις έως έξι παράγοντες επιτυχίας που βοηθούν τον οργανισμό να επιτύχει [32]. Το επίσημο Γλωσσάριο ITIL<sup>10</sup> ορίζει τους CSFs ως “κάτι που πρέπει να συμβεί ώστε μια διαδικασία, ένα έργο, ένα σχέδιο, ή μια υπηρεσία να πετύχει”.

<sup>9</sup> Περισσότερες πληροφορίες στην δημοσίευση του Eckerson, [http://www.cbpp.uaa.alaska.edu/afef/gauge\\_your\\_data\\_warehouse\\_maturi.htm](http://www.cbpp.uaa.alaska.edu/afef/gauge_your_data_warehouse_maturi.htm)

<sup>10</sup> Περισσότερες πληροφορίες στον παρακάτω σύνδεσμο [https://www.axelos.com/Corporate/media/Files/Glossaries/ITIL\\_2011\\_Glossary\\_GB-v1-0.pdf](https://www.axelos.com/Corporate/media/Files/Glossaries/ITIL_2011_Glossary_GB-v1-0.pdf)

Πληθώρα μελετών έχουν παρουσιαστεί ανά τα έτη σχετικά με τους CSFs στα Συστήματα Πληροφοριών (και ως εκ τούτου και της EN) που καθορίζουν την επιτυχία ή αποτυχία της εφαρμογής τους. Μια μελέτη με μεγάλη απήχηση και επιρροή προτείνει την αντιμετώπιση Συστημάτων Πληροφοριών ως πολύπλευρο κατασκεύασμα [33] ωστόσο δεν υπάρχουν κοινά αποδεκτά κριτήρια για την εφαρμογή τέτοιων συστημάτων και την πραγματοποίηση έργων Συστημάτων Πληροφορικής [30].

Στην δημοσίευση των William Yeoh και Aleš Popovic στο Journal of the Association for Information Science and Technology τον Ιανουάριο του 2016, αντλώντας από προηγούμενες μελέτες, αναλύεται μια προσέγγιση για την αξιολόγηση εφαρμογής EN. Καταλήγοντας σε συγκεκριμένους CSFs, με στόχο την καλύτερη κατανόηση, λαμβάνονται υπόψιν δύο βασικές διαστάσεις (παρακάτω γράφημα): η απόδοση της διαδικασίας (δηλαδή πόσο πετυχημένη χαρακτηρίζεται η διαδικασία της εφαρμογής ενός συστήματος EN), και η επίδοση των υποδομών (δηλαδή, την ποιότητα του αποτελέσματος). Αυτές οι δύο διαστάσεις δεν είναι αλληλένδετες, είναι αρκετά πιθανό για παράδειγμα ένα έργο με τελικό κόστος μεγαλύτερο του προβλεπόμενου καθώς και εκτός χρονοδιαγράμματος να αποδίδει υψηλής ποιότητας πληροφορίες, αντιστρόφως ένα έργο εντός κοστολογίου και χρονοδιαγράμματος να παρέχει κακής ποιότητας αναφορές.

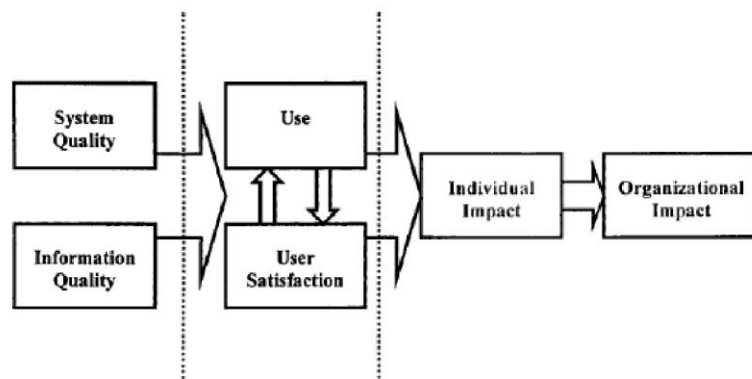


Figure 1. D&M IS Success Model.

Reprinted by permission, W. DeLone and E. McLean, Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3(1), 1992, pp. 60–95. Copyright 1992, The Institute of Management Sciences (now INFORMS), 901 Elkridge Landing Road, Suite 400, Linthicum, MD 21090 USA.

### Σχήμα 2.3: Παράγοντες επιτυχίας Συστημάτων Πληροφοριών κατά τους DeLone και McLean

Η επίδοση των υποδομών έχει ομοιότητες με τις τρεις κύριες διαστάσεις του μοντέλου επιτυχίας των Πληροφοριακών Συστημάτων [33], δηλαδή, την ποιότητα του συστήματος, την ποιότητα της πληροφορίας και χρήση του συστήματος.

- Η ποιότητα του συστήματος αντανακλά τα χαρακτηριστικά απόδοσης της επεξεργασίας πληροφοριών και το ίδιο το σύστημα. Αξιολογείται μέσω της ευελιξίας ενός συστήματος, της επεκτασιμότητας και της ικανότητας να ενσωματώσουν δεδομένα.
- Η ποιότητα πληροφοριών αναφέρεται στην ακρίβεια, την πληρότητα, την επικαιρότητα, την συνάφεια, την συνοχή, και τη χρησιμότητα των πληροφοριών που παρέχονται από το σύστημα.
- Η χρήση του συστήματος θεωρείται ως η κατανάλωση ενός τελικού χρήστη των παρεχόμενων αναφορών του συστήματος.

Οι Yeoh και Poronic μετά από ανάλυση διαφόρων οργανισμών και μελετώντας την εφαρμογή EN από τρεις διαφορετικές σκοπιές (οργανισμός, διαδικασία και τεχνολογία) κατέληξαν στο συμπέρασμα ότι για την απόφαση ενός οργανισμού για εφαρμογή EN υπάρχουν ποικίλα κίνητρα σχετικά με την διευκόλυνση της διαχείρισης επιδόσεων. Ένα αξιοσημείωτο κίνητρο εξίσου σημαντικό και κοινό σε όλες τις περιπτώσεις ήταν η ανάγκη για μια κοινή εκδοχή της πραγματικότητας μεταξύ διαφόρων χρηστών (διοίκηση, πωλήσεις, marketing, συνεργάτες κ.α.). Η ορολογία που υπάρχει στον κλάδο των επιχειρήσεων για να περιγραφεί η κοινή εκδοχή της πραγματικότητας αναφέρεται ως “single version of the truth (SVOT)” [30]. Συμπερασματικά και μετά από ανάλυση των περιπτώσεων η αναφερθείσα έρευνα καταλήγει στους παρακάτω CSFs:

- Δέσμευση και υποστήριξη από την Διοίκηση του οργανισμού
- Ξεκάθαρο όραμα και δομημένη αντίληψη των αναγκών
- Επιχειρηματοκεντρική και ισορροπημένη σύνθεση ομάδας έργου
- Καθοδήγηση βάση αναγκών του οργανισμού και ανάπτυξη σε βάθος χρόνου
- Διαχείριση αλλαγών με γνώμονα τον χρήστη του συστήματος
- Επιχειρηματοκεντρικό, επεκτάσιμο και ευέλικτο τεχνικό πλαίσιο
- Βιώσιμη ποιότητα και ακεραιότητα δεδομένων

Τα αποτελέσματα της έρευνας υποδεικνύουν πως περιπτώσεις στις οποίες ένας οργανισμός έβαλε σε προτεραιότητα τις επιχειρηματικές ανάγκες πάνω από τις επιμέρους ανάγκες των μερών του έργου είχε μεγαλύτερες πιθανότητες επιτυχίας ανάπτυξης του συστήματος. Είναι δηλαδή ιδιαίτερα σημαντική η προσοχή έτσι ώστε να μην λαμβάνονται υπόψιν κυρίως τα τεχνολογικά θέματα αλλά τα θέματα που αφορούν τους στόχους του οργανισμού και την εκτέλεση της διαδικασίας. Εν κατακλείδι οι ηγέτες του οργανισμού και του έργου παίζουν σπουδαίο ρόλο στην διαχείριση της ανάπτυξης των κύριων σημείων ενδιαφέροντος και στην συνεχή προσφορά εκπαίδευσης, καθοδήγησης, υποστήριξης, αναθεώρησης του έργου καθώς και στην δέσμευση του προσωπικού των επιμέρους συνεργαζόμενων τομέων [30].

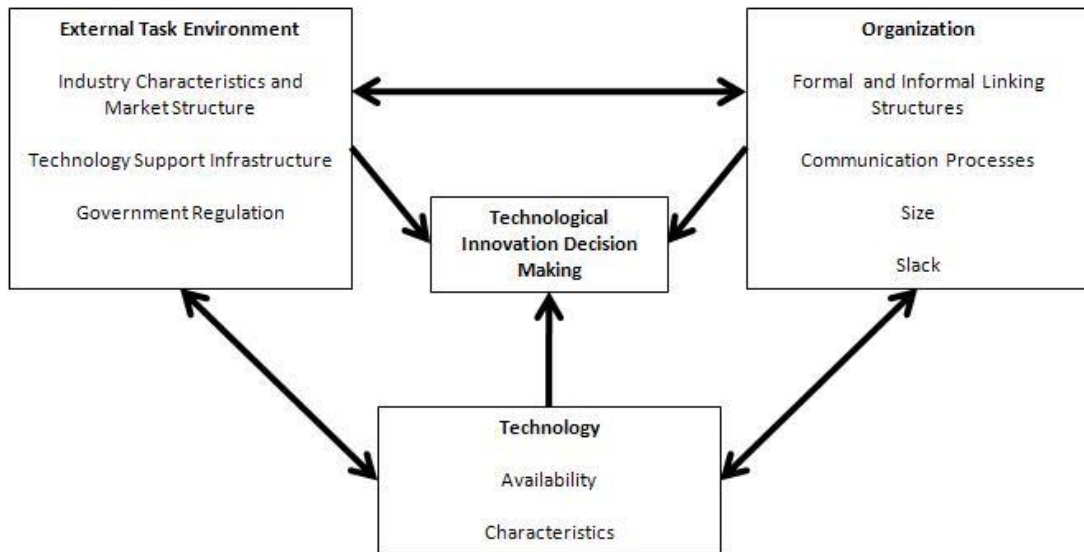
#### **B.4 Παράγοντες Υιοθέτησης Καινοτομιών**

Παρά το γεγονός ότι η EN αποτελεί σημαντική καινοτομία για έναν οργανισμό, πολλές μελέτες έχουν διεξαχθεί για την αξιολόγηση των καθοριστικών παραγόντων για την ορθή εφαρμογή της καινοτομίας. Πολλοί οργανισμοί προσπαθούν να επικεντρωθούν σε στοιχεία που πρέπει να συμπεριλάβουν στην προοπτική της καινοτομίας [34].

Οι Tornatzky και Fleischer μελέτησαν την διαδικασία της καινοτομίας και παρουσίασαν το 1990 το πλαίσιο Τεχνολογία-Οργανισμός-Περιβάλλον (Technology-Organization-Environment, TOE, Framework). Το πλαίσιο αυτό προβλέπει κυρίως ότι, όταν ένας οργανισμός αποφασίζει να εγκρίνει ή να εφαρμόσει μια νέα τεχνολογία πληροφοριών, η διαδικασία λήψης αποφάσεων θα πρέπει να επηρεαστεί από την τεχνολογία, την οργάνωση, και το περιβάλλον [35].

- **Τεχνολογία:** περιλαμβάνει τη σχετική τεχνολογία εντός και εκτός του οργανισμού όπως ο εξοπλισμός (software και hardware) και οι διαδικασίες.
- **Οργανισμός:** περιλαμβάνει επιχειρηματικά χαρακτηριστικά και τους πόρους, όπως το μέγεθος της επιχείρησης και ο βαθμός κεντρικοποίησης, η διοικητική δομή, το ανθρώπινο δυναμικό, και οι διασυνδέσεις μεταξύ των εργαζομένων.
- **Περιβάλλον:** περιλαμβάνει ολόκληρη τη δομή της αγοράς, το μέγεθος της αγοράς, τον ανταγωνισμό, το μακροοικονομικό πλαίσιο, και το κανονιστικό περιβάλλον.



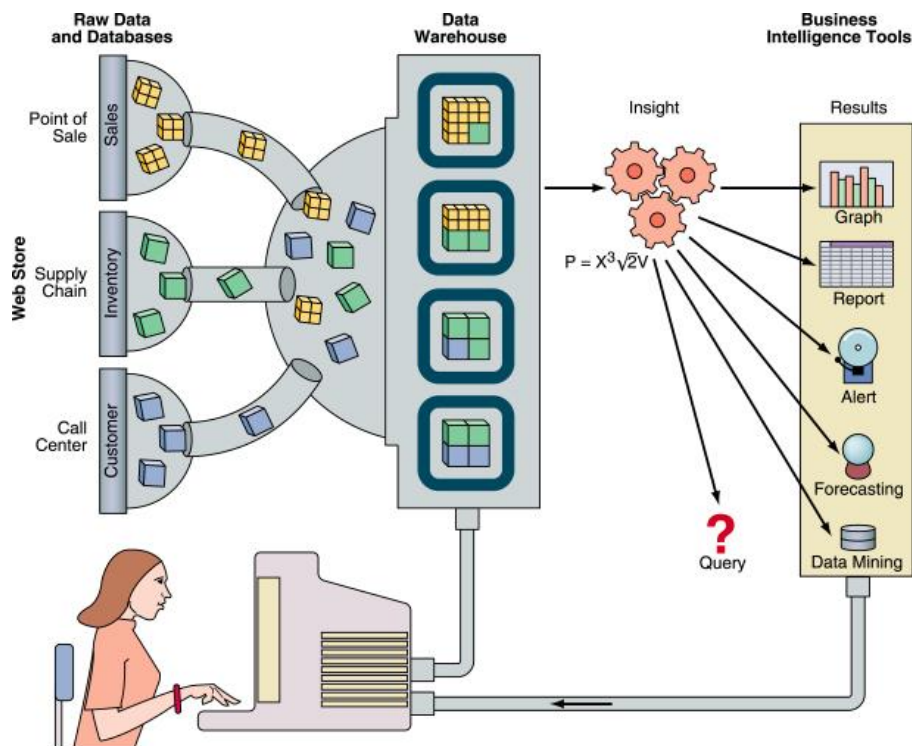


**Σχήμα 2.4:** Συσχετίσεις μεταξύ των παραγόντων υιοθέτησης καινοτομιών

Οι δύο επιστήμονες στην δημοσίευσή τους εξηγούν ότι αν και αυτά τα τρία πλαίσια δημιουργούν κάποιες ευκαιρίες για τους οργανισμούς να υιοθετήσουν την τεχνολογική καινοτομία, φέρνει επίσης κάποιους περιορισμούς. Ως εκ τούτου, είναι απαραίτητο να γίνει μια αυτό-αξιολόγηση πριν ληφθεί η απόφαση υιοθέτησης καινοτομίας ώστε να υπάρχει μια σαφή εικόνα και να προσαρμοστεί η λειτουργική δομή της και να μεγιστοποιηθεί το ανταγωνιστικό πλεονέκτημα [36]. Το παραπάνω πλαίσιο αποτελεί ένα σημαντικό εργαλείο στην διάθεση των οργανισμών για την σωστή προσέγγιση της εφαρμογής την EN ως καινοτομία.

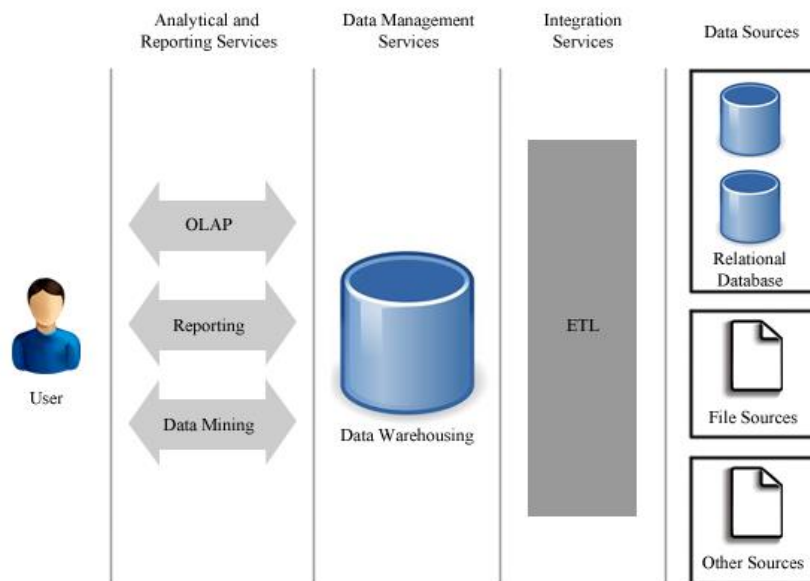
## ΚΕΦΑΛΑΙΟ Γ – ΕΠΙΠΕΔΟ ΕΦΑΡΜΟΓΩΝ

Καθώς η ΕΝ αποτελεί ένα σύνολο λύσεων και προτάσεων προς οργανισμούς για την διαχείριση των δεδομένων είναι εύκολα κατανοητό πως έχει αναπτυχθεί ένας ιστός από τεχνολογίες, εργαλεία και στοιχεία σε θεωρητικό και πρακτικό επίπεδο ο οποίος εξελίσσεται όσο αυξάνονται οι ανάγκες και ο ανταγωνισμός. Καθώς όλα αυτά τα στοιχεία που αποτελούν πλέον την ΕΝ είναι αλληλένδετα, είναι αναγκαίο να ορίσουμε τα ποια βασικά εξ αυτών με σκοπό την εις βάθος κατανόηση του επιπέδου εφαρμογής. Το παρακάτω σχήμα [37] παρουσιάζει απλουστευμένα το σύστημα της ΕΝ καθώς επίσης και τον τρόπο με τον οποίο δεδομένα από ανομοιογενής πηγές εξαγονται και αποθηκεύονται με σκοπό την ανάκτηση για ανάλυση τους.



**Σχήμα 3.1:** Το σύστημα ΕΝ από την εξαγωγή δεδομένων έως την ανάκτηση και ανάλυση

Οι κύριες δραστηριότητες περιλαμβάνουν τη συλλογή, την ενσωμάτωση, την διαχείριση και αποθήκευση, και την ανάλυση των δεδομένων. Οι διάφορες πηγές δεδομένων συλλέγονται, μετασχηματίζεται, καθαρίζονται, φορτώνονται και αποθηκεύονται σε μια Αποθήκη Δεδομένων. Ένα ορθά δομημένο σύστημα ΕΝ εκμεταλλεύεται πλήρως τα δεδομένα σε κάθε φάση εξέλιξης [37], από την διαχείριση των πηγών μέχρι τις υπηρεσίες αναφοράς. Οι πληροφορίες που διατίθενται στα διάφορα επιχειρησιακά περιβάλλοντα δημιουργούνται από τα ακατέργαστα δεδομένα.



**Σχήμα 3.2:** Υπηρεσίες ΕΝ στο επίπεδο εφαρμογών

Η πρόκληση στον σχεδιασμό του επιπέδου εφαρμογών είναι η όσο το δυνατόν η ορθότερη δημιουργία συνδέσεων μεταξύ των διαφόρων υπηρεσιών. Στο παραπάνω σχήμα φαίνονται όλες οι υπηρεσίες ΕΝ στο επίπεδο εφαρμογών, με την ανάλυση των οποίων ασχολείται το κεφάλαιο αυτό.

## Γ1. Πηγές Δεδομένων

Ως πηγές δεδομένων αναφέρονται όλες οι παραγωγικές βάσεις δεδομένων, τα ιστορικά στοιχεία πριν την μηχανογράφηση, τα εκτός οργανισμού δεδομένα (internet metrics, social media data, έρευνες-ερωτηματολόγια κ.α.) καθώς και δεδομένα πραγματικού χρόνου (live data).

Οι πηγές των δεδομένων μπορεί να είναι σχεσιακές βάσεις δεδομένων ή οποιαδήποτε άλλη δομή δεδομένων που υποστηρίζει τη γραμμή των επιχειρηματικών εφαρμογών. Μπορούν επίσης να βρίσκονται σε πολλές διαφορετικές πλατφόρμες και μπορεί να περιέχουν δομημένες πληροφορίες, όπως πίνακες ή λογιστικά φύλλα, ή αδόμητες πληροφορίες, όπως αρχεία κειμένου ή εικόνων και άλλων πληροφοριών πολυμέσων.

### *Γ1.1 Δεδομένα Πληροφορία και Γνώση*

Θα μπορούσαμε να πούμε πως η EN μετατρέπει τα δεδομένα σε πληροφορίες και γνώση για έναν οργανισμό. Ακαδημαϊκά ο διαχωρισμός δεδομένων, πληροφοριών, γνώσεων και σοφίας είναι γνωστός και ως DIKW Pyramid (Data Information Knowledge Wisdom Pyramid). Διάφοροι ακαδημαϊκοί έχουν ερευνήσει και δημοσιεύσει τα δικά τους συμπεράσματα και θεωρίες αλλά και διαφορετικές εκδοχές αυτού του μοντέλου. Συνοπτικά μπορούμε να ορίσουμε [38]:

- Δεδομένα είναι βασικά επιμέρους στοιχεία αριθμητικής ή άλλης μορφής τα οποία συγκεντρώνονται μέσω κάποιου είδους παρατήρησης
- Πληροφορία είναι το αποτέλεσμα ανάλυσης και ερμηνείας των δεδομένων λαμβάνονται ενδεχόμενος και υπόψιν το πλαίσιο στο οποίο δημιουργήθηκαν τα δεδομένα
- Γνώση είναι η γενική κατανόηση και συνειδητοποίηση που παρέχεται από την συσσώρευση πληροφοριών
- Σοφία είναι η μορφή της γνώσης σε ηθική κρίση και πράξεις



**Σχήμα 3.3:** Η πυραμίδα της γνώσης, DIKW Pyramid (Data Information Knowledge Wisdom Pyramid)

Για τη μετάβαση από τα δεδομένα, στη πληροφορία, στη γνώση και τέλος στη σοφία, το πρώτο και βασικό στοιχείο είναι τα δεδομένα και η πηγή από όπου προέρχονται. Τα δεδομένα μπορεί να προέρχονται από τις παρακάτω πηγές

- Εργαλεία διαχείρισης συναλλαγών
- Ενδοεταιρική πηγή κύριων δεδομένων
- Μη αυτόματες πηγές
- Εξωτερικά συνεργαζόμενες πηγές

#### *Γ1.2 Δεδομένα διαχείρισης συναλλαγών (transactional sources)*

Στις μέρες μας υπάρχουν πολλοί τρόποι συναλλαγών οι οποίοι πραγματοποιούνται μέσα από πληθώρα συστημάτων με ειδικά σχεδιασμένα ή τροποποιημένα software και hardware (TPS, transactional processing systems). Παραδείγματα είναι η ηλεκτρονική κράτηση εισιτηρίων, η αγορά εφαρμογών στις ηλεκτρονικές αγορές των κινητών τηλεφώνων, η ανάληψη από ATM, η αγορά σε κάποιο κατάστημα με ηλεκτρονική καταγραφή συναλλαγών κ.α. Η πρώτη επιτυχής προσπάθεια εργαλείων διαχείρισης συναλλαγών μας πηγαίνει πίσω στο 1960 και το Sabre<sup>11</sup>, το πρώτο σύστημα κεντρικής διαχείρισης κρατήσεων αεροπορικών

---

<sup>11</sup> Περισσότερες πληροφορίες για το Sabre στον παρακάτω σύνδεσμο  
<http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/sabre/>

εισιτηρίων που κατασκευάστηκε από την IBM για λογαριασμό της American Air Lines και υποστήριζε μέχρι και 83,000 συναλλαγές ημερησίως.

Η εφαρμογή είναι ευρεία και δεν περιορίζεται μόνο σε συναλλαγές μεταξύ του οργανισμού και πελατών αλλά επεκτείνεται και εσωτερικά όπως για παράδειγμα στην ηλεκτρονική απογραφή καταστημάτων ή στη διαχείριση αποθήκης (Logistics). Όπως γίνεται κατανοητό έχει αναπτυχθεί μια ολόκληρη αγορά στην οποία έχουν στηριχθεί για την ανάπτυξη συστημάτων διαχείρισης συναλλαγών (TPS) αρκετές εταιρίες και εκ του αποτελέσματος διαφορετικά προβλήματα απορρέουν από την ανάπτυξή τους.

### *Γ1.3 Ενδοεταιρική πηγή κύριων δεδομένων (Enterprise Master Sources)*

Σε κάθε οργανισμό υπάρχουν τα βασικά επιχειρηματικά δεδομένα που χρησιμοποιούνται σε όλα τα συστήματα, τις εφαρμογές και τις διαδικασίες. Σε αυτά τα δεδομένα στηρίζεται όλη η διαδικασία της EN καθώς περιγράφουν τις δομές, τις βασικές πληροφορίες και τις λεπτομέρειες των βασικών στοιχείων του οργανισμού. Αυτά τα βασικά δεδομένα μπορεί να είναι metadata, master data, ή reference data. Η διάκριση μεταξύ των τριών είναι ιδιαίτερα συγκεχυμένη. Είναι πέρα από τους σκοπούς αυτής της εργασίας να κατατεθεί μια εμπειριστατωμένη άποψη και ανάλυση αυτού του θέματος. Σε επόμενη ενότητα γίνεται αναφορά στα metadata οπότε για λόγους ευκολίας θα χρησιμοποιήσουμε τον όρο master data ώστε να αναφερθούμε σε όλες τις περιγραφικές πληροφορίες που είναι χρήσιμες για τους χρήστες ή τα συστήματα που υπάρχουν σε έναν οργανισμό [39].

Παραδείγματα τυπικών κατηγοριών master data είναι οι βασικές πληροφορίες προϊόντων, πελατών, εργαζομένων, προμηθευτών, πωλητών, και πληροφορίες πωλήσεων. Αυτά τα δεδομένα είναι αρκετά σημαντικά καθώς ουσιαστικά αποτελούν μια εσωτερική συμφωνία σχετικά με τις έννοιες και την χρήση των όρων. Σε πολλές επιχειρήσεις, τα master data διατηρούνται σε πολλά συστήματα που συχνά επικαλύπτονται και ανανεώνονται μέσα από διαφορετικές διαδικασίες κάτι που προκαλεί σύγχυση και έλλειψη συνοχής.

Για αυτό τον σκοπό έχουν αναπτυχθεί τα συστήματα διαχείρισης master data (Master Data Management, MDM Systems) και αποτελούν την ενδοεταιρική πηγή κύριων δεδομένων για ένα σύστημα EN. Ο στόχος ενός συστήματος MDM είναι να παρέχει έγκυρα κύρια δεδομένα σε έναν οργανισμό. Η επίτευξη του ιδανικού MDM

συστήματος μπορεί να είναι δύσκολη, κάτι το οποίο οφείλεται σε διάφορους παράγοντες είτε επιχειρηματικούς, είτε οικονομικούς είτε και νομικών περιορισμών [39].

Πολλές εταιρίες έχουν αναπτύξει λύσεις τέτοιων συστημάτων, όπως οι Oracle, IBM, SAP κ.α. Ο συγκεκριμένος τομέας γίνεται όλο και πιο ανταγωνιστικός καθώς οι λύσεις συγκεκριμενοποιούνται με στόχο την αγορά στην οποία δραστηριοποιείτε ο οργανισμός (πχ Τηλεπικοινωνίες, Δημόσιος Τομέας, Καταναλωτικά Προϊόντα) ή το τμήμα το οποίο αφορούν (πχ Πωλήσεις, Marketing, Ανθρώπινου Δυναμικού).

#### *Γ1.4 Μη αυτόματες πηγές δεδομένων (manual/user input)*

Υπάρχουν περιπτώσεις κατά τις οποίες η εισαγωγή των δεδομένων γίνεται από μη αυτόματες πηγές όπου χρήστες αναλαμβάνουν την ευθύνη της εισαγωγής. Το πιο σύνηθες είναι τα αρχεία φυλλομετρητή (spreadsheets) τα οποία ως εργαλεία κοινής χρήσης είναι πλέον αρκετά διαδεδομένα και οι χρήστες μπορούν να εκτελέσουν απλές και σύνθετες λειτουργίες. Η χρήση spreadsheets ή άλλων μη αυτόματων τρόπων έχει αρκετά μειονεκτήματα, ωστόσο γίνεται κατά κόρον σε μικρές ή και μεγάλες εταιρίες. Οι κύριοι λόγοι είναι είτε διότι ο όγκος των δεδομένων φαντάζει μικρός, είτε επειδή υπάρχει έλλειψη πόρων για την κάλυψη των αναγκών αυτοματοποίησης (προσωπικό IT, επενδύσεις σε λογισμικά νέας τεχνολογίας κ.α.), είτε υπάρχουν τεχνικοί ή λογικοί περιορισμοί (π.χ. δεδομένα από προμηθευτές που χρησιμοποιούν διαφορετικά λογισμικά και απαιτείται αρχικοποίηση δεδομένων) είτε πολύ απλά γίνεται λάθος η διαδικασία της ΕΝ (βλ. Ενότητα Β.2) Σε κάθε περίπτωση είναι σχεδόν αναπόφευκτο σε μεγάλα συστήματα να μην υπάρχει η ανάγκη για μη αυτόματες πηγές.

Τα πλεονεκτήματα της μη αυτόματης εισαγωγής είναι:

- Αρχικοποίηση και μετατροπή δεδομένων με απλές φόρμουλες όπου υπάρχουν αλληλεπικαλυπτόμενα συστήματα
- Έλεγχος χρόνου και όγκου εισαγωγής δεδομένων
- Ανθρώπινη παρέμβαση όπου απαιτείται λογική σκέψη και αποφάσεις

Τα μειονεκτήματα αφορούν:

- Πιθανότητα εισαγωγής ασαφών ή/και εσφαλμένων πληροφοριών.
- Κοστοβόρα συνεχής εκπαίδευση προσωπικού.
- Η πηγή εξαρτάται από τις ικανότητες διαφόρων ατόμων.
- Χρονοβόρα και δαπανηρή διαδικασία μέχρι την παραγωγή αναφορών.
- Διακινδυνευμένη ασφάλεια συστήματος και δεδομένων
- Συχνή επανάληψη και επικάλυψη δεδομένων.

Η μη αυτόματη εισαγωγή ασκεί πίεση στους ανθρώπους να είναι σωστοί σε όλες τις λεπτομέρειες της εργασίας τους ανά πάσα στιγμή. Το επίπεδο των υπηρεσιών εξαρτάται από άτομα και αυτό απαιτεί συνεχή κατάρτιση του προσωπικού καθώς και συνεχή κίνητρα και να διασφαλιστεί ότι οι διαδικασίες ακολουθούνται όπως έχουν σχεδιαστεί. Είναι πολύ εύκολο από κάποια λεπτομέρεια να καταλήξει το σύστημα σε ασυνέπεια στην εισαγωγή δεδομένων. Αυτό έχει ως αποτέλεσμα όχι μόνο να προκαλέσει προβλήματα στην διάθεση των πληροφοριών αλλά κυρίως καθιστά το σύστημα αναξιόπιστο δεν μπορεί να χρησιμοποιηθεί για αναφορές ή την λήψη αποφάσεων. Το πρόβλημα μπορεί να είναι περίπλοκο να λυθεί ή ακόμη χειρότερα να είναι δύσκολο να ανιχνευτεί. Όλα τα παραπάνω δείχνουν την ανάγκη για αυτοματοποίηση και αποφυγή όπου και όσο είναι δυνατόν της χειροκίνητης εισαγωγής, ιδιαίτερα για μεγάλο όγκο δεδομένων.

