

Big Data in Retailing

Μεγάλα Δεδομένα και Λιανεμπόριο

ΠΑΡΟΥΣΙΑΣΗ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Χρήστος Βασιλούδης
Μεταπτυχιακός Φοιτητής

Επιβλέπων Καθηγητής:
Κος Κωνσταντίνος Ψάννης

Συνεξεταστές: Κος Μαντάς Μιχαήλ &
Κος Ξυνόγαλος Στυλιανός

Τμήμα Εφαρμοσμένης Πληροφορικής

Πρόγραμμα Μεταπτυχιακών Σπουδών

Ειδίκευση: Επιστήμη και Τεχνολογία Η/Υ
(Computer Science and Technology)

Πανεπιστήμιο Μακεδονίας

Θεσσαλονίκη, Νοέμβριος 2021

Τι ακριβώς εννοούμε με τον όρο ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ.....?

➤ Τεράστια σύνολα δεδομένων

Η συνολική κίνηση δεδομένων μέσω κινητών εφαρμογών αναμένεται να φθάσει τα 30,6 exabytes(10^{18} bytes) μέχρι το 2021 (!!!)

➤ Ανομοιογενή μεταξύ τους

☐ Communication data : business, customers (bank transactions etc...)

☐ Internet data : video, images etc...

☐ Social Networks: Facebook, Instagram etc...

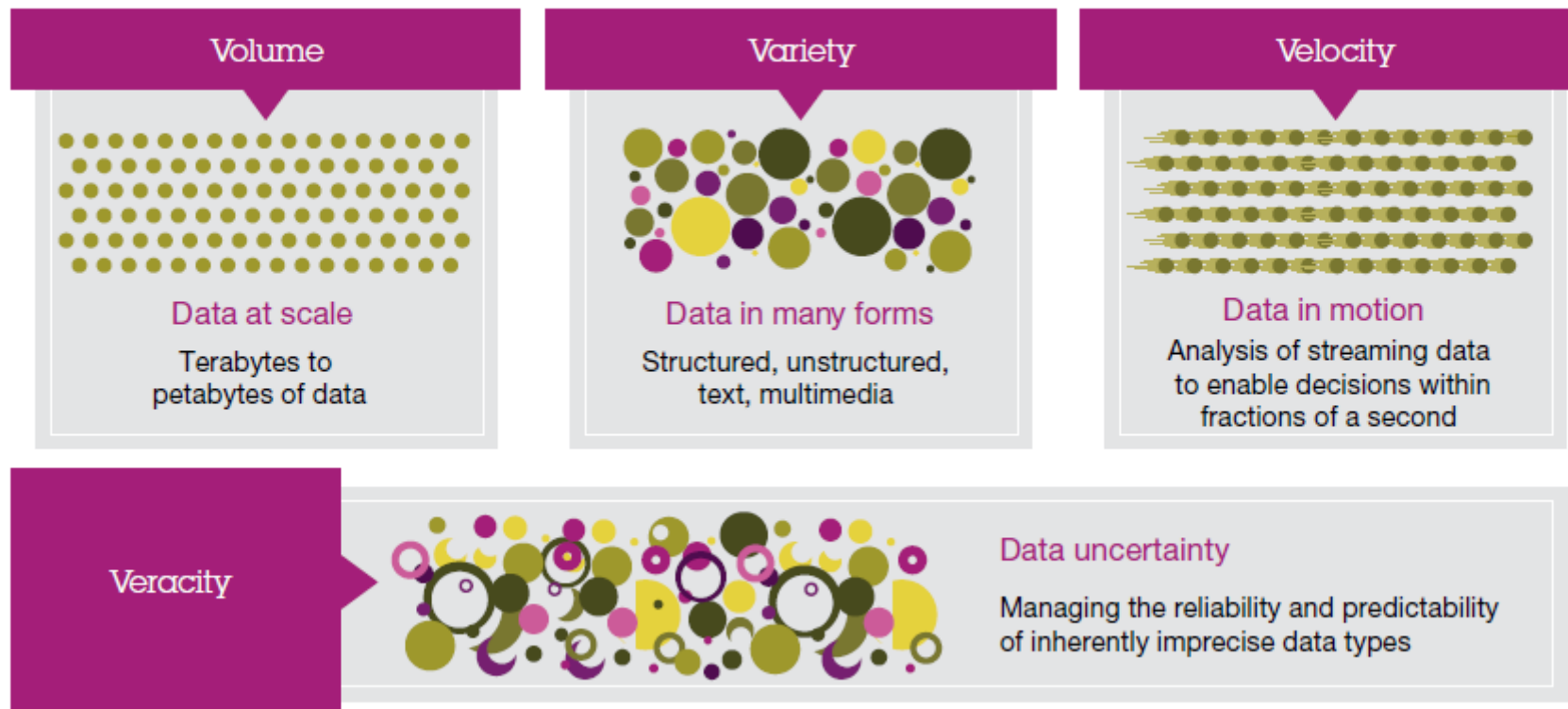
➤ Πολλές διαφορετικές πηγές

Τα χαρακτηριστικά των Μεγάλων Δεδομένων-3V

- Volume : Πλέον ή κίνηση είναι περίπου 35 zetta bytes (10^{21} bytes)!!!
- Variety (Ανομοιογένεια)
 - Δομημένα Δεδομένα (structured data)
 - Ημι ή Μη-Δομημένα Δεδομένα (semi or non-structured data)
- Velocity : Η ποιότητα των υπηρεσιών (QoS) θα πρέπει να καλύψει τις real-time απαιτήσεις των εφαρμογών.

