Decision Support, Data Warehousing, and OLAP

교과목 : 데이타베이스시스템

Contents

- Terminology : OLAP vs. OLTP
- Data Warehousing Architecture
- Technologies
- References

Decision Support and OLAP

- Information technology to help knowledge worker(executive,manager,analyst) make faster and better decisions

  >> What were the sales volumes by region and product category for the last year?

  >> Which orders should we fill to maximize revenues?

  >> Will a 10% discount increase sales volume sufficiently?

- ***On-line analytical processing(OLAP)*** is an element of ***decision support system(DSS).***

Evolution

- 60’s : Batch reports
  >> hard to find and analyze information
  >> inflexible & expensive, reprogram every new report
- 70’s : Terminal-based DSS and EIS(executive information system)
  >> still inflexible, not integrated with desktop tools
- 80’s : Desktop data access and analysis tools
  >> query tools, spreadsheets, GUIs
  >> easier to use, but only access operational databases
- 90’s : Data warehousing with integrated OLAP engines and tools

What is OLAP?

- ***A category of software technology*** that enables analysts,managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user

  - from The OLAP Council's Definition

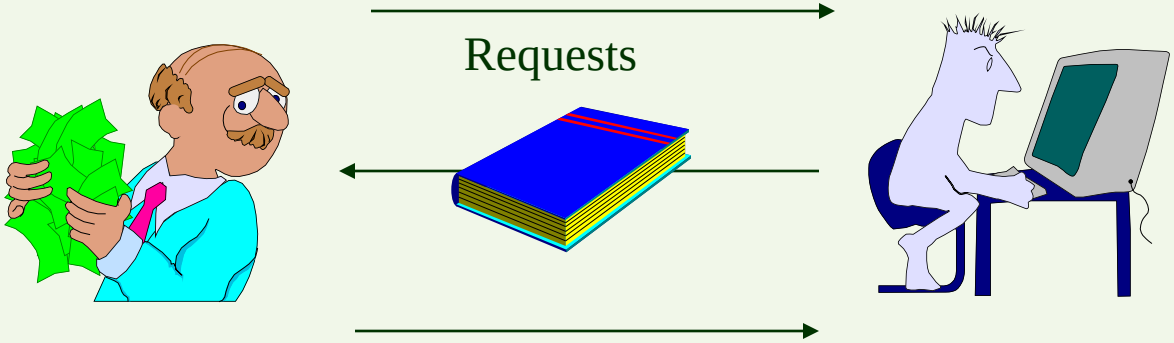
- ***Goal of OLAP is to support ad-hoc querying for the business analyst***
- ***Multidimensional view*** of data is the foundation of OLAP

OLTP vs. OLAP

	OLTP	OLAP
User	Clerk, IT professional	Knowledge worker
Function	Day to day operations	Decision support
DB design	Application-oriented (ER-based)	Subject-oriented (Star, Snowflake)
Data	Current, Isolated	Historical, Consolidated
View	Detailed, Flat relational	Summarized, Multidimensional
Usage	Structured,Repetitive	Ad hoc
Unit of work	Short,simple transaction	Complex query
Access	Read/Write	Read mostly
Operations	Index/hash on prim.key	Lots of scans
# Records accessed	Tens	Millions
#Users	Thousands	Hundreds
Db size	100MB-GB	100GB-TB
Metric	Trans.throughput	Query.throughput, response

DSS(Decision Support System)

- A system used ***to support managerial decisions.*** Usually DSS involves the analysis of many units of data in a heuristic fashion. As a rule, DSS processing does not involve the update of data



