**Πληροφοριακα Συστήματα 20/11 8ο**

**Επιχειρησιακή Ευφυία: Μεγάλα δεδομένα και Αναλυτική**

**(Μεγάλα) Δεδομένα**

**Είδη Δεδομένων**

Δομημένα Δεδομένα

* Τα δεδομένα μπορούν να παρουσιαστούν σε διάφορα είδη δομών, μορφών και μέσων.
* Ο όρος δομημένα δεδομένα αναφέρεται σε δεδομένα με υψηλό επίπεδο οργάνωσης, όπως στις σχεσιακές βάσεις δεδομένων
* Εύκολα προσβάσιμα, αποθηκεύσιμα και εύκολο να γίνουν ερωτήματα και να αναλυθούν (π.χ. με χρήση SQL)

Αδόμητα Δεδομένα

* Ο όρος περιλαμβάνει όλα τα είδη δεδομένων που δεν μπορούν να χαρακτηριστούν ή να ταιρίαξουν σε ένα απλό μοντέλο.
* π.χ. φωτο, βίντεο, Blods, wikis, posts, pdf, ppt. Τα tweets?

Μεταδεδομένα

* Δεδομένα για τα δεδομένα - Βοηθούν στην εύρεση και την ανάκτηση
* Ξεκίνησαν από τις βιβλιοθήκες

**Οργάνωση δεδομένων**

Δομημένα Δεδομένα

* Πίνακες δεδομένων→λογιστικά φύλλα, π.χ. OpenCalc, Excel
* Σχεσιακές βάσεις δεδομένων→mySQL, MS Access, Oracle
* Αποθήκες δεδομένων→Cognos by IBM, Oracle Database OLAP Option

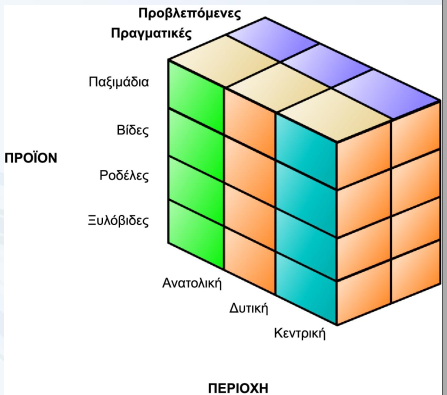
Αδόμητα Δεδομένα

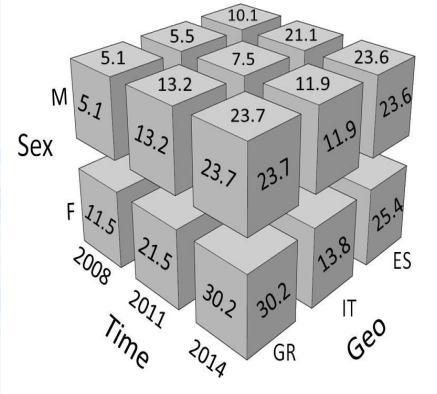
* Μη σχεσιακές βάσεις δεδομένων(NoSQL βάσεις )→mongoDB

Μεγάλα δεδομένα (big data)

**Σχεσιακές ΒΔ και ΠΣ**

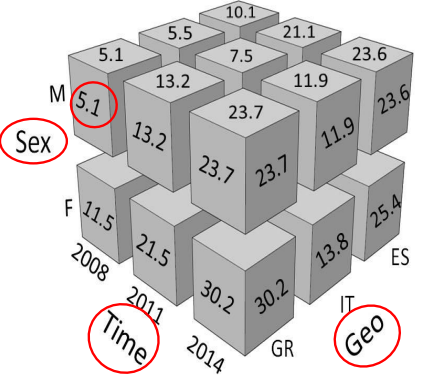
* Τα ΠΣ διαχειρίζονται δεδομένα, συνεπώς τα δεδομένα αυτά πρέπει να αποθηκεύονται κάπου
  + Άρα τα ΠΣ εξ ορισμού χρησιμοποιούν ΒΔ
* Τα περισσότερα ΠΣ χρησιμοποιούν σχεσιακές ΒΔ για την αποθήκευση των δεδομένων τους
  + Π.χ. το WordPress χρησιμοποιεί το ΣΔΒΔ mySQL, μια ΒΔ ανοικτού λογισμικού
  + Π.χ. το Salesforce χρησιμοποιεί Oracle, μια εμπορική ΒΔ που είναι ιδιαίτερα ισχυρή (και ακριβή)
  + Π.χ. το SAP S/4 χρησιμοποιεί HANA, μια ΒΔ που δημιούργησε η ίδια η SAP ώστε να καλύψευ καλύτερα τις ανάγκες της

**Αποθήκη δεδομένων**

* Βοηθά στη δημιουργία και την εργασία με Επιχ. Ευφυία
* Πολυδιάσταση
  + Γραμμές (rows), στήλες (columns), και επίπεδα (layers)
* Υποστηρίζει τη λήψη αποφάσεων και όχι την επεξεργασία συναλλαγών (transaction processing)
  + Περιέχει περιλήψεις των πληροφοριών
  + Όχι κάθε λεπτομέρεια

**Μοντελοποίηση Πολυδιάστατων Δδεομένων**

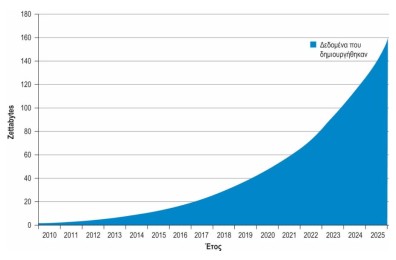
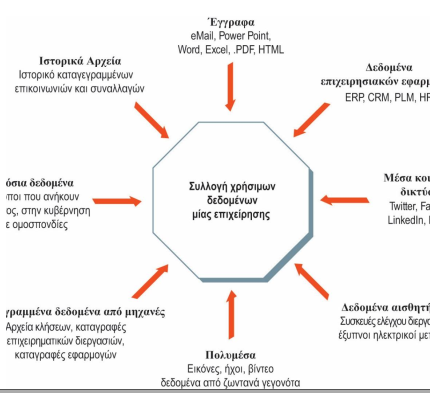
Συχνά οργανώνονται ως κύβοι όπου κάθε κελί περιέχει μια μέτρηση (measure) που περιγράφεται από ένα αριθμό από διαστάσεις (dimensions), οι οποίες οργανώνονται σε ιεραρχίες (hierarchies)

**Παράδειγμα Data Cube**

* Measure: unemployment rate
* Unit of measurement: %
* Value for (G=Greece, T=2008, S-Males) = 5.1