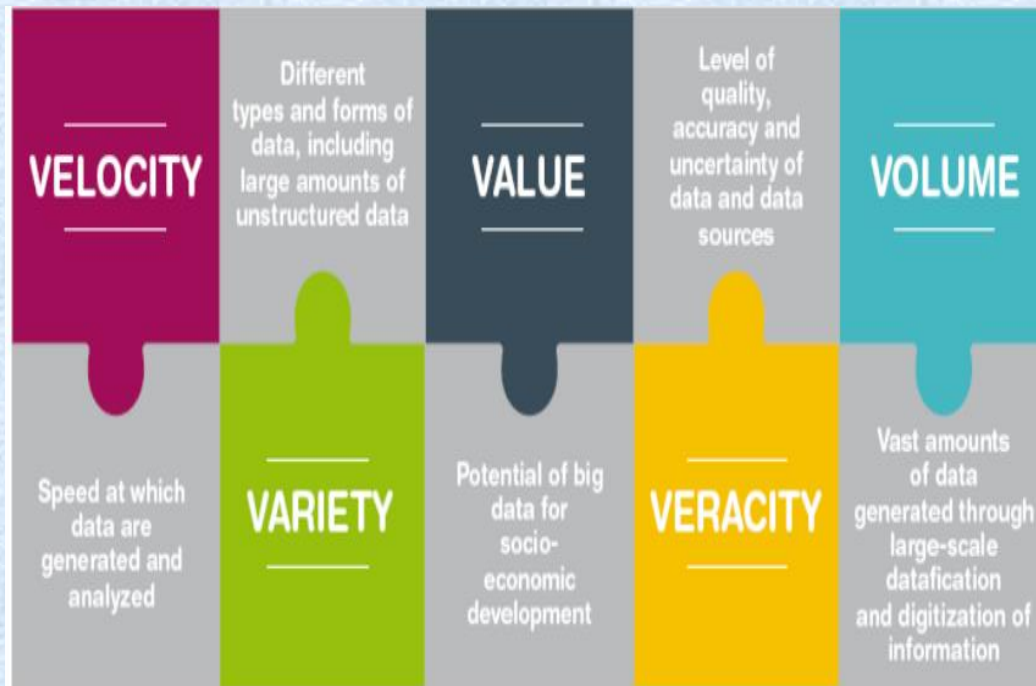
Πρόσθεση ενός ακόμη “V” (Veracity) – 4V

Big data in dimensions

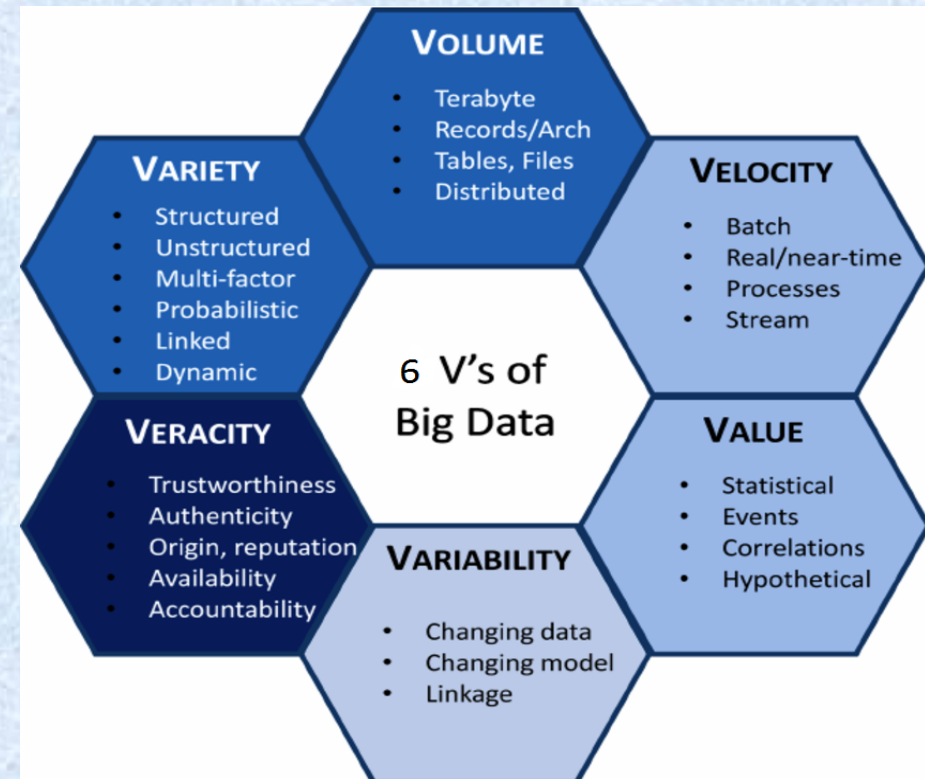


Επέκταση του μοντέλου σε 5V & 6V

• Value

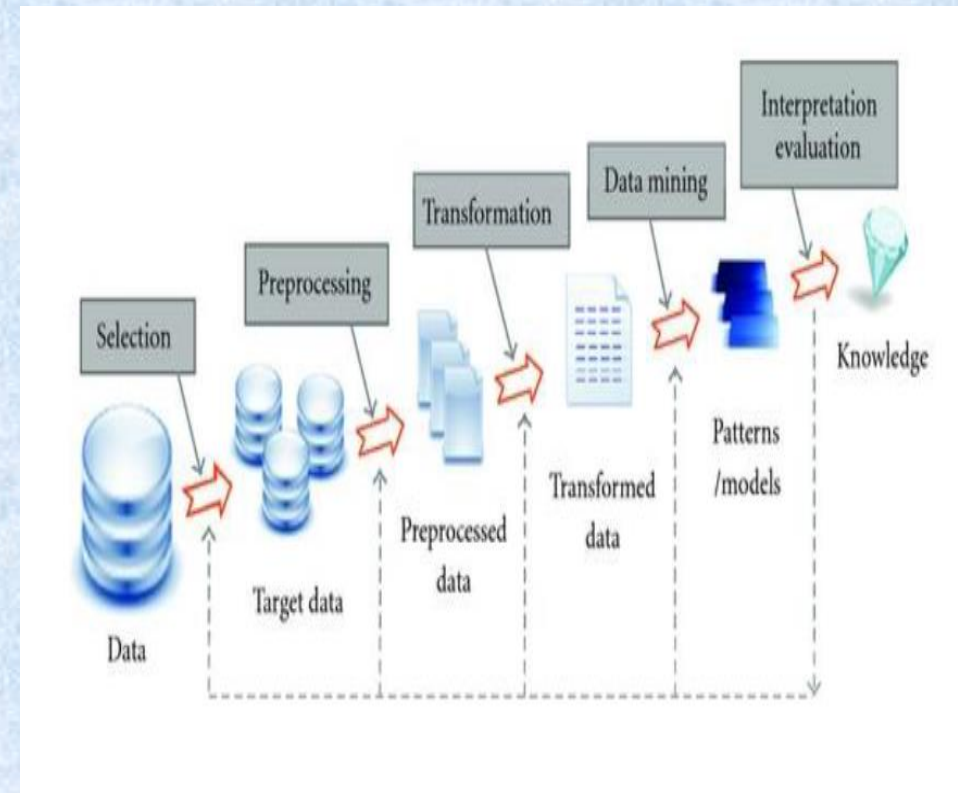


• Variability



Εξόρυξη Γνώσης από τα Μεγάλα Δεδομένα

- Data Mining (DM) or Knowledge Discovery in Data bases (KDD)
- Διαδικασία για την λήψη της τελικής πληροφορίας
 - ✓ Data cleaning
 - ✓ Data integration
 - ✓ Data Selection
 - ✓ Data Transformation
 - ✓ Data Mining
 - ✓ Pattern evaluation
 - ✓ Knowledge representation



Η αξιοποίηση της εξόρυξης γνώσης στο Λιανεμπόριο

- Πρόβλεψη (predictive data mining tasks)
- Αναγνώριση (ύπαρξη μιας δραστηριότητας ή ενός γεγονότος)
- Περιγραφή (descriptive data mining tasks)
- Ταξινόμηση πελατών
- Βελτιστοποίηση πόρων (χρόνος, χρήμα, χώρος κτλ)

αποτέλεσμα...

