

Σύγκριση συστημάτων OLTP, OLAP και Data Mining

On Line Transaction Processing – OLTP

Αποτελεί σύστημα βάσεως δεδομένων που υποστηρίζει την καθημερινή λειτουργία της επιχείρησης

Κύρια χαρακτηριστικά:

- Μικρά και απλά transactions
- Συχνά updates
- Τα Transactions κάνουν access σε ένα μικρό κομμάτι της database.
- OLTP queries είναι απλά queries, που μπορεί να αφορούν δεδομένα banking ή airline systems.

On Line Analytic Processing – OLAP

Τα OLAP tools χρησιμοποιούν information σε databases για να κατευθύνουν τις strategic decisions.

Κύρια χαρακτηριστικά:

- Complex queries
- Λίγα updates
- Τα Transactions έχουν πρόσβαση σε ένα σημαντικό κομμάτι της database.
- Τα Data δεν χρειάζεται να ενημερώνονται με κάθε μεμονωμένη transaction, αλλά αυτή η εργασία γίνεται συγκεντρωτικά (batch process).
- Συνήθως σύνθετα queries που απαιτούν πολύ ώρα.
- Ονομάζονται επίσης decision-support queries
- Τα OLAP tool επιτρέπουν σε analysts, managers, και executives να αποκτήσουν πρόσβαση σε data διαμέσου διαδραστικής πρόσβασης σε μια μεγάλη ποικιλία δεδομένων τα οποία έχουν διαμορφωθεί έτσι ώστε να αναδεικνύουν το dimensionality τους.

OLAP queries:

- Τα OLAP queries χρησιμοποιούν data τα οποία συγκεντρώνονται από την καθημερινή λειτουργία της επιχείρησης μέσω των OLTP systems.
- Τα Queries είναι ad hoc και συνήθως εξυπηρετούν τις ανάγκες των managers
- Τα OLAP χρησιμοποιούνται για marketing, budgeting, forecasting κ.λπ. Οι databases που χρησιμοποιούνται για OLAP είναι configured για complex και ad-hoc queries

OLAP vs OLTP

OLAP: (online analytical processing) είναι η εξεύρεση τρόπων για την υποβολή query σε multi-dimensional databases.

OLTP: Τα online operational Database Systems εκτελούν transaction και query processing, δηλαδή υποστηρίζουν την καθημερινή λειτουργία ενός οργανισμού σε τομείς όπως προμήθειες, inventory, manufacturing, banking, payroll και accounting.

Ένα OLTP δουλεύει με τρέχοντα data ενός οργανισμού, χωρίς τη χρήση historical data από εξωτερικές πηγές.

Ένα OLAP σύστημα διαχειρίζεται information που προέρχεται από διάφορους organizations και information από διάφορες πηγές τόσο εσωτερικές της επιχείρησης, όσο και εξωτερικές.

Concepts	OLTP for Operation system	OLAP for Data Warehouse
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and <u>decision support</u>
What the data	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data

Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch <u>data refreshes</u> and complex queries may take many hours; query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than OLTP

Backup and Recovery	Backup religiously; operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery method
---------------------	---	--

Data Mining

Data Mining είναι η προσπάθεια εξαγωγής knowledge από μία database.

Το Data mining ασχολείται με την εξαγωγή μοτίβων (patterns) από μεγάλα σύνολα data. Χρησιμοποιεί μεθόδους από artificial intelligence, statistics και database management.

Τα OLAP και Data mining χρησιμοποιούνται για την επίλυση διαφορετικού είδους προβλημάτων ανάλυσης:

Τα OLAP παρέχουν αθροίσματα δεδομένων (data) σε συνδυασμό με μοντέλα για να κάνει κάποιο projection ή simulation.

π.χ. Τα OLAP δίνουν απαντήσεις σε ερωτήματα όπως:

Πώς ήταν οι πωλήσεις αυτοκινήτων το τελευταίο τρίμηνο αυτού του έτους συγκρινόμενες με το αντίστοιχο τρίμηνο που προηγούμενου έτους?