**Μεγάλα Δεδομένα**

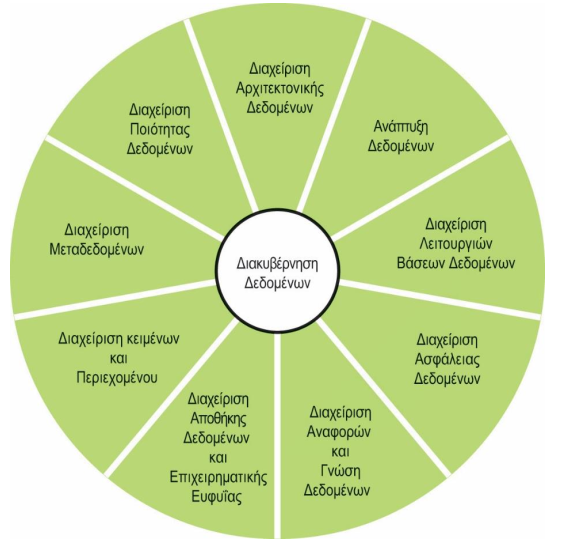
* Τεράστια (terabytes ή περισσότερο) Πολύπλοκα (από δεδομένα αισθητήρων ως και δεδομένα που προέρχονται από μέσα κοινωνικής δικτύωσης)
* Τα παραδοσιακά συστήματα διαχείρισης λογισμικού και υλικού και οι παραδοσιακές διαδικασίες ανάλυσης δεν έχουν τη δυνατότητα να τα διαχειριστούν
* Βασικά χαρακτηριστικά (3V, 4V, 5V)
  + Όγκος (Volume)
  + Ταχύτητα (Velocity)
  + Αξία (Value)
  + Ποικιλία (Variety)
  + Εγκυρότητα (Validity)
* Οι επιχειρήσεις χρησιμοποιούν τα μεγάλα δεδομένα:
  + Για καθημερινές λειτουργίες
  + Για σχεδιασμό
  + Για λήψη αποφάσεων

**Μεγάλα Δεδομένα και Πηγές Δεδομένων**

**Διαχείριση δεδομένων**

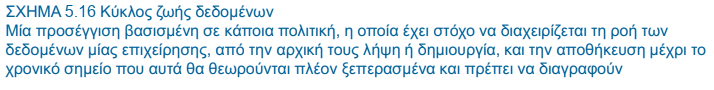
* **Διαχείριση δεδομένων (data management)**
  + Ένα ολοκληρωμένο σύνολο λειτουργιών,
  + Προσδιορίζει τις διαδικασίες με τις οποίες λαμβάνονται τα δεδομένα, πιστοποιούνται ως κατάλληλα για χρήση, αποθηκεύονται, ασφαλίζονται και υφίστανται επεξεργασία
  + Εξασφαλίζεται ότι η προσβασιμότητα, η αξιοπιστία και η επικαιρότητά τους ανταποκρίνονται στις ανάγκες των χρηστών
* **Διακυβέρνηση δεδομένων (data governance)**
  + Ορίζει τους ρόλους, τις ευθύνες και τις διαδικασίες
    - Εξασφαλίζει ότι τα δεδομένα είναι αξιόπιστα και μπορούν να χρησιμοποιηθούν από ολόκληρη την επιχείρηση
    - Προσδιορίζει τα άτομα που είναι υπεύθυνα για την επιδιόρθωση και αποτροπή προβλημάτων που μπορεί να εμφανιστούν σε αυτά

**Διακυβέρνηση δεδομένων**



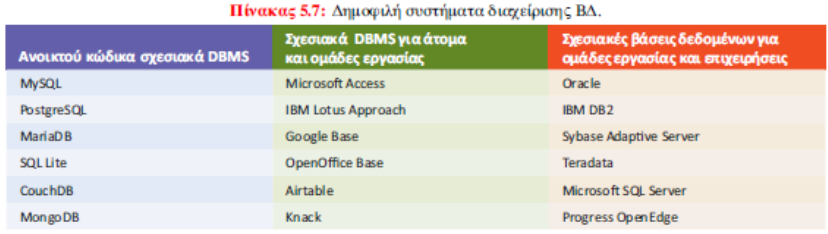
**Κύκλος ζωής δεδομένων**

* Διαχειριστής ΒΔ (DBA)
* Επιστάτης δεδομένων
* Διαχείριση του κύκλου ζωής των δεδομένων (DLM)



**Τεχνολογίες για (Μεγάλα) Δεδομένα**

**Δημοφιλή Συστήματα Διαχείρισης Σχεσιακών ΒΔ**

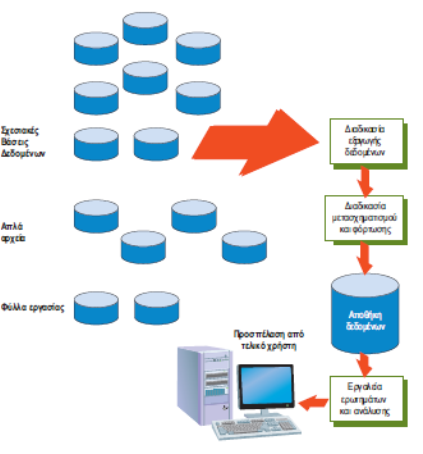


**ΒΔ ως Υπηρεσία (DAAS)**

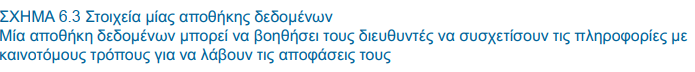
* H βάση δεδομένων αποθηκεύεται στους εξυπηρέτες του παρόχου και προσπελάζεται από τον συνδρομητή μέσω διαδικτύου
* O πάροχος της υπηρεσίας αναλαμβάνει τη διαχείριση της βάσης
* Πλεονέκτημα της DaaS
  + Εξαλείφει τις διαδικασίες εγκατάστασης, συντήρησης και  
    παρακολούθησης των βάσεων δεδομένων

**Τεχνολογίες που Χρησιμοποιούνται για τη Διαχείριση και Επεξεργασία των Μεγάλων Δεδομένων**

* Αποθήκες δεδομένων
* Πρατήρια δεδομένων
* Λίμνες δεδομένων
* ΒΔ NoSQL
* Hadoop
* Βάσεις Δεδομένων με Αποθήκευση Εντός Μνήμης

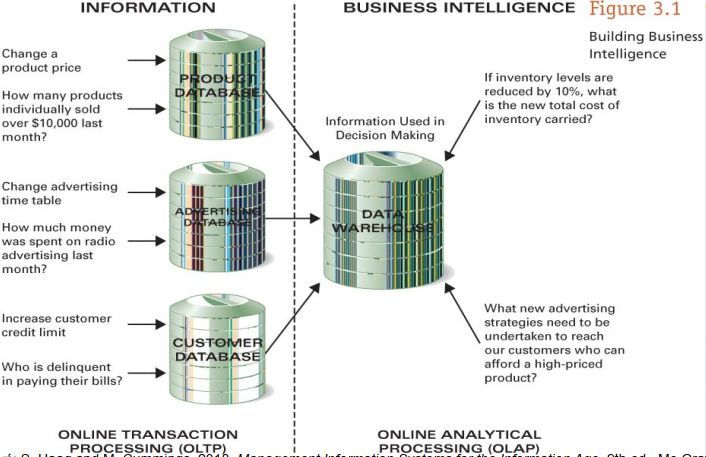
**Αποθήκες Δεδομένων**

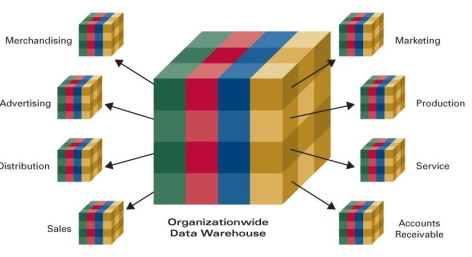
* Μέθοδος ETL για τη δημιουργία Αποθηκών:
  + Extract Information
  + Transform Information
  + Load Information