- ✓ *cross-channel marketing* σε πραγματικό χρόνο
- ✓ *Customer Experience & πρόβλεψη ζήτησης*
- ✓ *υποστήριξη των προσπάθειών των πελατών να εντοπίσουν και να προλάβουν πιθανή απάτη*
- ✓ *Χρήση Social Media για την διασφάλιση του brandname*

Εφαρμογές στο Λιανεμπόριο:

❖ Direct mail Marketing

❖ Category Management and Inventory Control

❖ Customer Relationship Management

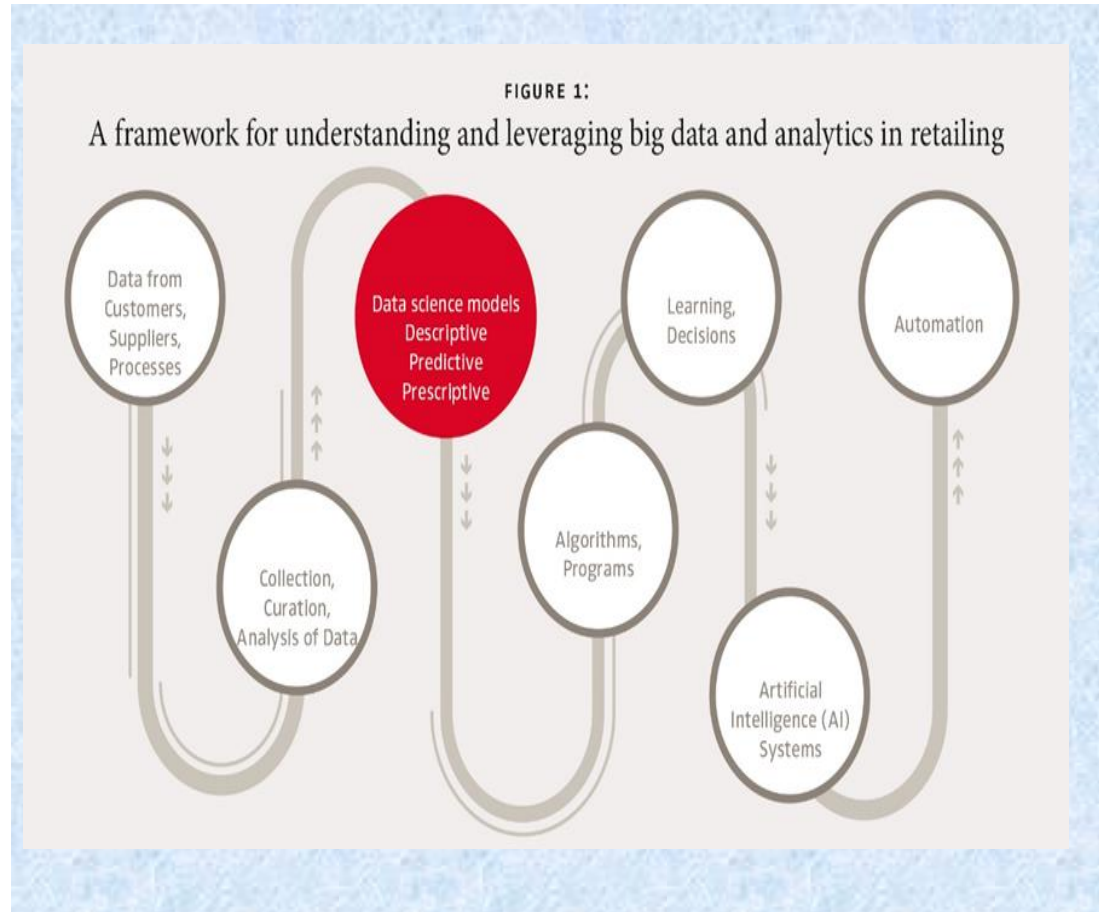
Retail Analytics & Big Data

- Είναι η νέα τεχνολογία στον κλάδο του Παγκόσμιου Λιανεμπορίου
- Τα BD είναι το ΑΤΟΥ για τις εταιρείες του κλάδου
- Αύξηση λειτουργίας τους έως και 60 %
- Αξιοποίηση τεράστιου όγκου πληροφοριών
- Το Λιανεμπόριο είναι εξορισμού βιομηχανία τεράστιων όγκων δεδομένων
- Ο κάθε καταναλωτής έχει γίνει ξεχωριστά μια κινητή μονάδα παραγωγής δεδομένων

Retail Analytics & Big Data

Μοντέλα της Επιστήμης των δεδομένων (Data Science Models)

- Descriptive
- Predictive
- Prescriptive



Τι προσφέρει η τεχνολογία της Στατιστικής Ανάλυσης Μεγάλων Δεδομένων στον κλάδο του Λιανεμπορίου?

- Πρόβλεψη κατανάλωσης
- Εξατομίκευση της εμπειρίας του καταναλωτή
- Κατάσταση προϊόντος και διαθεσιμότητα
- Διαχείριση καταστάσεων σε περίπτωση απάτης ή εξαπάτησης του πελάτη
- Προγνωστική Στατιστική Ανάλυση (Predictive Analytics)
- Δυναμική τιμολογιακή πολιτική
- Πρόβλεψη ζήτησης στο Λιανεμπόριο

Big Data Analytics (BDA) - Business Intelligence (BI)

- BI = DSS (Decision Support Systems)
- Data collection -> Data store -> Data Analysis ->

Knowledge management -> Decision making

- Καλύτερη κατανόηση πελατών, αγορών, ανταγωνιστών, προμηθειών και πόρων.
- Τροφοδότηση της διοίκησης με τη σωστή πληροφόρηση, την κατάλληλη στιγμή και με τον κατάλληλο τρόπο.
- Βελτίωση της ποιότητας των αποφάσεων
- Συμβολή στη διαμόρφωση των στρατηγικών στόχων-> αύξηση ανταγωνιστικότητας
- Δυνατότητες αύξησης της κερδοφορίας, μείωσης του κόστους και βελτίωσης της αποδοτικότητας.
- Επιτυχημένα συστήματα ΕΕ συμβάλλουν στην αύξηση των επιδόσεων και της κερδοφορίας.
- Αύξηση της απόδοσης της επένδυσης σε τεχνολογίες πληροφορικής.