< Legacy DSS cycle >

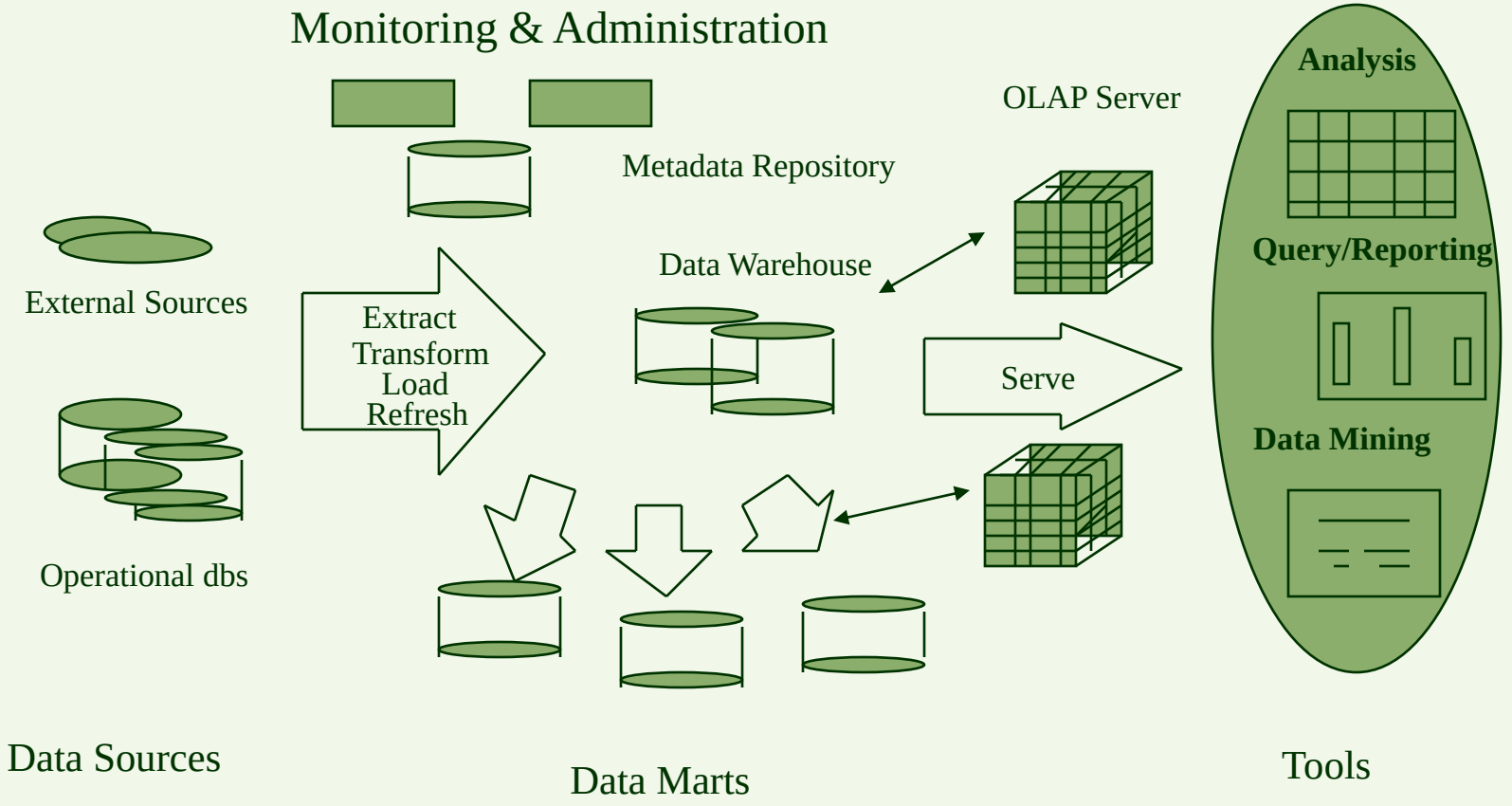
Data Warehouse

- A ***decision support database*** that is maintained separately from the organization's operational databases
- A data warehouse is a
  - subject-oriented
  - integrated
  - time-varying
  - non-volatile

  collection of data that is used primarily in organizational decision making

- W.H.Inmon, Building the Data Warehouse,1992

Data Warehousing Architecture



Three-Tier Architecture

- Warehouse database server
  >> almost always a relational DBMS ; rarely flat files
- OLAP servers
  >> Relational OLAP(ROLAP) : extended relational DBMS that maps operations on multidimensional data to standard relational operations
  >> Multidimensional OLAP(MOLAP) : special purpose server that directly implements multidimensional data and operations
- Clients
  >> Query and reporting tools
  >> Analysis tools
  >> Data mining tools(e.g., trend analysis, prediction)