**OLTP και OLAP**

* **Online transaction processing (OLTP):** συγκέντρωση και επεξεργασία πληροφοριών καθώς και επικαιροποίηση υπαρχουσών πληροφοριών ώστε να αντανακλούν τις επεξεργασμένες πληροφορίες
  + Υποστηρίζει τακτική επεξεργασία π.χ. καθημερινές εργασίες
  + Τα δεδομένα σχετίζονται με συναλλαγές ρουτίνας π.χ. πωλήσεις, παραγγελίες κλπ.
* **Online analytical processing (OLAP):** χειρισμός των πληροφοριών για την υποστήριξη της λήψης αποφάσεων
  + Βοηθά στη δημιουργία επιχειρηματικής ευφυΐας
  + Υποστηρίζεται από αποθήκες δεδομένων (data warehouses) και εργαλεία εξόρυξης δεδομένων (data-mining)

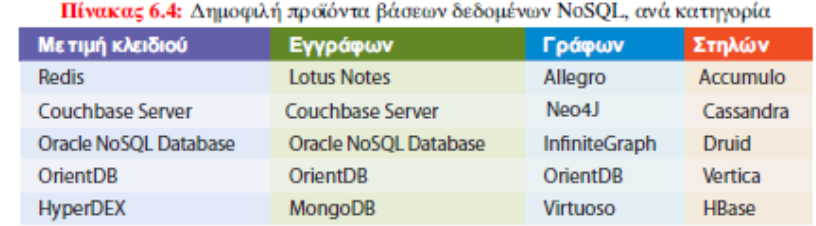
**OLTP, OLAP και ΕΕ**

**Πρατήρια και Λίμνες Δεδομένων**

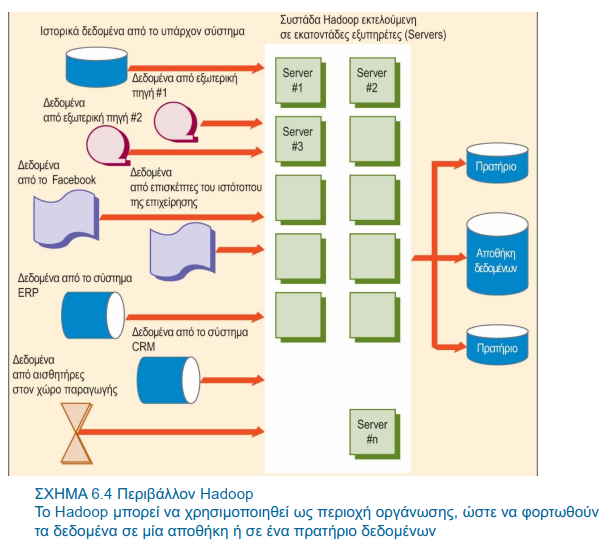
* + - **Πρατήριο Δεδομένων**
      * Ένα υποσύνολο μίας αποθήκης δεδομένων.
      * Χρησιμοποιούνται σε μικρές και μεσαίες επιχειρήσεις και σε τμήματα μεγάλων επιχειρήσεων
      * Υποστηρίζουν τη λήψη αποφάσεων
    - **Λίμνη δεδομένων**
      * Ακολουθεί μία προσέγγιση «καθολικής αποθήκευσης» των μεγάλων δεδομένων
      * Αποθηκεύει όλα τα δεδομένα σε ακατέργαστη μορφή

**Βάσεις Δεδομένων NoSQL**

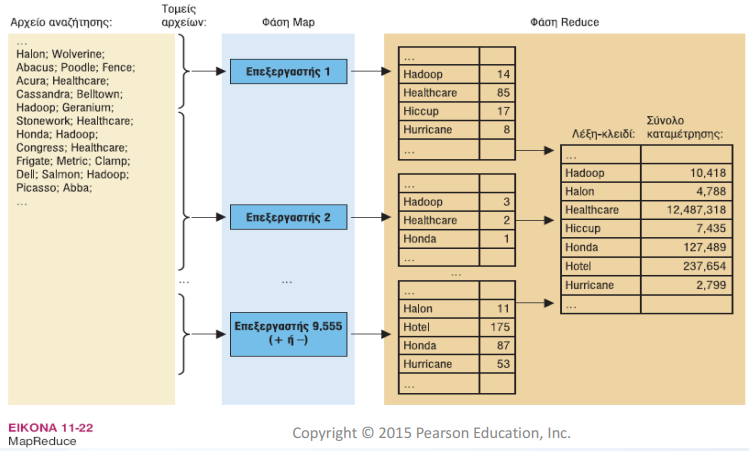
* NoSQL database
  + Διαφέρει από μία σχεσιακή βάση δεδομένων
    - Παρέχει έναν τρόπο αποθήκευσης και ανάκτησης των δεδομένων ο οποίος μοντελοποιείται με κάποιον τρόπο διαφορετικό από τις απλές σχέσεις δισδιάστατων πινάκων που χρησιμοποιούν οι σχεσιακές βάσεις δεδομένων
    - Χρησιμοποιεί οριζόντια κλιμάκωση
    - Δεν απαιτεί προκαθορισμένο σχήμα
* Δομές που χρησιμοποιούν οι ΒΔ NoSQL
  + Πιο ευέλικτες από τις αντίστοιχες των κλασικών ΒΔ
  + Βελτιώνουν την ταχύτητα προσπέλασης και τον πλεονασμό
* **Τέσσερις κατηγορίες**
  + Με τιμή κλειδιού
    - Δύο στήλες (“κλειδί” and “τιμή”)
* Εγγράφων
  + χρησιμοποιούνται για την αποθήκευση, ανάκτηση και διαχείριση πληροφοριών εγγράφων
* Γράφων
  + Για την ανάλυση διασυνδέσεων
* Στηλών
  + Αποθήκευση δεδομένων κατά στήλες



**Hadoop**

* Εφαρμογή ανοικτού κώδικα, η οποία συμπεριλαμβάνει αρκετά προγράμματα που παρέχουν έναν τρόπο αποθήκευσης και επεξεργασίας υπερβολικά μεγάλων συνόλων δεδομένων
* Κατανεμημένο σύστημα αρχείων (HDFS)
  + Κατανεμημένο σύστημα αρχείων
  + Για αποθήκευση δεδομένων
  + Χωρίζει δεδομένα σε υποσύνολα
  + Διανέμει τα υποσύνολα σε διαφορετικούς
  + εξυπηρετητές για να γίνει η επεξεργασία τους
    - **Πρόγραμμα MapReduce**
      * Σύνθετο πρόγραμμα
      * Δύο στοιχεία  
        • διαδικασία Map, η οποία φιλτράρει και ταξινομεί τα δεδομένα (για παράδειγμα ταξινόμηση  
        • μέθοδος Reduce, η οποία εκτελεί μία λειτουργία σύνοψης
* Περιορισμοί Hadoop
  + Εκτελεί μόνο επεξεργασία κατά δέσμες