Big Data Analytics (BDA) - Business Intelligence (BI)



Τεχνολογικά εργαλεία αξιοποίησης των BD στο Λιανεμπόριο

➤ Πρόβλημα: οι παραδοσιακές βάσεις δεδομένων δεν μπορούν να υποστηρίξουν τόσο μεγάλους όγκους δεδομένων

❖ Ανάγκη για δημιουργία νέων συστημάτων βάσεων -> βελτιστοποίηση απόδοσης,

Θα επιτρέπεται:

❖ Επιπλέον Hardware

❖ Εντοπισμός συγκεκριμένων δεδομένων στους κόμβους σε πραγματικό χρόνο

❖ Αξιοπιστία (ανάκτηση δεδομένων)

❖ Διαχείριση σφαλμάτων

➤ SQL DB??? Μη δυνατότητα υποστήριξης δεδομένων μεγάλου όγκου και ανομοιογένειας

✓ λύση -> **NoSQL databases** (**MongoDB**, Neo4j, CouchBase, DynamoDB, HBase και Cassandra)

Μοντέλο MapReduce

- Jeffrey Dean και Sanjay Ghemawat (Google)
- επεξεργάζεται γρήγορα και παράλληλα τεράστιες ποσότητες δεδομένων σε συστοιχίες (Clusters) υπολογιστών

1. **Map:** ζεύγη κλειδιού-τιμής (παίρνει δεδομένο εισόδου και παράγει ένα σύνολο ενδιάμεσων υποσυνόλων)

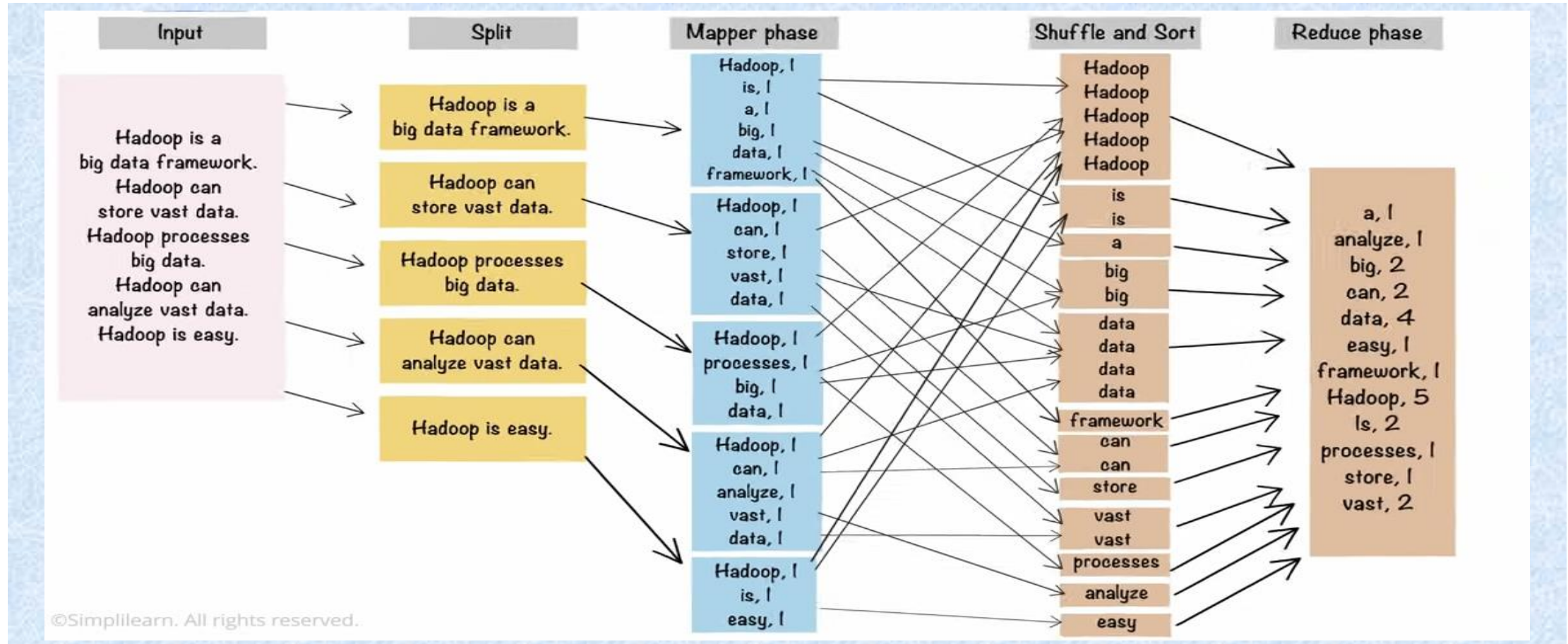
Η βιβλιοθήκη του MapReduce ομαδοποιεί όλα τα ενδιάμεσα υποσύνολα που συνδέονται με το ίδιο ενδιάμεσο κλειδί και τα προωθεί στην συνάρτηση Reduce.