OLAP Server Requirements

- Data Types
  >> text, calendar, numeric
- Calculation Engine
  >> domain-specific computations(calendar)
  >> rich library of aggregate functions
- Data at different granularities
  - estimated and actual sale for each product
- Data Load and Refresh
  >> write must update precomputed aggregates
  >> write/load must be accompanied by data cleaning

Data Mart & Data Mining

- ***Enterprise warehouse*** : collects all information about subjects (customers, products,personnel) that span the ***entire organization***

  >> Requires extensive business modeling

  >> May take years to design and build

- ***Data Marts*** : ***Departmental subsets*** that focus on selected subjects : Marketing data mart : customers, products, sales.

  >> Faster roll out, but complex integration in the long run

- ***Data Mining*** : New technology that ***recognizes patterns*** in the data to help you describe existing data and ***predict future behaviors*** based on current characteristic

Metadata Repository

- ***Administrative metadata***
  - **source databases and their contents**
  - **warehouse schema, view&derived data definition**
  - **dimensions, hierarchies**
  - **data extraction, cleansing, transformation rules, defaults**
  - **security : user authorization, access control**
  - **etc...**
- ***Business data***
  - **business terms and definitions**
  - **ownership of data**
  - **charging policies**
- ***Operational metadata***
  - **data lineage : history of migrated data and sequence of transformations applied**
  - **currency of data :active, archived , purged**
  - **monitoring information : warehouse usage statistics, error reports, audit trails**

Multidimensional Data Model

- Database is a set of ***facts(points)*** in a multidimensional space
- A fact has a ***measure*** dimension

  >> quantity that is analyzed, e.g., sales, budget
- A set of ***dimensions*** wrt which data is analyzed

  >> e.g., store, product, date, associated with a sales amount
- Dimensions form a sparsely populated coordinate system
- Each dimension has a set of ***attributes***

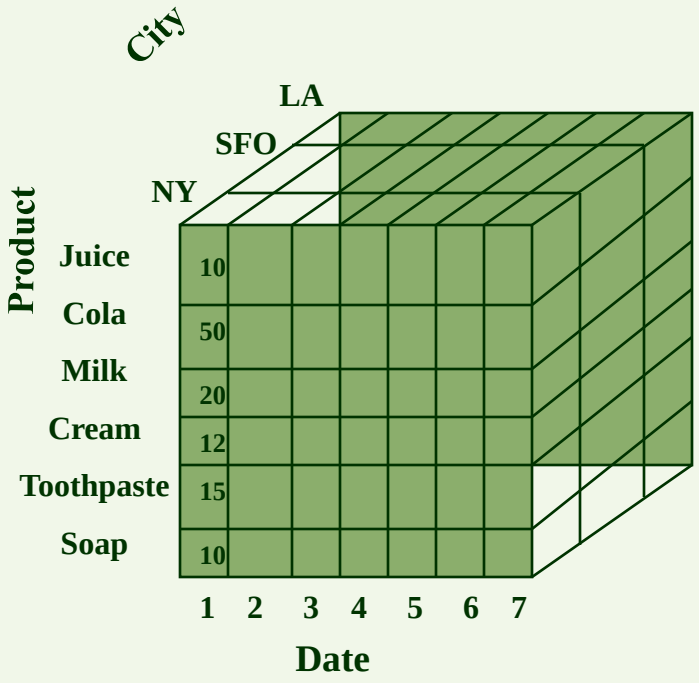
  >> e.g., owner, city and country of store
- Attributes of a dimension may be related by partial order