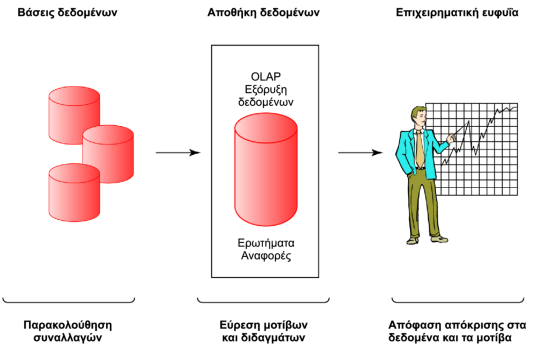
**Τεχνική MapReduce**



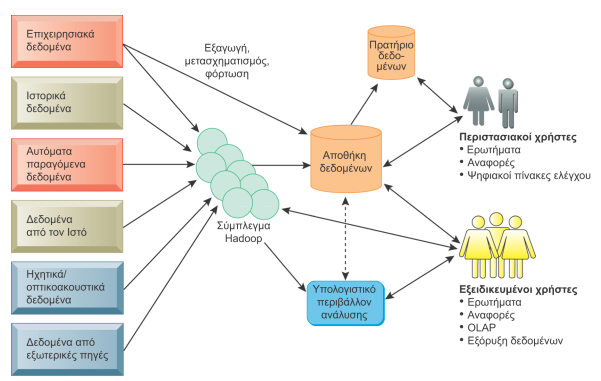
**Βάσεις Δεδομένων με Αποθήκευση Εντός Μνήμης**

* + - Αποθηκεύει ολόκληρη τη βάση στη (RAM)
    - Ταχύτερη προσπέλαση δεδομένων
    - Επιτρέπει την ανάλυση των μεγάλων δεδομένων και των δεδομένων από άλλες εφαρμογές που είναι δύσκολα στην επεξεργασία
    -  Εφικτότητα λόγω Αύξησης χωρητικότητας/ μείωση κόστους της RAM

**Επιχειρηματική Ευφυία**

* Διεργασία απόκτησης, ανάλυσης και δημοσίευσης δεδομένων, με σκοπό την ανάδειξη προτύπων δεδομένων που θα λειτουργούν ως πηγή πληροφόρησης για τα στελέχη των επιχειρήσεων

Μια σειρά από αναλυτικά   
εργαλεία που  
εφαρμόζονται σε  
δεδομένα αποθηκευμένα  
σε βάσεις δεδομένων για  
να εντοπίσουν μοτίβα και  
διδάγματα που βοηθούν  
τους μάνατζερ και τους  
εργαζόμενους να  
παίρνουν καλύτερες  
αποφάσεις για τη  
βελτίωση της απόδοσης  
του οργανισμού.

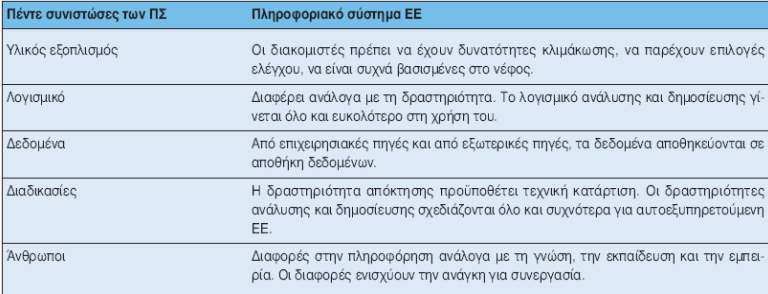
**Σύγχρονη υποδομή Επιχειρηματικής Ευφυίας**

Η σύγχρονη υποδομή επιχειρηματικής ευφυΐας περιλαμβάνει δυνατότητες και εργαλεία για τη διαχείριση και ανάλυση μεγάλου όγκου και  
διαφορετικών τύπων δεδομένων από πολλές πηγές. Έχει επίσης εύχρηστα εργαλεία υποβολής ερωτημάτων και δημιουργίας αναφορών τα οποία απευθύνονται στους λιγότερους εξειδικευμένους τελικούς χρήστες, αλλά και πιο εξεζητημένα πακέτα ανάλυσης για τους πιο πεπειραμένους χρήστες

**Τι είναι η επιχειρηματική ευφυΐα και γιατί είναι σημαντική για τις επιχειρήσεις;**

**Παραδείγματα ΕΕ**

**Κύριες συνιστώσες ενός συστήματος ΕΕ**



**Αναλυτική και Επιχειρηματική Ευφυία**

* Επιχειρηματική Ευφυία (BI)
  + Συμπεριλαμβάνει ένα μεγάλο εύρος εφαρμογών, πρακτικών και τεχνολογιών
  + Εξαγωγή, μετασχηματισμός, συνένωση, οπτικοποίηση, ανάλυση, ερμηνεία και παρουσίαση δεδομένων
  + Βελτίωση της λήψη αποφάσεων
* Αναλυτική
  + Εκτενής χρήση δεδομένων και ποσοτικής ανάλυσης
  + Υποστηρίζει τη λήψη δεδομένων βάσει γεγονότων μέσα σε μία επιχείρηση

**Οφέλη της ΕΕ και της Αναλυτικής**

**•** Ανίχνευση απάτης **•** Βελτίωση προβλέψεων  
**•** Αύξηση των Πωλήσεων **•** Βελτιστοποίηση λειτουργιών **•** Μείωση κόστους

**Ο Ρόλος του Επιστήμονα Δεδομένων**

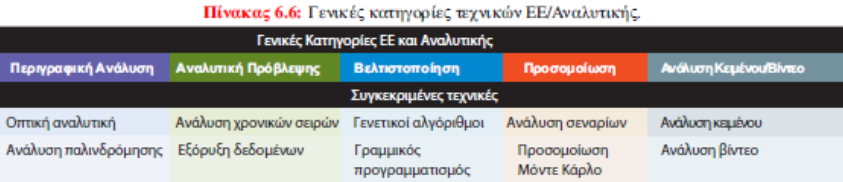
* + - **Επιστήμονας δεδομένων**
      * Συνδυάζει πολλές δεξιότητες  
        • Ισχυρό επιχειρηματικό δαιμόνιο  
        • Βαθιά κατανόηση της αναλυτικής  
        • Υγιή εκτίμηση των περιορισμών των δεδομένων
      * Βοηθά στη βελτίωση της λήψης αποφάσεων
      * Απαιτήσεις εκπαίδευσης
      * Σημαντικές προοπτικές εξέλιξης

**Στοιχεία που Απαιτούνται για Αποτελεσματική ΕΕ και Αναλυτική**

* + - Τρία βασικά στοιχεία
      * Ύπαρξη ενός εύρωστου προγράμματος διαχείρισης των δεδομένων  
        • Συμπεριλαμβάνεται η διακυβέρνηση
      * Δημιουργικοί επιστήμονες δεδομένων
      * Προσήλωση στη λήψη αποφάσεων βάσει των δεδομένων