Ποια είναι η πρόβλεψη για τις πωλήσεις αυτοκινήτων το επόμενο τρίμηνο?

Το Data mining ανακαλύπτει κρυφά patterns σε data.

Data mining λειτουργεί σε επίπεδο λεπτομέρειας αντί σε επίπεδο αθροισμάτων.

π.χ. Το Data mining δίνει απάντηση σε ερωτήσεις όπως:

Ποιος είναι πιθανό να αγοράσει αυτοκίνητο το επόμενο τρίμηνο και ποια είναι τα χαρακτηριστικά αυτού του αγοραστή?

Ουσιαστικά τα OLAP και data mining συμπληρώνει το ένα το άλλο.

- π.χ. το OLAP μπορεί να αναδείξει ένα πρόβλημα στις πωλήσεις αυτοκινήτων σε ένα συγκεκριμένο νομό.
- Data mining θα μπορούσε να χρησιμοποιηθεί για την καλύτερη κατανόηση της καταναλωτικής συμπεριφοράς των πιθανών αγοραστών σε αυτήν την περιοχή.
- Επίσης Data Mining θα μπορούσε να χρησιμοποιηθεί για να ταυτοποιηθούν οι σημαντικότερες παράμετροι (κριτήρια) που επηρεάζουν τις πωλήσεις αυτοκινήτων και αυτά τα κριτήρια θα μπορούσαν να χρησιμοποιηθούν για να δημιουργηθεί ένα μοντέλο στο OLAP.

Αν και **Data mining** και **OLAP** είναι ίδια από την άποψη ότι χρησιμοποιούν data για την απόκτηση της απαιτούμενης πληροφορίας, **η κύρια διαφορά τους προέρχεται από τον τρόπο που χρησιμοποιούν τα data.**

- Τα OLAP tools παρέχουν multidimensional data analysis και παρέχουν αθροίσματα δεδομένων.
- Εν αντιθέσει το data mining επικεντρώνεται σε ratios, patterns και influences σε ένα σύνολο δεδομένων.

Τα OLAP αφορούν αθροίσματα δεδομένων από πολλαπλές databases σε σύνθετους tables. Τα OLAP tools ασχολούνται με αθροίσματα δεδομένων.

- OLAP μπορεί να παράσχει πληροφόρηση για τις πωλήσεις μιας εταιρείας ανα νομό.
- Το Data mining μπορεί να παράσχει πληροφόρηση για τους παράγοντες που επηρεάζουν τις πωλήσεις τις εταιρείας ανά νομό. Από αυτή την άποψη το data mining δεν είναι άθροισμα δεδομένων αλλά διαίρεση αυτών σε υποκατηγορίες ανάλογα τα ιδιαίτερα χαρακτηριστικά του κάθε υποσυνόλου.

OLAP και data mining databases συχνά αποθηκεύονται σε ειδικούς servers που επονομάζονται data warehouses

Τα data warehouses μπορούν να διαχειριστούν το τεράστιο μέγεθος δεδομένων που παράγονται από τα OLTP systems.

Πολλές OLAP tools βασίζονται σε fact tables.

π.χ. μια συναλλαγή σε super market θα μπορούσε να αποτυπωθεί ως:

Sale (market_id, product_id, time_id)