  >> ***Hierarchy*** : e.g.,street ->> country ->> city

  >> ***Lattice*** : e.g., date -> month -> year, date -> week->year

Multidimensional Data

- Sales volume as a function of product, time, and geography



Fact data : Sales volume in $100

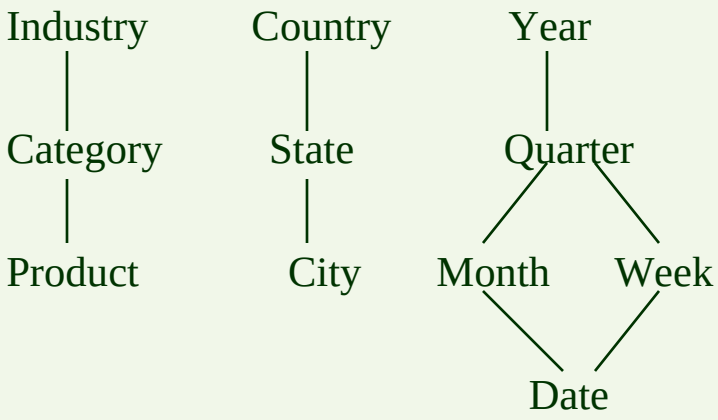
Dimensions

Product, City, Date

Attributes

Product(upc, price, package)

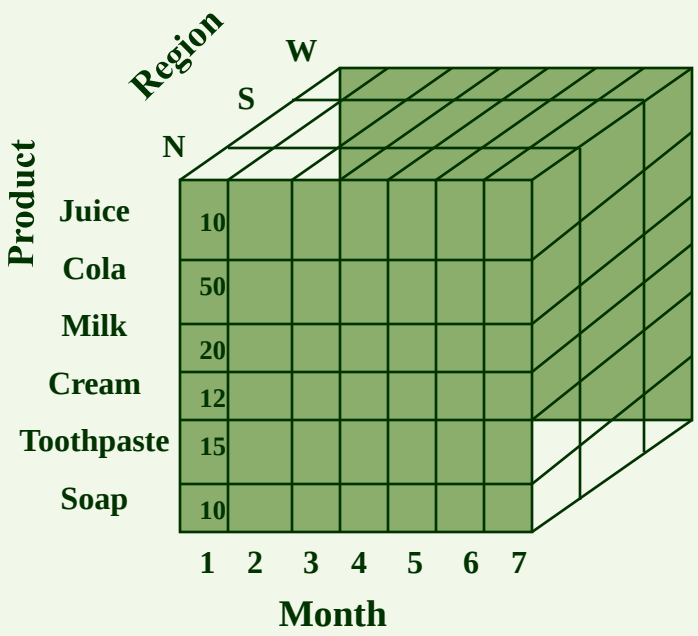
Attribute Hierarchies and Lattice

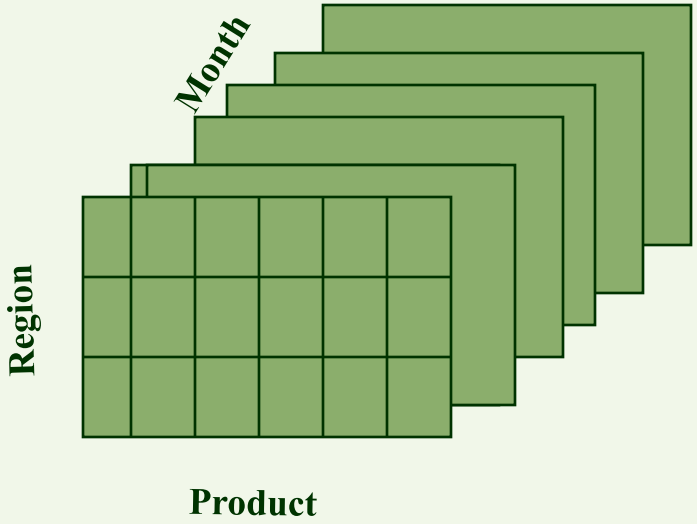


Operations in Multidimensional Data Model

- Aggregation (***roll-up***):
  - dimension reduction : e.g., total sales by city
  - summarization over aggregate hierarchy : e.g., total sales by city and year -> total sales by region and by year
- Selection (***slice***) defines a subcube
  - e.g., sales where city = Palo Alto and date = 1/15/96
- Navigation to detailed data (***drill-down***)
  - e.g., show supporting data for total sales figure for NW region
- Calculation and ranking
  - e.g., (sales-expense) bt city, top 3% of cities bt average income
- Visualization Operations(e.g., Pivot)

A Visualization Operation : Pivot (Rotate)