1. **Reduce:** παράγει καμία ή μία τιμή εξόδου (Η συνάρτηση αυτή δέχεται ένα ενδιάμεσο κλειδί και τα υποσύνολα που σχετίζονται με αυτό. Η συνάρτηση αυτή συγχωνεύει μαζί με αυτά τα υποσύνολα και το κλειδί για να σχηματίσουν ένα πιθανώς μικρότερο σύνολο τιμών)

Μοντέλο MapReduce - πλεονεκτήματα

- ✓ Το μοντέλο είναι εύκολο στην χρήση
- ✓ Παρέχει ανεξαρτησία αποθήκευσης του συστήματος, δεδομένου ότι δεν απαιτεί ιδιόκτητα συστήματα αρχείων δεδομένων ή προκαθορισμένα μοντέλα δεδομένων
- ✓ Τα δεδομένα αποθηκεύονται σε αρχεία απλού κειμένου και το σύστημα δεν είναι υποχρεωμένο να συμμορφώνεται στα σχεσιακά συστήματα δεδομένων ή οποιαδήποτε άλλη δομή.
- ✓ χρησιμοποιεί δεδομένα που έχουν μια αυθαίρετη μορφή, (Unstructured), έχει ανοχή σε σφάλματα, είναι διαθέσιμη για γλώσσες προγραμματισμού υψηλού επιπέδου και επιτρέπει χειρισμό με μεγάλη ταχύτητα

Μοντέλο MapReduce – Εικονικό παράδειγμα



Apache Hadoop using MapReduce tech

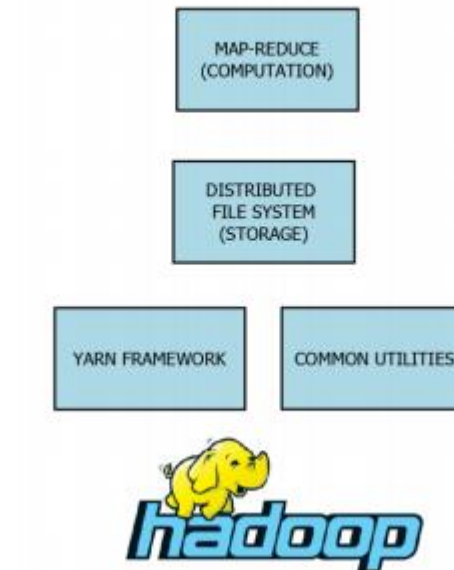
❖ **Apache Hadoop (open-source software)**

- Hadoop Common Utilities(βιβλιοθήκες και βασικές λειτουργίες)
- Hadoop Distributed File System (HDFS) (αποθήκευση κατανεμημένων δεδομένων)
- Hadoop YARN Framework(πλατφόρμα διαχείρισης πόρων)
- Hadoop Map-Reduce(κατανεμημένη επεξεργασία μεγάλης κλίμακας δεδομένων)

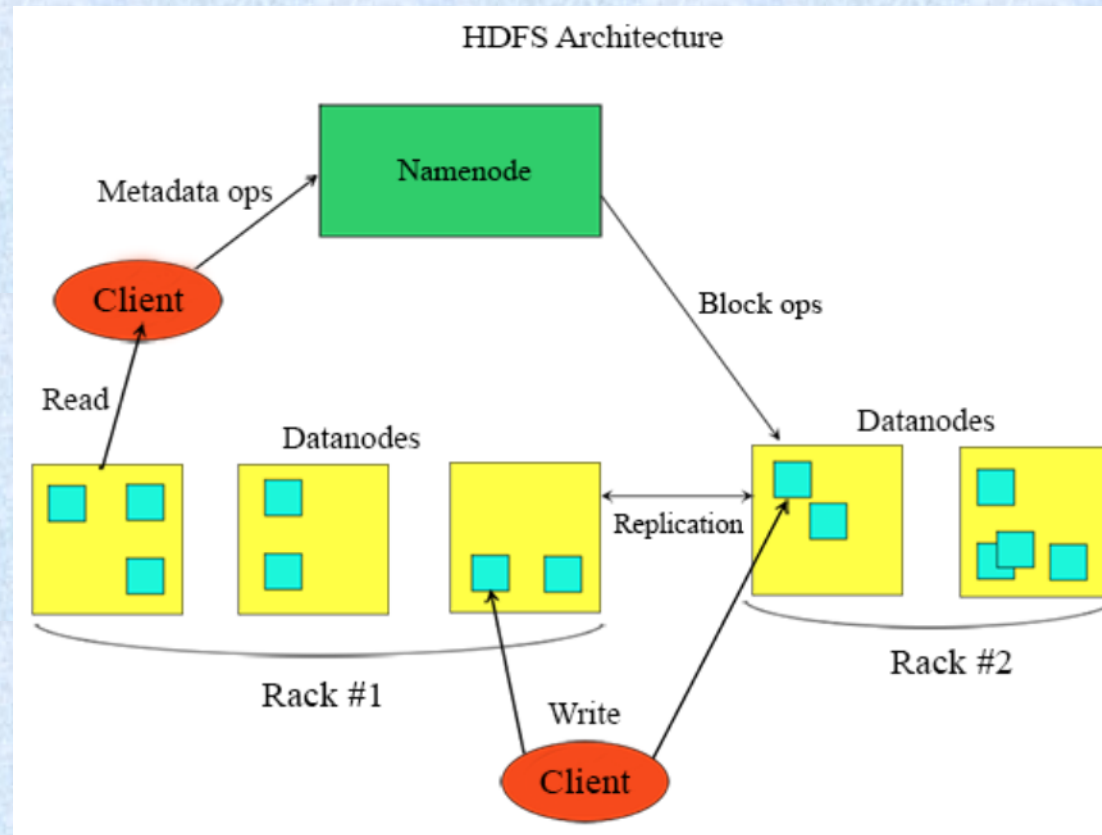
Αρχιτεκτονική του Hadoop

❖ Apache Hadoop (open-source software)

Δομή Apache Hadoop



Hadoop – HDFS Architecture



Πλεονεκτήματα του HDFS(Hadoop Distributed File System)

- Το HDFS συνεχώς ερευνά για λάθη και αυτόματα ανακτά τα δεδομένα από αυτά.
- Συνεχής ροή πρόσβασης δεδομένων
- Διαχείριση Μεγάλου συνόλου δεδομένων (TB)-πολλοί κόμβοι σε ενιαίο Cluster
- παράγεται μεγάλη ποσότητα «έργου» από έναν σταθμό εξυπηρέτησης, στη μονάδα του χρόνου
- παρέχει διεπαφές για τις εφαρμογές, ώστε να μπορούν οι τελευταίες να μετακινούνται κοντά στα δεδομένα
- Φορητότητα μεταξύ διαφορετικού υλικού και λογισμικού

Το Hadoop μπορεί να μεταμορφώσει την βιομηχανία του Λιανεμπορίου...

- Υποστηρίζει ένα βήμα παραπέρα την σχετική τεχνολογία για την ανάλυση-επεξεργασία των Μεγάλων Δεδομένων
- Διευκολύνει τον χειρισμό και την στατιστική ανάλυση για μη δομημένα δεδομένα όπως το Facebook, Instagram, Pinterest ...Η NLP
- Μπορεί να ανακαλύψει νέες αγοραστικές συμπεριφορές των καταναλωτών...