#### *Γ1.5 Εξωτερικά συνεργαζόμενες πηγές δεδομένων (External sources)*

Πέρα από τα προαναφερόμενα ένας οργανισμός ενδέχεται να χρειάζεται δεδομένα από εξωτερικές πηγές με σκοπό να δημιουργήσει το λεγόμενο και ως enterprise mashup. Ο όρος enterprise mashup βρίσκει τους ακαδημαϊκούς σε συζητήσεις ωστόσο θα μπορούσαμε να πούμε πως ένα enterprise mashup είναι ένας συνδυασμός υφισταμένων πόρων από παραπάνω από μια πηγή με σκοπό την ενδυνάμωση των χρηστών στην δημιουργία και προσαρμογή πληροφοριών και συνθηκών [40]. Για παράδειγμα η συλλογή πληροφοριών από υπηρεσίες διαχείρισης δεδομένων κοινωνικών δικτύων και η προβολή τους σε διαδραστικό χάρτη εν συνάρτηση των πωλήσεων στην εκάστοτε περιοχή.



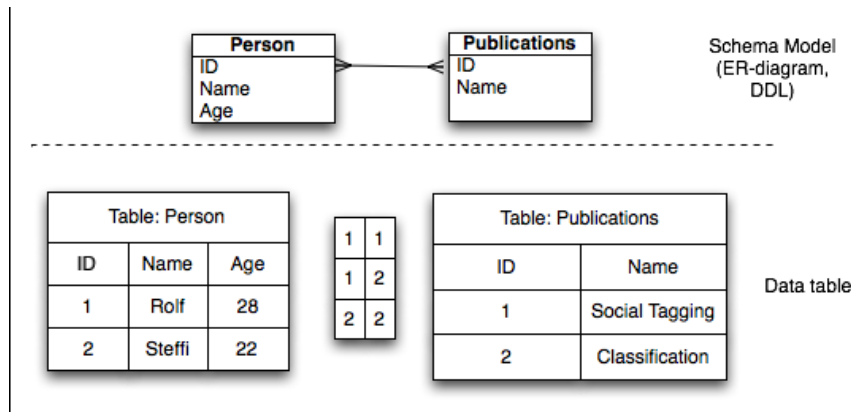
Τα δεδομένα από εξωτερικές πηγές διαμορφώνονται κατάλληλα στο επόμενο στάδιο, από τις υπηρεσίες ενσωμάτωσης δεδομένων (Enterprise Data Integration - EDI Services) όπου και συγκεντρώνονται μέσω διαφόρων διεπαφών (interfaces).

Εκτός της πηγής, τα δεδομένα αξίζει να εξεταστούν ως προς την δομή τους. Τα δεδομένα σύμφωνα με την δομή τους διαχωρίζονται σε τρεις βασικές κατηγορίες, πλήρως δομημένα, αδόμητα και ημιδομημένα. κάτι το οποίο είναι εξαιρετικά σημαντικό να αναλυθεί καθώς στο επόμενο στάδιο οι υπηρεσίες ενσωμάτωσης δεδομένων ενδέχεται να έχουν περιορισμούς στην δομή ή να απαιτούν κάποια συγκεκριμένη εξ αυτών.

#### *Γ1.6 Δομημένα, Ημι-δομημένα και Αδόμητα δεδομένα (Structured and Unstructured Data)*

##### *Πλήρως Δομημένα Δεδομένα*

Τα πλήρως δομημένα δεδομένα ακολουθούν ένα προκαθορισμένο σχήμα, τηρούν δηλαδή κάποια προδιαγραφή [41]. Ένα τυπικό παράδειγμα πλήρως δομημένων δεδομένων είναι μια σχεσιακή βάση δεδομένων. Η διαδικασία σχεδίασης του σχήματος μιας βάσης δεδομένων είναι μια περίπλοκη διαδικασία, καθώς ένα σχήμα πρέπει να οριστεί πριν από την δημιουργία του περιεχόμενου και του όγκου δεδομένων. Το σχήμα, ως γνωστόν, καθορίζει τον τύπο και τη δομή των δεδομένων καθώς και τις σχέσεις αυτών. Παρακάτω απεικονίζεται ένα διάγραμμα οντοτήτων-συσχετίσεων (ER-διάγραμμα) πινάκων σε σύστημα σχεσιακής βάσης δεδομένων (RDBMS). Μια τέτοια δομή διαχείρισης των πλήρως δομημένων δεδομένων επιτρέπει την αποτελεσματική αποθήκευση, επεξεργασία και διερεύνηση του περιεχομένου με απλές εντολές SQL [42].



**Σχήμα 3.4:** Διάγραμμα οντοτήτων-συσχετίσεων (ER-διάγραμμα) πινάκων σε σύστημα σχεσιακής βάσης δεδομένων

Το κόστος για την υψηλή απόδοση είναι η ευελιξία και επεκτασιμότητα καθώς υπάρχει δυσκολία να επεκταθεί ένα σχήμα το οποίο έχει ήδη οριστεί και διαθέτει περιεχόμενο. Για παράδειγμα, αν σε μία εγγραφή σε έναν πίνακα προκύψει η ανάγκη για μια επιπλέον ιδιότητα, θα πρέπει να δημιουργηθεί μια στήλη στο σύνολο του πίνακα η οποία θα επηρεάζει και τις υπόλοιπες εγγραφές όσον αφορά το κόστος αποθήκευσης, τους χρόνους αποδοτικότητας (ανάκτησης, επεξεργασίας, διαγραφής κλπ.). Μια τέτοια περίπτωση είναι ασύμφορη για πίνακες που περιέχουν χιλιάδες άλλες σειρές που δεν χρειάζονται άλλη ιδιότητα. Στα πλεονεκτήματα των δομημένων δεδομένων ανήκει η ευκολία στην διαχείριση, την αποθήκευση και την εξαγωγή συμπερασμάτων καθώς μπορούν να αναπτυχθούν περίπλοκες ερωτήσεις (SQL) στην βάση συσχετίζοντας πολλούς διαφορετικούς πίνακες. Επιπλέον, πολλές εφαρμογές και πρακτικά εργαλεία στηρίζονται στην ύπαρξη σχεσιακών βάσεων δεδομένων είτε με σκοπό την αποθήκευση (πχ σύστημα παραγγελιοληψίας), είτε την ανάκτηση και διαχείριση δεδομένων (πχ σύστημα αναφοράς πωλήσεων).

### *Αδόμητα Δεδομένα*

Σύμφωνα με τους Blumberg και Atre σε άρθρο τους που δημοσιεύθηκε το 2003 στο περιοδικό "Information Management", ο όρος «αδόμητα» χαρακτηρίζει το γεγονός μη ύπαρξης αναγνωρίσιμου μοντέλου δομής το οποίο να περιγράφει τα δεδομένα. Αδόμητα δεδομένα χαρακτηρίζονται επίσης τα δεδομένα τα οποία δεν μπορούν να αποθηκευτούν σε σειρές και στήλες σε μία σχεσιακή βάση δεδομένων. Η αποθήκευση των δεδομένων σε αδόμητη μορφή, ορίζεται χωρίς σχήμα. Παραδείγματα μη δομημένων δεδομένων είναι ένα έγγραφο κειμένου που αρχειοθετείται σε ένα φάκελο αρχείων, ένα βίντεο ή μια εικόνα [43].

Το πλεονέκτημα των αδόμητων δεδομένων είναι, ότι δεν είναι απαραίτητη καμία πρόσθετη προσπάθεια για την ταξινόμηση τους. Μια εικόνα μπορεί να αποθηκευτεί στο σκληρό δίσκο κάποιου υπολογιστή και από εκεί να μεταφερθεί απλά σε κάποιον άλλο. Δεν ισχύει το ίδιο και για τα δεδομένα από έναν πίνακα μιας σχεσιακής βάσης καθώς θα πρέπει να μεταφερθεί από ένα σύστημα διαχείρισης βάσεων δεδομένων σε ένα άλλο, μια διαδικασία με περισσότερους παραμέτρους όπως το σχήμα της βάσης, το λογισμικό διαχείρισης κλπ.

Αντίστοιχα, ένας περιορισμός αυτού του είδους των δεδομένων είναι, ότι είναι σχεδόν αδύνατη η ελεγχόμενη πλοήγηση στο αδόμητο περιεχόμενο. Για παράδειγμα η εύρεση μιας φράσης ή μιας λέξης σε ένα βίντεο. Αν και υπάρχουν εφαρμογές και τεχνολογίες που μπορούν να μας προσφέρουν τέτοιες δυνατότητες, ακολουθούν μια διαδικασία επεξεργασίας και μετατροπής των δεδομένων σε άλλες μορφές η οποία είναι πολύπλοκη, κοστοβόρα και με περιορισμένες δυνατότητες [42].

### *Ημι-δομημένα δεδομένα*

Τα ημι-δομημένα δεδομένα συχνά περιγράφονται ως δεδομένα αυτοπεριγραφούμενα, χαρακτηρισμός που υποδηλώνει ότι δεν υπάρχει ξεχωριστή περιγραφή του τύπου ή της δομής των δεδομένων. Αυτό δεν σημαίνει ότι ο ορισμός του σχήματος δεν είναι δυνατός, θα λέγαμε καλύτερα ότι είναι μάλλον προαιρετικός καθώς υπάρχουν και περιπτώσεις αλλαγών στο σχήμα ή

κληρονομίας του ορισμού με βάση κάποιο άλλο αρχείο. Μια ακόμα πιθανότητα είναι τα στιγμιότυπα των δεδομένων να έχουν περισσότερο από έναν τύπο. Αυτό το ιδιαίτερο χαρακτηριστικό είναι και ένα από τα πλεονεκτήματα των ημιδομημένων δεδομένων, η δυνατότητα δηλαδή να χαρακτηρίζονται από διαφορετικούς τύπους σε διαφορετικά στιγμιότυπα κάτι που σημαίνει ότι μπορεί να δημιουργηθούν δεδομένα σύμφωνα με χαρακτηριστικές προδιαγραφές [44].

Τυπικό παράδειγμα ημι-δομημένων δεδομένων είναι τα αρχεία XML, τα οποία είναι αρχεία σε μια γλώσσα σήμανσης, που καθορίζει ένα σύνολο κανόνων για την κωδικοποίηση των εγγράφων σε μορφή που είναι τόσο αναγνώσιμη από τον άνθρωπο και τα αναγνώσιμα από μηχανήματα<sup>12</sup>.

### *Γ1.7 Μεταδεδομένα (Metadata)*

Ο ακριβής ορισμός του όρου «μεταδοδομένα» είναι διφορούμενος, και ο πλήρης καθορισμός τους αποτελεί ένα πολύ δύσκολο έργο [45]. Σε ελεύθερη μετάφραση ορίζεται ως τα δεδομένα που περιγράφουν δεδομένα ωστόσο αυτός ο χαρακτηρισμός αν και χαριτωμένος και εύκολος στη μνήμη, δεν είναι πολύ ακριβής. [46]. Ο όρος metadata χρησιμοποιείται με διαφορετικό τρόπο σε διαφορετικές κοινότητες ωστόσο θα μπορούμε να τα ορίσουμε ως δομημένες πληροφορίες που περιγράφουν, εξηγούν, εντοπίζουν, ή γενικά καθιστούν ευκολότερο να ανακτηθεί, χρησιμοποιηθεί, ή διαχειριστεί μια πηγή πληροφοριών. Τα metadata αποτελούν ένα πολύ σημαντικό στοιχείο στα συστήματα EN καθώς σε πολλές περιπτώσεις αποσκοπούν στην κατανόηση των σχέσεων των δεδομένων από διαφορετικές πηγές όπως επίσης και η κοινή γλώσσα μερικών τεχνικών όρων μεταξύ διαφορετικών τμημάτων ή ομάδων. Η αξία των metadata φαίνεται πιο ξεκάθαρα στο κεφάλαιο Δ.

### *Γ1.8 Στοιχεία ενδιαφέροντος σχετικά με τις πηγές δεδομένων:*

Στο πλαίσιο της EN περισσότερο υπάρχουν σημαντικά και καιρία ζητήματα όσον αφορά τα δεδομένα καθώς ένα σφάλμα στην πηγή των δεδομένων καταστεί έως και άχρηστη την όλη διαδικασία. Παρακάτω αναλύονται μερικά από τα ζητήματα που αφορούν τις πηγές δεδομένων.

---

<sup>12</sup> Περισσότερες πληροφορίες για τα XML αρχεία βρίσκονται στο documentation <http://www.w3pdf.com/W3cSpec/XML/2/REC-xml11-20060816.pdf>

### *Ασφάλεια δεδομένων (Data Security)*

Από τα πιο καίρια ζητήματα, ιδιαίτερα όσον αφορά συναλλαγές χρηματικού περιεχομένου, είναι η ασφάλεια δεδομένων καθώς υπάρχει το ενδεχόμενο η συναλλαγή να είναι στόχος υποκλοπής προσωπικών δεδομένων είτε εταιρικών δεδομένων είτε γενικά μετατροπής αυτών. Ο κατασκευαστής ή ο διαχειριστής θα πρέπει να έχει επιμεληθεί για την συμμόρφωση του συστήματος με τα κατάλληλα πρότυπα ασφαλείας όπως για παράδειγμα τα The Payment Card Industry Data Security Standard (PCI DSS) τα οποία διαμορφώθηκαν τον Σεπτέμβρη του 2006 από το Payment Card Industry Security Standards Council (PCI SSC), έναν ανεξάρτητο φορέα που δημιουργήθηκε από τις μεγάλες μάρκες καρτών πληρωμής (Visa, MasterCard, American Express, Discovery και JCB). Τα PCI DSS αφορούν πρότυπα με έμφαση στη βελτίωση της ασφαλείας του λογαριασμού πληρωμών σε όλη τη διαδικασία της συναλλαγής και σκοπό έχουν να εξασφαλίσουν ότι όλες οι εταιρείες που επεξεργάζονται, αποθηκεύουν ή διαβιβάζουν πληροφορίες, διατηρούν ένα ασφαλές περιβάλλον<sup>13</sup>.

### *Ανθεκτικότητα δεδομένων σε σφάλματα (Resilience to data corruption faults)*

Καθότι μια συναλλαγή αποτελεί μια διαδικασία συγκεκριμένων βημάτων, σε περίπτωση αστοχίας, είτε από την πλευρά του λογισμικού είτε άλλων στοιχείων, η συναλλαγή δεν ολοκληρώνεται (το ίδιο συμβαίνει όταν για παράδειγμα υπάρχει πτώση τάσης ή διακοπή σύνδεσης στο δίκτυο) και δεδομένα δεν θα πρέπει να μεταφέρονται. Είναι πολύ σημαντικό ο κατασκευαστής να έχει προβλέψει όλες τις αστοχίες και να έχει φροντίσει ώστε ημιτελείς ή ακυρωμένες συναλλαγές να μην μεταφέρουν δεδομένα.

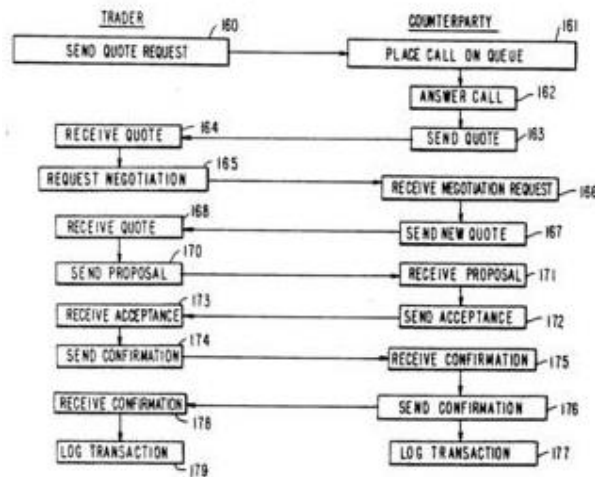
Για παράδειγμα μπορούμε να αναφέρουμε το πρωτόκολλο SoftX<sup>14</sup>, ένα πρωτόκολλο χαμηλής επεμβατικότητας για την υποστήριξη της εκτέλεσης των εφαρμογών συναλλαγών στηριζόμενη σε αφοσιωμένα νήματα (dedicated threads)

---

<sup>13</sup> Όλες οι πληροφορίες για τα PCI DSS και το PCI SSC στον παρακάτω σύνδεσμο <https://www.pcicomplianceguide.org/pci-faqs-2/#1>

<sup>14</sup> Περισσότερες πληροφορίες στο σύνδεσμο: <http://www.ssrgece.vt.edu/papers/nca14.pdf>

και την ευρεσιτεχνία της Telerate Systems Inc.<sup>15</sup> (Δεκέμβριος 1992), ένα σύστημα που προβλέπει τη διεκπεραίωση των συναλλαγών μεταξύ μιας ομάδας εμπορικών σταθμών (π.χ. ένα τερματικό αγορών και βάση δεδομένων) σε ένα δίκτυο επικοινωνίας μέσω μιας σειράς δομημένων σταδίων.



**Σχήμα 3.5:** πρωτόκολλο SoftX: διεκπεραίωση συναλλαγών μέσω σειράς δομημένων σταδίων

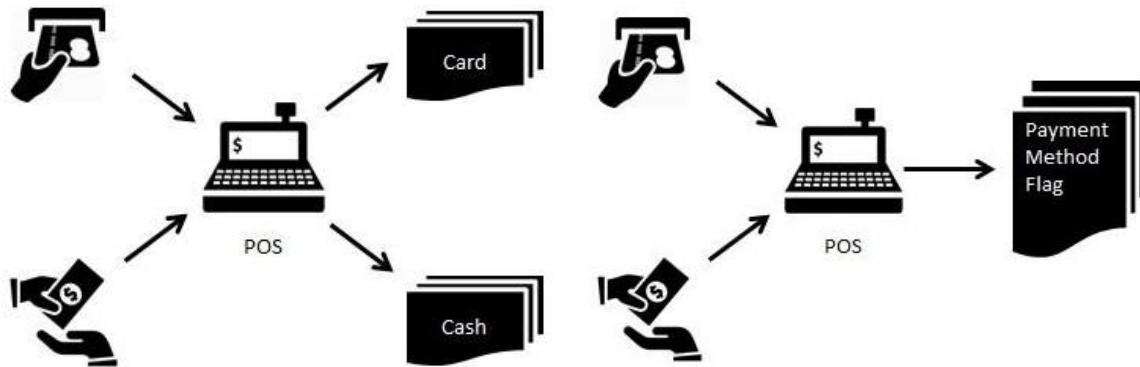
### *Μέγεθος και δομή εξαγωγίμων δεδομένων*

Κάθε μία από τις διάφορες εμπορικά διαθέσιμες λύσεις για την διαχείριση των συναλλαγών παρέχει όχι μόνο διαφορετικές πληροφορίες κατά την ολοκλήρωση μιας συναλλαγής αλλά συχνά και σε διαφορετική δομή.

Ένα λογισμικό που βρίσκεται εγκατεστημένο στο τερματικό της ταμειακής μηχανής ενός καταστήματος μπορεί να κρατάει σε ένα log ιδιαίτερες πληροφορίες σχετικά με τις συναλλαγές ενώ κάποιο άλλο λογισμικό δεν αποθηκεύει την ίδια πληροφορία (για παράδειγμα την μέθοδος αγοράς). Ο τρόπος που το λογισμικό αποθηκεύει την πληροφορία για την μέθοδο αγοράς μπορεί να βρίσκεται σε κάποιο πεδίο στο πακέτο της πληροφορίας ή να δημιουργεί διαφορετικό πακέτο πληροφοριών για συναλλαγές με συγκεκριμένη μέθοδο αγοράς. Αυτή η διαφορετική προσέγγιση φαίνεται στο παρακάτω σχήμα και καθορίζει όχι μόνο την

<sup>15</sup> Η κατάθεση της ευρεσιτεχνίας είναι δημοσίως διαθέσιμη στο παρακάτω σύνδεσμο:  
<http://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&d=PALL&RefSrch=yes&Query=PN/5168446>

δομή αλλά και το μέγεθος των εξαγωγίμων δεδομένων το οποίο επηρεάζει με την σειρά του τις υπηρεσίες διαχείρισης δεδομένων.



**Σχήμα 3.6.α:** διαφορετικά πακέτα δεδομένων ανά μέθοδο συναλλαγής

**Σχήμα 3.6.β:** ενιαίο πακέτο δεδομένων με χρήση flag ανά μέθοδο συναλλαγής

Μεταφορά δεδομένων σε περιοδικό ή πραγματικό χρόνο (batch time data processing, real time data processing)

Σύμφωνα με τον Michael Walker σε άρθρο του το 2003 στην πλατφόρμα Data Science Portal, η μεταφορά δεδομένων σε περιοδικό χρόνο (batch data processing) είναι ένας αποτελεσματικός τρόπος για την επεξεργασία μεγάλου όγκου δεδομένων όταν μια ομάδα συναλλαγών συλλέγονται κατά τη διάρκεια μιας χρονικής περιόδου. Τα δεδομένα συλλέγονται, εισάγονται, υφίστανται επεξεργασία και στη συνέχεια διαμορφώνεται το αποτέλεσμα σε batch datasets [47].

Η μεταφορά σε περιοδικό χρόνο απαιτεί ειδικά προγράμματα για την εισροή, την επεξεργασία και την παραγωγή δεδομένων. Σε αντίθεση, η επεξεργασία δεδομένων σε πραγματικό χρόνο περιλαμβάνει μια συνεχή εισροή, επεξεργασία και παραγωγή δεδομένων. Τα δεδομένα πρέπει να υποβάλλονται σε επεξεργασία σε ένα μικρό χρονικό διάστημα (ή σχεδόν πραγματικό χρόνο).

Οι περισσότεροι οργανισμοί χρησιμοποιούν μεταφορά δεδομένων σε περιοδικό χρόνο για το μεγαλύτερο μέρος των αναγκών τους, μερικές φορές χρειάζεται ή και ίσως επιβάλλεται η μεταφορά δεδομένων σε πραγματικό χρόνο. Σε ένα ταμείο για παράδειγμα (Point of Sales - POS) η μεταφορά δεδομένων σε πραγματικό χρόνο βοηθάει στην ενημέρωση των αποθεμάτων, παρέχει ιστορικό απογραφής, και λεπτομέρειες για τις πωλήσεις σε ένα συγκεκριμένο σημείο. Όλα τα παραπάνω

επιτρέπουν σε έναν οργανισμό να εκτελεί πληρωμές και να παίρνει αποφάσεις ( πχ διαχείριση της εφοδιαστικής αλυσίδας) σε πραγματικό χρόνο.

Και οι δύο μέθοδοι μεταφοράς δεδομένων, έχουν πλεονεκτήματα και μειονεκτήματα. Η απόφαση για την επιλογή εξαρτάται από πολλούς και διαφορετικούς παράγοντες (κόστος, διαθέσιμη τεχνολογία και τεχνογνωσία κ.α.) κυρίως όμως από το κατά πόσο απαιτείται η δυνατότητα άμεσης δράσης σε συγκεκριμένα γεγονότα.

## **G2. Υπηρεσίες Ενσωμάτωσης**

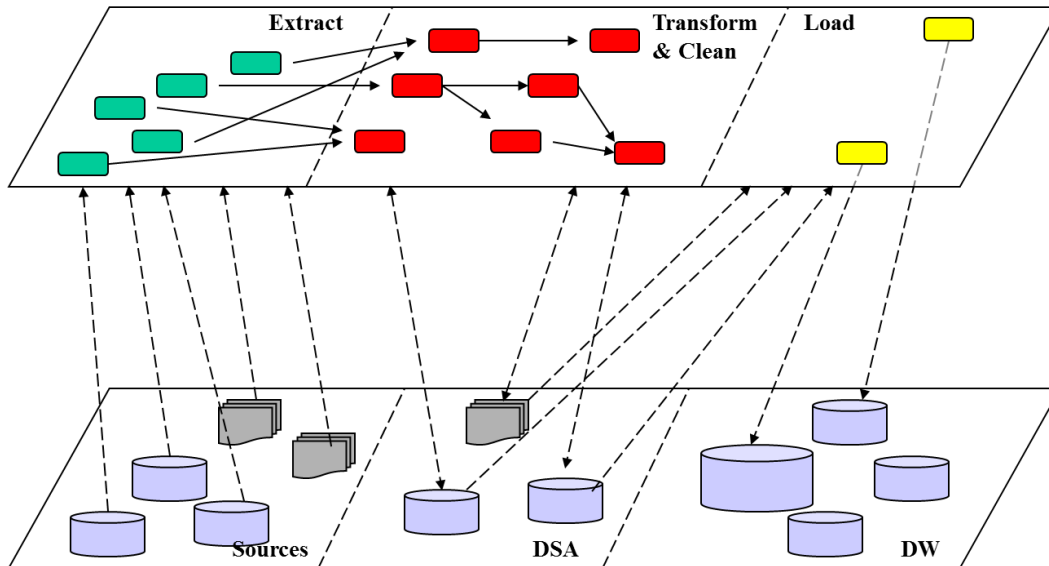
### *G2.1 Η Διαδικασία Εξαγωγής-Μετατροπής-Φόρτωσης δεδομένων (Extract-Transform-Load Process)*

Η διαδικασία Εξαγωγής-Μετατροπής-Φόρτωσης δεδομένων (Extract-Transform-Load, ETL) είναι μια κατηγορία των εξειδικευμένων εργαλείων με σκοπό της αντιμετώπιση θεμάτων ομοιογένειας, καθαρισμού και φόρτωσης δεδομένων σε μια Αποθήκη Δεδομένων (Data Warehouse, DW) [48]. Η Αποθήκη Δεδομένων, εν συντομία DW, είναι μια συλλογή ολοκληρωμένων αντικειμενοστραφών βάσεων δεδομένων που σχεδιάζονται για την υποστήριξη της διαδικασίας λήψης αποφάσεων [49].

Η ETL διαδικασία είναι ο ακρογωνιαίος λίθος μιας Αποθήκης Δεδομένων, δεδομένου ότι παρέχει τα απαραίτητα ολοκληρωμένα και ομοιογενή δεδομένα από τις διάφορες κατακευματισμένες πηγές. Η διαδικασία ETL, και ειδικότερα ο σχεδιασμός της, εξακολουθεί να θεωρείται ως μια διαδικασία πολύπλοκη, επιρρεπής σε σφάλματα, και χρονοβόρα. Για παράδειγμα, η χαμηλή ποιότητα δεδομένων μπορεί να προκαλέσει πολλά προβλήματα στη DW, όπως η ανεπιτυχής φόρτωση λόγω παραβιάσεων περιορισμών στους τύπους δεδομένων, ή ασυνέπεια στα δεδομένα που διέπονται από επιχειρησιακούς κανόνες που πρέπει να εφαρμοστούν. Οι παραπάνω χαρακτηρισμοί οφείλονται στο γεγονός ότι οι διαδικασίες ETL είναι συνήθως σχεδιασμένες βάση μιας συγκεκριμένης τεχνολογίας και λογισμικού DW και ανάλυσης και αναφορών. Καθαυτό τον τρόπο είναι δύσκολο να μοιράζονται και να επαναχρησιμοποιούνται μεθοδολογίες και πρακτικές μεταξύ των έργων ακόμα και εντός του ίδιου οργανισμού [50].



Στο παρακάτω γράφημα βλέπουμε μια αναπαράσταση της ETL διαδικασίας όπου και ξεχωρίζουν τα τρία διαφορετικά επίπεδα, Πηγές Δεδομένων, Data Staging Area και Data Warehouse [48].



**Σχήμα 3.7:** Τα τρία διακριτά στάδια της ETL διαδικασίας

### Γ2.2 Σχεδιασμός ETL διαδικασίας.

Ακαδημαϊκές αναφορές και εμπειρίες από την αγορά προσεγγίζουν τον σχεδιασμό της ETL διαδικασίας με ιδιαίτερη προσοχή καθώς, όπως αναφέρθηκε, διαφορετικές τεχνολογίες και προμηθευτές καθορίζουν ακόμα και τις μελλοντικές δυνατότητες επέκτασης. Το σημαντικότερο στοιχείο είναι η DW για τον σχεδιασμό της οποίας αρχικά υπάρχουν δύο ζητούμενα. Το πρώτο ζητούμενο περιλαμβάνει τη συλλογή των απαιτήσεων από την πλευρά των χρηστών. Το δεύτερο περιλαμβάνει την ανάλυση της δομής και του περιεχομένου των υφιστάμενων πηγών δεδομένων και τη χαρτογράφηση τους στο μοντέλο της DW [48].

Όσον αφορά τις γλώσσες προγραμματισμού που συμβάλλουν στη δημιουργία και τη διαχείριση της ETL διαδικασίας, εκτός από τις παραδοσιακές οπτικές γλώσσες προγραμματισμού (π.χ. Entity/Relationship σχήματα, E/R schemas), υπάρχουν ανάγκες που καλύπτονται και από γλώσσες τέταρτης γενιάς, (4GL) είτε μέσω μιας διαδικαστικής γλώσσας (π.χ. Oracle Metabase, OMB ή Visual Basic) [50].

### *Γ2.3 Διαδικασία Εξαγωγής από Πηγές Δεδομένων*

Τα ανεπεξέργαστα δεδομένα που προέρχονται από διάφορες συστημικές πηγές συνήθως γράφονται απευθείας σε κάποια μορφή αποθήκευσης (δίσκο, αδόμητα αρχεία flat files) με κάποια ελάχιστη αναδιάρθρωση [45]. Εξαρτώμενα από την τεχνολογία των πηγών, τα δεδομένα μπορεί να είναι κατανοητά στον άνθρωπο ή μόνο στην μηχανή αποτελούμενα μόνο από αδόμητες συμβολοσειρές. Τα αρχικά δεδομένα μπορούν να διαβαστούν πολλές φορές από διάφορα συστήματα αν αυτό είναι δυνατόν ανάλογα με τις ανάγκες. Σε μερικές περιπτώσεις, με σκοπό την οικονομία σε χώρο και επεξεργαστική ισχύ, τα αρχικά δεδομένα απορρίπτονται μετά το στάδιο μετατροπής. Σε άλλες περιπτώσεις, τα δεδομένα διατηρούνται, για όσο κρίνει ο διαχειριστής, σε εφεδρική αρχειοθέτηση (log) με σκοπό την ανασυγκρότηση σε περίπτωση κατάρρευσης του συστήματος ή τον έλεγχο αν και όποτε αυτός απαιτηθεί.