**Εργαλεία Επιχειρηματικής Ευφυίας (Stair & Raynolds)**

**Εργαλεία ΕΕ και Αναλυτικής**



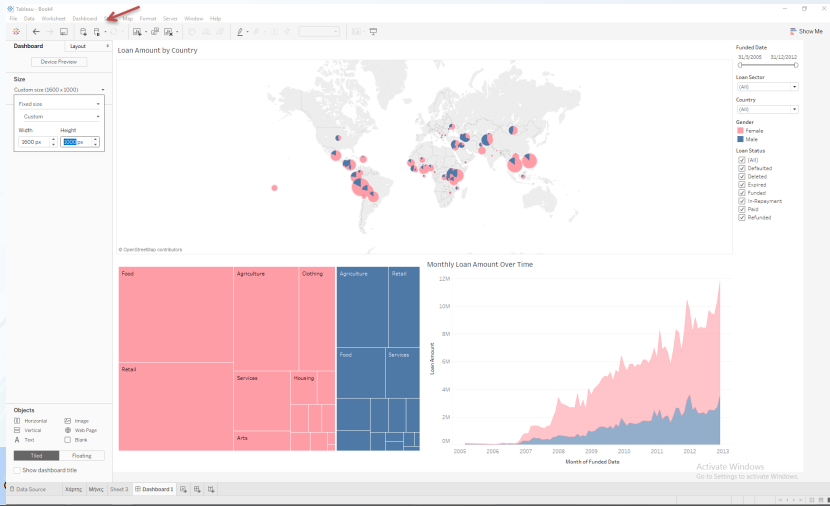
**Προσοχή:  
• Αυτή είναι η προσέγγιση των Stair & Reynolds  
• Δεν υπάρχει μια παγκοσμίως αποδεκτή προσέγγιση  
• Άλλοι υιοθετούν διαφορετικές προσεγγίσεις  
• Κάποιες θα δείτε σε επόμενη ενότητα!**

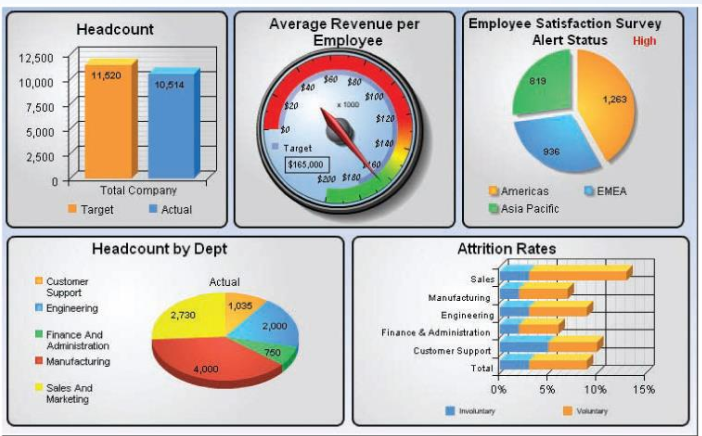
**Περιγραφική Ανάλυση**

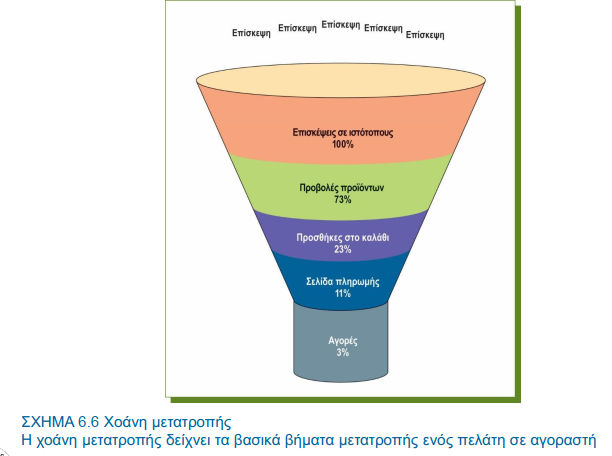
* Περιγραφική Ανάλυση
  + Βασικό στάδιο επεξεργασίας των δεδομένων
  + Αναγνώριση προτύπων στα δεδομένα
  + Απάντηση ερωτημάτων  
    • Ποιος, τι, πότε, πού και σε ποιον βαθμό
  + Δύο τύποι  
    • Οπτική ανάλυση  
    • Ανάλυση παλινδρόμησης
* **Οπτική ανάλυση**
  + Παρουσίαση των δεδομένων σε μορφή εικόνας ή γραφικών
  + **Λεξονέφος**  
    • Λεκτική απεικόνιση ενός συνόλου λέξεων  
    • Λέξεις που ομαδοποιούνται
    - Βάσει της συχνότητας εμφάνισής τους
* **Χοάνη μετατροπής**  
  • Γραφική αναπαράσταση  
  • Παράδειγμα: τα βήματα που ακολουθεί ένας καταναλωτής για να λάβει την απόφαση να αγοράσει ένα προϊόν και να γίνει πελάτης..

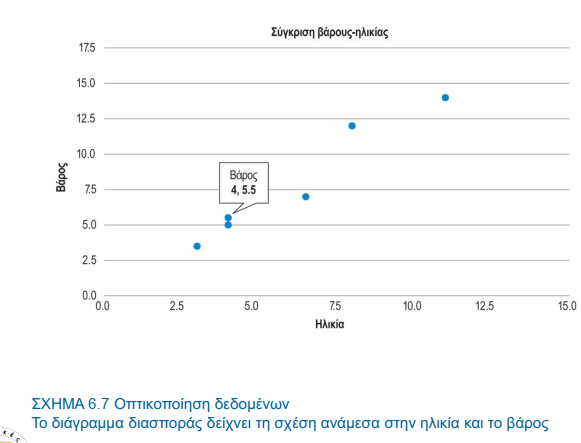
**Οπτικοποίηση Δεδομένων**

* Οι τεχνικές ανάλυσης δεδομένων μπορούν να μας δώσουν χρήσιμες τιμές, μπορεί όμως να χρειάζεται να χρησιμοποιηθεί η σωστή οπτικοποίηση για την ευκολότερη ερμηνεία
* Η οπτικοποίηση βοηθά στην κατανόηση των χαρακτηριστικών των δεδομένων και προσφέρει χρήσιμες πληροφορίες.
* Ανακάλυψη νέων φαινομένων
* Υπάρχει καλύτερη αίσθηση του τι μας προσφέρουν τα δεδομένα
* Η μέθοδος επικοινωνίας ανάμεσα σε αναλυτές δεδομένων, ατόμων που λαμβάνουν αποφάσεις, παρόχους υπηρεσιών κτλ...

**Παράδειγμα Οπτικοποίησης στο Tableau**

**Digital Dashboard**



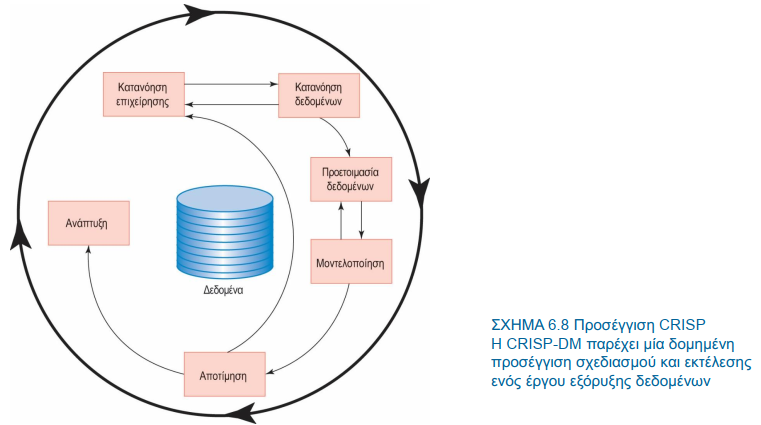


* **Ανάλυση παλινδρόμησης**
  + Προσδιορίζει τη σχέση ανάμεσα σε μία εξαρτημένη μεταβλητή και μία ή περισσότερες ανεξάρτητες
  + Παράγει μία εξίσωση παλινδρόμησης  
    • Οι συντελεστές αναπαριστούν τη σχέση ανάμεσα σε κάθε ανεξάρτητη μεταβλητή και την εξαρτημένη  
    • Χρήση για προβλέψεις