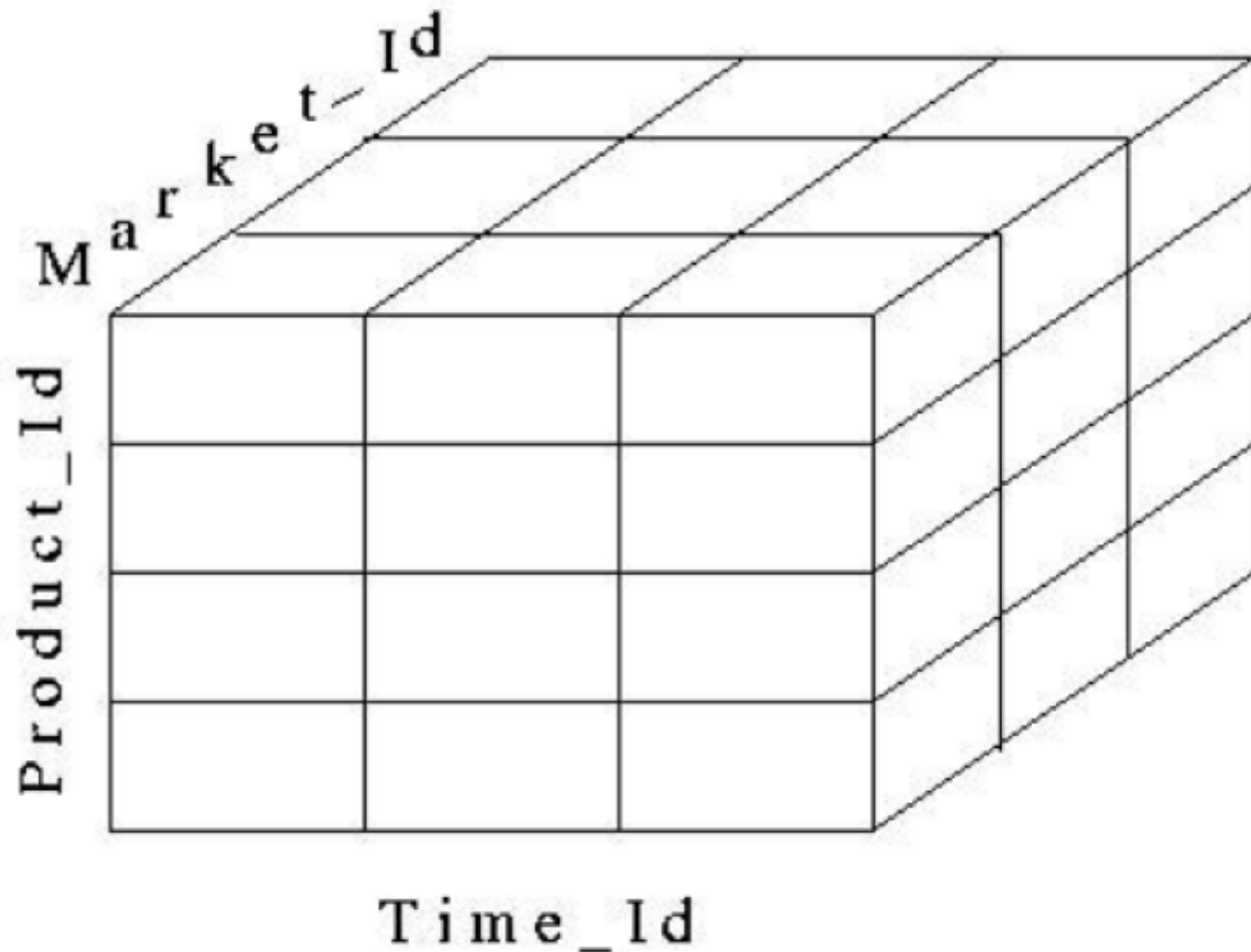
Όπου: market_id, product_id, time_id είναι οι διαστάσεις του πίνακα (table)

market_id: αντιπροσωπεύει συγκεκριμένο super market

product_id: προϊόν

time_id: χρόνος αγοράς

Αυτοί οι πίνακες (fact tables) μπορεί να θεωρηθούν ως N-dimensional data cubes (3-dimensional στο συγκεκριμένο παράδειγμα)



Οι dimensions ενός fact table μπορούν να αναλυθούν περαιτέρω μέσω ενός dimension table.