### *Γ2.4 Στάδιο Μετατροπής και Φόρτωση Δεδομένων*

Το staging area level είναι ένα ενδιάμεσο προσωρινό στάδιο το οποίο αποτελείται από διάφορες μορφές αποθήκευσης δεδομένων. Σε αυτό το στάδιο φτάνουν τα δεδομένα από τις πηγές και υπόκεινται σε διάφορες τεχνικές ώστε να αποφευχθούν αποτυχημένες εγγραφές και λανθασμένα δεδομένα στο DW. Διάφοροι προμηθευτές παρέχουν διάφορες λύσεις για την δημιουργία του staging area όπως για παράδειγμα η Oracle στην οποία μπορεί να δημιουργηθεί ένα DW αλλά και Staging Area επίπεδο από όπου μέσω της 3GL γλώσσας προγραμματισμού PL/SQL δημιουργούνται πακέτα κώδικα που εκτελούν την διαδικασία μετατροπής.

Καθότι στο Staging Area γίνονται όλες οι μετατροπές δεδομένων, εκεί εφαρμόζονται και ελέγχονται όλοι οι κανόνες που τίθενται από τον σχεδιασμό του DW. Οι εν λόγω κανόνες χρησιμοποιούνται για να διορθώσουν μη έγκυρες τιμές, να σηματοδοτήσουν εσφαλμένες εγγραφές. Υπάρχουν επίσης κανόνες που μπορούν να εφαρμοστούν στο επίπεδο DW όπως προτείνονται από τους Kimball και Caserta στο βιβλίο «The Data Warehouse ETL Toolkit, 2004». Οι κανόνες στο Staging Area μπορούν να χωριστούν σε τρεις κύριες κατηγορίες [51]:

- Ακεραιότητας Βάσης Δεδομένων: κανόνες που βασίζονται σε περιορισμούς στους τύπους δεδομένων.
- Κανόνες Αντιστοίχισης, (matching) και Συγχώνευσης, (merging): κανόνες που αφορούν την αντιστοίχιση και συγχώνευσή εγγραφών από τις διάφορες πηγές.
- Επιχειρηματικοί κανόνες: κανόνες που ορίζονται από τον οργανισμό και τους χρήστες και αφορούν λογικά σφάλματα όπως για παράδειγμα αρνητικές πωλήσεις.

Καθώς μεγάλες ποσότητες δεδομένων σε διάφορα χρονικά περιθώρια περνούν από το staging level είναι κρίσιμη η διαδικασία που ακολουθείται για την ενημέρωση των διαχειριστών για τυχόν παραβίαση κανόνων. Η παραβίαση των κανόνων μπορεί να διορθωθεί με μια απλή παρέμβαση ή να θέσει την διαδικασία σε κατάσταση προσωρινής διακοπής. Μορφές ενημέρωσης των διαχειριστών είναι οι παρακάτω [51]:

- Διακοπή Διαδικασίας: ορισμένες παραβιάσεις κανόνων είναι αρκετά σοβαρές ώστε να πρέπει να σταματήσει η διαδικασία και να γίνουν οι κατάλληλες ενέργειες για την αποκατάσταση.
- Σφάλματα: εγγραφές που χαρακτηρίστηκαν ως σφάλματα εμποδίζονται από την φόρτωση στο DW επίπεδο και επιθεωρούνται από τους διαχειριστές προκειμένου να αποφασιστεί η διόρθωση ή η απόρριψή τους.
- Προειδοποίηση: εγγραφές που χαρακτηρίστηκαν ως προειδοποιήσεις δεν αποτελούν σφάλματα αλλά ζητούν της προσοχής των διαχειριστών

Ο σχεδιασμός της διαχείρισης σφαλμάτων εξαρτάται επίσης από τις τεχνολογίες που χρησιμοποιούνται. Στο κεφάλαιο Δ, την Πρακτική Εφαρμογή, αναλύεται ένα παράδειγμα κατά το οποίο οι διαχειριστές ενημερώνονται για την παραβίαση κανονισμών και ύστερα αλλά όλες οι παραβιάσεις αποθηκεύονται σε μία περιοχή προσωρινής αποθήκευσης συμβάντων (logging area).

Από την στιγμή που τα δεδομένα έχουν περάσει από την διαδικασία ETL και βρίσκονται πλέον στο DW, είναι διαθέσιμα για τους σκοπούς των υπηρεσιών ανάλυσης και αναφορών οι οποίες παράγουν τα τελικά αποτελέσματα (πίνακες, γραφήματα, αναφορές, αναλύσεις κ.λπ.) προς τον χρήστη.

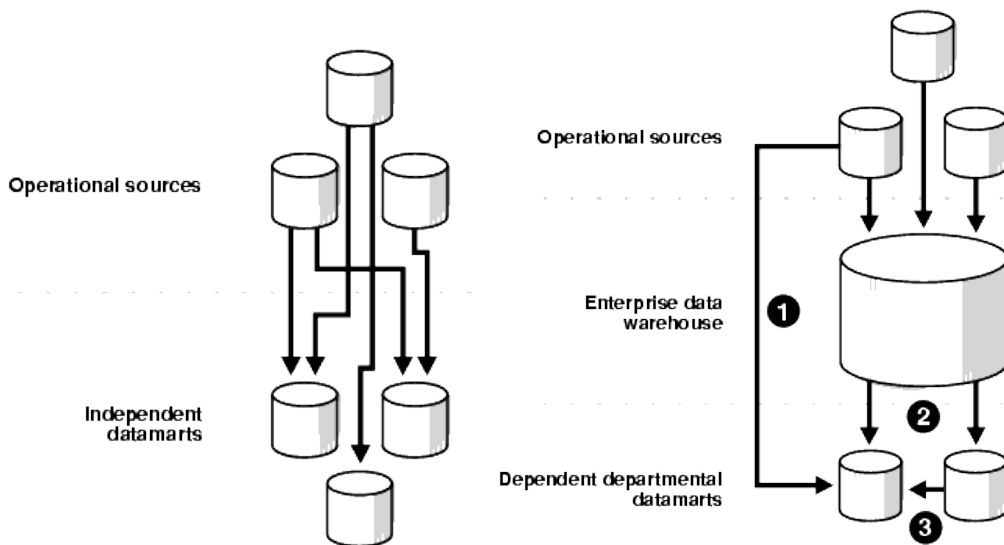
### Γ3. Υπηρεσίες Διαχείρισης Δεδομένων

Ένα πολύ σημαντικό στοιχείο στην διαδικασία της EN είναι η ορθή αρχιτεκτονική αποθήκευσης δεδομένων. Οι Αποθήκες Δεδομένων παρέχουν πρόσβαση σε δεδομένα που ενσωματώθηκαν από την ETL διαδικασία όσο και σε ιστορικά δεδομένα.

Έχουν αποδοθεί στην συγκεκριμένη ορολογία αρκετοί ορισμοί. Ο πιο δημοφιλής εξ αυτών ανήκει στον Bill Inmon ο οποίος περιγράφει την Αποθήκη Δεδομένων (DW) ως μια συλλογή δεδομένων προσανατολισμένη προς ένα θέμα (subject-oriented), ολοκληρωμένη (integrated- ενοποίηση ετερογενών δεδομένων), χρονικά μεταβαλλόμενη (time variant) που δεν διαγράφεται (non-volatile) με σκοπό την υποστήριξη λήψης αποφάσεων [49]. Υποσύνολα των DW είναι τα data marts τα οποία συνήθως αφορούν ένα συγκεκριμένο τμήμα ή μια συγκεκριμένη ομάδα ενός οργανισμού. Σε μερικές υλοποιήσεις κάθε τμήμα ή επιχειρηματική μονάδα θεωρείται ο ιδιοκτήτης του data mart συμπεριλαμβανομένων του υλικού (hardware), λογισμικού (software) αλλά και των δεδομένων που περιέχονται. Αυτό δίνει την δυνατότητα πολυδιάστατης χρήσης των DW χωρίς να επηρεάζεται η πληροφορία που αφορά διαφορετικά τμήματα του οργανισμού.

#### *Γ3.1 Data Mart*

Ένα data mart δεν είναι τίποτα άλλο παρά μια συλλογή από περιοχές δεδομένων που διοργανώνονται για την υποστήριξη αποφάσεων με βάση τις ανάγκες ενός συγκεκριμένου τμήματος. Για παράδειγμα, το Οικονομικό Τμήμα έχει το δικό του data mart, το Τμήμα Μάρκετινγκ και το Τμήμα Πωλήσεων έχει το δικό του αντίστοιχα. Τα μεταξύ τους data marts αποτελούν ιδιοκτησία του κάθε τμήματος οπότε επιτρέπεται στα τμήματα να παρακάμψουν οποιοδήποτε μέσο ελέγχου που μπορεί να συντονίζει τα δεδομένα που βρίσκονται στα διαφορετικά τμήματα. Η δομή τους συνήθως στεγάζεται σε πολυδιάστατη τεχνολογία η οποία ευνοείται για λόγους ευελιξίας ωστόσο δεν είναι η βέλτιστη δομή για υπερμεγέθους ποσότητες δεδομένων. Υπάρχουν δύο είδη data marts, ανεξάρτητα (Σχήμα 3.8.α) και εξαρτώμενα (Σχήμα 3.8.β). Ένα ανεξάρτητο data mart είναι εκείνο του οποίου η πηγή είναι περιβάλλοντα εφαρμογών. Από την άλλη πλευρά σε ένα εξαρτώμενο data mart τα δεδομένα προέρχονται από μια αποθήκη δεδομένων [52].



**Σχήμα 3.8.α:** Αρχιτεκτονική ανεξάρτητων data marts

**Σχήμα 3.8.β:** Αρχιτεκτονική εξαρτώμενων data marts

### Γ3.2 Αρχιτεκτονικές Αποθήκης Δεδομένων (Data Warehouse):

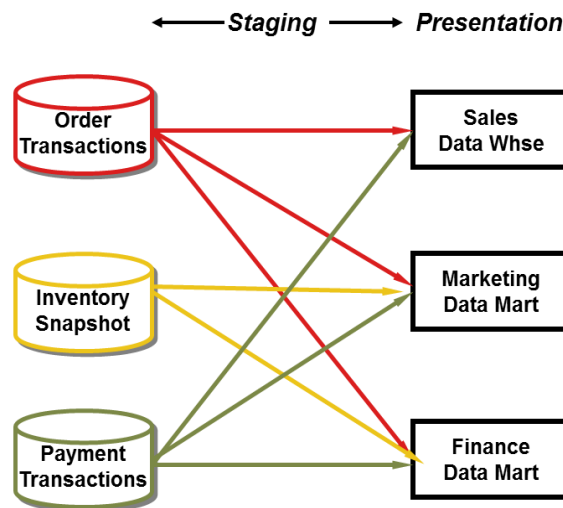
Πολύς λόγος γίνεται σχετικά με την το ερώτημα «ποια είναι η καλύτερη αρχιτεκτονική». Υπάρχουν διάφορες αρχιτεκτονικές δομές των DW η κάθε μία εκ των οποίων προσφέρει διαφορετικά πλεονεκτήματα και η επιλογή μίας εξ αυτών προς υλοποίηση διαφέρει ανάλογα με τον προβλεπόμενο σκοπό. Άξιο αναφοράς είναι το γεγονός ότι δύο μεγάλοι αναγνωρισμένοι επιστήμονες, οι Bill Inmon και Ralph Kimball υποστηρίζουν διαφορετικές οπτικές ως προς το ποια είναι η καλύτερη αρχιτεκτονική. Ο Bill Inmon υποστηρίζει την hub-and-spoke αρχιτεκτονική ενώ ο Ralph Kimball τάσσεται υπέρ της data mart bus αρχιτεκτονικής με συμμορφούμενες διαστάσεις<sup>16</sup> [53].

Όλες οι απόψεις και οι θεωρίες συμφωνούν σε κάποια πράγματα που σχετίζονται με την αποθήκευση δεδομένων. Κατ' αρχάς, στο πιο στοιχειώδες επίπεδο, όλοι οι οργανισμοί επωφελούνται από τη δημιουργία μιας αποθήκης δεδομένων και ενός περιβάλλοντος ανάλυσης για την υποστήριξη της λήψης

<sup>16</sup> Συμμορφούμενη είναι μια διάσταση που έχει το ίδιο νόημα για κάθε γεγονός (fact) με το οποίο σχετίζεται. Οι συμμορφούμενες διαστάσεις επιτρέπουν την κατηγοριοποίηση και περιγραφή των γεγονότων και των μετρήσεων (measures) με τον ίδιο τρόπο σε πολλά data marts.

αποφάσεων. Δεύτερον, ο στόχος κάθε περιβάλλοντος αποθήκης δεδομένων είναι να δημοσιεύσει τα δεδομένα ακαριαία, δίχως απώλειες πληροφορίας και να υπάρχει ευκολία πρόσβασης στους υπεύθυνους λήψης αποφάσεων.

Οι δύο προσεγγίσεις, των Inmon και Kimball συμφωνούν ότι είναι συνετό να δομηθεί ένα περιβάλλον αποθήκης δεδομένων με κριτήρια τη μακροπρόθεσμη ενσωμάτωση και επεκτασιμότητα. Αν και υποσύνολα της αποθήκης δεδομένων πιθανόν να υλοποιηθούν σε φάσεις κατά τη διάρκεια του χρόνου, είναι ωφέλιμο να υπάρχει ένας ολοκληρωμένος τελικός στόχος κατά τον σχεδιασμό. Επίσης αποτελεί κοινή παραδοχή πως ο σχεδιασμός αυτόνομων data marts ή DW (Σχήμα 3.9) χαρακτηρίζετε προβληματικός και ασύμφορος.



**Σχήμα 3.9:** Σχεδιασμός αυτόνομων data marts ή DW

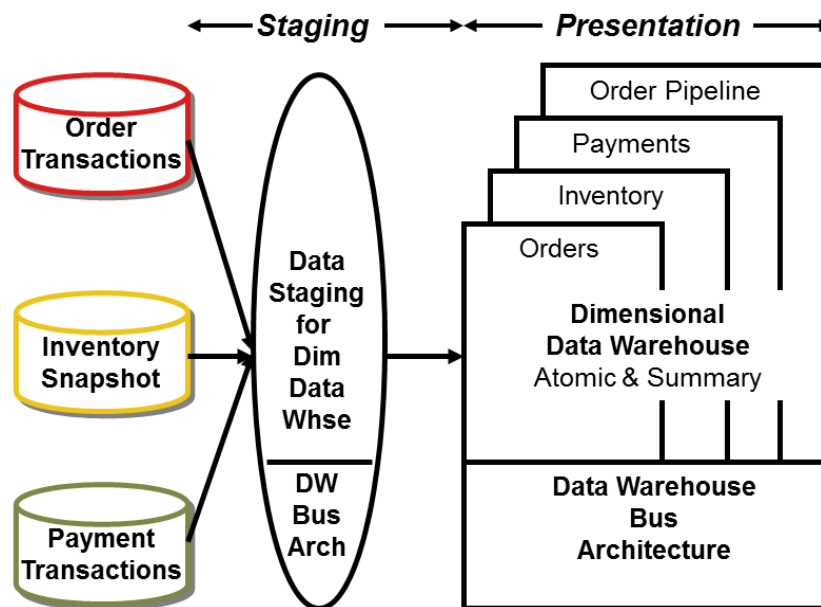
Αυτές οι ανεξάρτητες δομές έχουν κατασκευαστεί για να ικανοποιήσουν συγκεκριμένες ανάγκες, χωρίς να λαμβάνονται υπόψη άλλες υφιστάμενες ή υπό σχεδίαση. Αφορούν ένα τμήμα ή και ομάδα ενός οργανισμού και τείνουν να είναι απλές μεν στη φύση τους, συχνά δε χαλαρά και πρόχειρα δομημένες. Σε βάθος χρόνου μια ανεξάρτητη προσέγγιση δεν αποτελεί βιώσιμη λύση όπως επίσης γίνεται σπατάλη χώρου και ισχύς για πολλαπλές, ασυντόνιστες διεργασίες από συχνά ίδιες λειτουργικές πηγές πετυχαίνοντας μια δομή αναποτελεσματική και σπάταλη που συχνά αντιμετωπίζεται με αβεβαιότητα και αμφιβολία ως προς την ποιότητα των δεδομένων που φτάνουν στον χρήστη.

Παρακάτω παρουσιάζονται περιληπτικά οι δύο διαφορετικές προσεγγίσεις.

*Αρχιτεκτονική διαύλου data mart με συμμορφούμενες διαστάσεις*

Η προσέγγιση Kimball φαίνεται στο Σχήμα 3.10. Τα ανεπεξέργαστα δεδομένα μετατρέπονται σε ευπαρουσίαστες πληροφορίες στην ενδιάμεση περιοχή (staging area), πάντα δίνοντας ιδιαίτερη βάση στην απόδοση και την ποιότητα. Η διαδικασία αρχίζει με συντονισμένη εξαγωγή δεδομένων από λειτουργικές πηγές δεδομένων. Μερικές εργασίες όπως η συντήρηση και αποθήκευση των δεδομένων κοινής αναφοράς είναι κεντριοποιημένες, ενώ άλλες μπορούν να διανεμηθούν.

Τα μοντέλα διαστάσεων έχουν σχεδιαστεί σύμφωνα με την επιχειρηματική διαδικασία και όχι σύμφωνα με τα διάφορα τμήματα. Για παράδειγμα, τα δεδομένα παραγγελιών αποθηκεύονται μια φορά στην κανονικοποιημένη αποθήκη δεδομένων αντί να αποθηκεύονται, δύο και τρεις φορές, σε διάφορα data marts. Καθ' αυτό τον τρόπο όλες οι θεμελιώδεις διεργασίες των επιχειρήσεων είναι διαθέσιμες στην αποθήκη δεδομένων. Η προσέγγιση αυτή ενισχύει τις σχέσεις μεταξύ των δεδομένων επιχειρηματικής διαδικασίας (γεγονότα ή facts) και τα περιγραφικά χαρακτηριστικά (διαστάσεις ή dimensions).

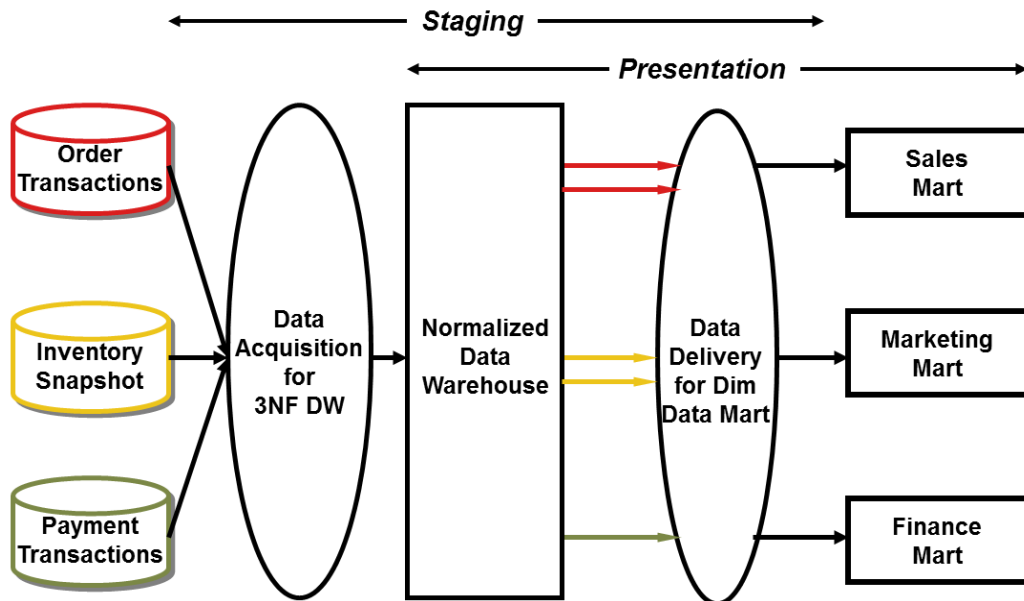


**Σχήμα 3.10:** Προσέγγιση Kimball: Αρχιτεκτονική διαύλου data mart με συμμορφούμενες διαστάσεις

## Hub-and-spoke αρχιτεκτονική

Το Σχήμα 3.11 απεικονίζει την hub-and-spoke αρχιτεκτονική προσέγγιση<sup>17</sup>. Η αρχική διαδικασία είναι ίδια όπως και στη προσέγγιση Kimball. Υπάρχει συντονισμένη εξαγωγή δεδομένων από λειτουργικές πηγές δεδομένων. Από εκεί φορτώνεται μια σχεσιακή βάση δεδομένων σε τρίτη κανονική μορφή<sup>18</sup> (3NF) που περιέχει τα ατομικά δεδομένα. Αυτή η κανονικοποιημένη αποθήκη δεδομένων χρησιμοποιείται για τη αποθήκευση επιπλέον δεδομένων παρουσίασης, συμπεριλαμβανομένων των αποθηκών ειδικού σκοπού για την εξερεύνηση και την εξόρυξη δεδομένων, καθώς και data marts.

Σε αυτή την προσέγγιση, τα data marts έχουν προσαρμοστεί με το κάθε τμήμα της επιχείρησής μέσα από πολυδιάστατες δομές δεδομένων. Τα ατομικά δεδομένα είναι προσβάσιμα μέσω της κανονικοποιημένης αποθήκη δεδομένων. Προφανώς, τα ατομικά δεδομένα έχουν διαρθρωθεί πολύ διαφορετικά από ότι τα συγκεντρωτικά.



Σχήμα 3.11: Προσέγγιση Inmon: Αρχιτεκτονική Hub-and-spoke

<sup>17</sup> Σε παλαιότερη βιβλιογραφία αναφέρεται και ως προσέγγιση EDW (Enterprise Data Warehouse)

<sup>18</sup> Ένας πίνακας θεωρείται ότι είναι σε τρίτη κανονική μορφή (3NF = 3<sup>rd</sup> Normal Form) αν κάθε ιδιότητα είναι ατομική και εξαρτάται από το κλειδί και από τίποτα άλλο εκτός από το κλειδί.

<http://www.businessintelligence-solutions.com/differences-1st-2nd-3rd-normal-form/>



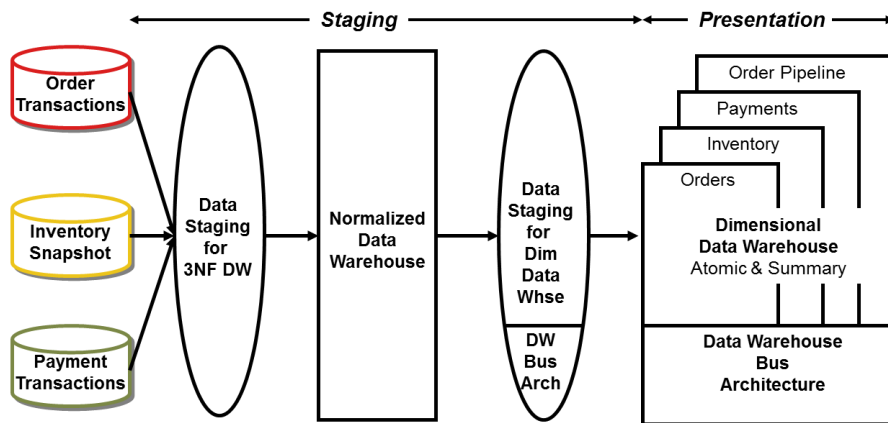
*Βασικές διαφορές και υβριδική εκδοχή*

Υπάρχουν δύο βασικές διαφοροποιήσεις μεταξύ των δύο προαναφερθέντων προσεγγίσεων. Η πρώτη αφορά την ανάγκη για μια κανονικοποιημένη δομή δεδομένων πριν από τη φόρτωση στην Αποθήκη Δεδομένων. Η δεύτερη βασική διαφορά αφορά την προσέγγιση των ατομικών δεδομένων. Οι χρήστες μπορεί να μην ενδιαφέρονται ή αντιλαμβάνονται τις λεπτομέρειες στην απόδοση ωστόσο, από τεχνικής πλευράς υπάρχουν λεπτομέρειες που καθορίζουν την επιλογή της προσέγγισης. Η κάθε προσέγγιση έχει τα δικά της πλεονεκτήματα και μειονεκτήματα απέναντι στην άλλη. Ο παρακάτω πίνακας συνοψίζει τις διαφορές μεταξύ των δύο προσεγγίσεων [54].

	<b>Kimball's Approach</b>	<b>Inmon's Approach</b>
1.	Everyone is allowed to fabricate their database according to their requirements and department structure. All these independent repositories can be integrated as and when required. This methodology is known as bottom up approach.	Supports a top down approach. Here no one is allowed to develop any database independently. The database for an organization should be planned and designed centrally. Every department within the organization will follow the centrally designed schema to fabricate their database.
2.	This structure is easier to build.	The structure proposed is very typical one to craft.
3.	It is a nimble approach.	Rigorous analysis and designing is required.
4.	Problematic to maintain as an enterprise resource.	Easier to maintain as an enterprise resource.
5.	Data is often redundant.	Redundancy is regulated to a great extent.
6.	Very difficult to integrate independent data marts with varying structure.	Integration of data marts is comparatively easier.
7.	This approach is flexible.	This approach is comparatively rigid.

**Πίνακας 3.1:** Βασικές διαφορές στις προσεγγίσεις Kimball και Inmon

Ορισμένοι οργανισμοί υιοθετούν μια υβριδική προσέγγιση (Σχήμα 3.13) στην οποία υπάρχει μια κανονικοποιημένη αποθήκη δεδομένων (hub-and-spoke) καθώς και μια πολυδιάστατη αποθήκη δεδομένων ατομικών και συγκεντρωτικών στοιχείων που βασίζονται στην αρχιτεκτονική διαύλου data mart με συμμορφούμενες διαστάσεις.



**Σχήμα 3.12:** Υβριδική προσέγγιση με στοιχεία βασιζόμενα και στις δύο κυρίαρχες αρχιτεκτονικές

Ως λογικό επακόλουθο έρχεται η ερώτηση για το ποια μπορούμε να θεωρήσουμε ως κριτήρια μέτρησης αποτελεσματικότητας και επιτυχίας μιας αρχιτεκτονικής αποθήκευσης δεδομένων. Απαντώντας σε αυτό το ερώτημα διεξήχθει η έρευνα των Ariyachandra και Watson για την καλύτερη κατανόηση των παραγόντων που

επηρεάζουν την επιλογή και την επιτυχία των διαφόρων αρχιτεκτονικών. Για τη μελέτη, οι συγγραφείς σχημάτισαν μια ομάδα 20 εμπειρογνομώνων για να βοηθήσουν στον εντοπισμό των αρχιτεκτονικών λύσεων προς μελέτη και τα μετρήσιμα κριτήρια επιτυχίας τους στη πράξη. Οι Bill Inmon και Ralph Kimball ήταν μεταξύ των συμμετεχόντων και τα κριτήρια που κατέληξαν είναι [53]:

1. Ποιότητα της πληροφορίας: που αφορά την ακρίβεια, την πληρότητα, και την συνέπεια των πληροφοριών. Να μην διακινδυνεύετε η απώλεια πληροφορίας κατά την διαδικασία που περνούν τα δεδομένα.
2. Την ποιότητα του συστήματος: περιλαμβάνει την ευελιξία του συστήματος, επεκτασιμότητα, και την ευκολία ένταξης του (integration).
3. Επιπτώσεις σε ατομικό επίπεδο: οι οποίες αναφέρονται στο κατά πόσο οι χρήστες μπορούν γρήγορα και εύκολα να έχουν πρόσβαση στα δεδομένα.
4. Επιπτώσεις σε επίπεδο οργανισμού: όταν πληροί τις απαιτήσεις των οργανισμών, διευκολύνει τη χρήση της ΕΝ και στηρίζει την επίτευξη των στρατηγικών στόχων.

Η έρευνα βρήκε τις δύο προσεγγίσεις με παρόμοια αποτελέσματα, εύρημα το οποίο δεν προκάλεσε έκπληξη καθώς κατά κάποιο τρόπο, οι αρχιτεκτονικές έχουν εξελιχθεί με την πάροδο του χρόνου και έχουν γίνει πιο παρόμοιες. Και οι δύο τονίζουν την ανάγκη να μικρών και ταχύτατων αποτελεσμάτων κατά την υλοποίηση έχοντας πολύ καλό μακροπρόθεσμο σχεδιασμό.

#### **Γ4. Υπηρεσίες Ανάλυσης και Αναφορών**

##### *Γ4.1 OLAP και OLTP*

Ο ορισμός που δίνεται από την πλατφόρμα ενημέρωσης σχετικά με τις τεχνολογίες OLAP<sup>19</sup>, αναφέρει πως η OLAP (Online Analytical Processing) είναι από τις τεχνολογίες εφαρμογών της EN. Αποτελεί μια δυνατή λύση για ανακάλυψη δεδομένων (data discovery) ενώ παρέχει απεριόριστες δυνατότητες για απεικόνιση και αναφορές (reporting). Επίσης δίνει την δυνατότητα για περίπλοκους αναλυτικούς υπολογισμούς καθώς και για αναλυτικά προγνωστικά “what if” σενάρια (π.χ. σχεδίαση προϋπολογισμού).

Σύμφωνα με την Oracle, οι Διαδικασίες Συναλλαγών σε Απευθείας Σύνδεση (Online Transaction Processing, OLTP) είναι εφαρμογές υψηλής απόδοσης και συχνής χρήσης λειτουργιών βάσεων δεδομένων. Οι εφαρμογές αυτές χαρακτηρίζονται από τον αυξανόμενο όγκο των δεδομένων καθώς αρκετές εκατοντάδες χρήστες ενδέχεται να έχουν πρόσβαση ταυτόχρονα. Τυπικές εφαρμογές OLTP είναι τα συστήματα κρατήσεων αεροπορικών εταιρειών, μεγάλες εφαρμογές παραγγελιοληψίας και τραπεζικές εφαρμογές. Οι βασικοί στόχοι OLTP εφαρμογών αφορούν την διαθεσιμότητα (ενίοτε 7 ημέρες / 24 ώρες), απόδοση (throughput) συγχρονισμού και ανάκτησης<sup>20</sup>.

##### *Γ4.2 Εξόρυξη Δεδομένων (Data Mining)*

Η Εξόρυξη Δεδομένων μετατρέπει έναν όγκο δεδομένων σε γνώση. Για αυτόν τον λόγο αναφέρεται σε παλαιότερη βιβλιογραφία και ως ανακάλυψη γνώσης (knowledge discovery from data, KDD) [55]. Ο όρος αναφέρεται στη διερεύνηση

<sup>19</sup> <http://olap.com/olap-definition/>

<sup>20</sup> Περισσότερες πληροφορίες στο παρακάτω σύνδεσμο  
[http://docs.oracle.com/cd/A87860\\_01/doc/server.817/a76992/ch3\\_eval.htm#2680](http://docs.oracle.com/cd/A87860_01/doc/server.817/a76992/ch3_eval.htm#2680)

και ανάλυση μεγάλων όγκων δεδομένων, προκειμένου να ανακαλυφτούν κανόνες και πρότυπα. Αν και για τους σκοπούς της εργασίας υποθέτουμε ότι ο στόχος της εξόρυξης δεδομένων είναι η υποστήριξη των στόχων των οργανισμών οι τεχνικές και τα εργαλεία εξόρυξης δεδομένων εφαρμόζονται εξίσου σε τομείς που κυμαίνονται από την αστρονομία μέχρι την ιατρική [56].