Εν ολίγοις, οι Εταιρείες στο Λιανεμπόριο κάνουν χρήση του Hadoop για να κατανοήσουν την νέα τάση των καταναλωτών όπως επίσης να μειώσουν τα κόστη και να αυξήσουν τα κέρδη τους.

Παράδειγμα υπολογισμού πωλήσεων προϊόντος ανά χώρα με βάση το Hadoop & Map Reduce /Java /εντολές σε Linux (1)

- δημιουργία νέου φακέλου με την εντολή

sudo mkdir MapReduceDirectory

- Δίνουμε άδεια μέσω της εντολής

sudo chmod -R 777 MapReduceDirectory

- Εξάγουμε το classpath με τις ακόλουθες εντολές

export CLASSPATH="\$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.2.0.jar:\$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.2.0.jar:\$HADOOP_HOME/share/hadoop/common/hadoop-common-2.2.0.jar:~/MapReduceTutorial/SalesCountry/*:\$HADOOP_HOME/lib/*"

Παράδειγμα υπολογισμού πωλήσεων προϊόντος ανά χώρα με βάση το Hadoop & Map Reduce /Java /εντολές σε Linux (2)

- Κάνουμε compiling στα αρχεί Java με την εντολή

```
javac -d . SalesMapper.java SalesCountryReducer.java SalesCountryDriver.java
```

- Δημιουργούμε ένα αρχείο κειμένου

```
sudo gedit Manifest.txt
```

Και μετέπειτα

```
Main-Class: SalesCountry.SalesCountryDriver
```

- Δημιουργούμε αρχείο Jar

```
jar cfm ProductSalePerCountry.jar Manifest.txt SalesCountry/*.class
```

Παράδειγμα υπολογισμού πωλήσεων προϊόντος ανά χώρα με βάση το Hadoop & Map Reduce /Java /εντολές σε Linux (3)

- Κάνουμε εκκίνηση του Hadoop

```
$HADOOP_HOME/sbin/start-dfs.sh
```

&

```
$HADOOP_HOME/sbin/start-yarn.sh
```

- Αντιγράφουμε το αρχείο με τα δεδομένα εισόδου στο φάκελο inputMapReduce και κατόπιν το inputMapReduce στο HDFS.

```
$HADOOP_HOME/bin/hdfs dfs -copyFromLocal ~/inputMapReduce /
```

&

```
$HADOOP_HOME/bin/hdfs dfs -ls /inputMapReduce
```

Παράδειγμα υπολογισμού πωλήσεων προϊόντος ανά χώρα με βάση το Hadoop & Map Reduce /Java /εντολές σε Linux (4)

- Τρέχουμε την εργασία MapReduce

```
$HADOOP_HOME/bin/hadoop jar ProductSalePerCountry.jar /inputMapReduce /mapreduce_output_sales
```

Αυτό θα δημιουργήσει ένα φάκελο εξόδου με το όνομα `mapreduce_output_sales` στο HDFS.

Τα περιεχόμενα αυτού του φακέλου θα είναι ένα αρχείο που θα εμπεριέχει τις πωλήσεις των προϊόντων ανά χώρα.

- Το αποτέλεσμα μπορεί να εμφανισθεί μέσω της ακόλουθης εντολής

```
$HADOOP_HOME/bin/hdfs dfs -cat /mapreduce_output_sales/part-00000
```

Συμπεράσματα-Σχόλια (1)

Η χρήση των Μεγάλων Δεδομένων στο Λιανεμπόριο οδηγεί σε

- ✓ Ανίχνευση συμπεριφοράς καταναλωτών
- ✓ Μελλοντικές πωλήσεις
- ✓ Στρατηγικές αποφάσεις
- ✓ Βελτιστοποίηση τιμών προϊόντων
- ✓ Πρόβλεψη ζήτησης και νέων τάσεων
- ✓ RoI (Return of Investments)

Συμπεράσματα-Σχόλια (2)

- Εγείρονται ερωτήματα περί ηθικής και παράβασης της Ιδιωτικότητας.
- Δικτατορία των Δεδομένων (Data Dictatorship)

ΛΥΣΕΙΣ:

1. ξεκάθαρη πολιτική με συμφωνημένους όρους
2. οφέλη της τεχνολογίας της Προγνωστικής Ανάλυσης στην καταναλωτική τους βάση
3. επιβράβευση της εμπιστοσύνης των πελατών
4. Εξισορρόπηση των εταιρειών - πολιτών

Βιβλιογραφία (1)

- Aeron, H., Kumar, & A.; Moorthy, J. (2012), "Data mining framework for customer lifetime value-based segmentation", Database Marketing & Customer Strategy Management, Vol. 19, 1, Page 17-30.
- Ahmed, F.D., Jaber, A.N., Majid, M.B.A., & Ahmad, M.S. (2015, August). Agent-based big data analytics in retailing: a case study. In 2015 4th International Conference on Software Engineering and Computer Systems (ICSECS) (pp. 67-72).
- Akter, S., & Wamba, S.F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. Electronic Markets, 26(2), 173-194.
- Asta Zelenkauskaitė, Bruno Simões (2014 February), "The Big Data User-Centric Model"
- Bentley, B. (2013), "Building User-Centric Platforms For Big Data", Marketing News, Vol. 47 Issue 3, Page 46-50.
- Biehn, Neil, (2013), "The Missing V's in Big Data: Viability and Value", Wired, 05/06/2013.
- C. Spurgeon and G. Goggin, "Mobiles into media: Premium rate sms and the adaptation of television to interactive communication cultures," Continuum: Journal of Media & Cultural Studies, vol. 21, no. 2, pp. 317–329, 2007
- Chen, H., Chiang, R.H., & Storey, V.C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188
- D. Laney. 3D data management: controlling data volume, velocity and variety [J]. META group research note, 2001, 6: 70.
- D. Laney, "3-d data management: Controlling data volume, velocity and variety," META Group," Research Note, February 2001, <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.