**Προβλεπτική Αναλυτική**

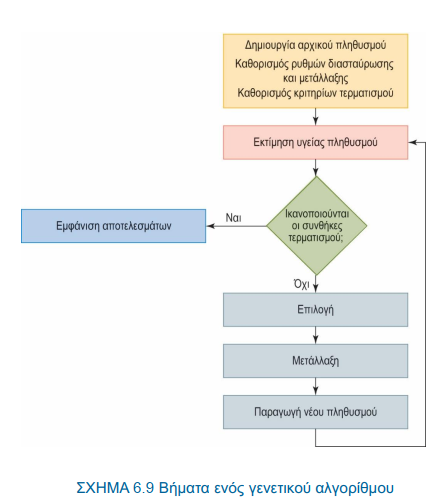
Σύνολο από τεχνικές που χρησιμοποιούνται για να αναλύονται τα τρέχοντα δεδομένα και να  
προσδιορίζονται μελλοντικές πιθανότητες και τάσεις, και για να προβλέπεται το μέλλο

* **Ανάλυση χρονοσειρών**
  + Η χρήση στατιστικών μεθόδων για την ανάλυση των δεδομένων χρονοσειρών και την εξαγωγή χρήσιμων στατιστικών και χαρακτηριστικών των δεδομένων
* **Εξόρυξη δεδομένων**
  + ένα εργαλείο ανάλυσης της ΕΕ
  + Χρησιμοποιείται για να εξερευνούμε μεγάλες ποσότητες δεδομένων για την εύρεση κρυμμένων μοτίβων
  + Πρόβλεψη μελλοντικών τάσεων και συμπεριφορών
  + Χρήση για τη λήψη αποφάσεων
  + Τρεις τεχνικές εξόρυξης  
    • Ανάλυση συσχέτισης  
    • Νευρωνική υπολογιστική  
    • Συλλογιστική βάσει περιπτώσεων



**Βελτιστοποίηση**

* + - Εκχώρηση πόρων
      * Μείωση κόστους ή αύξηση κερδών
    - Γενετικός Αλγόριθμος
      * Τεχνική που χρησιμοποιεί μία διαδικασία παρόμοια μετη φυσική επιλογή
* Βρίσκει προσεγγιστικές λύσεις σε προβλήματα βελτιστοποίησης και αναζήτηση Γραμμικός προγραμματισμός
  + Τεχνική εύρεσης της βέλτιστης τιμής (μεγαλύτερης ή μικρότερης
  + Υπολογισμός με βάση την τιμή ενός συνόλου μεταβλητών απόφασης, που υπόκεινται σε ένα σύνολο περιορισμών  
    • Οι μεταβλητές υπόκεινται σε ένα σύνολο περιορισμών



**Προσομοίωση**

* Εξομοιώνει τις δυναμικές αποκρίσεις ενός πραγματικού συστήματος σε διάφορες εισόδους
* **Ανάλυση σεναρίων**
  + Διαδικασία πρόβλεψης των μελλοντικών τιμών βάσει συγκεκριμένων πιθανών γεγονότων
* **Προσομοίωση Monte Carlo**
  + Επιτρέπει να διαπιστώσει κανείς ένα φάσμα από χιλιάδες πιθανά αποτελέσματα
  + Λαμβάνει υπόψη όχι μόνο τις πολλές μεταβλητές που εμπλέκονται
  + Λαμβάνει υπόψη το εύρος των πιθανών τιμών για καθεμία από αυτές τις μεταβλητές

**Ανάλυση κειμένου και βίντεο**

* + - Παρέχει γνώσεις και δεδομένα που συνδέονται με τη λήψη αποφάσεων
    - Ανάλυση κειμένου
      * Διαδικασία εξαγωγής τιμών από μεγάλες ποσότητες αδόμητου κειμένου
* Ανάλυση βίντεο
  + Διαδικασία λήψης πληροφοριών και γνώσης μέσα από αμοντάριστα πλάνα βίντεο

**Αναλυτική Αυτοεξυπηρέτησης**

* Συμπεριλαμβάνει την εκπαίδευση, τις τεχνικές και τις διαδικασίες που ενισχύουν τους τελικούς χρήστες, ώστε να υλοποιούν τις δικές τους αναλύσεις χρησιμοποιώντας ένα σύνολο εργαλείων
* Πλεονεκτήματα
  + Εξαλείφει τις καθυστερήσεις στη λήψη αποφάσεων
  + Επιταχύνει τη λήψη αποφάσεων
  + Παρέχει λύση όταν υπάρχει έλλειψη σε επιστήμονες δεδομένων

**Είδη Επιχειρηματικής Ευφυίας – άλλες προσεγγίσεις**

**Δραστηριότητες ΕΕ και ΠΣ**

* Απόκτηση
* Ανάλυση
* Δημοσίευση



**Δραστηριότητες ΕΕ και ΠΣ:** **Απόκτηση**

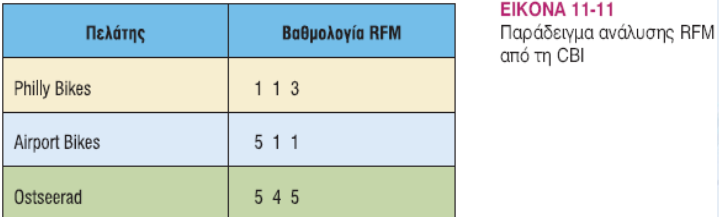
* Απόκτηση δεδομένων
* Καθορισμός δεδομένων
* Οργάνωση και συσχέτιση δεδομένων
* Δημιουργία δεδομένων

**Δραστηριότητες ΕΕ και ΠΣ: Ανάλυση**

* Αναφορά και εξόρυξη δεδομένων

**Δραστηριότητες ΕΕ και ΠΣ: Αναφορά**

* Μη διαδραστικη
  + Χρήση προκαθορισσμένης δομής εκροών (ταξινόμηση, φιλτάρισμα, ομαδοποίηση, υπολογισμός)
  + Ανάλυση RFM (recently- χρόνος, frequency- συχνότητα, money- χρήμα ): βαθμολογία 1-5(1: υψηλότερη βαθμολογία)



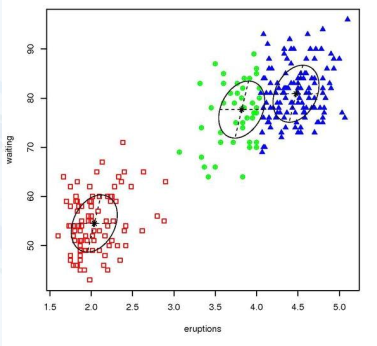
* **﻿**Διαδραστική
  + Οι χρήστες μπορούν να αλλάξουν τόσο την ανάλυση (π.χ. αλλαγή μέτρησης) όσο και τη δομή (π.χ. αλλαγή διάστασης) της αναφοράς
  + OLAP – άμεση αναλυτική επεξεργασία
    - Οριζόντιος και κάθετος τεμαχισμός OLAP (slice, dice)
    - Αναλυτική κάθοδος (drill-down, roll-up)