Η εξόρυξη δεδομένων εφαρμόζεται με δύο διαφορετικές λογικές, την κατευθυνόμενη και μη κατευθυνόμενη. Η κατευθυνόμενη εξόρυξη δεδομένων επιχειρεί να εξηγήσει ή να κατηγοριοποιήσει κάποιο συγκεκριμένο πεδίο ή ιδιότητα όπως για παράδειγμα ποσά πωλήσεων. Η μη κατευθυνόμενη εξόρυξη δεδομένων σκοπό έχει την εξεύρεση συσχετίσεων ή μοτίβων σε μεγάλες σχεσιακές βάσεις δεδομένων χωρίς τη χρήση ενός συγκεκριμένου πεδίου ή ιδιότητας [56].

Η επιλογή ενός συγκεκριμένου συνδυασμού τεχνικών για την εφαρμογή εξόρυξης δεδομένων εξαρτάται από τη φύση των δεδομένων, καθώς και το επίπεδο γνώσης των διαχειριστών της διαδικασίας. Πολλές φορές υπάρχει ανάγκη για γνώσεις στατιστικής, machine learning, data warehouses κ.α. [55]. Κατά το συνέδριο IEEE International Conference of Data Mining (ICDM) το 2006 παρουσιάστηκε η λίστα των 10 αλγορίθμων με την πιο σημαντική επιρροή<sup>21</sup>:

1. C4.5
2. k-means
3. SVM (Support Vector Machine)
4. Apriori
5. EM (Expectation Maximization)
6. PageRank
7. AdaBoost
8. kNN (k-nearest neighbors)
9. Naïve Bayes
10. CART

Οι αλγόριθμοι αυτοί καλύπτουν διαδικασίες ταξινόμησης, ομαδοποίησης, παλινδρόμησης, ανάλυσης σύνδεσης, και ανάλυση εργασίας. Οι περισσότεροι από αυτούς τους αλγόριθμους είναι διαθέσιμοι σε εμπορικά και ανοικτού κώδικα συστήματα εξόρυξης δεδομένων.

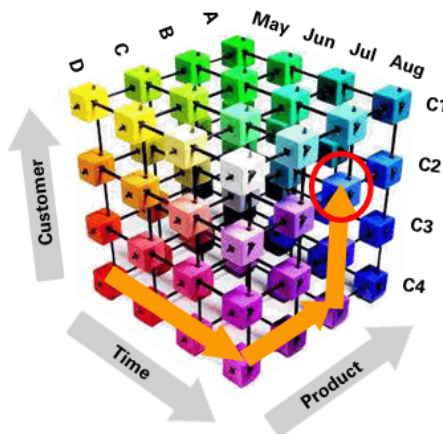
---

<sup>21</sup> Τα συμπεράσματα από το συνέδριο IEEE ICDM σχετικά με τους αλγόριθμους data mining βρίσκονται στον παρακάτω σύνδεσμο

<http://www.cs.uvm.edu/%7Eicdm/algorithms/index.shtml>

### Γ4.3 Κύβος Δεδομένων (Data Cube)

Στα πλαίσια της επιστήμης δεδομένων, ένας κύβος δεδομένων (Data cube) είναι ένας πίνακας τιμών τριών (ή περισσότερων) διαστάσεων, που χρησιμοποιείται συνήθως για να περιγράψει μια σειρά από δεδομένα συσχετίσεων. Κάθε διάσταση αντιπροσωπεύει ένα νέο χαρακτηριστικό στη βάση δεδομένων και κάθε κελί του κύβου αντιπροσωπεύει το μέγεθος που μας ενδιαφέρει. Οι εφαρμογές ανάλυσης δεδομένων συνήθως συγκεντρώνουν δεδομένα σε πολλές διαστάσεις ψάχνοντας για ανωμαλίες ή ασυνήθιστα σχήματα. Τέτοιες εφαρμογές χρειάζονται την πολυδιάστατη συσχέτιση, θα μπορούσαμε να πούμε πως η καινοτομία είναι ότι κύβοι είναι οι σχέσεις μεταξύ πινάκων [57]. Η μελέτη των κύβων δεδομένων αποτελεί ένα μεγάλο κεφάλαιο σε συνδυαζόμενες επιστήμες όπως η άλγεβρα, η στατιστική, η επιστήμη δεδομένων κ.α. έχοντας ως αποτέλεσμα την δυσκολία ορισμού και απλής επεξήγησης του όρου δίχως την παρουσίαση τεχνικών λεπτομερειών. Στο παρακάτω Σχήμα (Σχήμα 3.13) παρουσιάζεται μια απεικόνιση κύβου δεδομένων 3-διαστάσεων (Πελάτης - Προϊόν - Χρονική Περίοδος) για την καλύτερη κατανόηση του όρου.



**Σχήμα 3.13:** Απεικόνιση κύβου δεδομένων 3-διαστάσεων

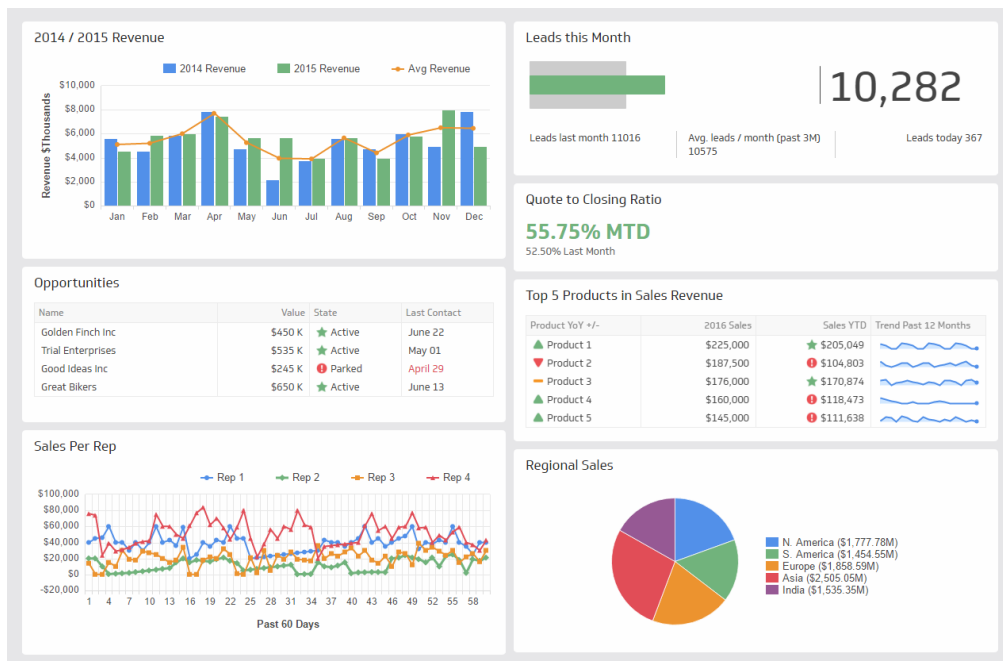
### Γ4.4 Προχωρημένα Συστήματα Αναλύσεων (Advanced Analytics)

Ως προχωρημένα συστήματα αναλύσεων ορίζεται η διαδικασία ανακάλυψης διαχείρισης και επικοινωνίας μοτίβων δεδομένων (data patterns). Αποτελεί ένα ευρέως χρησιμοποιούμενο εργαλείο καθώς παρέχει την δυνατότητα εξαγωγής

συμπερασμάτων καθώς και αποφάσεων σε όλους τους τομείς μιας επιχειρήσης (πωλήσεις, marketing, οικονομικό, διαχείριση αποθήκης κ.α). Τα Advance Analytics απαιτούν την συνεργασία της μαθηματικής, στατιστικής και πληροφορικής επιστήμης καθώς διαχειρίζονται μεγάλο όγκο δεδομένων (big data) με μεγάλες απαιτήσεις σε υπολογιστική ισχύ και μικρό περιθώριο σφαλμάτων.

#### Γ4.5 Οπτικοποίηση δεδομένων (Data visualization)

Η οπτικοποίηση των δεδομένων (data visualization) αποτελεί τη διαδικασία αναπαράστασης δεδομένων με τη μορφή γραφημάτων, η οποία μπορεί να συμβάλει στην καλύτερη και βαθύτερη κατανόηση της σημασίας των δεδομένων και των μεταβλητών ή των μονάδων που συνιστούν τα δεδομένα αυτά. Λόγω των τεράστιων και συνεχώς αυξανόμενων ποσοτήτων και πηγών πληροφορίας, η ανάγκη για οπτικοποίηση είναι μεγάλη για αυτό τον λόγο έχουν αναπτυχθεί πολυάριθμες εμπορικές εφαρμογές που παρέχουν αυτή την δυνατότητα. Εφαρμόζοντας διάφορα μέσα οπτικοποίησης τα δεδομένα εξετάζονται μαζικά και γρήγορα από χρήστες με διαφορετικό επιστημονικό και επαγγελματικό υπόβαθρο. Παραδείγματα εμπορικών λύσεων αποτελούν το SAP Business Objects και το MicroStrategy τα οποία αναλύονται στο επόμενο κεφάλαιο.



**Σχήμα 3.14:** Παραδείγματα γραφημάτων κατά την απεικόνιση δεδομένων. Η εικόνα των δεδομένων όπως παρουσιάζεται στον τελικό χρήστη.



## ΚΕΦΑΛΑΙΟ Δ – ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ (CASE STUDY)

Σε αυτό το κεφάλαιο παρουσιάζεται μια πρακτική εφαρμογή της EN όπως αυτή αναπτύχθηκε σε κυρίαρχη πολυεθνική εταιρία. Τα ονόματα και οι πληροφορίες, καθώς και τα νούμερα έχουν παραποιηθεί σε βαθμό έτσι ώστε να μην διαρρεύσει καμία ευαίσθητη πληροφορία για την επιχείρηση δίχως να διαβρωθεί καθόλου η αξία και η ουσία της εφαρμογής EN.

### **Δ1. Εισαγωγή**

#### *Δ1.1 Σύγχρονες ανάγκες της αγοράς*

Μεγάλες εταιρίες και οργανισμοί έρχονται συχνά αντιμέτωποι με ένα πρόβλημα στο οποίο η λύση πρέπει να βρεθεί γρήγορα, αποτελεσματικά και κυρίως έχοντας ένα μακροπρόθεσμο πλάνο. Μια λύση της οποίας η εφαρμογή έχει ημερομηνία λήξης ή δεν θα μπορεί να βελτιωθεί και κλιμακωθεί σε άλλα επίπεδα αποτελεί κακή επένδυση πόρων. Η επικοινωνία μιας ανάγκης πολλές φορές γίνεται από - ή προς - αναρμόδια άτομα τα οποία έχουν μερικές φορές λανθασμένη εικόνα για το πρόβλημα, ή ακόμα χειρότερα, για την λύση.

Όπως έχει αναλυθεί και στο θεωρητικό κομμάτι, ως EN δεν εννοείται μια απλώς τεχνολογική λύση αλλά η τομή των αναγκών των χρηστών, με τους στόχους της επιχείρησης και τις τεχνολογικές δυνατότητες αυτής. Για αυτό τον λόγο, παρότι δεν είναι το κεντρικό θέμα που μας απασχολεί, μια καλή ανάλυση και εισαγωγή θα πρέπει να γίνει και στις επιχειρηματικές ανάγκες. Επίσης σημαντικό είναι να αναλυθούν οι συνθήκες της εταιρίας και της αγοράς καθώς όσο καλή και να είναι μια λύση σε θεωρητικό επίπεδο, όταν έρχεται σε σύγκρουση με την πραγματικότητα τα πράγματα τείνουν να είναι λίγο διαφορετικά.

Παρακάτω αναλύονται οι δεδομένες συνθήκες και οι ανάγκες όπως αυτές προέκυψαν από τα υπόλοιπα τμήματα προς το Τμήμα Πληροφορικής. Ακολουθεί μια ανάλυση του περιβάλλοντος, του πλάνου καθώς και του επιθυμητού αποτελέσματος.



### *Δ1.1 Η ανάγκη για EN*

Η εταιρία είναι μία από τους κυρίαρχους στον κλάδο της ένδυσης και υπόδησης στον τομέα της χονδρικής αλλά και λιανικής πώλησης ( στα Αγγλικά γνωστά ως Wholesale και Retail αντίστοιχα). Έχει δηλαδή ως πελάτες, εταιρίες (Wholesale) όσο και καταναλωτές (Retail). Για λόγους ευκολίας και συνεννόησης εφ' εξής θα αναφέρεται η εταιρία αυτή ως Εταιρία, οι πελάτες χονδρικής ως απλά Πελάτες ή Πελάτης και οι καταναλωτές ως Καταναλωτές ή Καταναλωτής. Για την ανάπτυξη της λύσης EN συνεργάστηκαν το Τμήμα Πωλήσεων Χονδρικής, το Τμήμα Προμηθειών, το Τμήμα Marketing καθώς και το Τμήμα Πληροφορικής (εφ' εξής αναφέρεται ως IT).

Ένα από τα πιο σημαντικά κομμάτια της Εταιρίας είναι οι αναφορές από τα διάφορα τμήματα. Όπως είναι εύκολα κατανοητό, από τις σημαντικότερες αναφορές είναι οι αναφορές πωλήσεων (Sales Reports). Μέσω του τμήματος IT είναι σχετικά εύκολο να παρακολουθούνται οι πωλήσεις λιανικής, δηλαδή τα κέρδη ανά προϊόν ή ανά κατάστημα. Δηλαδή στην απλή ερώτηση «η καινούργια μπλούζα που πουλάμε πόσο καλά αποτελέσματα φέρνει;» μπορεί να απαντηθεί εύκολα και αξιόπιστα προς όλους τους ενδιαφερόμενους σε ένα Retail Sales Report.

Σε επίπεδο χονδρικής από την άλλη πλευρά τα πράγματα είναι περισσότερο περίπλοκα. Από πλευράς οικονομικών υπάρχουν λογιστικές και χρηματοοικονομικές διαδικασίες ώστε να φαίνονται οι συνολικές πωλήσεις προς τους Πελάτες ανά χρηματοοικονομική περίοδο. Ωστόσο ένα λεπτό ερώτημα είναι «πόσες και πότε γίνονται οι πωλήσεις προς τους Καταναλωτές αναλυτικά;». Ο τρόπος που για χρόνια λειτουργούσε η Εταιρία ήταν μεταφέροντας την ερώτηση προς τους Πελάτες. Ωστόσο, κατά πόσο μια τέτοια απάντηση είναι έγκαιρη, αξιόπιστη και αναλυτική;

Την σήμερον ημέρα, με τα εργαλεία που μπορεί να προσφέρει το IT, αναπτύχθηκε ένα ερώτημα για το πώς θα μπορέσει η εταιρία να έχει μεγαλύτερο έλεγχο των πληροφοριών αυτών. Με μία πρόταση θα θέλαμε να παρακολουθούμε τις αγορές των Καταναλωτών από τα καταστήματα των Πελατών όσο εύκολα και αξιόπιστα μπορούμε να το παρακολουθούμε από τα καταστήματα της Εταιρίας.

### *Δ1.2 Επιθυμητό Αποτέλεσμα*

Το τελικό αποτέλεσμα δεν είναι τίποτα παραπάνω από ένα Retail Sales Report για Wholesale Customers όπου θα μπορεί να γίνει ανάλυση πωλήσεων και σε επίπεδο προϊόντος αλλά και σε επίπεδο Πελατών (μια καλύτερη εικόνα για την διεύθυνση αγοράς). Απαντήσεις που θέλουμε να δίνονται μέσα στο Report αφορούν τις πωλήσεις ανά προϊόν εντός και εκτός season (πχ ένα καλοκαιρινό ρούχο να πωληθεί το καλοκαίρι ή τον χειμώνα), την ζήτηση προϊόντων, την αποτελεσματικότητα του Τμήματος Πωλήσεων, την αποτελεσματικότητα των Πελατών και την συγκεντρωτική εικόνα πωλήσεων.

Η λύση ζητείται να είναι διαθέσιμη προς διάφορους χρήστες. Την Διοίκηση, τα διάφορα εμπλεκόμενα τμήματα καθώς και προς τους εργαζομένους αυτών έχοντας αναλόγως διαφορετικά επίπεδα πρόσβασης.

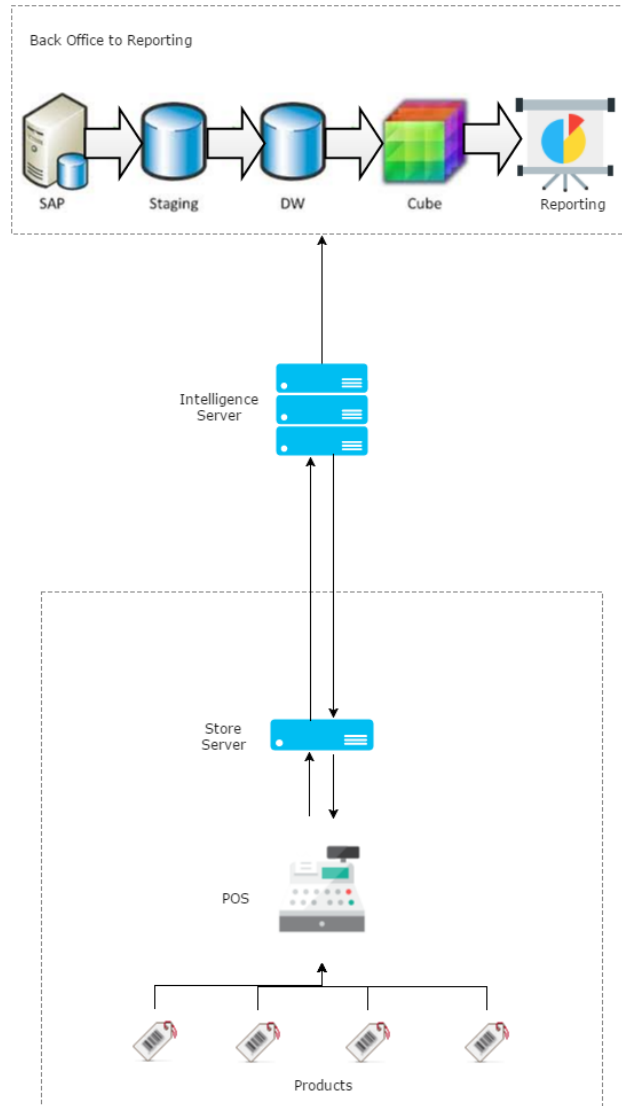
Το IT συμφώνησε πως τα παραπάνω μπορούν να επιτευχθούν μέσα από Sales Report το οποίο θα έχει διάφορα γραφήματα τα οποία αναπαριστούν τα δεδομένα. Όπως είναι λογικό το σύστημα που θα υλοποιήσει την ΕΝ θα πρέπει να δομηθεί ξεκινώντας από την Πηγή Δεδομένων καταλήγοντας στις υπηρεσίες ανάλυσης και διανομής πληροφοριών συνδυάζοντας υπάρχοντα και νέα συστήματα. Για παράδειγμα, τα master data υπάρχουν ήδη στην Βάση Δεδομένων, τα data marts μας δίνουν όλα τα διαθέσιμα χαρακτηριστικά που ενδιαφέρουν τα διάφορα τμήματα. Ωστόσο ακόμα και αν δεν υπάρχει διαθέσιμη πηγή για τα sales data των Πελατών στην Εταιρία, αυτό θα μπορούσε να δημιουργηθεί με την βοήθεια των Τμημάτων Πληροφορικής των Πελατών με πληροφορίες από τα POS (ένα πρόβλημα που προκύπτει είναι ότι ο κάθε Πελάτης έχει διαφορετικές POS back office λύσεις).

## **Δ2. Ανάλυση δεδομένων συνθηκών και αναγκών**

### *Δ2.1 Εργαλεία Επιχειρησιακής Νοημοσύνης*

Το αποτέλεσμα της EN είναι η εξαγωγή χρήσιμων συμπερασμάτων μέσα από προβολή των δεδομένων προς τα άτομα που είναι υπεύθυνα για την λήψη αποφάσεων. Σε μια επιχείρηση όπως η Εταιρία είναι πολύ σημαντικό να υπάρχουν διαφορετικές προβολές των δεδομένων ώστε να υπάρχει ανάλυση από διαφορετικές σκοπιές. Ο ακρογωνιαίος λίθος του συστήματος δεν είναι άλλος από τα δεδομένα ξεκινώντας από την εξαγωγή αυτών έως την δημιουργία αναφορών οι οποίες μετατρέπουν τα απλά δεδομένα (raw data) σε χρήσιμες πληροφορίες. Σε αυτό το case study μας ενδιαφέρουν οι πωλήσεις λιανικής των Πελατών χονδρικής οπότε τα δεδομένα ακολουθούν μια μεγάλη και δαιδαλώδη διαδικασία έως ότου εμφανιστούν ως έγκυρα, σύγχρονα και αξιόπιστα Sales Reports.

Η διαδικασία για την εξαγωγή ενός Sales Report είναι απλή όσον αφορά πωλήσεις λιανικής, δηλαδή από τα καταστήματα της Εταιρίας. Όταν γίνεται μια πώληση δύο λογισμικά συνεργάζονται. Ένα λογισμικό Front Office το οποίο είναι εγκατεστημένο στο κατάστημα (τερματικό ταμιακής και server), και ένα ή περισσότερα λογισμικά Back Office τα οποία είναι εγκατεστημένα στους τοπικούς και κεντρικούς servers. Το ένα λειτουργικό δηλαδή λειτουργεί στο προσκήνιο και το άλλο στο παρασκήνιο. Στο παρακάτω σχήμα αναπαρίσταται απλοϊκά η διαδρομή των δεδομένων για ένα Retail Sales Report.



**Σχήμα 4.1:** Διαδρομή των δεδομένων για ένα Retail Sales Report

Ένα Front Office λογισμικό είναι λογισμικό που χρησιμοποιεί η ταμιακή μηχανή στα καταστήματα. Έχει ως κύρια είσοδο ένα scanner το οποίο σαρώνει το barcode του προϊόντος και αυτόματα όλες οι πληροφορίες (metadata) έρχονται στην οθόνη του τερματικού πωλήσεων μέσω μιας κλήσης προς τον κεντρικό Intelligence Server. Στην περίπτωση που κλείσει μια πώληση, δηλαδή έχουμε την δημιουργία απόδειξης ή τιμολογίου, το Back Office σύστημα στέλνει τα δεδομένα της πώλησης στον τοπικό server όπου γίνεται ο εμπλουτισμός των δεδομένων με χαρακτηριστικά του καταστήματος καθώς και διάφορα άλλα νομικώς απαραίτητα χαρακτηριστικά (τρόπος πληρωμής, αριθμός απόδειξης και ταμιακής μηχανής κ.α.). Ο τοπικός server έπειτα στέλνει τα δεδομένα στον Intelligence Server ο

οποίος ενεργοποιεί κλήσεις προς τα διάφορα συστήματα Back Office και Βάσεις Δεδομένων για ενημέρωση.

Το Back Office σύστημα μετά τον συγκερασμό των δεδομένων αποστέλλει σήμα ορθώς ή λανθασμένο προς τους διαχειριστές και δημιουργεί τα Δεδομένα Συναλλαγών τα οποία και αποθηκεύονται στην κεντρική Βάση Δεδομένων. Με αυτόν τον τρόπο μπορεί να γίνει έλεγχος των πωλήσεων. Για παράδειγμα σε περίπτωση που ένα προϊόν έχει σφάλμα στην τιμολόγηση. Από εκεί και πέρα ένα λογισμικό Back Office παρέχει διάφορες δυνατότητες προς τους χρήστες (εξαρτάται από τον πάροχο). Ύστερα και μέσω διαφόρων κλήσεων συστήματος, και σε συνεργασία με την Βάση Δεδομένων, περίπλοκα SQL queries εκτελούνται και τα αποτελέσματα τροφοδοτούν τις εφαρμογές που εμφανίζουν τα Sales Reports σε πίνακες και γραφήματα διαθέσιμα για διάφορους χρήστες.

Στην Εταιρία χρησιμοποιείται για Front Office η λύση SAP JK και για Back Office οι λύσεις SAP HANA, SAP ERP και MicroStrategy η κάθε μία καλύπτοντας διαφορετικές ανάγκες και με διαφορετικές δυνατότητες οι οποίες θα αναλυθούν παρακάτω.

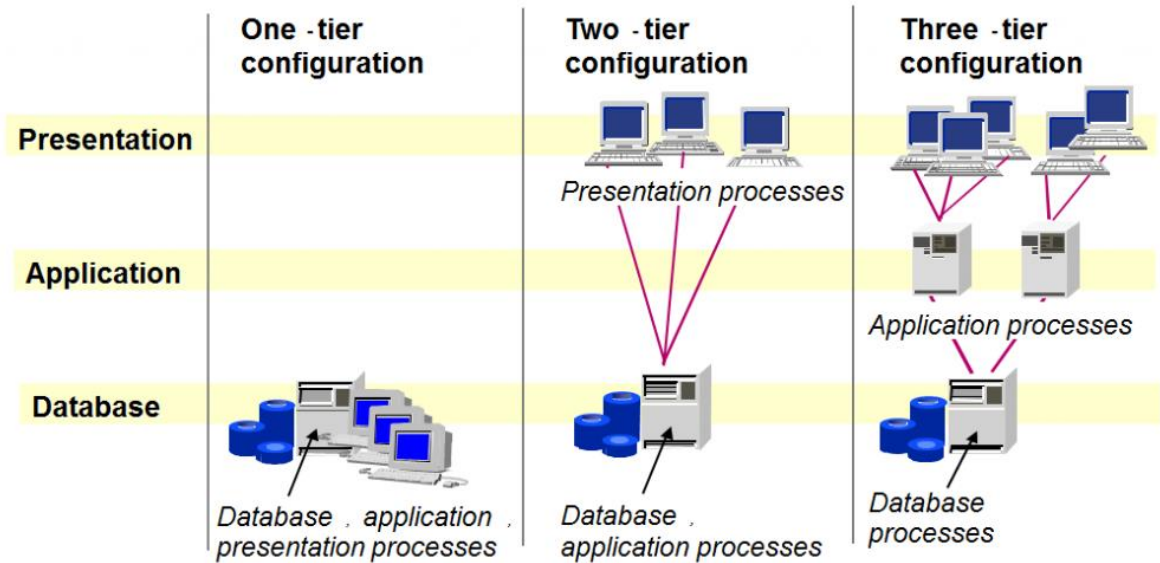
### *Εργαλείο SAP<sup>22</sup>*

Τα προϊόντα του λογισμικού SAP παρέχουν ισχυρά εργαλεία για τις εταιρείες ώστε να διαχειρίζονται τους τομείς οικονομικών, logistics, ανθρώπινων πόρων και διάφορους άλλους. Το σημαντικότερο προϊόν που προσφέρει το λογισμικό SAP είναι το SAP ERP System το οποίο είναι από τα πιο προηγμένα συστήματα Enterprise Resource Planning (ERP) διαθέσιμα στην αγορά ωστόσο δεν είναι το μοναδικό προϊόν. Το SAP ERP παρέχει δυνατότητες για την ενσωμάτωση δεδομένων από διαφορετικές εφαρμογές και είναι δομημένο σε 3-tier αρχιτεκτονική client/server. Η αρχιτεκτονική τριών επιπέδων συνιστάτε διότι επιτρέπει υψηλή επεκτασιμότητα και ευελιξία, ωστόσο υπάρχει δυνατότητα ανάπτυξης σε 2-tier και 1-tier αρχιτεκτονικές. Το παρακάτω Σχήμα παρέχει μια γραφική απεικόνιση των διαφορών μεταξύ αυτών των αρχιτεκτονικών.

---

<sup>22</sup> Πηγή πληροφοριών καθώς και περισσότερες λεπτομέρειες στον παρακάτω σύνδεσμο:

<http://sap-certification.info/how-does-sap-work/>



**Σχήμα 4.2:** Διαφορετικές αρχιτεκτονικές συστήματος SAP

Στην αρχιτεκτονική SAP τριών επιπέδων, το επίπεδο παρουσίασης παρέχει τη διεπαφή με τον χρήστη, το επίπεδο εφαρμογών εκτελεί τις επιχειρηματικές εργασίες και το επίπεδο της Βάσης Δεδομένων αποθηκεύει τα δεδομένα.

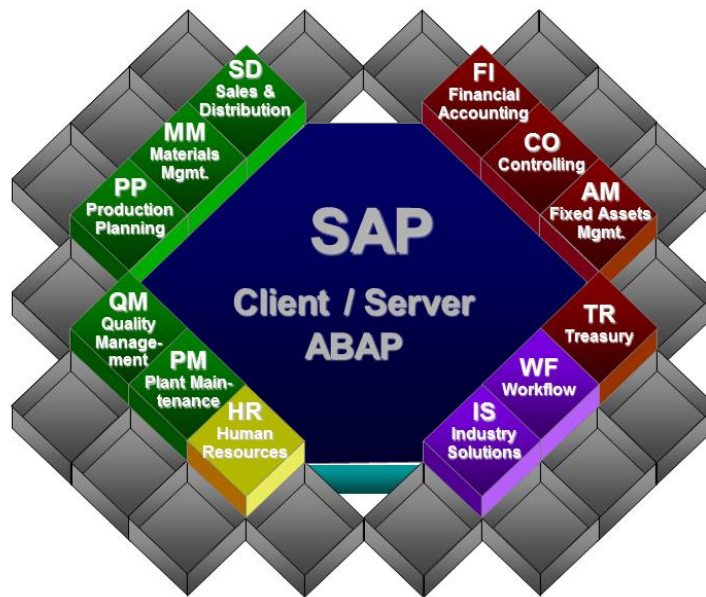
Η σειρά παρουσίασης βρίσκεται στα τερματικά των χρηστών μέσω της εφαρμογής SAP GUI η οποία μπορεί να εγκατασταθεί σε οποιοδήποτε υπολογιστή που τρέχει MS Windows ή Mac OS και παρέχει τη διεπαφή για την επικοινωνία μεταξύ του χρήστη και του συστήματος SAP ERP.

Το επίπεδο εφαρμογής, η καρδιά του συστήματος SAP ERP, βρίσκεται εγκατεστημένο σε servers και είναι υπεύθυνο για την επεξεργασία συναλλαγών, τις εργασίες εκτύπωσης, την δημιουργία εκθέσεων, τον συντονισμό και πρόσβαση στη βάση δεδομένων, και τη διασύνδεση με άλλες εφαρμογές.

Η Βάση Δεδομένων χρησιμοποιείται για την αποθήκευση δύο τύπου αντικειμένων: τα στοιχεία των επιχειρήσεων (fact data) και τα προγράμματα εφαρμογών SAP. Τα στοιχεία των επιχειρήσεων αντιπροσωπεύουν δεδομένων που δημιουργούνται ως μέρος των διαφόρων επιχειρηματικών διαδικασιών που εκτελούν οι χρήστες (για παράδειγμα, οι παραγγελίες πωλήσεων) ενώ τα προγράμματα εφαρμογών SAP είναι ρουτίνες γραμμένες σε ABAP (ειδική γλώσσα προγραμματισμού που χρησιμοποιείται στο SAP). Είναι δυνατή η χρήση Βάσεων Δεδομένων από διάφορους προμηθευτές (για παράδειγμα, η Oracle ή Microsoft)

και είναι στην ευχέρεια της επιχείρησης να αποφασίσει τον τύπο της Βάσης Δεδομένων

Στηριζόμενη σε αυτά τα επίπεδα, η αρχιτεκτονική SAP μπορεί να υποδιαιρεθεί περαιτέρω σε διάφορες ενότητες της εφαρμογής. Κάθε ενότητα περιέχει τα προγράμματα εφαρμογών SAP και τα επιχειρηματικά δεδομένα που σχετίζονται με ένα συγκεκριμένο επιχειρηματικό τομέα, όπως χρηματοοικονομικής λογιστικής (FI) ή διαχείρισης υλικών (MM). Στο παρακάτω σχήμα φαίνονται οι μονάδες της εφαρμογής που περιλαμβάνονται στο σύστημα SAP ERP.