Βιβλιογραφία (2)

- Diebold, Francis X., On the Origin(s) and Development of the Term 'Big Data' (September 21, 2012). PIER Working Paper No. 12-037. Available at SSRN: <http://ssrn.com/abstract=2152421> or <http://dx.doi.org/10.2139/ssrn.2152421>
- Eric T. Bradlow, Manish Gangwar, Praveen Kopalle, Sudhir Voleti, (2016 November), "The Role of Big Data and Predictive Analytics in Retailing", Forthcoming, Journal of Retailing
- F. Provost and T. Fawcett, Data science for Business. Sebastopol, CA: O'Reilly Media, 2013. J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," Technical report, IDC, Tech. Rep., 2012.
- Hall, J. M., Kopalle, P. K., & Krishna, A. (2010), "Retailer dynamic pricing and ordering decisions: category management versus brand-by-brand approaches", Journal of Retailing, 86(2), 172-183.
- M. Chen, S. W. Mao, Y. H. Liu. Big data: a survey [J]. Mobile networks and applications, 2014, 19(2): 171-209.
- Mayer-Schönberger and Kenneth Cukier (2013) "Big Data-A Revolution That Will Transform How We Live,Work and Think")
- Nunan, D., & Domenico, M. (2013), "Market research and the ethics of big data", International Journal of Market Research,Vol. 55, No. 4, Page 2 -13.
- Shankar, V. (2019). Big Data and Analytics in Retailing. NIM Marketing Intelligence Review, 11(1), 36-40.
- X. D. Wu, X. Q. Zhu, G. Q. Wu, et al. Data mining with big data [J]. IEEE transactions on knowledge and data engineering, 2014, 26(1): 97-107.
- Zhen Wang, Guofu Wei, Yaling Zhan, Yanhuan Sun. Big Data in telecommunications operators: data, platform and practices. 2017, 1-7
- Zikopoulos, P., & Eaton, C. (2011). Understanding big data: Analytics for enterprise-class Hadoop and streaming data. McGraw-Hill Osborne Media.
- Hadoop MapReduce. <http://hadoop.apache.org/mapreduce>

Appendix (data input, Java files used)

1	Transaction_date	Product	Price	Payment_Type	Name	City	State	Country	Account_Create	Last_Login	Latitude	Longitude
2	1/2/09 6:17	Product1	1200	Mastercard	carolina	Basildon	England	United Kingdom	1/2/09 6:00	1/2/09 6:08	51.5	-1.1166667
3	1/2/09 4:53	Product1	1200	Visa	Betina	Parkville	MO	United States	1/2/09 4:42	1/2/09 7:49	39.195	-94.68194
4	1/2/09 13:08	Product1	1200	Mastercard	Federica e Andrea	Astoria	OR	United States	1/1/09 16:21	1/3/09 12:32	46.18806	-123.83
5	1/3/09 14:44	Product1	1200	Visa	Gouya	Echuca	Victoria	Australia	9/25/05 21:13	1/3/09 14:22	-36.1333333	144.75
6	1/4/09 12:56	Product2	3600	Visa	Gerd W	Cahaba Heights	AL	United States	11/15/08 15:47	1/4/09 12:45	33.52056	-86.8025
7	1/4/09 13:19	Product1	1200	Visa	LAURENCE	Mickleton	NJ	United States	9/24/08 15:19	1/4/09 13:04	39.79	-75.23806
8	1/4/09 20:11	Product1	1200	Mastercard	Fleur	Peoria	IL	United States	1/3/09 9:38	1/4/09 19:45	40.69361	-89.58889
9	1/2/09 20:09	Product1	1200	Mastercard	adam	Martin	TN	United States	1/2/09 17:43	1/4/09 20:01	36.34333	-88.85028
10	1/4/09 13:17	Product1	1200	Mastercard	Renee Elisabeth	Tel Aviv	Tel Aviv	Israel	1/4/09 13:03	1/4/09 22:10	32.0666667	34.7666667
11	1/4/09 14:11	Product1	1200	Visa	Aidan	Chatou	Ile-de-France	France	6/3/08 4:22	1/5/09 1:17	48.8833333	2.15
12	1/5/09 2:42	Product1	1200	Diners	Stacy	New York	NY	United States	1/5/09 2:23	1/5/09 4:59	40.71417	-74.00639
13	1/5/09 5:39	Product1	1200	Amex	Heidi	Eindhoven	Noord-Brabant	Netherlands	1/5/09 4:55	1/5/09 8:15	51.45	5.4666667
14	1/2/09 9:16	Product1	1200	Mastercard	Sean	Shavano Park	TX	United States	1/2/09 8:32	1/5/09 9:05	29.42389	-98.49333
15	1/5/09 10:08	Product1	1200	Visa	Georgia	Eagle	ID	United States	11/11/08 15:53	1/5/09 10:05	43.69556	-116.35306
16	1/2/09 14:18	Product1	1200	Visa	Richard	Riverside	NJ	United States	12/9/08 12:07	1/5/09 11:01	40.03222	-74.95778
17	1/4/09 1:05	Product1	1200	Diners	Leanne	Julianstown	Meath	Ireland	1/4/09 0:00	1/5/09 13:36	53.6772222	-6.3191667
18	1/5/09 11:37	Product1	1200	Visa	Janet	Ottawa	Ontario	Canada	1/5/09 9:35	1/5/09 19:24	45.4166667	-75.7
19	1/6/09 5:02	Product1	1200	Diners	barbara	Hyderabad	Andhra Pradesh	India	1/6/09 2:41	1/6/09 7:52	17.3833333	78.4666667
20	1/6/09 7:45	Product2	3600	Visa	Sabine	London	England	United Kingdom	1/6/09 7:00	1/6/09 9:17	51.52721	0.14559
21	1/2/09 7:35	Product1	1200	Diners	Hani	Salt Lake City	UT	United States	12/30/08 5:44	1/6/09 10:52	40.76083	-111.89028
22	1/6/09 12:56	Product1	1200	Visa	Jeremy	Manchester	England	United Kingdom	1/6/09 10:58	1/6/09 13:31	53.5	-2.2166667
23	1/1/09 11:05	Product1	1200	Diners	Janis	Ballynora	Cork	Ireland	12/10/07 12:37	1/7/09 1:52	51.8630556	-8.58
24	1/5/09 4:10	Product1	1200	Mastercard	Nicola	Rodepoort	Gauteng	South Africa	1/5/09 2:33	1/7/09 5:13	-26.1666667	27.8666667
25	1/6/09 7:18	Product1	1200	Visa	asuman	Chula Vista	CA	United States	1/6/09 7:07	1/7/09 7:08	32.64	-117.08333
26	1/2/09 1:11	Product1	1200	Mastercard	Lena	Kuopio	Itä-Suomen Laani	Finland	12/31/08 2:48	1/7/09 10:20	62.9	27.6833333
27	1/1/09 2:24	Product1	1200	Visa	Lisa	Sugar Land	TX	United States	1/1/09 1:56	1/7/09 10:52	29.61944	-95.63472
28	1/7/09 8:08	Product1	1200	Diners	Bryan Kerrene	New York	NY	United States	1/7/09 7:39	1/7/09 12:38	40.71417	-74.00639
29	1/2/09 2:57	Product1	1200	Visa	chris	London	England	United Kingdom	1/3/08 7:23	1/7/09 13:14	51.52721	0.14559
30	1/1/09 20:21	Product1	1200	Visa	Maxine	Morton	IL	United States	10/24/08 6:48	1/7/09 20:49	40.61278	-89.45917
31	1/8/09 0:42	Product1	1200	Visa	Family	Los Gatos	CA	United States	1/8/09 0:28	1/8/09 3:39	37.22667	-121.97361
32	1/8/09 3:56	Product1	1200	Mastercard	Katherine	New York	NY	United States	1/8/09 3:33	1/8/09 6:19	40.71417	-74.00639
33	1/8/09 3:16	Product1	1200	Mastercard	Linda	Miami	FL	United States	1/8/09 3:06	1/8/09 6:34	25.77389	-80.19389
34	1/8/09 1:59	Product1	1200	Mastercard	SYLVIA	Vesenz	Geneve	Switzerland	11/28/07 11:56	1/8/09 7:20	46.2333333	6.2
35	1/3/09 9:03	Product1	1200	Diners	Sheila	Brooklyn	NY	United States	1/3/09 8:47	1/8/09 10:38	40.65	-73.95
36	1/5/09 13:17	Product1	1200	Mastercard	Stephanie	Badhoevedorp	Noord-Holland	Netherlands	1/5/09 12:45	1/8/09 11:45	52.3333333	4.7833333
37	1/6/09 7:46	Product1	1200	Amex	Kelly	Reston	VA	United States	1/6/09 7:30	1/8/09 12:40	38.96861	-77.34139
38	1/5/09 20:00	Product2	3600	Visa	James	Burpengary	Queensland	Australia	12/10/08 19:53	1/8/09 17:58	-27.1666667	152.95
39	1/8/09 16:24	Product1	1200	Visa	jennifer	Phoenix	AZ	United States	1/8/09 15:57	1/8/09 18:30	33.44833	-112.07333
40	1/9/09 6:39	Product1	1200	Mastercard	Anneli	Houston	TX	United States	1/9/09 5:09	1/9/09 7:11	29.76306	-95.36306
41	1/6/09 22:19	Product2	3600	Amex	Ritz	Pittsfield	VT	United States	1/6/09 12:00	1/9/09 10:05	43.77222	-72.81333
42	1/6/09 23:00	Product2	3600	Amex	Sylvia	Pittsfield	VT	United States	1/6/09 12:00	1/9/09 10:05	43.77222	-72.81333
43	1/7/09 7:44	Product1	1200	Mastercard	Marie	Ball Ground	GA	United States	1/7/09 5:35	1/9/09 10:52	34.33806	-84.37667
44	1/3/09 13:24	Product1	1200	Visa	Mehmet Fatih	Helsingør	Frederiksborg	Denmark	1/3/09 12:47	1/9/09 11:14	56.0333333	12.6166667
45	1/7/09 15:12	Product2	3600	Visa	Anabela	Flossmoor	IL	United States	1/5/09 19:55	1/9/09 16:03	41.54278	-87.68472
46	1/7/09 20:45	Product1	1200	Amex	Nicole	Houston	TX	United States	1/7/09 17:47	1/9/09 20:00	29.76306	-95.36306