**Δραστηριότητες ΕΕ και ΠΣ: εξόρυξη Δεδομέων**

* Παλιδρόμηση ( προσοχή στη διαφορά από Stair & Raynolds)
* Ανάλυση καλαθιού αγορών
* Ανάλυση δεδομένων με επίβλεψη και χωρίς επίβλεψη
* Ανάλυση καλαθιού αγορών
  + Προσδιορισμός προτύπων πωλήσεων (πληροφορίες για προϊόντα που συνήθως αγοράζονται μαζί)
  + Π.χ. σε ένα σούπερ μάρκετ όσοι αγοράζουν ψωμί συνήθως αγοράζουν και γάλα κατά την ίδια επίσκεψη.
  + Η πληροφορία χρησιμεύει για προσφορές ή διαμόρφωση χώρου
* Επιβλεπόμενη και μη επιβλεπόμενη ανάλυση δεδομένων
* Ανάλυση με επίβλεψη σημαίνει ότι έχει δημιουργηθεί ένα μοντέλο ανάλυσης πριν από τη διενέργειά της.
* Εάν είναι άνευ επίβλεψης, τότε οι αναλυτές δεν δημιουργούν κάποιο μοντέλο ούτε προβαίνουν σε υποθέσεις πριν από τη διενέργεια της ανάλυσης.
  + Μοντέλα και υπόθεση

**Τύποι Πληροφοριών από την Εξόρυξη Δεδομέων (Laudon & Laudon)**

* **Συσχετίσεις ή εξαρτήσεις (dependencies):** Περιστατικά που συνδέονται με (ή εξαρτώνται από) μοναδικό γεγονός, π.χ. σε ένα supermarket όσοι αγοράζουν pizza αγοράζουν επίσης cola (χρήση σε εκπτωτικά κουπόνια)
* **Ακολουθίες (sequences):** Χρονική σύνδεση γεγονότων
* **Ταξινομήσεις (classification**): Μοτίβα που περιγράφουν την ομάδα στην οποία ανήκει ένα είδος (συνάγονται από ιστορικά δεδομένα - αλλά όχι μόνο!)
* **Ομαδοποιήσεις (clustering):** Ανακάλυψη αταξινόμητων ακόμη ομάδων που έχουν κάτι κοινό (βλ. βιβλίο Freakonomics)
* **Προβλέψεις (predictions):** Χρησιμοποιεί σειρά τιμών για την πρόβλεψη μελλοντικών τιμών

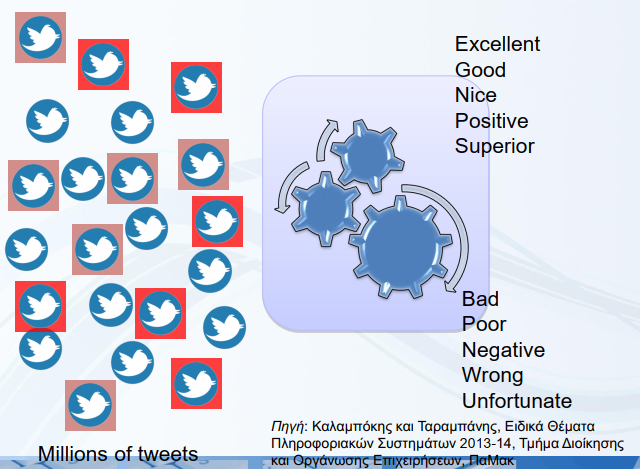
**Ομαδοποίηση(Clustering)**

* Η ομαδοποίηση ομάδων δεδομένων με τρόπο τον οποίο τα δεδομένα που ανήκουν στην ίδια ομάδα (cluster) είναι πιο παρόμοια (με κάποιον συγκεκριμένο τρόπο) από αντικείμενα που ανήκουν σε άλλες ομάδες
* Ο πιο γνωστός σχετικός αλγόριθμος είναι των k- μέσων (k-means)
* Βασίζεται στη μέτρηση της «απόστασης» κάθε σημείου από το μέσον κάθε ομάδας

**Ταξινόμηση (Classification)**



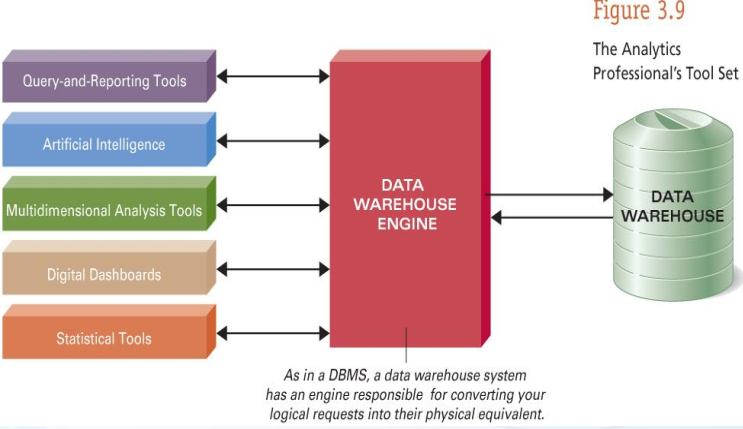
**Lexicon based classification (positive vs negative)**

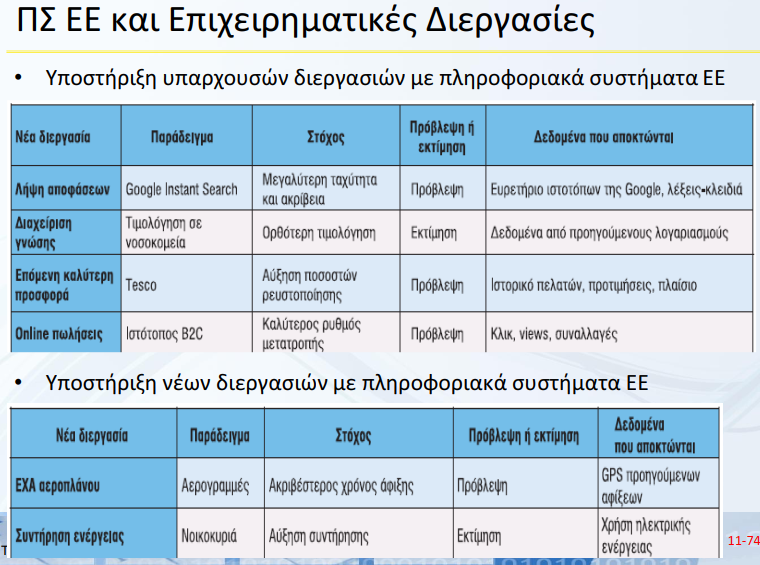


**Μηχανική Μαθηση**

* Ένα σύνολο δεδομένων δίνεται στον Η/Υ μαζί με τη «σωστή λύση» Το σύνολο αυτό ονομάζεται training set
* Στη συνέχεια δίνουμε τον Η/Υ ένα δεύτερο σύνολο δεδομένων χωρίς τη λύση
* Αυτό ονομάζεται test set
* Ο Η/Υ κάνει πρόβλεψη των λύσεων στο test set
* Το ποσοστό (%) των σωστών προβλέψεων καθορίζει την επιτυχία του αλγόριθμου
* Π.χ. Antivirus software