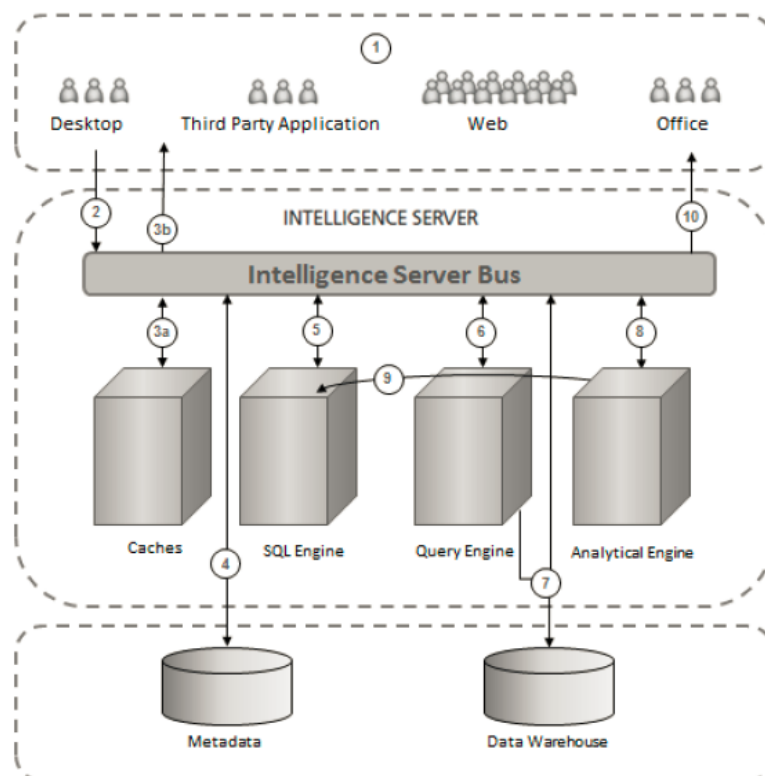


**Σχήμα 4.3:** Οι εφαρμογές του συστήματος SAP ERP (Application Modules)

Η Εταιρία χρησιμοποιεί το SAP ERP δομημένο σε 3-tier αρχιτεκτονική με τις ενότητες FiCo, SD, MM και HR. Οι δύο κύριες ενότητες που αφορούν την εφαρμογή EN σε αυτό το case study είναι τα MM και SD καθώς εκεί αποθηκεύονται αντίστοιχα όλα τα δεδομένα των προϊόντων και τα δεδομένα πωλήσεων χονδρικής σε συνδυασμό με τα στοιχεία των Πελατών. Ο λεπτομερής και τεχνικός τρόπος λειτουργίας και δομής του SAP ERP ξεφεύγει από τους σκοπούς της εργασίας ωστόσο αξίζει να αναφερθεί το γεγονός ότι οι Πωλητές μέσω διαφόρων εντολών στο επίπεδο παρουσίασης έχουν την δυνατότητα να εξάγουν δεδομένα σχετικά με τις πωλήσεις χονδρικής με βάση τον Πελάτη, το Προϊόν ή χαρακτηριστικά αυτών σε μορφή αρχείου φυλλομετρητή.

### Υπηρεσία Ανάλυσης και Αναφορών SAP Business Objects.

Η υπηρεσία ανάλυσης και αναφορών η οποία λειτουργεί στην εταιρία είναι το Business Objects το οποίο είναι ένα ακόμα προϊόν της SAP. Η εγκατάσταση της υπηρεσίας έχει γίνει στον κεντρικό Intelligence Server, έναν server ο οποίος ονομάζεται έτσι καθώς παρέχει δυνατότητες εκτέλεσης ερωτημάτων και αναλύσεων με σκοπό την τελική δημιουργία Κύβων ή Μικροκύβων δεδομένων, εκτελεί δηλαδή την OLAP διαδικασία<sup>23</sup>. Η διαδικασία εκτέλεσης ερωτημάτων αναλύεται στο παρακάτω σχήμα.



**Σχήμα 4.4:** Διαδικασία εκτέλεσης ερωτημάτων στο SAP Business Objects

#### Διαδικασία εκτέλεσης ερωτημάτων Reporting

1. Ένας χρήστης βρίσκεται στην εφαρμογή αναφορών, Reporting
2. Καλείται ένα Sales Report από το τερματικό. Στον Intelligence Server ο πυρήνας δέχεται την εντολή και συντονίζει τις δραστηριότητες

<sup>23</sup> Περισσότερες πληροφορίες στον παρακάτω σύνδεσμο:  
<http://www.element61.be/e/resourc-detail.asp?ResourceId=547>

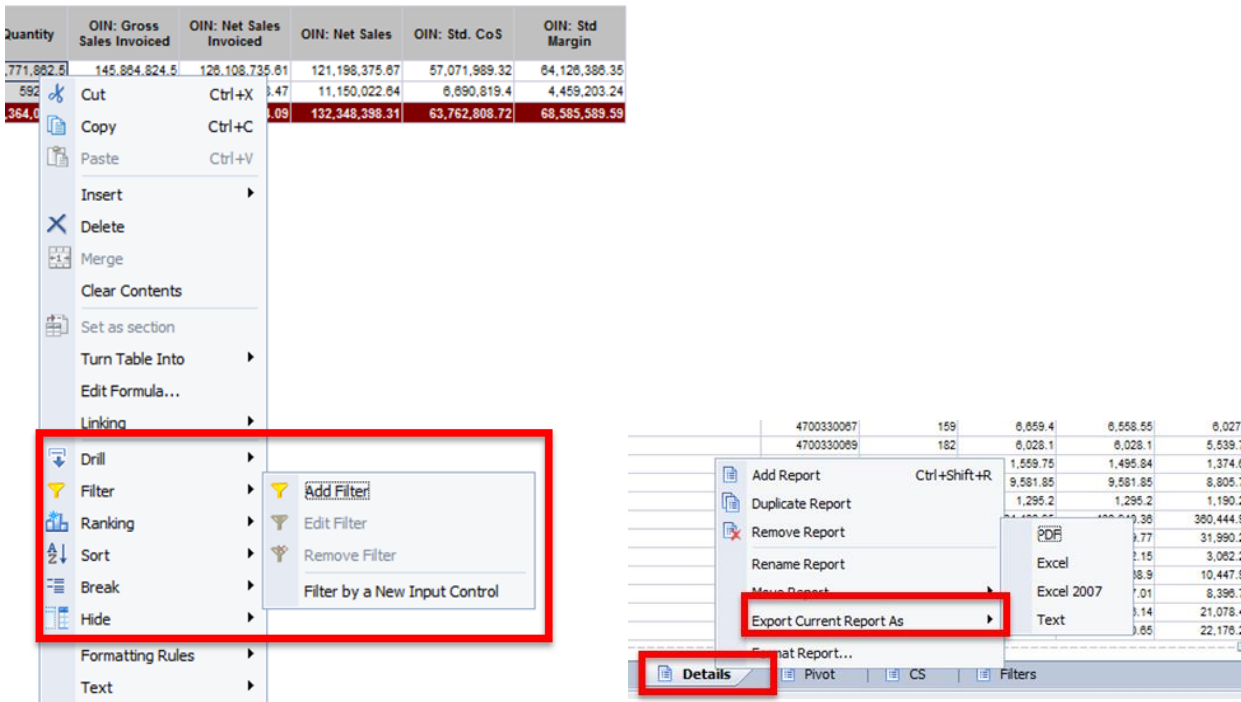


3. Σε περίπτωση που το Sales Report είναι διαθέσιμο στην μνήμη cache, καλείται αμέσως και τα αποτελέσματα αποστέλλονται πίσω στο τερματικό του χρήστη.
4. Αν το Sales Report δεν είναι διαθέσιμο στην μνήμη cache, τα στοιχεία που ζητήθηκαν ανακτώνται από την Βάση Δεδομένων που περιέχει τα metadata
5. Η SQL Engine δημιουργεί τα κατάλληλα SQL ερωτήματα
6. Η Query Engine δέχεται και εκτελεί το SQL ερώτημα
7. Η SQL εκτελείται στο Data Warehouse και υπάρχει ροή αποτελεσμάτων πίσω στον Πυρήνα.
8. Ενδεχομένως να γίνει κλήση του Μηχανισμού Ανάλυσης για την εκτέλεση περαιτέρω υπολογισμών.
9. Εκτελούνται πρόσθετες επαναλήψεις μεταξύ των βημάτων 5 και 8 για όσο αυτό είναι αναγκαίο
10. Τα αποτελέσματα του ερωτήματος τελικά παραδίδονται στον χρήστη.

Η είσοδος των χρηστών στο Business Objects γίνεται μέσω browser και κατά την είσοδο απαιτείται κωδικός και όνομα χρήστη. Υπάρχουν διάφορα δικαιώματα χρήσης και πρόσβασης ανάλογα με τις ανάγκες του οργανισμού και σύμφωνα με το IT τμήμα που διαχειρίζεται το BO. Ο χρήστης στο περιβάλλον εισόδου έχει στην διάθεσή του διάφορα εργαλεία διαχείρισης και διαμόρφωσης όπως επίσης πλοήγησης για το σύνολο των αναφορών. Η κάθε αναφορά μπορεί να αρχειοθετηθεί σε φακέλους ενώ δίνετε να περιέχει πίνακες, γραφήματα και εικόνες, στοιχεία στα οποία μπορεί να γίνει επεξεργασία και τοπικά αποθήκευση. Είναι σημαντικό να αναφέρουμε ότι δικαίωμα μόνιμων αλλαγών και επεξεργασίας αναφορών έχουν μόνο οι διαχειριστές, ωστόσο ο κάθε χρήστης έχει δικαίωμα είτε απλής ανάγνωσης και εκτύπωσης της αναφοράς είτε τοπικών αλλαγών και αποθήκευσης.

Οι δυνατότητες διαχείρισης και διαμόρφωσης των αναφορών αποτελούν ένα ισχυρό εργαλείο στην ευχέρεια του κάθε χρήστη. Το Business Objects δίνει την δυνατότητα επέμβασης του χρήστη στην αναφορά, εισάγοντας εντολές απλές και περίπλοκες συμπεριλαμβανομένων αριθμητικών, λογικών, σχετικές με ημερομηνία, κείμενο και πολλές άλλες. Ένα επίσης σημαντικό χαρακτηριστικό του BO είναι η δυνατότητα διαμοιρασμού (είτε σε μορφή φυλλομετρητή, είτε pdf) και

ανανέωσης των αναφορών ανά τακτά χρονικά διαστήματα όπως αποφασίσει ο διαχειριστής Χάριν ευκολίας επίσης ένας χρήστης, ενός των αναφορών, έχει γρήγορη πρόσβαση στις παρακάτω ενέργειες όπως φαίνονται στο στιγμιότυπο του σχήματος.



**Σχήμα 4.5.α:** Δυνατότητες ανάλυσης δεδομένων (drilling, filtering, break)

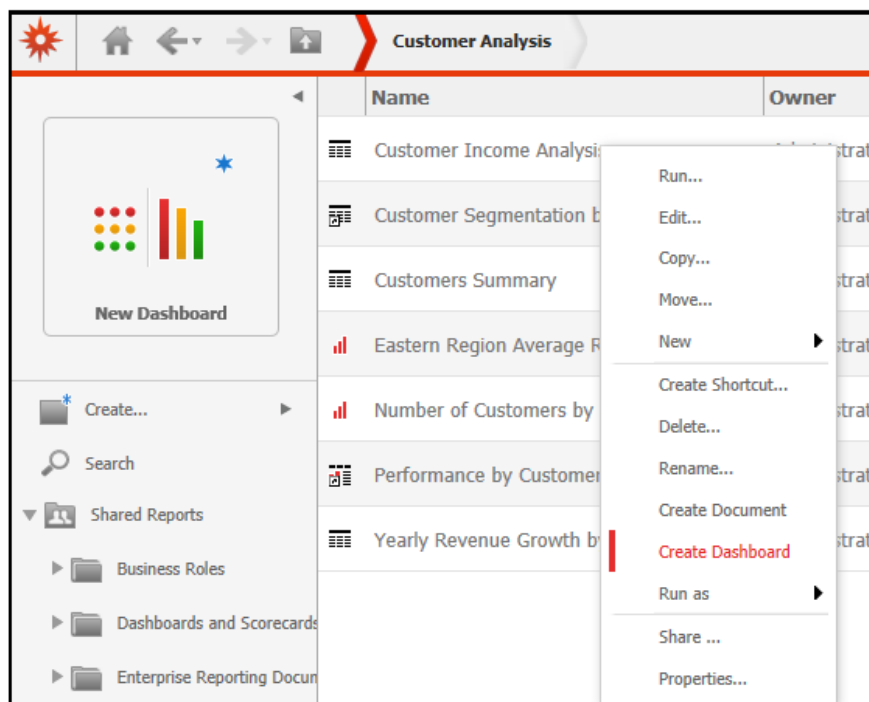
**Σχήμα 4.5.β:** Δυνατότητες εξαγωγής αναφορών σε αρχεία pdf, excel και text

1. Ανάλυση δεδομένων που βρίσκονται σε χαμηλότερο επίπεδο σε έναν ιεραρχικά δομημένο πίνακα (drill down)
2. Εισαγωγή φίλτρου
3. Κατάταξη κατά το επιθυμητό μέγεθος
4. Ταξινόμηση
5. Διαχωρισμός πίνακα με βάση το επιθυμητή διάσταση.
6. Απόκρυψη αντικειμένου από πίνακα
7. Εισαγωγή μαθηματικών συναρτήσεων στα κελιά των μεγεθών.
8. Δυνατότητα εξαγωγής αναφοράς σε μορφή αρχείου φυλλομετρητή είτε κειμένου (text/pdf)

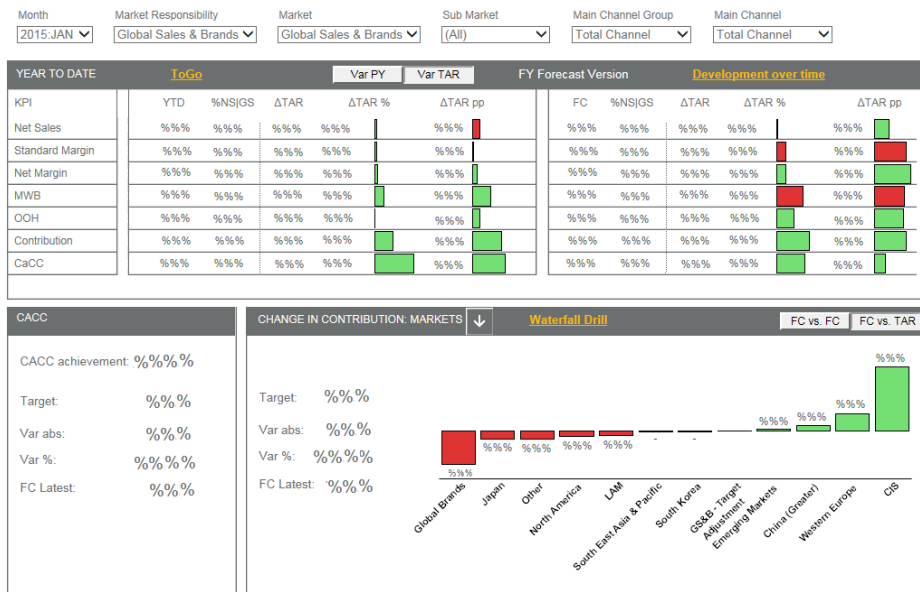
Το Business Objects εργαλείο της SAP όπως καταλήγει το συμπέρασμα, συνυπολογίζοντας όλα τα παραπάνω και μετά από κατάλληλη εκπαίδευση των χρηστών, αποτελεί ένα δυνατό εργαλείο για την Εταιρία καθώς ανεξάρτητα ο κάθε χρήστης του κάθε τμήματος έχει την δυνατότητα πρόσβασης, διαχείρισης και διαμόρφωσης των αναφορών ωστόσο παράλληλα επαφίεται η ασφάλεια δεδομένων και η διαχείριση στο το τμήμα IT το οποίο καλύπτει νέες ανάγκες και διαθέτει απλές και σύνθετες αναφορές όπου υπάρχει η ανάγκη.

### *MicroStrategy*

Ως εργαλείο EN στην εταιρία έχει εγκατασταθεί και η λύση της MicroStrategy σε υψηλό επίπεδο. Πρακτικά αυτό σημαίνει ότι γίνεται μια κεντρική διαχείριση δεδομένων σε έννοιες και μεγέθη πολυεθνικά. Ως αποτέλεσμα, τα διαφορετικά εργαλεία που παρέχονται σε υψηλό επίπεδο παρέχουν μια συγκεντρωτική αξιόπιστη αλλά δύσκαμπτη εικόνα για τα γεγονότα σε τοπικό επίπεδο. Ωστόσο, καθώς οι δομές είναι αναλυτικές, περίπλοκες και αξιόπιστες, μπορούν να αποτελέσουν και εργαλείο ελέγχου και εισόδου δεδομένων στο τοπικό σύστημα EN, ως ενδείκνυται από την κεντρική διαχείριση της Εταιρίας.



**Σχήμα 4.6:** Περιβάλλον δημιουργίας αναφορών στο εργαλείο MicroStrategy



**Σχήμα 4.7:** Παράδειγμα αναφοράς στο εργαλείο MicroStrategy

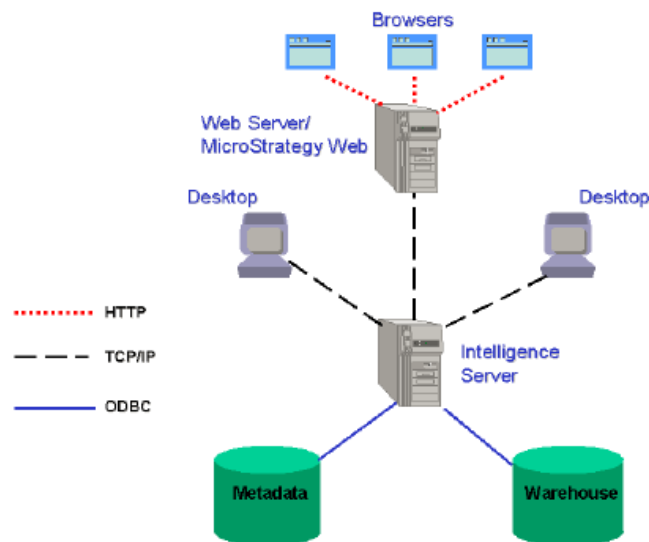
Στο εργαλείο MicroStrategy επίσης υπάρχει διαχωρισμός στους χρήστες ανάλογα με τις ανάγκες και τα δικαιώματα και η είσοδος γίνεται μέσω Web browser. Προς όλους τους χρήστες δίνεται η δυνατότητα παρακολούθησης διαφόρων ειδών αναφορών με πρόσβαση σε εργαλεία διαχείρισης των δεδομένων (όχι ωστόσο διαχείριση και πρόσβαση στην πηγή αυτών) και τοπική αποθήκευση για ατομικούς σκοπούς. Επίσης δίνεται, ανάλογα με το επίπεδο χρήστη, περιορισμένη δυνατότητα σε τοπική δημιουργία αναφορών. Τα αντικείμενα που περιέχονται στο MicroStrategy είναι τα εξής:

- Αντικείμενα σχήματος (Schema Objects): λογικά αντικείμενα που αντιστοιχούν σε στήλες και πίνακες στη Data Warehouse όπως Attributes και Facts.
- Αντικείμενα Εφαρμογών (Application Objects): αντικείμενα που χρησιμοποιούνται για την κατασκευή των αναφορών (οι αναφορές αποτελούν επίσης αντικείμενα εφαρμογών). Αποθηκεύονται στο φάκελο "Public Objects". Παραδείγματα αυτών: Filters, Prompts και Metrics.
- Αντικείμενα Διαμόρφωσης: αντικείμενα που βρίσκονται στο διαχειριστικό επίπεδο του συστήματος (system layer) και ως εκ τούτου μπορούν να χρησιμοποιηθούν σε όλα τα έργα. Παραδείγματα: Users, User Groups, Schedules και Database Logins.

Επίσης διαφορετικά είδη αναφορών είναι διαθέσιμα για τους χρήστες, το κάθε ένα με τον δικό του σκοπό και χρήση. Διαθέσιμες αναφορές είναι:

- Standard Dashboard– Απλές αναφορές προς χρήση από ανώτατα στελέχη και συνεργάτες
- Ad-Hoc Flexible Reporting– Δυνατότητα εύχρηστης διάθεσης (one-click download) αναφορών βάση οποιουδήποτε στοιχείου επιλογής του χρήστη
- Exception Reporting– Υποστήριξη επιχειρησιακών διαδικασιών τονίζοντας λανθασμένα ή λειψά δεδομένα
- Real-time Reporting – Συγχρονισμός δεδομένων γεγονότος σε σύντομο χρονικό διάστημα.

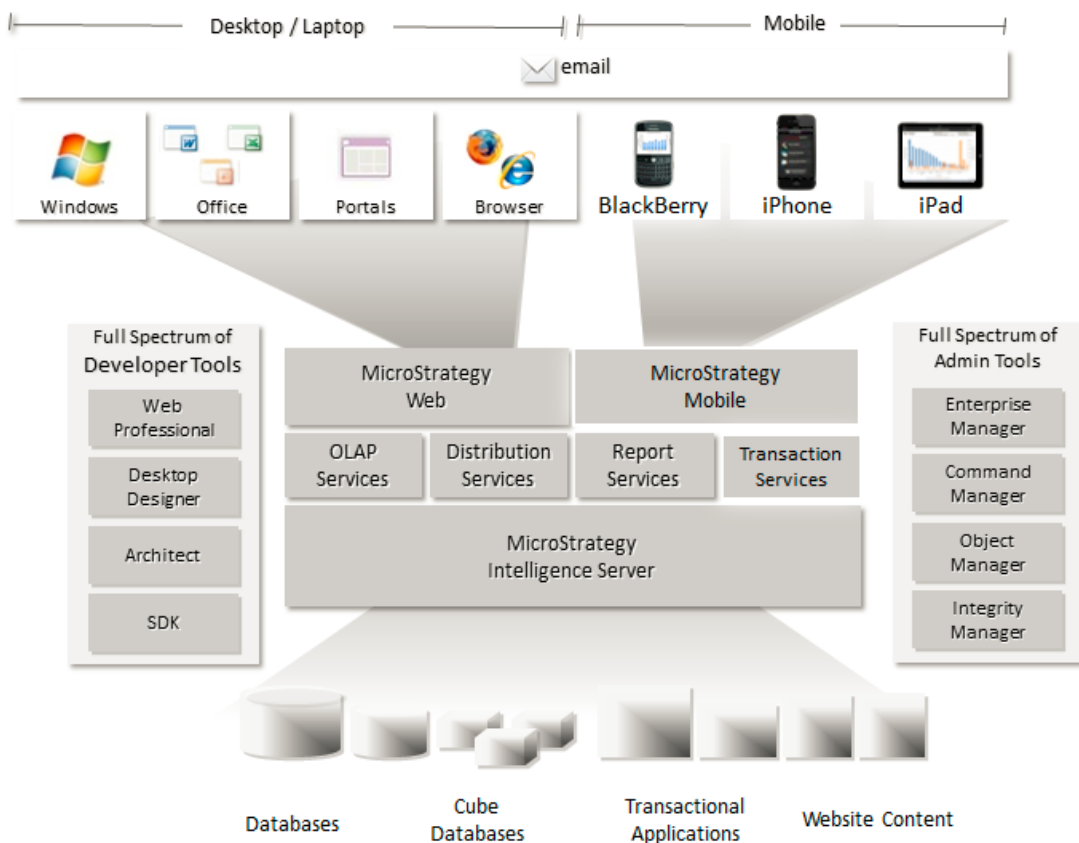
Η αρχιτεκτονική του MicroStrategy είναι μια 4-tier αρχιτεκτονική (Εικόνα 4.8). Ο σχεδιασμός είναι παρόμοιος με τον σχεδιασμό της λύσης της SAP ωστόσο αναφέρουμε πως στην 4-tier αρχιτεκτονική ο Intelligence Server και ο Web Server φιλοξενούνται σε διαφορετικά μηχανήματα για λόγους απόδοσης. Επίσης θα μπορούσε να εφαρμοστεί και μια αρχιτεκτονική 3-tier, ωστόσο δεν θα ήταν δυνατή η πρόσβαση μέσω Web Browser και φορητών συσκευών.



**Σχήμα 4.8:** Απεικόνιση της 4-tier αρχιτεκτονικής του εργαλείου MicroStrategy

Η 4-tier αρχιτεκτονική δίνει την δυνατότητα εύκολης πρόσβασης στους χρήστες (μέσω συσκευών που συνδέονται στο διαδίκτυο) και κυρίως την άμεση και ταχύτατη επεξεργασία δεδομένων και ανανέωση των πληροφοριών των αναφορών. Μια συνολική εικόνα για τα επιμέρους χαρακτηριστικά του συστήματος του MicroStrategy φαίνεται στο παρακάτω γράφημα η οποία περιέχει:

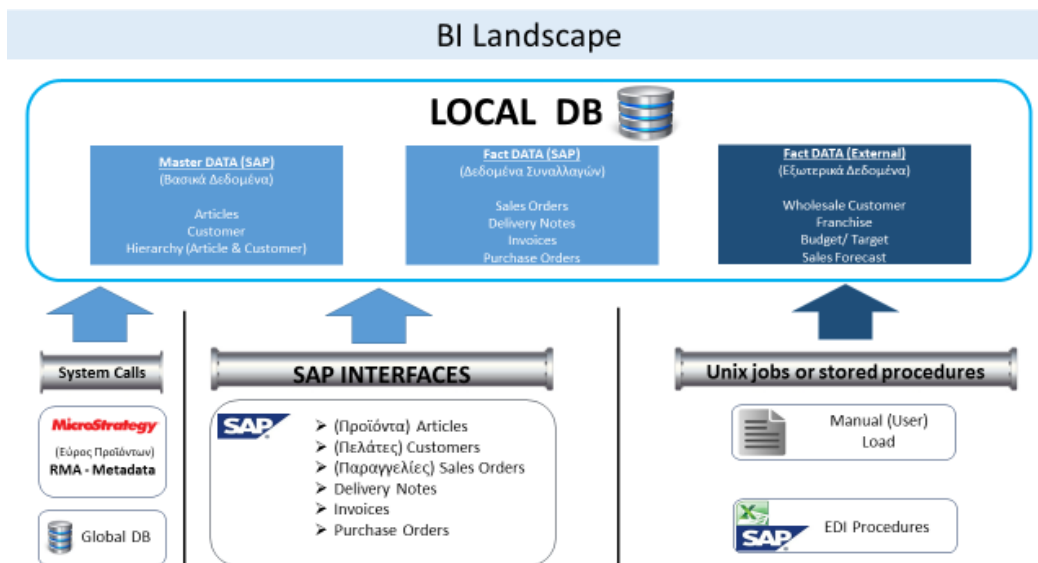
- Τον Intelligence Server
- Τις Υπηρεσίες Διανομής
- Τις Υπηρεσίες Συναλλαγών
- Την δυνατότητα MicroStrategy Web
- Την δυνατότητα MicroStrategy Mobile
- Τα εργαλεία χρηστών/προγραμματιστών
- Τα εργαλεία διαχειριστών



**Σχήμα 4.9:** Τα επιμέρους χαρακτηριστικά του συστήματος του MicroStrategy

## Δ2.2 Σχεδιασμός τοπίου Επιχειρησιακής Νοημοσύνης

Όπως αναφέρθηκε σε προηγούμενο κεφάλαιο, ο σχεδιασμός του τοπίου EN σχεδιάζεται σύμφωνα με τις επιμέρους τεχνολογίες. Μέχρι στιγμής έχουμε αναλύσει τις τεχνολογίες της SAP, SAP BO και MicroStrategy. Σύμφωνα με όλα τα παραπάνω το τοπίο EN για την εταιρία μοιάζει με το γράφημα. Δεδομένα από 3 διαφορετικές πηγές βρίσκονται στην τοπική βάση δεδομένων, από τα SAP interfaces, το MicroStrategy καθώς και εξωτερικά δεδομένα από μη-αυτόματες πηγές.



**Σχήμα 4.10:** Το τοπίο Επιχειρησιακής Νοημοσύνης στο case study

Αξίζει να δοθεί περισσότερο έμφαση στα δεδομένα και ιδιαίτερα στα metadata και τα fact data (δεδομένα συναλλαγών) μέσω SAP αλλά και μη-αυτόματων πηγών με σκοπό την ορθή κατανόηση των σχημάτων που ύστερα αναπτύσσονται.

### Metadata

Πληροφορίες που έρχονται από τους κεντρικούς servers της Εταιρίας και καμία μετατροπή αυτών δεν είναι δυνατή. Από τα Metadata και σύμφωνα με τις τοπικές ανάγκες δημιουργούνται πίνακες κατάτμησης και ιεραρχίας στο Data Warehouse. Για παράδειγμα metadata που αφορούν τα προϊόντα συνδυάζονται με δεδομένα Πελατών και πίνακες ιεραρχίας Πελατών-Προϊόντων δημιουργούνται. Δηλαδή

μπορούμε να κατηγοριοποιήσουμε τα προϊόντα ανάλογα με την δυναμική κάποιου πελάτη.

### *Fact Data (SAP)*

Όλες οι πληροφορίες που αποθηκεύονται μετά από κάποιο γεγονός στην βάση από το Back Office λογισμικό (SAP ERP). Είναι το λογισμικό στο οποίο γίνονται όλες οι παραγγελίες από το Τμήμα Πωλήσεων προς τους Πελάτες. Από εκεί μπορούμε να δούμε πότε, με τι κόστος και σε ποια ποσότητα έχει γίνει η αρχική αγορά εμπορεύματος.

### *Fact Data (External)*

Κάνοντας χρήση εξωτερικών και μη-αυτόματων πηγών, μπορούμε να έχουμε τα δεδομένα που μας απασχολούν. Γίνεται εισαγωγή δεδομένων σε έναν παράλληλο Unix Server ο οποίος λειτουργεί ως ενδιάμεσος server με σκοπό να γίνονται εκεί οι μετατροπές και αρχικοποιήσεις δεδομένων και λιγότεροι πόροι να χρησιμοποιούνται από τον Intelligence Server.