Fact table:

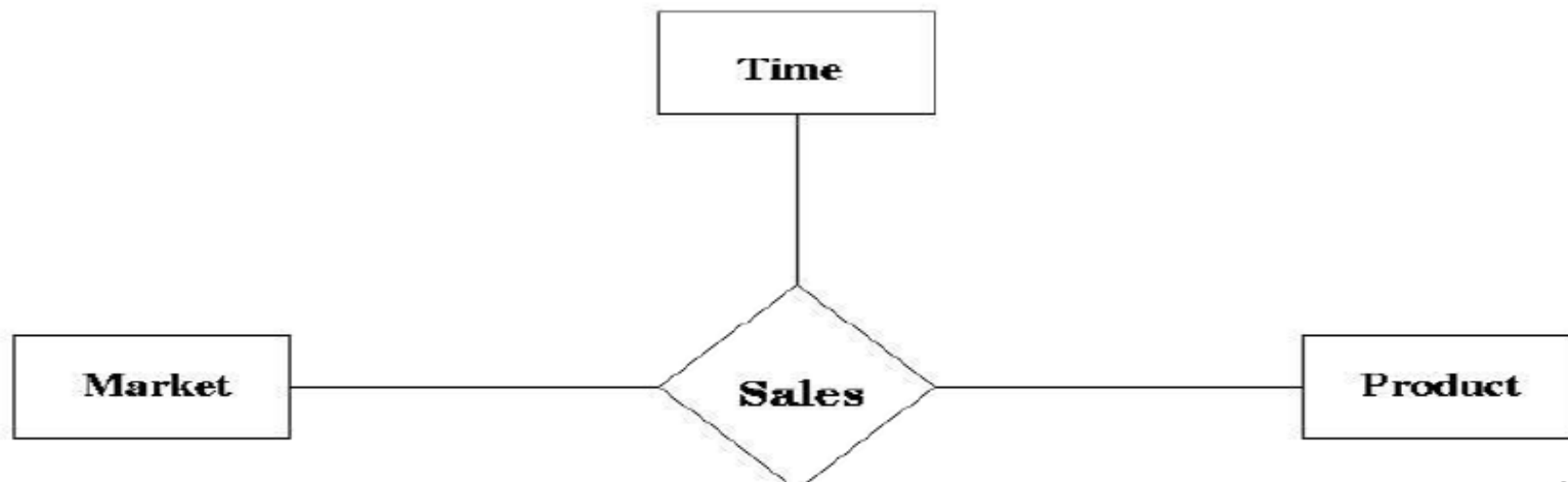
Sales (Market_id, Product_Id, Time_Id)

Dimension Tables:

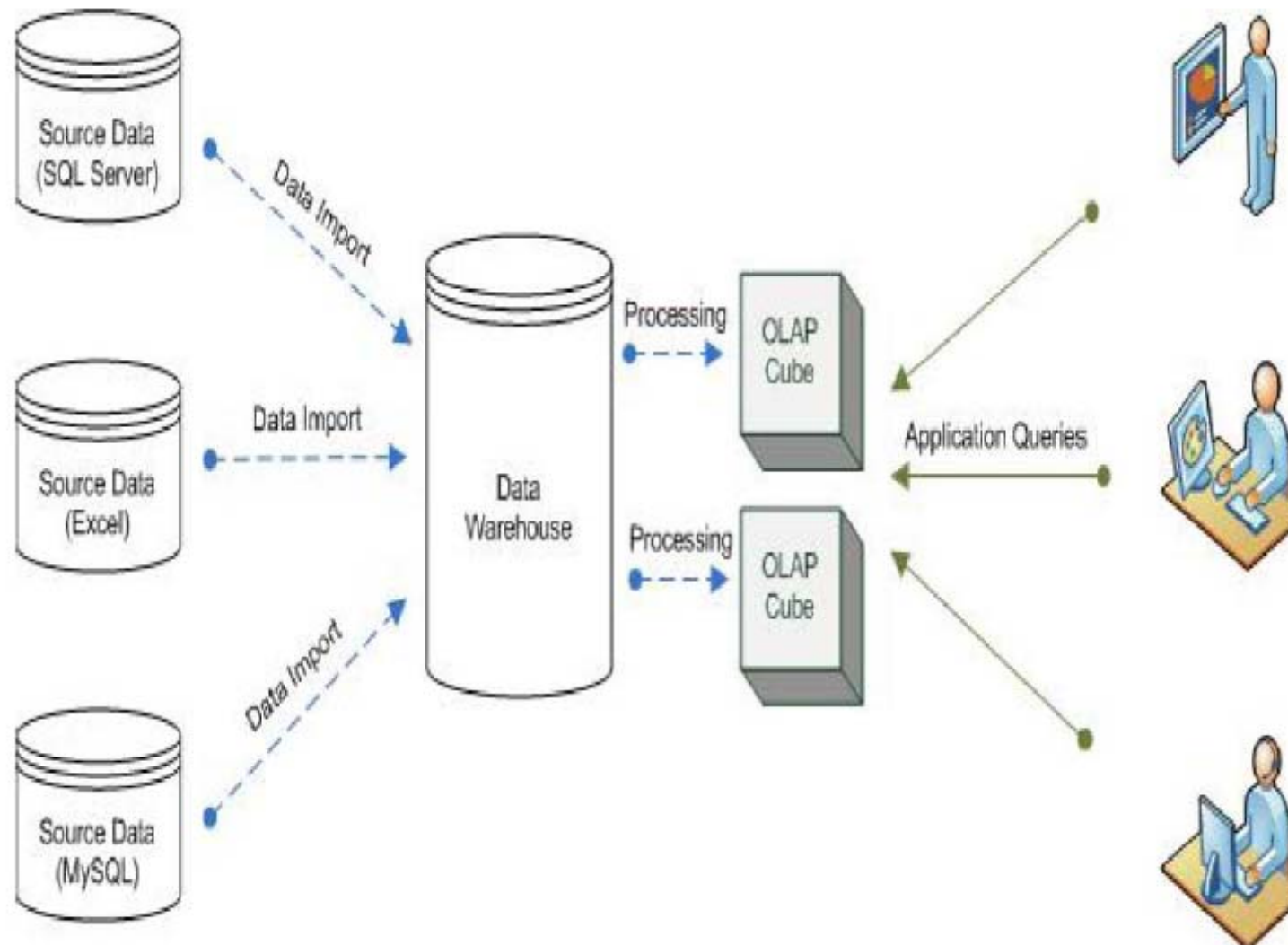
Market (Market_Id, City, State, Region)

Product (Product_Id, Name, Category, Price)