**ΕΕ (Haang and Xummings )**



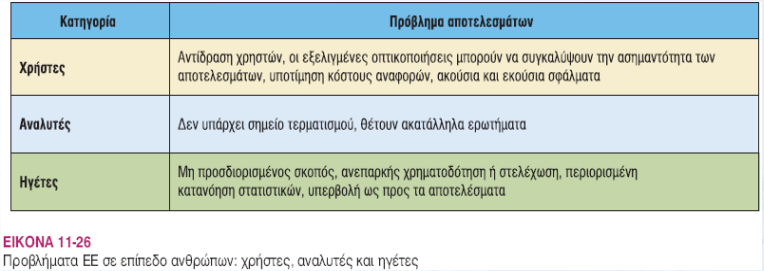


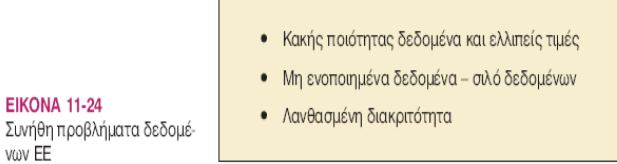
**Παράδειγμα χρήσης big data**

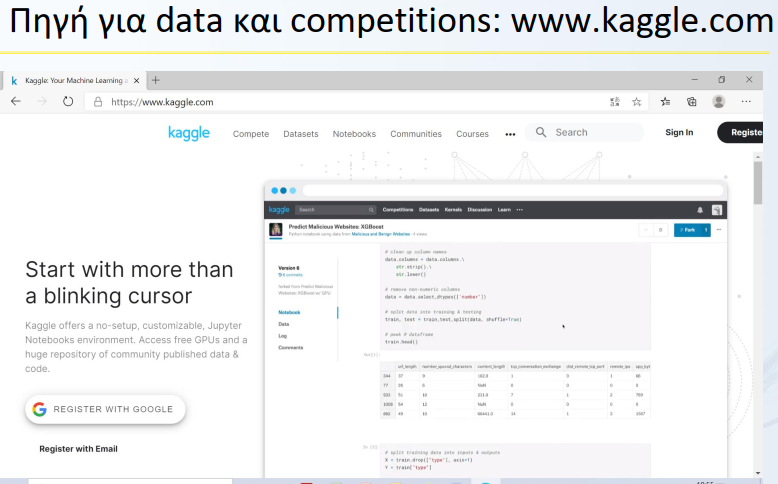


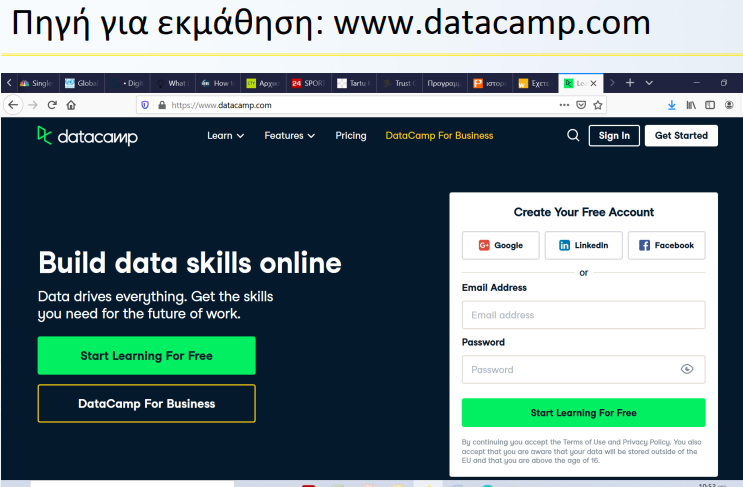
**Διαχείριση Κινδύνων ΕΕ**

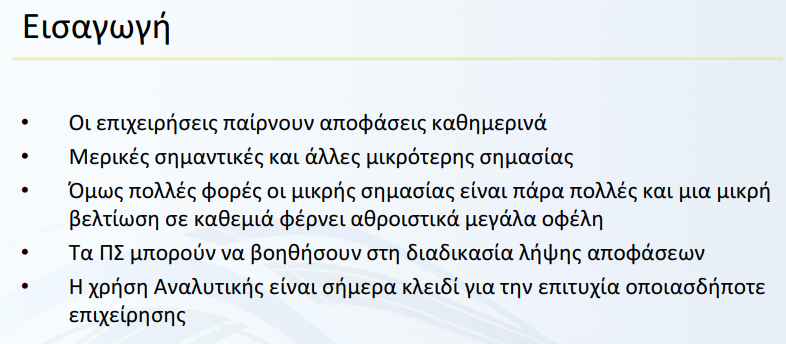
* Κίνδυνοι επιχειρηματικής ευφυϊας
  + Πρόβληματα δεδομένων
  + Προβλήματα ανθρώπων

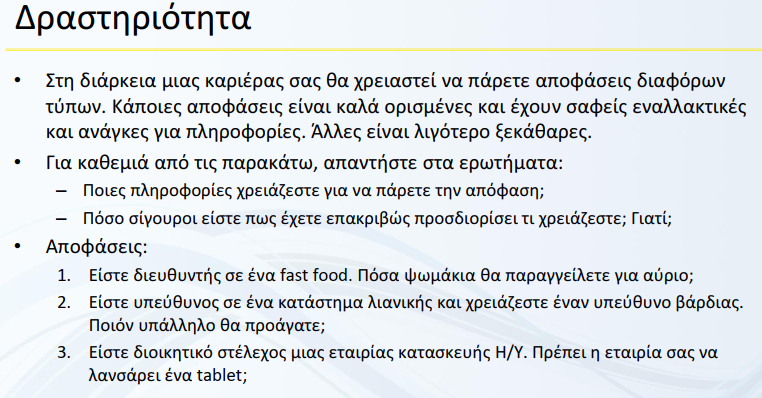


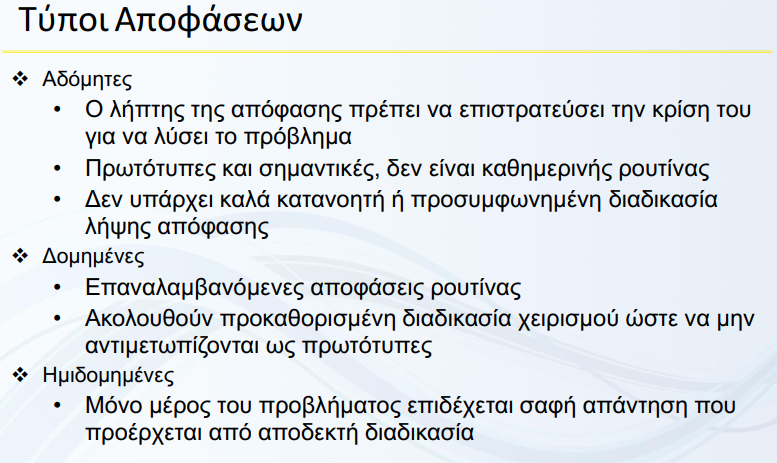
* + Προβήματα δεδομέων (80% χρόνου για προετοιμασία δεδομέων)





**ΠΣ Υποστήριξης Λήψης Αποφάσεων**





**Στάδια της Λήψης Αποφάσεων**