### *Επίπεδα Αποθήκης δεδομένων*

Όπως είδαμε στο Κεφάλαιο Γ, σε μια Αποθήκη Δεδομένων ακρογωνιαίος λίθος αποτελεί η διαδικασία ETL η οποία παρέχει τα απαραίτητα ολοκληρωμένα και ομοιογενή δεδομένα από τις διάφορες πηγές. Υπάρχουν τρία επίπεδα στην Αποθήκη Δεδομένων για αυτόν τον σκοπό (external level, staging level, data warehouse level).

### *External Level<sup>24</sup>*

Το external level αποτελεί ένα μέσω ενσωμάτωσης δεδομένων από μη-αυτόματες πηγές με την βάση δεδομένων. Μέσω του SQL Loader (διαδικασία που καλείται από τα stored procedures του Unix Server) η βάση συνδέεται με αρχεία τύπου csv τα οποία έχουν μορφοποίηση που υποστηρίζεται από την Oracle για την δημιουργία εξωτερικού πίνακα.

---

<sup>24</sup> Πηγή: [https://docs.oracle.com/cd/B19306\\_01/server.102/b14215/et\\_concepts.htm](https://docs.oracle.com/cd/B19306_01/server.102/b14215/et_concepts.htm)  
Λάμπρος Ι. Βότσης



Ο SQL Loader<sup>25</sup> είναι μια υπηρεσία φόρτωσης (bulk loader utility) που χρησιμοποιείται για τη μετακίνηση δεδομένων από εξωτερικά αρχεία στη βάση δεδομένων της Oracle. Η σύνταξη του υποστηρίζει διάφορες μορφές φόρτωσης όπως επιλεκτική φόρτωση, και φόρτωση πολύ-επίπεδων πινάκων. Με την κλήση της εντολής sqlldr στον Unix Server θα εκτελεστεί το control file που έχουμε θέσει ως όρισμα

```
sqlldr username/password@server control=sales_loader.ctl
```

Στο παραπάνω παράδειγμα το sales\_loader.ctl είναι ένα αρχείο το οποίο όταν καλείται εκτελεί τις παρακάτω Unix εντολές.

```
load data
infile 'c:\data\mysalesdata.csv'
into table sales
fields terminated by "," optionally enclosed by '"'
(customer, period, sold_qty)
```

Το αρχείο mysalesdata.csv είναι το αρχείο που περιέχει τα δεδομένα από τις μη αυτόματες πηγές. Η δομή του αρχείου έχει τρία πεδία, Πελάτης, Περίοδος Πώλησης, Ποσότητα Τεμαχίων Πώλησης. Η μορφή του είναι η εξής:

```
"Top Sports","JAN-16", 1000
"Top Sports","FEB-16", 500
```

Για την σύνδεση του εξωτερικού αρχείου, θα πρέπει να είναι βέβαιο πως η δομή ακολουθεί τις οδηγίες όπως αυτές ορίστηκαν κατά την δημιουργία του εξωτερικού πίνακα στην Oracle. Παρακάτω φαίνεται η διαδικασία δημιουργίας εξωτερικού πίνακα στην Oracle.

```
SQL> CREATE TABLE ext_sellout
2 (customer_name CHAR(20),
3 period CHAR(20),
4 sold_qty CHAR(5))
5 ORGANIZATION EXTERNAL
6 (TYPE ORACLE_LOADER
```

<sup>25</sup> [http://www.orafaq.com/wiki/SQL\\*Loader\\_FAQ](http://www.orafaq.com/wiki/SQL*Loader_FAQ)

```
7      DEFAULT DIRECTORY def_dir1
8      ACCESS PARAMETERS
9          (RECORDS DELIMITED BY NEWLINE
10         FIELDS (customer_name          CHAR(20) ,
11                period                  CHAR(20) ,
12                sold_qty                 CHAR(5) )
13         )
14      LOCATION ('mysalesdata.csv')
15  );
```

Αξίζει να σημειωθεί ότι σε εξωτερικούς πίνακες, η Oracle έχει φροντίσει ώστε να μην μπορούν να εκτελεστούν όλες οι SQL εντολές. Όπως γίνεται κατανοητό δεν μπορεί να εκτελεστούν οι truncate και delete καθώς τα δεδομένα είναι σε αρχείο και όχι στην βάση. Απλά στην βάση επί της ουσίας μπορούμε να δούμε και να διαχειριστούμε μια προβολή αυτών των δεδομένων. Τέλος, τα αριθμητικά δεδομένα ορίζονται ως CHAR(n) καθώς σε αυτό το επίπεδο δεν είναι απαραίτητος ο έλεγχος των δεδομένων. Το επίπεδο στο οποίο μας απασχολούν τα αριθμητικά ορίσματα είναι το staging level όπου και πραγματοποιείται το data transformation.

### *Staging Level*

Σε αυτό το επίπεδο πραγματοποιείται μέσω πακέτων PL/SQL η ETL διαδικασία. Σε αυτό το επίπεδο βρίσκονται πίνακες αρχικοποίησης και άλλοι διάφοροι πίνακες που χρησιμεύουν στις διαδικασίες διαμόρφωσης δεδομένων. Τα δεδομένα μεταφέρονται σε αυτό το επίπεδο σε πίνακες οι οποίοι διαγράφονται και να διαμορφώνονται κατά το δοκούν ώστε να υπάρχει μια σωστή και αξιόπιστη μορφή των δεδομένων. Σε αυτό το επίπεδο έχουμε εισαγωγή από τους εξωτερικούς πίνακες και έξοδο προς τους παραγωγικούς πίνακες στο Data Warehouse.

### *Data Warehouse Level*

Στο τέλος των διαδικασιών τροποποίησης τα δεδομένα αποθηκεύονται στην επίπεδο Data Warehouse. Είναι το επίπεδο στο οποίο εκτελούνται τα SQL ερωτήματα από το SAP BO για την δημιουργία των Data Cubes και τελική προβολή σε πίνακες και γραφήματα. Στο Data Warehouse βρίσκονται δεδομένα από διάφορες πηγές όπως τα εξωτερικά δεδομένα (external fact data),

συνδυαστικοί πίνακες κατάτμησης και ιεραρχίας (segmentation & hierarchy data) καθώς και τα fact data από το SAP και το MicroStrategy.

### **Δ3. Εφαρμογή επιλεχθέντος λύσης**

Στις προηγούμενες ενότητες αναλύθηκε η διαδικασία μέσω της οποίας δημιουργούνται τα Sales Reports όταν οι πωλήσεις αφορούν δεδομένα από την ίδια Εταιρία. Για την δημιουργία ωστόσο των αναφορών πωλήσεων των πελατών χονδρικής η διαδικασία γίνεται αρκετά περίπλοκη καθώς δεν έχουμε πρόσβαση στο Front Office λογισμικό και στα δεδομένα από τους Πελάτες με αυτόματο τρόπο.

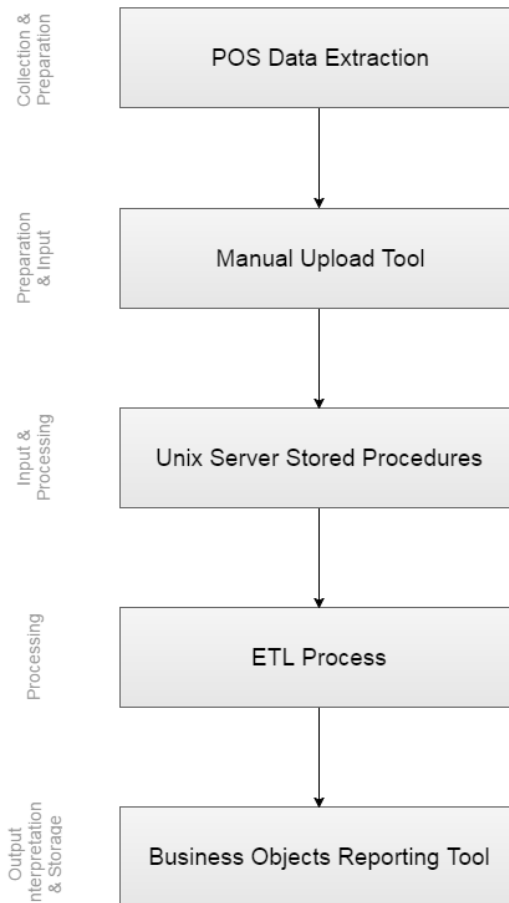
Το πρόβλημα ξεκινάει από την βάση, κατά πόσο μπορούμε να βρούμε τα δεδομένα. Επίσης απαντήσεις θα πρέπει να δοθούν σε ερωτήματα σχετικά με την ασφάλεια, την ανθεκτικότητα σε σφάλμα, το μέγεθος, την δομή, τις ανάγκες σε πόρους καθώς και τον τρόπο μεταφοράς των δεδομένων.

Ο τρόπος με τον οποίο έχει επιλεχθεί να λυθεί το πρόβλημα πρόσβασης των δεδομένων αφορά συνεργασία μεταξύ των τμημάτων IT Εταιρίας-Πελάτη και κατά αυτόν τον τρόπο ο κύκλος επεξεργασίας δεδομένων αποκτά τον πιο κρίσιμο ρόλο του έργου. Μέσω διαφόρων φίλτρων και ελέγχων λύνεται το πρόβλημα της αξιοπιστίας των δεδομένων. Όταν τα δεδομένα είναι αναξιόπιστα, αυτομάτως όσο καλά δομημένα είναι η αρχιτεκτονική, τα αποτελέσματα θα είναι εσφαλμένα. Όπως αναφέρθηκε και στην εισαγωγή του κεφαλαίου, μια θεωρία φαινομενικά είναι άψογη ωστόσο στην επαφή με την πραγματικότητα η πρακτική λύση ενδέχεται να διαφέρει.

#### *Δ3.1 Κύκλος Επεξεργασίας Δεδομένων*

Ο κύκλος επεξεργασίας δεδομένων πραγματοποιείται από τρία διαφορετικά μέρη, τους Πελάτες, του Πωλητές και το Τμήμα IT. Τα δεδομένα λιανικής συλλέγονται από τους Πελάτες, τα δεδομένα χονδρικής συλλέγονται από τους Πωλητές και το τμήμα IT είναι υπεύθυνο μέσα σε αυτόν τον κύκλο επεξεργασίας

να καθαρίσει<sup>26</sup> τα δεδομένα, να εισάγει τους επιχειρηματικούς κανόνες, να συσσωματώσει και να επιβεβαιώσει την ορθότητα των δεδομένων. Η διαδικασία φαίνεται στο παρακάτω γράφημα και τα στάδιά της αναφέρονται παρακάτω και αναλύονται στις ενότητες που ακολουθούν.



**Σχήμα 4.11:** Κύκλος Επεξεργασίας Δεδομένων

#### *Εξαγωγή Δεδομένων από το τερματικό πωλήσεων (POS Data)*

Το πρώτο στάδιο είναι μια συλλογή δεδομένων από την πλευρά του Πελάτη. Τα δεδομένα (raw data) εξάγονται από το από το Back Office σύστημά σύμφωνα με προσυμφωνημένη δομή (data requirements) και αποστέλλονται στο Τμήμα

<sup>26</sup> Ο όρος καθαρισμός δεδομένων (Data Cleansing) παρότι δεν έχει οριστεί ακαδημαϊκά αποτελεί χρησιμοποιείται κατά κόρον για την περιγραφή όλων των απαραίτητων ενεργειών ανίχνευσης, διόρθωσης, αφαίρεσης ή αντικατάστασης διεφθαρμένων και ανακριβή δεδομένων [58]

Πωλήσεων σε μορφή csv αρχείων. Τα δεδομένα που υπάρχουν μετά από αυτό το πρώτο στάδιο είναι οι πωλήσεις λιανικής ανά ημέρα/ ανά Πελάτη.

### *Excel Upload Tool*

Κατά το δεύτερο στάδιο πραγματοποιείται ένας συνδυασμός συλλογής (master data και wholesale fact data), προετοιμασίας και εισαγωγής δεδομένων. Για την πραγματοποίηση αυτών των σταδίων έχει δημιουργηθεί από το Τμήμα IT της Εταιρίας ένα ειδικά διαμορφωμένο Data Upload Tool σε Windows Excel με βοήθεια αυτοματοποιημένων macros σε γλώσσα VBA. Οι Πωλητές εισάγουν τα δεδομένα csv από τον Πελάτη μαζί με τα master data των Πελατών στο Excel Upload Tool όπου λαμβάνουν μέρος μερικοί λογικοί έλεγχοι στην δομή των csv αρχείων. Έπειτα το αρχείο csv που δημιουργείται μέσω του Excel Upload Tool μεταφέρετε στον Unix Server ώστε να μπορέσει να γίνει κλήση του SQL Loader όπως αναφέρθηκε παραπάνω. Η πρόσβαση στον Unix Server είναι αποκλειστικά τοπική και μόνο το τμήμα IT έχει την δυνατότητα πρόσβασης. Σε αυτό το σημείο σταματάει η μη-αυτόματη διαδικασία.

### *UNIX Server Stored Procedures*

Στον Unix Server όπου και αποθηκεύονται τα δεδομένα, εκτελούνται διαδικασίες μεταφοράς και αρχικοποίησης για την δημιουργία του συγκεντρωτικού αρχείου (από διαφορετικά αρχεία των Πωλητών). Επίσης καλείται και ο SQL Loader για την σύνδεση του αρχείου με το εξωτερικό επίπεδο της βάσης δεδομένων, εκεί δηλαδή που θα εκτελεστεί η ETL διαδικασία. Επίσης αποθηκεύονται τα αρχεία σε χώρο δεσμευμένο για δημιουργία back up. Οι stored procedures εκτελούνται ανά ορισμένα τακτά χρονικά διαστήματα (δύο φορές την ημέρα) για την οικονομία πόρων και καθώς οι ανάγκη δεν αφορά σύγχρονη ανάλυση (real time analysis)

### *ETL Process*

Στην βάση δεδομένων πραγματοποιείται η ETL διαδικασία. Ξεκινώντας από το εξωτερικό επίπεδο και προχωρώντας στο επίπεδο staging, τα δεδομένα διαμορφώνονται με τέτοιο τρόπο ώστε στο Data Warehouse επίπεδο, και σε συνδυασμό με τους πίνακες αρχικοποίησης των πελατών και των προϊόντων, να

ευνοείται η δημιουργία Data Cubes από τον Intelligence Server για τους σκοπούς των BO Reports.

### *BO Reports*

Στο τελευταίο αυτό στάδιο ο Intelligence Server έχει δημιουργήσει μέσω SQL queries, Data Cubes οι οποίοι στο εργαλείο Business Objects μεταφράζονται σε μορφή πινάκων και γραφημάτων στις αναφορές. Στην συνέχεια γίνεται ο προγραμματισμός του διαμοιρασμού και ανανέωσης των αναφορών σε τακτά διαστήματα συμβατά με την ανανέωση της Αποθήκης Δεδομένων. Στην συνέχεια του Κεφαλαίου αναλύονται τα βήματα

#### *Δ3.2 Εισαγωγή Δεδομένων*

##### *Αρχική αποστολή δεδομένων (POS data)*

Το πρώτο στάδιο είναι μια συλλογή δεδομένων από την πλευρά του Πελάτη. Τα δεδομένα (raw data) εξάγονται από το από το Back Office σύστημά σύμφωνα με προσυμφωνημένη δομή (data requirements) και αποστέλλονται στο Τμήμα Πωλήσεων σε μορφή csv αρχείων. Τα δεδομένα που υπάρχουν μετά από αυτό το πρώτο στάδιο είναι οι πωλήσεις λιανικής ανά ημέρα/ ανά Πελάτη. Σκοπός των Πωλητών είναι η συμπλήρωση του παρακάτω αρχείου Data Upload Tool.

##### *Data Upload Tool*

Κατά το δεύτερο στάδιο πραγματοποιείται ένας συνδυασμός συλλογής (master data και wholesale fact data), προετοιμασίας και εισαγωγής δεδομένων. Για την πραγματοποίηση αυτών των σταδίων έχει δημιουργηθεί από το Τμήμα IT της Εταιρίας ένα ειδικά διαμορφωμένο Data Upload Tool σε Windows Excel με βοήθεια αυτοματοποιημένων macros σε γλώσσα VBA.

Οι Πωλητές εισάγουν δεδομένα από τα csv αρχεία του Πελάτη μαζί με τα master data των Πελατών<sup>27</sup> και τα fact data των Πελατών<sup>28</sup> στο Excel Upload Tool. Με το

---

<sup>27</sup> Υπενθυμίζεται πως τα Master Data των Πελατών είναι τα χαρακτηριστικά που αφορούν τον Πελάτη (κλειδιά ιεραρχίας, δημογραφικά χαρακτηριστικά κ.α.)

<sup>28</sup> Υπενθυμίζεται πως τα Fact Data των Πελατών είναι ουσιαστικά τα δεδομένα παραγγελιών. Οι ποσότητες δηλαδή που έχει προμηθευτεί ο κάθε πελάτης. Στόχος είναι ο λογικός έλεγχος, δηλαδή οι ποσότητες που πωλήθηκαν να μην είναι μεγαλύτερες από τις ποσότητες που προμηθευτήκαν. Έτσι αντιλαμβανόμαστε λογικά σφάλματα από την πλευρά του Πελάτη.

πάτημα ενός κουμπιού τα macros αναλαμβάνουν μερικούς λογικούς ελέγχους στα δεδομένα και στην δομή του αρχείου. Έπειτα δημιουργείτε ένα άλλο αρχείο csv το οποίο μεταφέρετε στον Unix Server. ώστε να μπορέσει να συνεχιστεί η διαδικασία (πλέον αυτόματα) με την κλήση του SQL Loader.

Το Data Upload Tool έχει 3 φύλλα UserForm, List, Correspondence List. τα δύο εξ αυτών (List και Correspondence List) είναι κρυφά για τους χρήστες, ενώ το κύριο φύλλο (UserForm) βρίσκεται προς διάθεση των χρηστών. Ωστόσο υπάρχει κλείδωμα ασφαλείας σε όλα τα κελιά εκτός αυτά που επιτρέπεται η εισαγωγή δεδομένων. Τα κύρια σημεία ενδιαφέροντος και η μορφή του κύριου φύλλου, UserForm φαίνεται στο παρακάτω στιγμιότυπο.



Customer	Period	Season	Article	Delivered QTY	Sold QTY	Comment	Customer Description	GUID	User	Error Description
----------	--------	--------	---------	---------------	----------	---------	----------------------	------	------	-------------------

**Σχήμα 4.12:** Πεδία του εργαλείου Data Upload Tool

1. Το μέρος του αρχείου το οποίο περιέχει τα δεδομένα. Οι πρώτες έξι στήλες (A-F) είναι προς διάθεση των χρηστών για εισαγωγή δεδομένων ενώ οι υπόλοιπες στήλες συμπληρώνονται αυτόματα κατά την κλήση της διαδικασίας ελέγχου εγκυρότητας.
2. Τα δύο μοναδικά αναγνωριστικά. Το ένα για το όνομα του χρήστη (User) και το δεύτερο για την κάθε χρήση του αρχείου (GUID). Και τα δύο συμπληρώνονται αυτόματα με εντολές VBA.
3. Κουμπιά αλλαγής Upload mode είτε σε μηνιαία φόρτωση είτε σε εβδομαδιαία.
4. Κουμπί που εκκινεί την διαδικασία ελέγχου εγκυρότητας μέσω εντολών VBA.
5. Κουμπί παραγωγής και αποστολής του csv αρχείου στον Unix Server.
- 6.

Στο φύλλο List βρίσκονται όλες οι λίστες με το εύρος τιμών για κάθε ένα από τα δεδομένων το οποίο δεν αποτελεί τίποτα άλλο παρά ένα στιγμιότυπο από τα Master Data που περιέχονται στην Βάση Δεδομένων. Σκοπός του φύλλου List είναι ο τοπικός έλεγχος των δεδομένων. Για παράδειγμα αν ο κωδικός του

προϊόντος ανταποκρίνεται στους κωδικούς της Εταιρίας και δεν έχει γίνει λάθος από τον Πελάτη με προϊόν άλλης εταιρίας.

H	I	J
Week		Calendar
1		1/1/2014
2		2/1/2014
3		3/1/2014
4		4/1/2014
5		5/1/2014
6		6/1/2014
7		7/1/2014
8		8/1/2014
9		9/1/2014
10		10/1/2014
11		11/1/2014
12		12/1/2014

**Σχήμα 4.13:** Φύλλο List του εργαλείου Data Upload Tool με την αντιστοιχία ημερομηνιών

Τέλος, στο φύλλο Correspondence List βρίσκονται οι αντιστοιχίες μήνα και εβδομάδος και κάθε έτος. Η συγκεκριμένη λίστα αποτελεί σημαντικό δεδομένο καθώς υπάρχει μικρή διαφοροποίηση στον τρόπο που καταγράφονται οι εβδομαδιαίες πωλήσεις εντός της Εταιρίας, χαρακτηριστικό το οποίο δεν μπορεί να γνωρίζει ο Πελάτης.



G41 :

Below our conventional week-month correspondence. Whenever weekly data is uploaded, the column "period" of the upload tool must be filled according to the following correspondence.

WEEK  WEEK

PERIOD  PERIOD

2015			2014		
WEEK	PERIOD	DATE (week end)	WEEK	PERIOD	DATE (week end)
1	2015-Jan	4/1/2015	1	2014-Jan	5/1/2014
2	2015-Jan	11/1/2015	2	2014-Jan	12/1/2014
3	2015-Jan	18/1/2015	3	2014-Jan	19/1/2014
4	2015-Jan	25/1/2015	4	2014-Jan	26/1/2014
5	2015-Jan	1/2/2015	5	2014-Jan	2/2/2014
6	2015-Feb	8/2/2015	6	2014-Feb	9/2/2014
7	2015-Feb	15/2/2015	7	2014-Feb	16/2/2014
8	2015-Feb	22/2/2015	8	2014-Feb	23/2/2014
9	2015-Feb	1/3/2015	9	2014-Feb	2/3/2014
10	2015-Mar	8/3/2015	10	2014-Mar	9/3/2014
11	2015-Mar	15/3/2015	11	2014-Mar	16/3/2014
12	2015-Mar	22/3/2015	12	2014-Mar	23/3/2014
13	2015-Mar	29/3/2015	13	2014-Mar	30/3/2014
14	2015-Apr	5/4/2015	14	2014-Apr	6/4/2014
15	2015-Apr	12/4/2015	15	2014-Apr	13/4/2014
16	2015-Apr	19/4/2015	16	2014-Apr	20/4/2014
17	2015-Apr	26/4/2015	17	2014-Apr	27/4/2014
18	2015-May	3/5/2015	18	2014-May	4/5/2014
19	2015-May	10/5/2015	19	2014-May	11/5/2014
20	2015-May	17/5/2015	20	2014-May	18/5/2014
21	2015-May	24/5/2015	21	2014-May	25/5/2014
22	2015-May	31/5/2015	22	2014-May	1/6/2014
23	2015-Jun	7/6/2015	23	2014-Jun	8/6/2014
24	2015-Jun	14/6/2015	24	2014-Jun	15/6/2014
25	2015-Jun	21/6/2015	25	2014-Jun	22/6/2014
26	2015-Jun	28/6/2015	26	2014-Jun	29/6/2014
27	2015-Jul	5/7/2015	27	2014-Jul	6/7/2014
28	2015-Jul	12/7/2015	28	2014-Jul	13/7/2014
29	2015-Jul	19/7/2015	29	2014-Jul	20/7/2014
30	2015-Jul	26/7/2015	30	2014-Jul	27/7/2014

UserForm List Correspondence List

**Σχήμα 4.14:** Φύλλο Correspondence List με τις αντιστοιχίες ημερομηνιών (εβδομάδα – περίοδο – εβδομάδα)

Το εργαλείο αυτό διαθέτει δύο διαφορετικά Upload Modes για δύο διαφορετικές ανάγκες. Τα δεδομένα για μερικούς πελάτες αφορούν πωλήσεις ένα περίοδο (μήνα) ωστόσο για μερικούς σημαντικούς Πελάτες οι οποίοι έχουν μεγάλο όγκο πωλήσεων και τζίρου μας ενδιαφέρουν τα δεδομένα ανά εβδομάδα. Ο χρήστης (Πωλητής) έχει την δυνατότητα επιλογής εισαγωγής δεδομένων για Monthly upload ή Weekly upload. Ως Week ορίζουμε την εβδομάδα του χρόνου την οποία αφορά η πώληση, άρα έχει τιμές από 1 έως 52, όσες δηλαδή και οι εβδομάδες ενός έτους. Κατά την εκτέλεση σε Monthly Upload Mode δημιουργείτε μια μηδενική στήλη Week κατά την δημιουργία του csv αρχείου. Τα δύο modes του αρχείου φαίνονται ως εξής

The image shows two screenshots of a data entry form. The top screenshot shows a form with columns: Customer, Period, Season, Article, Delivered QTY, Sold QTY, Comment, Customer Description, GUID, User, and Error Description. It includes 'Validate' and 'Switch to Weekly Mode' buttons. The bottom screenshot is similar but includes a 'Week' column.

**Σχήμα 4.15.α:** εικόνα εργαλείου για εισαγωγή δεδομένων ανά μήνα

**Σχήμα 4.15.β:** εικόνα εργαλείου για εισαγωγή δεδομένων ανά εβδομάδα

Τα δεδομένα που περιέχονται σε αυτό το κεντρικό φύλλο, UserForm, είναι στην μορφή που θα διαβαστούν από την βάση δεδομένων και θα μεταφερθεί στην Αποθήκη Δεδομένων. Η δομή είναι η εξής

- Customer: Το ID του Πελάτη. Αυτό το δεδομένο είναι μοναδικό και διαθέσιμο από την καρτέλα του Πελάτη στο SAP. Το Customer\_ID είναι ένας δεκαψήφιος αριθμητικός κωδικός, σε περίπτωση λανθασμένου κωδικού επιστρέφεται μήνυμα σφάλματος.
- Period: Η χρονική περίοδος για την οποία εισάγουμε δεδομένα. Η μορφή είναι YYYY-MMM, για παράδειγμα 2016-JUL για τον Ιούλιο του 2016. Οι χρήστες έχουν την δυνατότητα επιλογής περιόδου από λίστα. Άλλη μορφή δεδομένων στο κελί επιστρέφει μήνυμα σφάλματος.
- Week (προαιρετικό πεδίο): Ο αύξον αριθμός της εβδομάδας κατά το τρέχον έτος για την οποία εισάγουμε δεδομένα.
- Season: Εδώ εισάγεται χαρακτηριστική τιμή για την Season που ανήκει το προϊόν. Το πεδίο συμπληρώνεται με τον κωδικό της season (δύο συγκεκριμένες τιμές SS ή FW), και τα τελευταία δύο ψηφία του έτους. Για παράδειγμα FW16 είναι μια έγκυρη τιμή. Οποιαδήποτε άλλη μορφή δεδομένων επιστρέφει μήνυμα σφάλματος.
- Article: Ο μοναδιαίος κωδικός του προϊόντος. Το Article\_ID είναι ένα κατά κόρον εξαψήφιο αλφαριθμητικό δεδομένο. Υπάρχουν ωστόσο μερικές περιπτώσεις όπου το Article\_ID έχει πέντε ή και τέσσερα ψηφία. Σημαντικό είναι να αναφέρουμε πως το Article\_ID αποτελεί τον κωδικό που περιέχεται σαν πληροφορία στο barcode του προϊόντος κατά την πώληση στον καταναλωτή.

- **Delivered QTY**: Η ποσότητα προμήθειας του Πελάτη από τους Πωλητές. Αυτό το δεδομένο βρίσκεται στις αναφορές πωλήσεων στο ΒΟ καθώς και στις παραγγελίες του Πελάτη στο SAP. Ιδιαιτερότητα του Delivered\_QTY καθώς και του Sold\_QTY είναι πως τα δεδομένα αυτά αφορούν το αθροιστικό σύνολο από την αρχή της εκάστοτε season.
- **Sold QTY**: Η ποσότητα πώλησης προς Καταναλωτές από τους Πελάτες για την περίοδο που ορίζεται από το Season (ή το Week) πεδίο. Αυτή η πληροφορία έρχεται από τους Πελάτες και αφορά αθροιστικό σύνολο από την αρχή της εκάστοτε season.
- **Comment**: Στα σχόλια εμφανίζονται μηνύματα σχετικά με την διαδικασία ελέγχου εγκυρότητας αλλά όχι μηνύματα σφαλμάτων.
- **Description**: Ως description έχουμε το πλήρες όνομα του Πελάτη και συμπληρώνεται αυτόματα με την διαδικασία ελέγχου εγκυρότητας. Η αντιπαράθεση γίνεται σύμφωνα με το φύλλο List. Ο σκοπός αυτού του πεδίου είναι ο οπτικός έλεγχος από τους χρήστες σχετικά με τους Πελάτες για τους οποίους θέλουν να εισάγουν δεδομένα.
- **GUID**<sup>29</sup>: Το GUID (Globally Unique Identifier) αποτελεί αναγνωριστικό διαχειριστή που αποτελείται από μία ομάδα 8 ψηφίων, ακολουθούμενη από τρεις ομάδες 4 ψηφίων και καταλήγει σε μια ομάδα 12 ψηφίων στο δεκαεξαδικό σύστημα (π.χ. 6B29FC40-CA47-1067-B31D-00DD010662DA). Το αναγνωριστικό παράγεται με εντολές VBA κατά την διαδικασία ελέγχου εγκυρότητας όπως φαίνεται στο παρακάτω στιγμιότυπο κώδικα.

```
Dim varGuid As String
varGuid = Mid$(CreateObject("Scriptlet.TypeLib").GUID, 2, 36)
Range("G2").Select
Selection.ClearContents
Range("G2").Value = varGuid
```

<sup>29</sup> Περισσότερες πληροφορίες για το GUID στον παρακάτω σύνδεσμο:

[https://msdn.microsoft.com/en-us/library/aa373931\(VS.85\).aspx](https://msdn.microsoft.com/en-us/library/aa373931(VS.85).aspx)

- User<sup>30</sup>: Το πεδίο αυτό συμπληρώνεται αυτόματα με εντολή VBA κατά την διάρκεια του ελέγχου εγκυρότητας (επιστρέφεται το όνομα του χρήστη) όπως φαίνεται στο στιγμιότυπο. Το πεδίο User σε συνδυασμό με το GUID μας παρέχει ευκολία διαχείρισης σφαλμάτων επαφιομένων στα αρχικά δεδομένα.

```
Range("G1").Value = Application.UserName
```

- Error Description: Σε αυτό το πεδίο εμφανίζονται μηνύματα σφαλμάτων ανάλογα με το δεδομένο που αφορούν. Τα σφάλματα είναι όλα κριτικά σφάλματα

### *Διαδικασία Εγκυρότητας Δεδομένων (Validation Process)*

Η Διαδικασία Εγκυρότητας Δεδομένων εκκινεί με την χρήση του κουμπιού Validate και από διάφορα VBA Modules. Μετά το πέρας του ελέγχου, και σε περίπτωση κατά την οποία δεν έχουν εμφανιστεί κρίσιμα σφάλματα, ο χρήστης δύναται με το κουμπί Save Data να μεταφέρει τα δεδομένα στον Unix Server. Σε περίπτωση κρίσιμων σφαλμάτων εμφανίζονται μηνύματα στο πεδίο Error Description τα οποία καθοδηγούν κατάλληλα τον χρήστη για την επίλυση τους. Η επιλογή φόρτωσης δεν είναι διαθέσιμη έως ότου λυθούν όλα τα ζητήματα. Τα εξής μηνύματα εμφανίζονται στους χρήστες:

- Για περιπτώσεις κατά τις οποίες τα δεδομένα δεν αντιστοιχίζονται στο φύλλο List ή έχει γίνει εσφαλμένη εισαγωγή δεδομένων, εμφανίζονται τα παρακάτω μηνύματα σφάλματος.

```
CUSTOMER NOT LISTED  
PERIOD INVALID  
OUT OF SEASON  
WEEK INVALID  
ARTICLE ID NOT LISTED
```

- Για περιπτώσεις κατά τις οποίες δεν έχει συμπληρωθεί το αντίστοιχο πεδίο τα παρακάτω μηνύματα εμφανίζονται:

---

<sup>30</sup> Περισσότερες πληροφορίες για το όνομα χρήστη στα MS Office στον παρακάτω σύνδεσμο:

<https://msdn.microsoft.com/en-us/library/office/ff822584.aspx>

CUSTOMER EMPTY  
PERIOD EMPTY  
SEASON NOT FOUND  
WEEK NOT FOUND  
ARTICLE ID EMPTY  
NO DELIVER QTY

- Τέλος, δύο διαφορετικά σφάλματα εμφανίζονται σχετικά με διπλές εγγραφές καθώς επίσης και το λογικό σφάλμα κατά το οποίο ένας πελάτης φαίνεται να έχει πωλήσει περισσότερα τεμάχια από όσα προμηθεύτηκε:

DUPLICATE ENTRIES  
SOLD QTY GREATER THAN DELIVERED QTY.