SalesMapper.java

```
• package SalesCountry;

• import java.io.IOException;

• import org.apache.hadoop.io.IntWritable;
• import org.apache.hadoop.io.LongWritable;
• import org.apache.hadoop.io.Text;
• import org.apache.hadoop.mapred.*;

• public class SalesMapper extends MapReduceBase implements Mapper <LongWritable, Text, Text, IntWritable> {
•     private final static IntWritable one = new IntWritable(1);

•     public void map(LongWritable key, Text value, OutputCollector <Text, IntWritable> output, Reporter reporter) throws IOException {

•         String valueString = value.toString();
•         String[] SingleCountryData = valueString.split(",");
•         output.collect(new Text(SingleCountryData[7]), one);
•     }
• }
```

SalesCountryReducer.java

```
• package SalesCountry;

•
• import java.io.IOException;
• import java.util.*;

• import org.apache.hadoop.io.IntWritable;
• import org.apache.hadoop.io.Text;
• import org.apache.hadoop.mapred.*;

• public class SalesCountryReducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {

•     public void reduce(Text t_key, Iterator<IntWritable> values, OutputCollector<Text,IntWritable> output, Reporter reporter) throws IOException {
•
•         Text key = t_key;
•         int frequencyForCountry = 0;
•         while (values.hasNext()) {
•
•             // replace type of value with the actual type of our value
•             IntWritable value = (IntWritable) values.next();
•             frequencyForCountry += value.get();
•
•         }
•         output.collect(key, new IntWritable(frequencyForCountry));
•     }
• }
```

SalesCountryDriver.java

```
• package SalesCountry;

• import org.apache.hadoop.fs.Path;
• import org.apache.hadoop.io.*;
• import org.apache.hadoop.mapred.*;

• public class SalesCountryDriver {
•     public static void main(String[] args) {
•         JobClient my_client = new JobClient();
•
•         JobConf job_conf = new JobConf(SalesCountryDriver.class);
•         job_conf.setJobName("SalePerCountry");
•
•         job_conf.setOutputKeyClass(Text.class);
•         job_conf.setOutputValueClass(IntWritable.class);
•
•         job_conf.setMapperClass(SalesCountry.SalesMapper.class);
•         job_conf.setReducerClass(SalesCountry.SalesCountryReducer.class);
```

```
•     job_conf.setInputFormat(TextInputFormat.class);
•     job_conf.setOutputFormat(TextOutputFormat.class);
•
•     FileInputFormat.setInputPaths(job_conf, new Path(args[0]));
•     FileOutputFormat.setOutputPath(job_conf, new Path(args[1]));
•
•     my_client.setConf(job_conf);
•     try {
•         // RUNNING MAPREDUCE JOB
•         JobClient.runJob(job_conf);
•     } catch (Exception e) {
•         e.printStackTrace();
•     }
• }
```

*Σας ευχαριστώ πολύ για τον
χρόνο σας και την υπομονή
σας...*