

1. discuss the design of Data Warehouse

A Business Analysis Framework

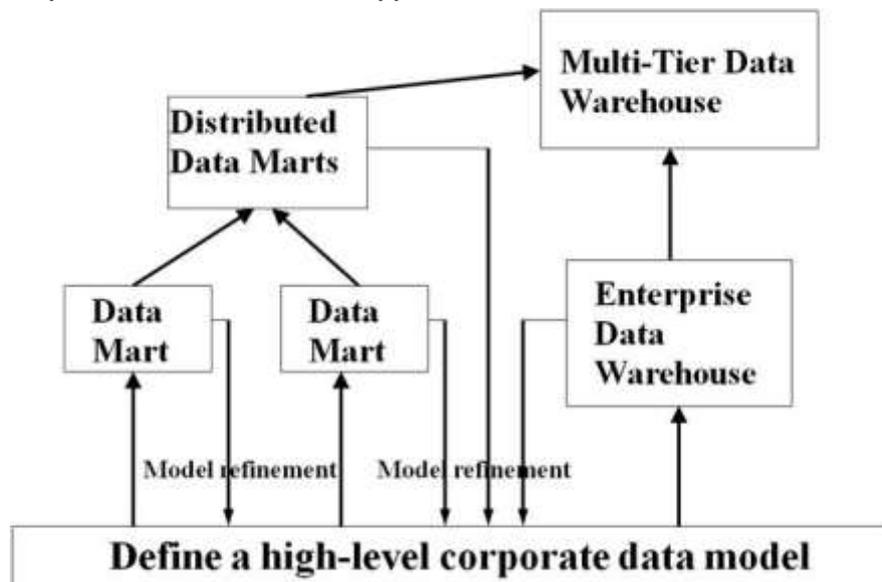
Four views regarding the design of a data warehouse

- **Top-down view** :allows selection of the relevant information necessary for the data warehouse
- **Data source view** :exposes the information being captured, stored, and managed by operational systems
- **Data warehouse view** :consists of fact tables and dimension tables
- **Business query view** :sees the perspectives of data in the warehouse from the view of end-user

Data Warehouse Design Process

- **Top-down, bottom-up** approaches or a combination of both
 - Top-down: Starts with overall design and planning (mature)
 - Bottom-up: Starts with experiments and prototypes (rapid)
- From software engineering point of view
 - Waterfall: structured and systematic analysis at each step before proceeding to the next
 - Spiral: rapid generation of increasingly functional systems, short turn around time, quick turn around
- Typical data warehouse design process
 - Choose a business process to model, e.g., orders, invoices, etc.
 - Choose the grain (*atomic level of data*) of the business process
 - Choose the dimensions that will apply to each fact table record
 - Choose the measure that will populate each fact table record

Data Warehouse Development: A Recommended Approach



2. On-Line Analytical Processing (OLAP) to On Line Analytical Mining (OLAM)

Or Why online analytical mining?

- **High quality of data in data warehouses**: DW contains integrated, consistent, cleaned data
- **Available information processing structure surrounding data warehouses** : ODBC, OLEDB, Web accessing, service facilities, reporting and OLAP tools
- **OLAP-based exploratory data analysis** :Mining with drilling, dicing, pivoting, etc.
- **On-line selection of data mining functions** : Integration and swapping of multiple mining functions, algorithms, and tasks

3. Data Warehouse Usage = data warehouse applications

- **Information processing** : supports querying, basic statistical analysis, and reporting using crosstabs, tables, charts and graphs
- **Analytical processing** (multidimensional analysis of data warehouse data) supports basic OLAP operations, slice-dice, drilling, pivoting
- **Data mining** (knowledge discovery from hidden patterns) supports associations, constructing analytical models, performing classification and prediction, and presenting the mining results using visualization tools