Καθ' όλη την διάρκεια της χρήσης του αρχείου οι χρήστες κατευθύνονται πάντα προς την ορθή και επιτυχή συμπλήρωση των πεδίων. Για αυτόν τον σκοπό εμφανίζονται διάφορα μηνύματα ενημέρωσης και επιλογής (Message Boxes).



**Σχήμα 4.16:** Μηνύματα ενημέρωσης και επιλογής του εργαλείου Data Upload Tool

1. Δεν βρέθηκαν καθόλου εγγραφές. Αποτρέπεται η αποστολή κενού αρχείου
2. Επιλογή του χρήστη σε περίπτωση που θέλει να εισάγει μηδενική τιμή σε κενά κελία
3. Μήνυμα ενημέρωσης εύρεσης σφαλμάτων
4. Βρέθηκαν διπλές εγγραφές.
5. Επιλογή χρήσης αυτόματου φίλτρου έτσι ώστε να είναι ορατές μόνο οι εσφαλμένες εγγραφές.
6. Μήνυμα ενημέρωσης ορθής εκτέλεσης διαδικασίας Validation
7. Αποδεικτικό αποστολής δεδομένων με ενημέρωση για το όνομα του αρχείου csv      α. Για Monthly Upload Mode      β. Για Weekly Upload Mode

Σε αυτό το σημείο έχει παρθεί μια σημαντική απόφαση σχετικά με την εισαγωγή δεδομένων. Το μέγεθος Delivered QTY θα μπορούσε να ανακληθεί μέσα από την Βάση Δεδομένων (SAP fact data). Ωστόσο αυτό αποτρέπει τον εντοπισμό λογικών σφαλμάτων μεταξύ Delivery\_QTY και Sold\_QTY σε αρχικό στάδιο οπότε τα σφάλματα θα ήταν εμφανή στο τελευταίο στάδιο της διαδικασίας EN, δηλαδή στις αναφορές. Κατά τον σχεδιασμό ενός συστήματος EN οι διαχειριστές θα πρέπει να είναι σίγουροι πως ένα λάθος μπορεί να εντοπιστεί όσο το δυνατόν πιο αποτελεσματικά ώστε να μην σπαταλούνται πόροι αναποτελεσματικά.

### *Δ3.3 Unix Server (Stored Procedures)*

Το επόμενο στάδιο προετοιμασίας και εισαγωγής πραγματοποιείται στον τοπικό Unix Server μέσω stored procedures. Σκοπός σε αυτό το στάδιο είναι η δημιουργία back up files αλλά κυρίως η δημιουργία ενός συγκεντρωτικού αρχείου από τις εισαγωγές των χρηστών. Το συγκεντρωτικό αρχείο csv συνδέεται στην βάση με την κλήση του SQL\_LOADER. Η δομή του ορίζεται από τον κώδικα δημιουργίας του εξωτερικού πίνακα.

Αναλυτικά, η διαδικασία που εκτελείται στον Unix Server ξεκινάει με την κλήση της store procedure sellout\_dwh.start η οποία παρουσιάζεται παρακάτω και στην ουσία δημιουργεί ένα log file για λόγους παρακολούθησης της διαδικασίας. Στο log file αποθηκεύεται μήνυμα σφάλματος ή ορθής εκτέλεσης. Τέλος η sellout\_dwh.start καλεί την sellout\_dwh.

```
#!/bin/ksh
echo "**** execution profile ****"
. /etc/profile
. /usr/local/bin/oraenv

mv $PLOG/sellout_dwh.log $PLOG/sellout_dwh.oldlog
echo "**** lancement de la procedure sellout ****"
echo "      $(date)\n"

cd $PPROC/sellout
sellout_dwh $* 1>>$PLOG/ sellout_dwh.log 2>&1
Store Procedure sellout_dwh.start
```

Η διαδικασία `sellout_dwh` περιλαμβάνει τρία βασικά βήματα ξεκινώντας από τις αρχικοποιήσεις μεταβλητών. Το πρώτο βήμα περιλαμβάνει την μετακίνηση των αρχείων από τον κατάλογο `UPLOAD`, όπου αρχικά αποθηκεύονται τα `csv` αρχεία, στον κατάλογο `TRF`. Με την χρήση της εντολής `CAT` δημιουργείτε το συγκεντρωτικό αρχείο και με την εντολή `RETCODE` γίνεται έλεγχος της σωστής δημιουργίας αυτού.

```
step1 () ### move files from upload to TRF
{
  echo " ==> step1: move files from upload to TRF  \n"
  mv $UPLOAD/ SELLOUT/*.csv $TRF/DWH
  rm $TRF/DWH/All_Sellout.csv
  cat $TRF/DWH/Sellout_*.csv > $TRF/DWH/All_Sellout.csv
  tail -n 5 $TRF/DWH/All_Sellout.csv
  RETCODE=$?
  if [ RETCODE -ne 0 ]
  then
    echo " *** copy files from UPLOAD to TRF not OK  ***"
    exit
  fi
  step2
}
```

Store Procedure sellout\_dwh Βήμα 1<sup>ο</sup>: Μετακίνηση αρχείων από τον κατάλογο UPLOAD στον TRF και δημιουργία συγκεντρωτικού αρχείου All\_Sellout.csv

Κατά το δεύτερο βήμα γίνεται κλήση του SQL LOADER με χρήση της εντολής SQLPLUS η οποία συνδέει τον Server με την Oracle Βάση Δεδομένων.

```
step2  () ### integration of SELLOUT files to the DB
{
  echo " ==> step2: upload files to the DB          \n"
  echo $NLS_LANG
  sqlplus $CONN_STRING @$PREQ/sellout_upload.sql
  RETCODE=$?
  if [ RETCODE -ne 0 ]
  then
    echo " *** error in sellout_upload.sql ***"
    exit
  fi
  step3
}
```

Store Procedure sellout\_dwh Βήμα 2<sup>ο</sup>: Κλήση SQL Loader για την δημιουργία εξωτερικού πίνακα. Στο τελευταίο βήμα έχουμε την μεταφορά των αρχείων στον κατάλογο DWH/SAVE για λόγους back up. Τέλος, η διαδικασία στον Unix Server καταλήγει με εμφάνιση ορθής εκτέλεσης στην κονσόλα.

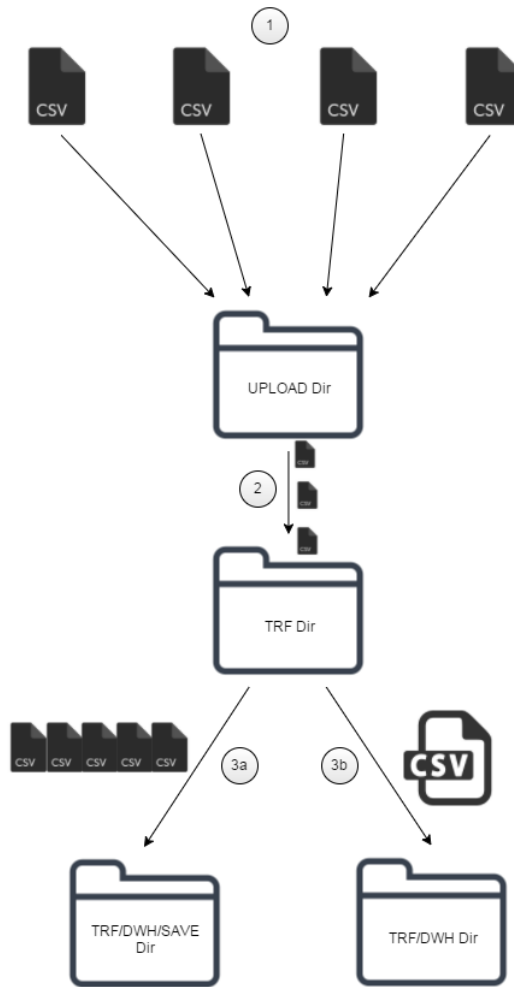
```
step3  () ### move SEE files from TRF to SAVE
{
  echo " ==> step3: move files from TRF to ARCHIVE  \n"
  mv $TRF/DWH/Sellout*.csv $TRF/DWH/SAVE/
  RETCODE=$?
  if [ RETCODE -ne 0 ]
  then
    echo " *** copy files from TRF to TRF/SAVE not OK ***"
    exit
  fi
  stepfin
}
stepfin ()
{
```



```
echo " ==> sellout_dwh sucessfull $(date) "  
exit  
}
```

Store Procedure sellout\_dwh Βήμα 3<sup>ο</sup>: Μεταφορά αρχείων από τον TRF στον SAVE (back up files).

Η διαδικασία συνολικά φαίνεται απλοϊκά στο παρακάτω γράφημα.



**Σχήμα 4.17:** Διαδικασία μετακίνησης αρχείων στον Unix Server

1. Τα αρχεία από τους διάφορους Πωλητές αποθηκεύονται στον Unix Server μετά από τις κατάλληλες εντολές VBA.
2. Τα μοναδικά αρχεία μεταφέρονται στον TRF κατάλογο. Αποφεύγονται τυχόν πολλαπλές φορτώσεις ίδιων αρχείων.
  - a. Τα αρχεία στον κατάλογο DWH/SAVE για λόγους back up.
  - b. Πραγματοποιείται η δημιουργία του συγκεντρωτικού αρχείου csv.

Επαναφέροντας το ζήτημα της σπατάλη πόρων, η διαδικασία δεν είναι σχεδιασμένη για σύγχρονη εξαγωγή πληροφοριών. Με στόχο την καλύτερη διαχείριση των επιχειρηματικών πόρων η κλήση των stored procedures γίνεται δύο φορές ανά ημέρα. Τον προγραμματισμό των stored procedures τον πετυχαίνουμε με την χρήση της Unix εντολής crontab. Τέλος, το συγκεντρωτικό csv αρχείο επαναδημιουργείται καθημερινά όταν και εφόσον όλα τα δεδομένα έχουν περάσει από την ETL διαδικασία και βρίσκονται στην Αποθήκη Δεδομένων. Σε περίπτωση ανάγκης υπάρχει η δυνατότητα να ανατρέξουμε στα αρχεία back up στον κατάλογο TRF/SAVE. Συνδυάζοντας τα παραπάνω πετυχαίνουμε αξιοπιστία με ταυτόχρονη οικονομία πόρων.

#### *Δ3.4 Επεξεργασία και αποθήκευση δεδομένων (ETL Process)*

Όπως αναφέρθηκε στο θεωρητικό μέρος, η διαδικασία ETL είναι σχεδιασμένη βάση της τεχνολογίας DW, στην συγκεκριμένη περίπτωση της Oracle DB. Στο προηγούμενο βήμα τα δεδομένα έχουν συγκεντρωθεί σε ένα αρχείο csv το οποίο αποτελεί εξωτερικό πίνακα στην Oracle Βάση Δεδομένων. Η λύση της Oracle δίνει την δυνατότητα εκτέλεσης πακέτων SQL εντολών για διάφορες εργασίες. Η διαδικασία ETL εκτελείται μέσω του πακέτου pack\_sellout\_south.process\_sellout\_load την κλήση του οποίου αναλαμβάνει η sellout\_upload.sql όπως φαίνεται στο παρακάτω στιγμιότυπο.

```
whenever sqlerror exit 20
set timing on
/*
  upload SELLOUT files
*/
execute pack_sellout.process_sellout_load;
exit
```

### *Ενέργειες στο Extract μέρος της διαδικασίας.*

Η διαδικασία ETL ξεκινάει με την εισαγωγή όλων των δεδομένων από τον εξωτερικό πίνακα σε πίνακα LOG του επιπέδου SAS μεταφέροντας όλα τα δεδομένα χαρακτήρων ως κεφαλαία με την εντολή UPPER(char)<sup>31</sup> καθώς επίσης εισάγεται το χρονικό στιγμιότυπο με την εντολή SYSDATE<sup>32</sup>. Ο σκοπός της χρήσης της UPPER δεν είναι άλλος παρά η ομοιογένεια και η αποφυγή λαθών (case sensitivity).

```
-- ***** EXTRACT part of the process. Operational level: EXT *****

-- +++ BEGINING OF LOAD_EXT_TO_SAS PROCEDURE +++

PROCEDURE LOAD_EXT_TO_SAS IS
  lv_table CONSTANT afrsas.LOG.LOG_TABLE%TYPE := 'SELLOUT';

-- Insert into SOUTH_SELLOUT_LOG from EXT
BEGIN
  INSERT INTO AFRSAS.SELLOUT_LOG
  SELECT CUSTOMER_ID,
         UPPER(PERIOD_DESC),
         UPPER(SEASON_ID),
         WEEK,
         UPPER(ARTICLE_ID),
         DELIVERED_QTY,
         SOLD_QTY,
         AUX1,
         AUX2,
         GUID_ID,
         USER_ID,
         sysdate
  FROM EXT_SELLOUT;
  COMMIT;
```

<sup>31</sup> [https://docs.oracle.com/cd/B19306\\_01/server.102/b14200/functions206.htm](https://docs.oracle.com/cd/B19306_01/server.102/b14200/functions206.htm)

<sup>32</sup> [https://docs.oracle.com/cd/B19306\\_01/server.102/b14200/functions172.htm](https://docs.oracle.com/cd/B19306_01/server.102/b14200/functions172.htm)

Πριν η διαδικασία προχωρήσει στο TRANSFORM μέρος, διαγράφεται ο πίνακας SELLOUT στο επίπεδο SAS ώστε να αποτρέπεται η επανάληψη της διαδικασίας σε δεδομένα που έχουν ήδη διαμορφωθεί κατάλληλα.

```
-- Truncate AFRSAS.SELLOUT
DELETE FROM AFRSAS.SELLOUT;
COMMIT;
```

Κατά το τέλος του EXTRACT μέρους της διαδικασίας μεταφέρονται τα μοναδιαία δεδομένα στον SELLOUT από τον EXT πίνακα κάνοντας χρήση ξανά της εντολής UPPER(char). Τα μοναδιαία δεδομένα διαλέγονται με τον όρο DISTINCT<sup>33</sup> έτσι ώστε να αποφευχθούν διπλές εγγραφές οι οποίες προέρχονται από διαφορετικά αρχεία καθώς όπως αναλύθηκε σε προηγούμενο μέρος του κεφαλαίου διπλές εγγραφές στο ίδιο αρχείο αποφεύγονται από το Upload Tool.

```
-- Insert into AFRSAS.SELLOUT
INSERT INTO AFRSAS.SELLOUT
(
    CUSTOMER_ID,
    PERIOD_DESC,
    SEASON_ID,
    WEEK,
    ARTICLE_ID,
    DELIVERED_QTY,
    SOLD_QTY
)
SELECT DISTINCT CUSTOMER_ID,
    UPPER(PERIOD_DESC),
    UPPER(SEASON_ID),
    WEEK,
    UPPER(ARTICLE_ID),
    DELIVERED_QTY,
    SOLD_QTY
FROM EXT_SELLOUT;
COMMIT;
```

<sup>33</sup> <http://www.techonthenet.com/oracle/distinct.php>

```
END LOAD_EXT_TO_SAS;  
-- +++ END OF LOAD_EXT_TO_SAS PROCEDURE +++
```

### *Ενέργειες στο Transform μέρος της διαδικασίας*

Το κύριο και πιο πολύπλοκο μέρος της ETL διαδικασίας αρχικά δημιουργεί πίνακα LOG έτσι ώστε να παρακολουθείτε και να εξυπηρετείται ο εντοπισμός σφαλμάτων και λαθών.

```
-- ***** TRANSFORM part of the process. Operational level: SAS *****  
PROCEDURE SAS_DATA_TRANSFORM IS  
  lv_table CONSTANT afrsas.LOG.LOG_TABLE%TYPE := 'SELLOUT';  
BEGIN
```

Αμέσως μετά πραγματοποιούνται μικρές αρχικοποιήσεις και μετατροπές των δεδομένων. Σχετικά με τους κωδικούς προϊόντων μετατρέπονται όλοι σε μήκους 6 χαρακτήρων αλφαριθμητικά μεγέθη.

```
-- Update where ARTICLE_ID starts with 00  
BEGIN  
  UPDATE AFRSAS.SELLOUT  
  SET ARTICLE_ID =SUBSTR(ARTICLE_ID, 3, 6)  
  WHERE ARTICLE_ID LIKE '00%';  
  COMMIT;  
END;
```

Η επόμενη μετατροπή δεδομένων αφορά τιμές οι οποίες έχουν φορτωθεί χρησιμοποιώντας το Monthly Upload Mode, (άρα το πεδίο WEEK έχει τιμή 0). Ο πίνακας SELLOUT\_TIME\_DIM περιέχει την αντιστοιχία μήνα-εβδομάδας για κάθε έτος με παρόμοιο τρόπο που περιέχεται στο εργαλείο φόρτωσης (φύλλο Correspondence List). Μετά και από αυτό το βήμα έχει πλέον επιτευχθεί η ομοιογένεια των δεδομένων.

```
-- Update week 0 entries based on correspondence table  
BEGIN  
  PACK_LOG.START_MESSAGE('  UPDATE  AFRSAS.SELLOUT  Week  Update  
,lv_table);
```

```
UPDATE AFRSAS.SELLOUT
SET SELLOUT.WEEK = (SELECT MAX(WEEK)
                    FROM AFRDWH.SELLOUT_TIME_DIM
                    WHERE UPPER(PERIOD) = SELLOUT.PERIOD_DESC)
WHERE SELLOUT.WEEK = 0;
COMMIT;
EXCEPTION
WHEN OTHERS THEN
    dbms_output.put_line('DWH: data manipulation (WEEK) on SELLOUT !:
'||SQLERRM);
    ROLLBACK;
END;
```

Έχει αναφερθεί πολλάκις σε διάφορα σημεία της εργασίας η σημασία της οικονομίας πόρων. Μια ενέργεια οικονομίας πόρων είναι η διαγραφή εγγραφών που περιέχονται ήδη από προηγούμενη εκτέλεση της διαδικασίας στο DW επίπεδο όπως επίσης και η διαγραφή διπλών εγγραφών στο επίπεδο SAS. Σημειώνεται πως διπλές εγγραφές μπορούν να υπάρξουν ύστερα από την ομοιογενοποίηση των δεδομένων.

```
-- Remove entries that exists on DWH
BEGIN
    PACK_LOG.START_MESSAGE('    UPDATE    AFRSAS.SELLOUT    Week    Update
',lv_table);
    DELETE FROM AFRSAS.SELLOUT s
    WHERE EXISTS
        (SELECT CUSTOMER_ID,
        PERIOD_DESC,
        SEASON_ID,
        WEEK,
        ARTICLE_ID,
        DELIVERED_QTY,
        SOLD_QTY FROM AFRDWH.SELLOUT d
        WHERE
            d.CUSTOMER_ID = s.CUSTOMER_ID AND
            d.PERIOD_DESC = s.PERIOD_DESC AND
            d.SEASON_ID = s.SEASON_ID AND
            d.WEEK = s.WEEK AND
```

```
d.ARTICLE_ID = s.ARTICLE_ID AND
d.DELIVERED_QTY = s.DELIVERED_QTY AND
d.SOLD_QTY = s.SOLD_QTY);
COMMIT;
EXCEPTION
WHEN OTHERS THEN
    dbms_output.put_line('DWH: data manipulation (delete existing
entries) on SELLOUT !: '||SQLERRM);
ROLLBACK;
END;
```

Για την διαγραφή διπλών εγγραφών σε έναν πίνακα δίχως χρήση δευτέρου βοηθητικού πίνακα, χρησιμοποιούνται τεχνικές όπως φαίνονται παρακάτω<sup>34</sup>. Σημαντικό ρόλο στον κώδικα του παρακάτω στιγμιότυπου έχει η εντολή **Count**<sup>35</sup>.

```
--Remove duplicate entries from SAS
BEGIN
    PACK_LOG.START_MESSAGE(' UPDATE AFRSAS.SELLOUT Remove Duplicates
',lv_table);
    DELETEFROM AFRSAS. SELLOUT
    WHERE ROWID IN
    (SELECT HID
    FROM
    (SELECT HID, CUSTOMER_ID, PERIOD_DESC, SEASON_ID, WEEK, ARTICLE_ID,
DELIVERED_QTY, SOLD_QTY, RANK() OVER
    (PARTITION BY CUSTOMER_ID, PERIOD_DESC, SEASON_ID, WEEK,
    ARTICLE_ID, DELIVERED_QTY, SOLD_QTY ORDER BY CUSTOMER_ID,
    PERIOD_DESC, SEASON_ID, WEEK, ARTICLE_ID, DELIVERED_QTY,
    SOLD_QTY
    DESC, ROWID) AS SeqNumber
    FROM
    (SELECT ROWID AS HID,
CUSTOMER_ID, PERIOD_DESC, SEASON_ID, WEEK,
    ARTICLE_ID, DELIVERED_QTY, SOLD_QTY FROM AFRSAS.SELLOUT
    WHERE (CUSTOMER_ID, PERIOD_DESC, SEASON_ID, WEEK, ARTICLE_ID,
```

<sup>34</sup> Περισσότερες πληροφορίες και τεχνικές για διαχείριση διπλών εγγραφών στον παρακάτω σύνδεσμο:

<http://www.xaprb.com/blog/2006/10/09/how-to-find-duplicate-rows-with-sql/>

<sup>35</sup> <https://www.techonthenet.com/oracle/functions/count.php>

```

DELIVERED_QTY, SOLD_QTY) IN
      (SELECT s.CUSTOMER_ID, s.PERIOD_DESC, s.SEASON_ID, s.WEEK,
            s.ARTICLE_ID, s.DELIVERED_QTY, s.SOLD_QTY
FROM AFRSAS.SELLOUT s
GROUP BY s.CUSTOMER_ID, s.PERIOD_DESC, s.SEASON_ID, s.WEEK,
         s.ARTICLE_ID, s.DELIVERED_QTY, s.SOLD_QTY
HAVING COUNT(*) > 1))
      WHERE SeqNumber > 1);

Pack_Log.end_message(TO_CHAR(SQL%rowcount) || ' UPDATE
AFRSAS.SOUTH_SELLOUT Remove Duplicates ');

COMMIT;

EXCEPTION
  WHEN OTHERS THEN
    Pack_Log.start_message('DWH: Data manipulation (Remove Duplicates)
!: ' || SQLERRM, lv_table, Pack_Log.gv_error_level);
    ROLLBACK;
  END;
END SAS_DATA_TRANSFORM;

```

### *Ενέργειες στο Load μέρος της διαδικασίας*

Φτάνοντας στο τελευταίο μέρος της ETL διαδικασίας, την φόρτωση δεδομένων στην τελική μορφή στην Αποθήκη Δεδομένων, συνεχίζουν να μας απασχολούν θέματα οικονομίας πόρων και κυρίως χρόνου. Για αυτόν τον σκοπό και για αποφυγή δημιουργίας διπλών εγγραφών, ενημερώνονται ήδη υπάρχουσες εγγραφές με τις πρόσφατες τιμές. Καθώς δεδομένα εισάγονται καθημερινά για εβδομαδιαίες ή μηνιαίες πωλήσεις υπάρχει περίπτωση διόρθωσης και προόδου.

```

-- ***** LOAD part of the process. Operational level: DWH
*****

PROCEDURE LOAD_SAS_TO_DWH IS
  lv_table CONSTANT afrsas.LOG.LOG_TABLE%TYPE := 'SELLOUT';
BEGIN
  --Update AFRDWH.SELLOUT with latest values where entry already exists
  BEGIN
    Pack_Log.start_message('UPDATE AFRDWH.SELLOUT', lv_table);
    dbms_output.put_line('UPDATE AFRDWH.SELLOUT ' || lv_table);
    UPDATE AFRDWH.SELLOUT d

```



```
SET
(
  d.CUSTOMER_ID,
  d.PERIOD_DESC,
  d.SEASON_ID,
  d.WEEK,
  d.ARTICLE_ID,
  d.DELIVERED_QTY,
  d.SOLD_QTY
) = (
  SELECT DISTINCT
    s.CUSTOMER_ID,
    s.PERIOD_DESC,
    s.SEASON_ID,
    s.WEEK,
    s.ARTICLE_ID,
    s.DELIVERED_QTY,
    s.SOLD_QTY
  FROM AFRSAS.SELLOUT s
  WHERE
    s.CUSTOMER_ID = d.CUSTOMER_ID AND
    s.PERIOD_DESC = d.PERIOD_DESC AND
    s.SEASON_ID = d.SEASON_ID AND
    s.WEEK = d.WEEK AND
    s.ARTICLE_ID = d.ARTICLE_ID
  )
WHERE
EXISTS
(
  SELECT 1
  FROM
    AFRSAS.SELLOUT s
  WHERE
    d.CUSTOMER_ID = s.CUSTOMER_ID AND
    d.PERIOD_DESC = s.PERIOD_DESC AND
    d.SEASON_ID = s.SEASON_ID AND
    d.WEEK = s.WEEK AND
    d.ARTICLE_ID = s.ARTICLE_ID
  );
```

```
Pack_Log.end_message(TO_CHAR(SQL%rowcount) || ' UPDATE AFRDWH.SELLOUT
');
dbms_output.put_line('UPDATE AFRDWH.SELLOUT
'||TO_CHAR(SQL%rowcount));
COMMIT;
EXCEPTION
WHEN OTHERS THEN
    Pack_Log.start_message('DWH: update SELLOUT !: ' || SQLERRM,
lv_table, Pack_Log.gv_error_level);
    dbms_output.put_line('DWH: update SELLOUT !: ' || SQLERRM);
    ROLLBACK;
END;
```

Πλησιάζοντας στο τέλος του Load μέρους, και της ETL διαδικασίας καθ' αυτής, φορτώνονται τα δεδομένα στο DW. Οι μοναδικές εγγραφές που εξαιρούνται είναι οι εγγραφές με μηδενικές τιμές στο Sold\_QTY αποφεύγοντας έτσι δεδομένα που δεν περιέχουν πληροφορίες πωλήσεων.

```
--Insert into AFRDWH.SELLOUT where QTYs are not 0
BEGIN
Pack_Log.start_message('Insertions',lv_table);
dbms_output.put_line('Insertions ' ||lv_table);
    INSERT INTO AFRDWH.SELLOUT
        (
            CUSTOMER_ID,
            PERIOD_DESC,
            SEASON_ID,
            WEEK,
            ARTICLE_ID,
            DELIVERED_QTY,
            SOLD_QTY
        )
    SELECT CUSTOMER_ID,
           PERIOD_DESC,
           SEASON_ID,
           WEEK,
           ARTICLE_ID,
           DELIVERED_QTY,
           SOLD_QTY
```

```

FROM AFRSAS.SELLOUT s
WHERE
  (DELIVERED_QTY <>0 or SOLD_QTY <>0) AND
  NOT EXISTS
    (SELECT 1
     FROM
       AFRDWH.SELLOUT d
     WHERE
       d.CUSTOMER_ID = s.CUSTOMER_ID AND
       d.PERIOD_DESC = s.PERIOD_DESC AND
       d.SEASON_ID = s.SEASON_ID AND
       d.WEEK = s.WEEK AND
       d.ARTICLE_ID = s.ARTICLE_ID
    );

Pack_Log.end_message(TO_CHAR(SQL%rowcount)||' insertion(s).');
dbms_output.put_line('Insertions ' ||TO_CHAR(SQL%rowcount));
COMMIT;

Pack_Log.end_message(TO_CHAR(SQL%rowcount)||' insertion(s).');
dbms_output.put_line('Insertions ' ||TO_CHAR(SQL%rowcount));
COMMIT;
EXCEPTION
  WHEN DUP_VAL_ON_INDEX THEN
    Pack_Log.start_message('DWH: update or insert SELLOUT !: ' ||
SQLERRM, lv_table,Pack_Log.gv_error_level);
    dbms_output.put_line('DWH: update or insert SELLOUT !:
' ||SQLERRM);
    ROLLBACK;
    END;

END LOAD_SAS_TO_DWH;
END PACK_SELLOUT;

```

Συνολικά η διαδικασία ETL, όπως αναλύθηκε παραπάνω, πετυχαίνει την καλύτερη δυνατή οικονομία πόρων ενώ μετατρέπει και μεταφέρει τα δεδομένα στην Αποθήκη Δεδομένων με ασφάλεια και αξιοπιστία. Στο παρακάτω γράφημα φαίνεται συνοπτικά η ETL διαδικασία με όλα τα στάδια που περιλαμβάνει για κάθε επίπεδο της βάσης δεδομένων.

### Δ3.5 Ανάλυση και Αναφορές Πωλήσεων στο Business Objects

Στα πλαίσια της ανάλυσης και του σχεδιασμού αναφορών τα περισσότερα θέματα τα αναλαμβάνει το Business Objects. Στο διαχειριστικό περιβάλλον δημιουργείτε ένα Universe, μια γραφική απεικόνιση συσχετίσεων μεταξύ πινάκων του DW από τους οποίους η μηχανή ερωτημάτων θα δημιουργήσει ένα data cube. Παρακάτω φαίνεται ένα παράδειγμα γραφικής απεικόνισης συσχετίσεων πινάκων παρόμοιο με το BO Universe (Entity-Relationship, ER Diagram).