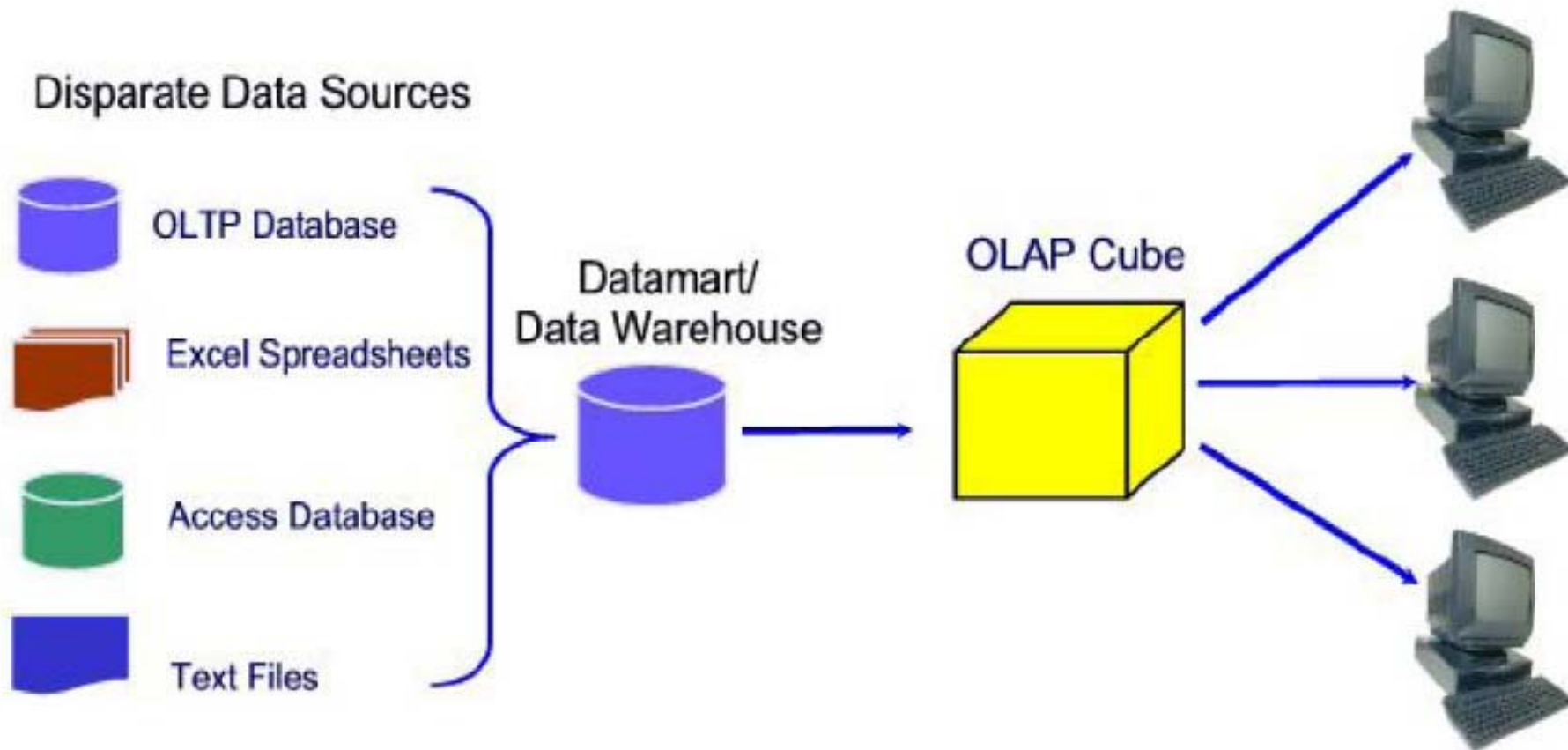
Time (Time_Id, Week, Month, Quarter)



The **CUBE operator** χρησιμοποιείται συχνά για να προϋπολογίσει αθροίσματα (aggregations) σε όλες τις dimensions ενός fact table τα οποία αποθηκεύονται για να επιταχύνει μελλοντικά queries.



Ένας multidimensional cube συνδυάζει δεδομένα από διάφορες πηγές και αποθηκεύει την σχετική πληροφορία με τρόπο ώστε να είναι χρήσιμη για τους επιχειρηματικούς σκοπούς.



Drilling Down και Rolling Up

Οι dimension tables δημιουργούν μια ιεραρχία (hierarchy):

market_id \implies city \implies state \implies region

- Εκτελώντας queries που κινούνται προς τα κάτω ιεραρχικά (π.χ. από region σε city) η διαδικασία ονομάζεται **drilling down**.
- Εκτελώντας queries που κινούνται προς τα πάνω ιεραρχικά (π.χ. από market_id σε state) η διαδικασία ονομάζεται **rolling up**.

Region	Sales variance
Africa	105%
Asia	57%
Europe	122%
North America	97%
Pacific	85%
South America	163%

Drill-down
(Group by Nation)

Roll-up
(group by Region)

Nation	Sales variance
China	123%
Japan	52%
India	87%
Singapore	95%