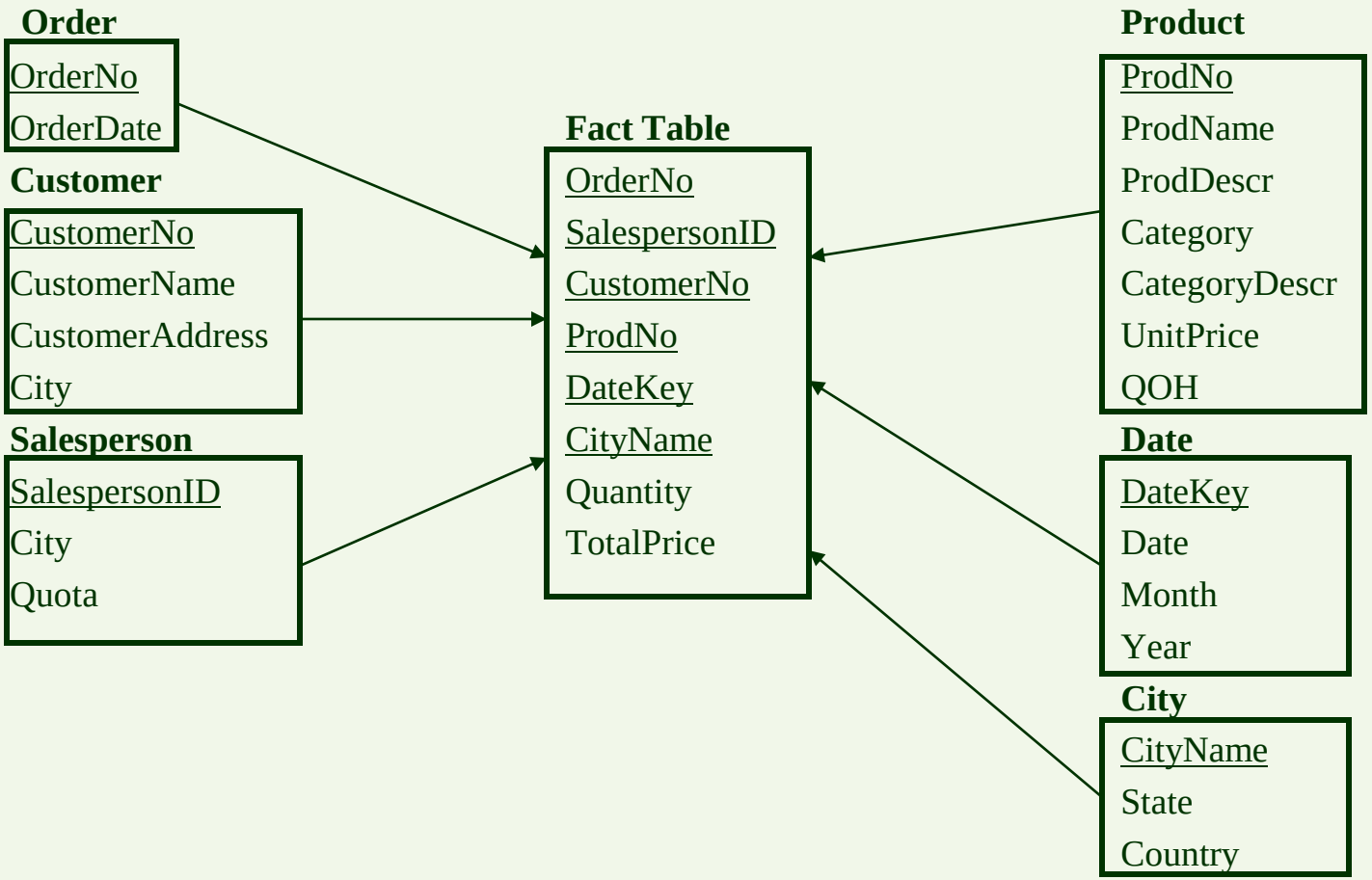


Fact data : Sales volume in $100

Warehouse Database Schema

- ER design techniques not appropriate
- Design should reflect multidimensional view
  - >> ***Star Schema***
  - >> ***Snowflake Schema***
  - >> ***Fact Constellation Schema***