Οι γραφικές απεικονίσεις είναι αποτελέσματα ενεργειών στο data cube και όχι στο DW. Το BO ανακτεί δεδομένα και παράγει metrics και dimensions τα οποία είναι αριθμητικά δεδομένα και χαρακτηριστικά αντίστοιχα (για παράδειγμα το Sold\_QTY είναι metric ενώ το Customer\_ID είναι dimension). Το παρακάτω στιγμιότυπο κώδικα δημιουργεί έναν κύβο δεδομένων 3-διαστάσεων όπου και περιέχει όλα τα δεδομένα των Πελατών που εισήχθησαν από τους Πωλητές.

```
SELECT
  LAST_DAY (TO_DATE (SELLOUT.PERIOD_DESC || '-01', 'YYYY/MON/DD')),
  SELLOUT_CUST_ATTR.STORECOACH,
  SELLOUT_CUST_ATTR.STORECOACH_NAME,
  Table__165.MAXWEEK,
  SELLOUT_TIME_DIM.MONTH,
  SELLOUT_TIME_DIM.FULL_DATE,
  to_number (substr (SELLOUT.PERIOD_DESC, 1, 4)),
  SELLOUT.PERIOD_DESC,
  SELLOUT.SEASON_ID,
  TO_CHAR (LAST_DAY (TO_DATE (SELLOUT.PERIOD_DESC || '-01',
  'YYYY/MON/DD')), 'YYYY.MM'),
  SELLOUT.WEEK,
  SELLOUT.DELIVERED_QTY,
  SELLOUT.SOLD_QTY,
  CUST_ATT_HIE.LEVEL1_DESCR,
  CUST_ATT_HIE.LEVEL1,
  CUST_ATT_HIE.LEVEL2_DESCR,
  CUST_ATT_HIE.LEVEL2,
  CUST_ATT_HIE.CUSTOMER_ATTRIBUTE9_DESCR,
  CUST_ATT_HIE.CUSTOMER_ATTRIBUTE9,
  ARTICLE.ARTICLE_ID,
```

```
ARTICLE.ARTICLE_DESCR,  
BS_HIERARCHY.BS_L1_DESCR,  
BS_HIERARCHY.BS_L2_DESCR,  
BS_HIERARCHY.BS_L3_DESCR,  
BUSINESS_SEGMENT.BUSINESS_SEGMENT_DESCR,  
DECODE (PRODUCT_DIVISION.PRODUCT_DIVISION_ID, '01', 'Footwear', '02', 'Appare  
l', '03', 'Hardware'),  
MARKETING_SEGMENT.MARKETING_SEGMENT_DESCR,  
SPORTS_CATEGORY.SPORTS_CATEGORY_DESCR,  
GENDER.GENDER_DESCR,  
AGE_GROUP.AGE_GROUP_DESCR,  
PRODUCT_DIVISION.PRODUCT_DIVISION_ID,  
HIE_SALES_ORG.COUNTRY,  
HIE_SALES_ORG.MARKET,  
HIE_SALES_ORG.REGION,  
MODEL.MODEL_DESCR,  
SELLOUT_BRAND.REFERENCE_TYPE  
FROM  
SELLOUT,  
SELLOUT_CUST_ATTR,  
(  
SELECT MAX(WEEK) as MaxWeek, YEAR,  
MONTH  
FROM SELLOUT_TIME_DIM  
GROUP BY YEAR, MONTH  
 ) Table__165,  
SELLOUT_TIME_DIM,  
CUSTOMER,  
CUST_ATT_HIE,  
ARTICLE,  
BS_HIERARCHY,  
BUSINESS_SEGMENT,  
PRODUCT_DIVISION,  
MARKETING_SEGMENT,  
SPORTS_CATEGORY,  
GENDER,  
AGE_GROUP,  
HIE_SALES_ORG,  
MODEL,  
BRAND,
```

```

CUSTOMER
WHERE
  ( CUSTOMER.CUSTOMER_ID=SELLOUT.CUSTOMER_ID )
  AND ( ARTICLE.ARTICLE_ID=SELLOUT.ARTICLE_ID )
  AND (
CUSTOMER.SALES_ORGANIZATION_ID=HIE_SALES_ORG.SALES_ORGANIZATION_ID )
  AND ( CUSTOMER.SALES_ORGANIZATION_ID=SELLOUT_BRAND.LONG_LABEL )
  AND ( SELLOUT.WEEK=SELLOUT_TIME_DIM.WEEK )
  AND ( SUBSTR(SELLOUT.PERIOD_DESC,1,4)=SELLOUT_TIME_DIM.YEAR )
  AND ( Table__165.YEAR=SELLOUT_TIME_DIM.YEAR )
  AND ( Table__165.MONTH=SELLOUT_TIME_DIM.MONTH )
  AND (
CUSTOMER.SALES_ORGANIZATION_ID=SELLOUT_CUST_ATTR.SALES_ORGANIZATION_ID(+
) )
  AND ( CUSTOMER.CUSTOMER_ID=SELLOUT_CUST_ATTR.CUSTOMER_ID(+ ) )
  AND ( ARTICLE.MODEL_ID=MODEL.MODEL_ID )
  AND (
PRODUCT_DIVISION.PRODUCT_DIVISION_ID=ARTICLE.PRODUCT_DIVISION_ID )
  AND (
MARKETING_SEGMENT.MARKETING_SEGMENT_ID=ARTICLE.MARKETING_SEGMENT_ID )
  AND (
MODEL.SPORTS_CATEGORY_ID=SPORTS_CATEGORY.SPORTS_CATEGORY_ID(+ ) )
  AND ( MODEL.GENDER_ID=GENDER.GENDER_ID(+ ) )
  AND ( MODEL.AGE_GROUP_ID=AGE_GROUP.AGE_GROUP_ID(+ ) )
  AND (
BUSINESS_SEGMENT.BUSINESS_SEGMENT_ID(+)=ARTICLE.BUSINESS_SEGMENT_ID )
  AND ( BS_HIERARCHY.BS_ID=BUSINESS_SEGMENT.BUSINESS_SEGMENT_ID )
  AND ( SELLOUT_BRAND.SHORT_LABEL = BS_HIERARCHY.BS_L1_ID AND
SELLOUT_BRAND.REFERENCE_TYPE = 'SELLOUT_BRAND' )

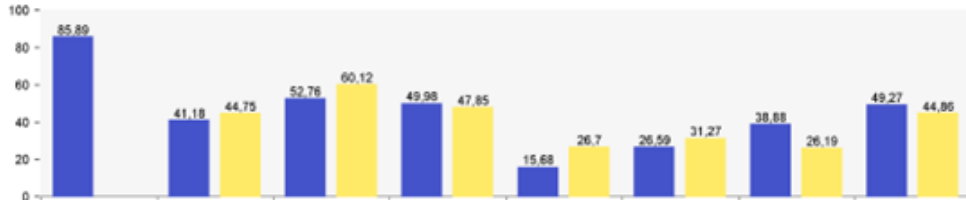
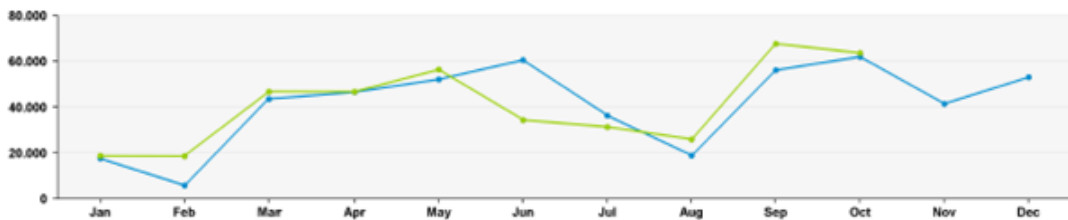
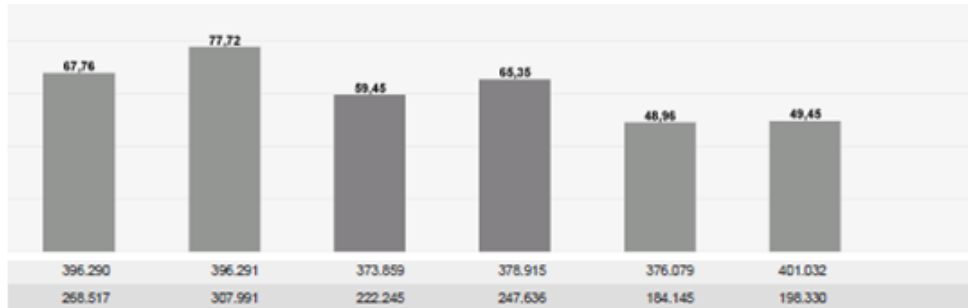
```

Οι δυνατότητες για γραφική απεικόνιση που διαθέσει το ΒΟ είναι αρκετά απλές στην χρήση τους χωρίς να περιορίζουν τους χρήστες σε περιπτώσεις περίπλοκων συσχετίσεων. Το τελικό αποτέλεσμα φαίνεται στο παρακάτω σχήμα 4.18 όπου μπορούμε να δούμε

α. Πωλήσεις ανά περίοδο σε συνδυασμό με τις ποσότητες προμήθειας, το οποίο μεταφράζεται ως συσχέτιση Sold\_QTY, Delivered\_QTY και Period.

β. Πωλήσεις ανά μήνα (ή εβδομάδα) πωλήσεων για κάθε έτος, το οποίο μεταφράζεται ως συσχέτιση Sold\_QTY, Period και Week.

γ. Προϊόντα με τις περισσότερες πωλήσεις σε συσχέτιση με τα προϊόντα με τις λιγότερες πωλήσεις για τον τρέχον μήνα, το οποίο μεταφράζεται ως συσχέτιση Sold\_QTY, Period και Article\_ID.



**Σχήμα 4.18:** Το τελικό στάδιο των αναφορών με παραδείγματα τελικών γραφημάτων στο περιβάλλον του SAP Business Objects μετά από ανάλυση των πωλήσεων του case study

## ΚΕΦΑΛΑΙΟ Ε – ΣΥΜΠΕΡΑΣΜΑΤΑ

Η διαδικασία η οποία αναλύθηκε στο case study αφορά μια περίπλοκη και αρκετά πολυδάπανη εφαρμογή EN παρόμοια της οποίας ένας οργανισμός χωρίς πόρους και υπάρχοντες δομές δεν δύναται να αναπαράγει. Δομές που προϋπήρχαν στήριξαν την αντιμετώπιση των τεχνολογικών και επιχειρηματικών προκλήσεων καθώς και την συνεργασία των τμημάτων. Οι πληροφορίες των αναφορών βοήθησαν την εύστοχη προσαρμογή στις ανάγκες της αγοράς μέσω της ενδυνάμωσης της πελατειακής βάσης.

Βάση των παραπάνω καταλήγουμε στο συμπέρασμα πως τα κύρια στοιχεία ισχυροποίησης της διαδικασίας EN σε έναν οργανισμό είναι (α) τα θέματα αξιοπιστίας δεδομένων και (β) οι τεχνικές απόδοσης. Τα παραπάνω ως αποτέλεσμα έχουν την ορθή σχεδίαση των παραμέτρων του τοπίου EN (επιλογή αρχιτεκτονικών DW, συστήματος ΒΔ, εργαλείου αναφορών κ.α.).

Η χαμηλή αξιοπιστία, δεν πρόκειται να εμποδίσει την λειτουργικότητα του συστήματος στις περισσότερες των περιπτώσεων. Οι πηγές δεδομένων μπορούν να διοχετεύουν το σύστημα με δεδομένα, οι αναφορές μπορούν να κατασκευαστούν και είναι δυνατή η εξαγωγή συμπερασμάτων. Ο στόχος της EN ωστόσο είναι τα συμπεράσματα να χρησιμοποιούνται για λήψη αποφάσεων και να κατευθύνουν ένα οργανισμό. Εάν τα δεδομένα είναι ανακριβή, οι πληροφορίες που θα οδηγήσουν σε λήψη αποφάσεων πιθανών να είναι επίσης εσφαλμένες.

Η θέσπιση κανόνων ποιότητας των δεδομένων είναι κρίσιμη για να αφαιρούνται ανακρίβειες και σφάλματα τα οποία προβάλλονται στις τελικές αναφορές. Τα περισσότερα ζητήματα ποιότητας των δεδομένων προέρχονται από ανθρώπινο λάθος. Αυτό οφείλεται στο γεγονός ότι πολλά στοιχεία σε έναν οργανισμό εξακολουθούν να βασίζονται στην καταχώριση στοιχείων από ανθρώπους. Στόχος των τεχνολόγων είναι να αναπτύξουν και προοδεύουν την εφαρμογή της τεχνολογίας. Κάτι τέτοιο είναι δυνατό, όπως αναφέρθηκε, με την χρήση του μοντέλου ωριμότητας. Επίσης, από την πλευρά των διαχειριστών, μια λεπτομερής ανάλυση με χρήση scenario testing κατά την ανάπτυξη προλαμβάνει τα σφάλματα πριν συμβούν.

Συμπεραίνουμε πως η θεμελιώδης πτυχή πίσω από την αντιμετώπιση των ζητημάτων ποιότητας των δεδομένων είναι να εξασφαλιστεί η διόρθωση από την



πηγή. Επίσης, σωστή αντιμετώπιση αποτελεί η συνεχής πρόβλεψη ζητημάτων (π.χ. με τεχνική scenario testing) σε κάθε στάδιο μεταφοράς και μετατροπής δεδομένων. Μια τέτοια νοοτροπία ανάπτυξης τεχνολογικών έργων, οικονομεί πόρους και κυρίως βελτιώνει αποτελεσματικά την απόδοση του συστήματος.

Είναι γεγονός πως κάθε μεγάλο έργο είναι απαιτητικό σε θέματα απόδοσης. Σε αυτή την κατηγορία εμπίπτει και η ανάπτυξη συστημάτων EN. Σε αρκετά σημεία έχει τονισθεί η σημασία της ταχύτητας εκτέλεσης και κυρίως της επιλογής του σημείου ελέγχου. Στο case study χρησιμοποιούνται διάφορες τεχνικές όπως για παράδειγμα η χρήση των Dictionary Object της VBA<sup>36</sup> για τον έλεγχο των εγγραφών στο User Upload Tool. Με την χρήση του Dictionary Object υπολογίστηκε βελτίωση του χρόνου εκτέλεσης περισσότερο από 370% σε σχέση με τον γραμμικό αλγόριθμο αναζήτησης σε λίστα.

Τα data requirements, η σύνδεση των δεδομένων με την υπάρχουσα δομή των Data Marts με το E-R diagram, η διαχείριση της υπολογιστικής διαθεσιμότητας του Intelligence Server, ο προγραμματισμός ανανέωσης και διανομής των αναφορών, είναι μερικά από τα σημαντικά ζητήματα τα οποία αρκετές φορές τέθηκαν υπό συζήτηση κατά την ανάπτυξη. Λαμβάνοντας υπόψιν την ποικιλομορφία στην ανάπτυξη, τις διαθέσιμες εμπορικές λύσεις, τους διαθέσιμους πόρους και τεχνογνωσία, κάθε έργο EN έρχεται αντιμέτωπο με τις ιδιαίτερες του απαιτήσεις σε τεχνικές απόδοσης. Η συνεχής εκπαίδευση, παρακολούθηση νέων τάσεων και βελτιστοποίηση σίγουρα αποτελούν τον δρόμο για ανακάλυψη νέων τεχνικών απόδοσης.

Κλείνοντας επισημάνετε το γεγονός πως όσο οι ανάγκες για διαχείριση του μεγάλου όγκου των δεδομένων γίνονται όλο και πιο σημαντικές, οι οργανισμοί αναμένεται να επενδύσουν στον εκσυγχρονισμό των τεχνολογιών και στην καλύτερη κατάρτιση των τελικών χρηστών. Στόχος είναι πλέον η δημιουργία κουλτούρας που ενθαρρύνει την εξερεύνηση των δεδομένων με βάση την λήψη αποφάσεων. Η Επιχειρησιακή Νοημοσύνη αποτελεί ένα ισχυρό εργαλείο σε μεγάλη ζήτηση και με τεράστια ανάπτυξη τα τελευταία χρόνια. Αφορά ένα μεγάλο εύρος επαγγελματιών και επιτάσσει την δημιουργία εκπαιδευτικών και ερευνητικών δομών που θα υποστηρίξουν την ταχύτερη αυτή ανάπτυξη των τεχνολογιών δεδομένων.

---

<sup>36</sup> [https://msdn.microsoft.com/en-us/library/x4k5wbx4\(v=vs.84\).aspx](https://msdn.microsoft.com/en-us/library/x4k5wbx4(v=vs.84).aspx)

## ΠΑΡΑΠΟΜΠΕΣ

[http://aisel.aisnet.org/ecis2013\\_cr/126](http://aisel.aisnet.org/ecis2013_cr/126)  
<http://blogs.gartner.com/svetlana-sicular/the-illusions-of-big-data/>  
<http://datawarehouse4u.info/OLTP-vs-OLAP.html>  
<http://dictionary.cambridge.org/dictionary/english/ad-hoc>  
[http://docs.oracle.com/cd/A87860\\_01/doc/server.817/a76992/ch3\\_eval.htm#2680](http://docs.oracle.com/cd/A87860_01/doc/server.817/a76992/ch3_eval.htm#2680)  
<http://DSSResources.COM/history/dsshistory.html>  
<http://dx.doi.org/10.1371/journal.pmed.0020267>  
<http://freakonomics.com/2008/02/25/hal-varian-answers-your-questions/>  
<http://olap.com/olap-definition/>  
<http://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&d=PALL&RefSrch=yes&Query=PN/5168446>  
<http://sap-certification.info/how-does-sap-work/>  
<http://simplystatistics.org/2013/12/12/the-key-word-in-data-science-is-not-data-it-is-science/>  
<http://www.1keydata.com/datawarehousing/data-warehouse-definition.html>  
<http://www.businessintelligence-solutions.com/differences-1st-2nd-3rd-normal-form/>  
[http://www.cbpp.uaa.alaska.edu/afef/gauge\\_your\\_data\\_warehouse\\_maturi.htm](http://www.cbpp.uaa.alaska.edu/afef/gauge_your_data_warehouse_maturi.htm)  
<http://www.cs.uvm.edu/%7Eicdm/algorithms/index.shtml>  
<http://www.datasciencecentral.com/profiles/blogs/batch-vs-real-time-data-processing>  
<http://www.dataversity.net/distinguishing-analytics-business-intelligence-data-science/>  
<http://www.economist.com/node/15557421>  
<http://www.economist.com/node/15579717>  
<http://www.economist.com/node/21531109>  
<http://www.element61.be/e/resourc-detail.asp?ResourceId=547>  
<http://www.forbes.com/sites/gilpress/2013/08/19/data-science-whats-the-half-life-of-a-buzzword/#3cab7e24a30>  
<http://www.information-management.com/infodirect/19991120/1675-1.html>  
<http://www.information-management.com/issues/20030201/6287-1.html>  
<http://www.jatit.org/volumes/research-papers/Vol9No1/9Vol9No1.pdf>  
<http://www.kimballgroup.com/2004/03/differences-of-opinion/>  
[http://www.mckinsey.com/Insights/MGI/Research/Technology\\_and\\_Innovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_Innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_Innovation)  
[http://www.orafaq.com/wiki/SQL\\*Loader\\_FAQ](http://www.orafaq.com/wiki/SQL*Loader_FAQ)  
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>  
<http://www.slideshare.net/AlmogRamrajkar/introduction-to-business-intelligence-33691369>  
[http://www.tdwi.eu/wissen/whitepaper/?no\\_cache=1&tx\\_mwknowledgebase\\_pi1%5BshowUid%5D=46](http://www.tdwi.eu/wissen/whitepaper/?no_cache=1&tx_mwknowledgebase_pi1%5BshowUid%5D=46)

<http://www.techonthenet.com/oracle/distinct.php>  
<http://www.xaprb.com/blog/2006/10/09/how-to-find-duplicate-rows-with-sql/>  
<http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/sabre/>  
[http://www-07.ibm.com/sg/events/blueprint/pdf/day1/Introduction\\_to\\_Business\\_Intelligence.pdf](http://www-07.ibm.com/sg/events/blueprint/pdf/day1/Introduction_to_Business_Intelligence.pdf)  
[https://docs.oracle.com/cd/B19306\\_01/server.102/b14200/functions172.htm](https://docs.oracle.com/cd/B19306_01/server.102/b14200/functions172.htm)  
[https://docs.oracle.com/cd/B19306\\_01/server.102/b14200/functions206.htm](https://docs.oracle.com/cd/B19306_01/server.102/b14200/functions206.htm)  
[https://docs.oracle.com/cd/B19306\\_01/server.102/b14215/et\\_concepts.htm](https://docs.oracle.com/cd/B19306_01/server.102/b14215/et_concepts.htm)  
<https://hbr.org/2006/01/competing-on-analytics>  
[https://msdn.microsoft.com/en-us/library/aa373931\(VS.85\).aspx](https://msdn.microsoft.com/en-us/library/aa373931(VS.85).aspx)  
[https://msdn.microsoft.com/en-us/library/ms711688\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/ms711688(v=vs.85).aspx)  
<https://msdn.microsoft.com/en-us/library/office/ff822584.aspx>  
[https://msdn.microsoft.com/en-us/library/x4k5wbx4\(v=vs.84\).aspx](https://msdn.microsoft.com/en-us/library/x4k5wbx4(v=vs.84).aspx)  
[https://www.axelos.com/Corporate/media/Files/Glossaries/ITIL\\_2011\\_Glossary\\_GB-v1-0.pdf](https://www.axelos.com/Corporate/media/Files/Glossaries/ITIL_2011_Glossary_GB-v1-0.pdf)  
<https://www.pcicomplianceguide.org/pci-faqs-2/#1>  
[https://www.researchgate.net/publication/228622273\\_Data\\_warehouse\\_architectures\\_factors\\_in\\_the\\_selection\\_decision\\_and\\_the\\_success\\_of\\_the\\_architectures](https://www.researchgate.net/publication/228622273_Data_warehouse_architectures_factors_in_the_selection_decision_and_the_success_of_the_architectures)  
<https://www.techonthenet.com/oracle/functions/count.php>

## ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Davenport T. 2006, "Competing on Analytics", January 2006 issue, (<https://hbr.org/2006/01/competing-on-analytics>).

[2] Power, D.J. 2007, "A Brief History of Decision Support Systems", DSSResources.COM, World Wide Web, version 4.0, March 10, 2007 (<http://DSSResources.COM/history/dsshhistory.html>).

[3] Ralph H. Sprague, Eric D. Carlson. 1982, "Building Effective Decision Support Systems".

[4] Alter, S.L. 1975, "A Study of Computer Aided Decision Making in Organizations", Ph.D. dissertation, M.I.T., 1975.

[5] Negash, S. 2004, "Communications of the Association for Information Systems: Business Intelligence".

[6] Gibson, M, Arnott, D, Jagielska, I, Melbourne A. 2004, "Evaluating the Intangible Benefits of Business Intelligence: Review & Research Agenda", Proceedings of the 2004 IFIP International Conference on Decision Support Systems.

[7] Azvine, B., Cui, Z., Nauck, D. & Majeed, B. 2006, "Real Time Business Intelligence for the Adaptive Enterprise", Proceedings of the 8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing, San Francisco, California, USA, 2006.

[8] Ballard, C., Farrell, D., M., Gupta, A., Mazuela, C., Vohnik, S. 2006, "Dimensional Modeling: In a Business Intelligence Environment".

[9] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H. 2011, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, ([http://www.mckinsey.com/Insights/MGI/Research/Technology\\_and\\_Innovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_Innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_Innovation)).

[10] Chen, H., Chiang, H., L., R., Storey, V., C. 2012, "Business Intelligence and Analytics: From Big Data to Big Impact", MIS Quarterly Vol.36 No.4, December 2012.

[11] Sallam, R., L., Richardson, J., L., Hagerty, J., Hostmann, B. 2011, "Magic Quadrant for Business Intelligence Platforms", Gartner report published: 27 January 2011.

[12] O'Reilly, T. 2005. "What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software," September 30, (<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>).

[13] Bitterer, A. 2011, "Hype Cycle for Business Intelligence", Gartner report published: 12 August 2011.

[14] Chen, H. 2011. "Design Science, Grand Challenges, and Societal Impacts," ACM Transactions on Management Information Systems.

[15] Zikopoulos, C., P., Eaton, C., deRoos, D., Deutsch, T., Lapis, G. 2012, "Understanding Big Data. Analytics for Enterprise class Hadoop and streaming data".

[16] Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B. 2014, "The current state of Business Intelligence in Academia: The arrival of Big Data".

[17] Sicular S. 2013, “The Illusions of Big Data” Gartner article published: 24 July, 2013, (<http://blogs.gartner.com/svetlana-sicular/the-illusions-of-big-data/>).

[18] Wieder, B., Ossimitz, M. 2015, “Impact of Business Intelligence on the quality of decision making - a mediation model”.

[19] Sommer, D., Sood, B. 2014, “Business Intelligence and Analytics Software, Worldwide 2013”, Gartner Research, Market Share Analysis published: April 28, 2014.

[20] Herrmann, KR. 2001, “Visualizing your business: Let graphics tell the story”, published by John Wiley & Sons 2001.

[21] Kowalczyk, M., Buxmann, P., Besier, J. 2013, “Investigating Business Intelligence and Analytics from a Decision Process perspective: A Structured Literature Review”, ECIS 2013 Completed Research. Paper 126, ([http://aisel.aisnet.org/ecis2013\\_cr/126](http://aisel.aisnet.org/ecis2013_cr/126)).

[22] Raden, N. 2004, “Analysis of the Features, Benefits and Costs of “Best of Breed” Integration vs. a Single Architecture”.

[23] Barney, JB. 1991, “Firm Resources and Sustained Competitive Advantage”, Journal of Management March 1991 Vol.17, No.1 99-120.

[24] Davenport, T., H., Harris, G., J., Morison, R. 2010, “ANALYTICS AT WORK: Smarter Decisions. Better Results”.

[25] Rajteric, I., H. 2010, “Overview of Business Intelligence Maturity Models”, Management: Journal of Contemporary Management, Vol.15, Issue June 2010.

[26] Humphrey, W.,S. 1989, Managing the Software Process”.

[27] Williams, S., Williams, N. 2007, “The profit impact of business intelligence”.

[28] Eric A., M. 2008, "Service-oriented architecture governance for the services driven enterprise".

[29] Valipour H., M., Amirzafari, B., Maleki, N., Kh., Valadbeigi, M., Daneshpour, N. 2009, "Concepts of Service Orientation in Software Engineering: A Brief Survey," MASAUM Journal Of Reviews and Surveys, Vol.1, No.3, November 2009. (extension of a previous paper presented in ICCSIT 2009).

[30] Yeoh, W., Popovic, A. 2015, "Extending the Understanding of Critical Success Factors for implementing Business Intelligence systems", Journal of the Association for Information Science and Technology, Vol.67, Issue 1 January 2016.

[31] Wixom, B., H., Watson, H., J., Marjanovic, O., Ariyachandra, T. 2010, "Educating the Next-Generation BI Workforce", Business Intelligence Journal, Vol.15, No.3, 2010.

[32] Daniel, D., R. 1961, "Management Information Crisis," Harvard Business Review, Sept.-Oct.1961.

[33] DeLone W., H., McLean, E., R. 2003, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update", Journal of Management Information Systems, Vol.19, Issue 4, No.4, 2003.

[34] Jamaludin, I., A., Mansor, Z. 2011, "Review on Business Intelligence (BI) Success Determinants in Project Implementation", International Journal of Computer Applications, Vol.33, No.8, November 2011.

[35] Tornatzky, L.G., Fleischer, M. 1990, "The Processes of Technological Innovation". Lexington Books, Lexington, Massachusetts, 1990.

[36] Angeles, R. 2013, “Using the Technology-Organization-Environment framework and Zuboff’s concepts for understanding environmental sustainability and RFID: Two case studies”. International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, Vol.7, No.11, 2013.

[37] Ranjan, J. 2009, “Business Intelligence: Concepts, Components, Techniques and Benefits”, Journal of Theoretical and Applied Information Technology (<http://www.jatit.org/volumes/research-papers/Vol9No1/9Vol9No1.pdf>)

[38] Zins, C. 2007, “Conceptual Approaches for Defining Data, Information, and Knowledge”, Journal of the American society for Information Science and Technology, 58(4):479–493, 2007.

[39] Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., Wolfson, P. 2008, “Enterprise Master Data Management, An SOA Approach to Managing Core Information” by International Business Machines Corporation.

[40] Hoyer, V., Stanoevska-Slabeva, K. 2008, “Design Principles of Enterprise Mashups”, SAP Global Network.

[41] Arasu A., Garcia-Molina, H. 2003, “Extracting Structured Data from Web Pages”, Proceedings of the 2003 ACM SIGMOD international conference on Management of data.

[42] Sint, R., Schaffert, S., Stroka, S., Ferstl, R. 2009, “Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis”, Proceedings of the 4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009), Hersonissos, Greece, June 1st, 2009.

[43] Blumberg, R., Atre, S. 2003, “The Problem with Unstructured Data”. Available at (<http://www.information-management.com/issues/20030201/6287-1.html>).



[44] Abiteboul, S., Buneman, P., Suci, D. 1999, “Data on the Web: From Relations to Semi structured Data and XML”, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

[45] Kimball, R., Caserta, J. 2004, “The Data Warehouse ETLToolkit”, Wiley Publishing Inc., Indianapolis, USA.

[46] Bargmeyer, B., E., Gillman, D., W. 2000 “Metadata standards and metadata registries: An overview”, In International Conference on Establishment Surveys II, 2000.

[47] Walker, M. 2003, “Batch vs. Real Time Data Processing”, publication of the Data Science Portal, available at (<http://www.datasciencecentral.com/profiles/blogs/batch-vs-real-time-data-processing>).

[48] Vassiliadis, P., Simitsis, A., Skiadopoulou, S. 2002, “Conceptual Modeling for ETL Processes”, Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP.

[49] Inmon, W. 2002, “Building the Data Warehouse”. Wiley, 2002.

[50] Akkaoui, Z., Zimanyi, E., Mazón, J., N., Trujillo, J. 2011, “A model-driven framework for ETL process development”, Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP, 2011.

[51] Rodic, J., Baranovic, M. 2009, “Generating data quality rules and integration into ETL process”, Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP.

[52] Inmon, W. 1999, “Data Mart does not equal Data Warehouse”, Information Management magazine article found at (<http://www.information-management.com/infodirect/19991120/1675-1.html>).

[53] Watson, H., J., Ariyachandra, T. 2010, “Key organizational factors in data warehouse architecture selection”, published in Journal Decision Support Systems archive, Vol.49, Issue 2, May, 2010.

[54] Singh, J. 2010, “Promulgation of contemporary testing techniques for effective engineering of data warehouses”, doctorate thesis, faculty of Engineering and Technology Punjabi University, India July, 2010.

[55] Han, J., Kamber, M., Pei, J. 2012, “Data Mining concepts and techniques”, 3rd Edition, published by Elsevier India.

[56] Berry, M., J.A., Linoff, G., S. 2011, “Data Mining Techniques for Marketing, Sales and Customer Relationship Management”, 3rd Edition, published by Wiley.

[57] Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F., Pirahesh, H. 1996, “Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals”, published in journal Data Mining and Knowledge Discovery, Vol.1, Issue 1, 1997.

[58] Broeck, J., Cunningham, S., A., Eeckels, R., Herbst, K. 2005, “Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities”, published: September 6, 2005 (<http://dx.doi.org/10.1371/journal.pmed.0020267>).