Example of Star Schema

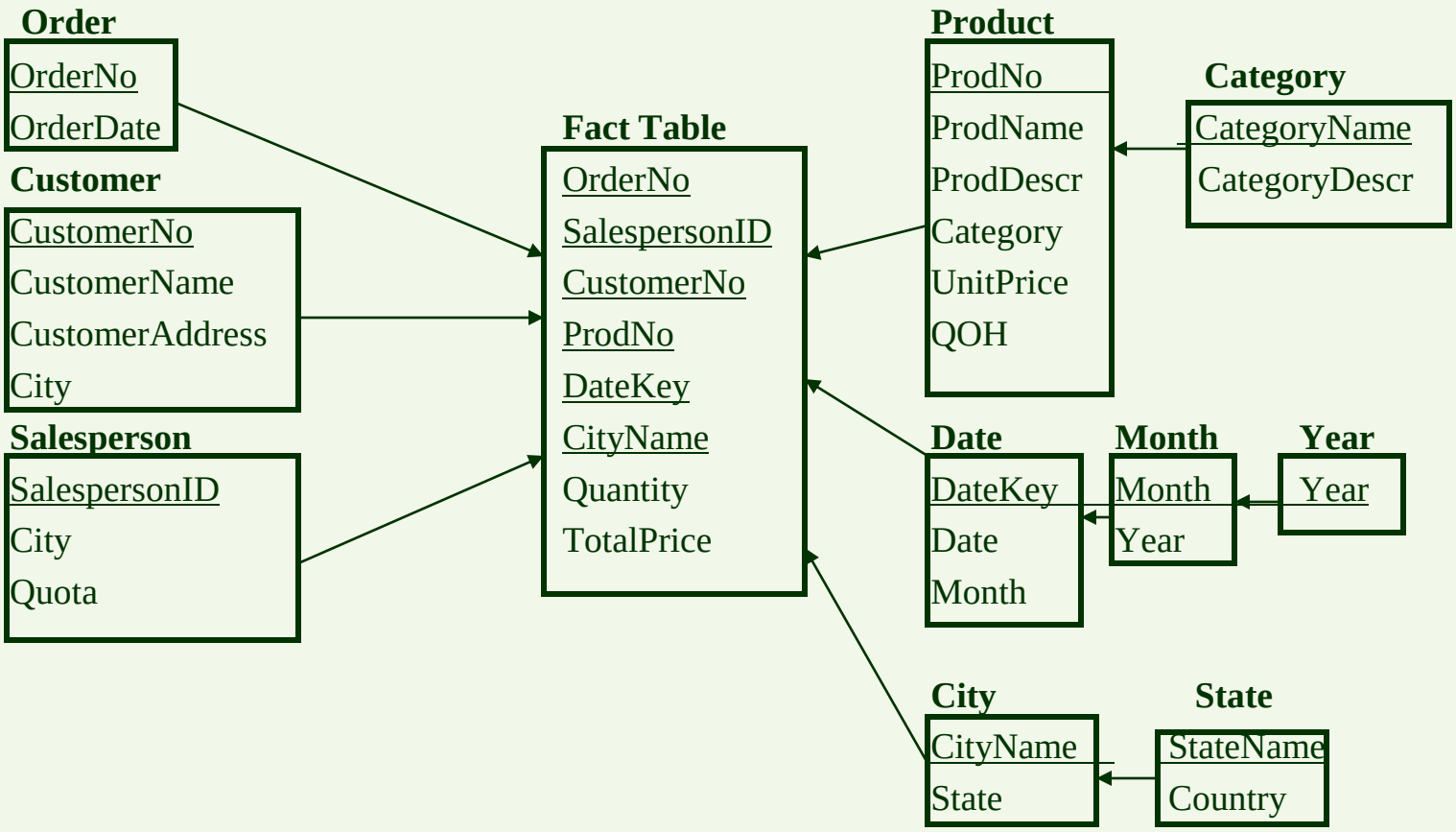


Star Schema

- A single fact table and a single table for each dimension
- Every fact points to one tuple in each of the dimensions and has additional attributes
- Does not capture hierarchies directly
- Generated keys are used for performance and maintenance reasons
- ***Fact Constellation***: Multiple Fact tables that share many dimension tables

  >> Example : Projected expense and actual expense may share dimension tables

Example of a Snowflake Schema



Snowflake Schema

- Represent dimensional hierarchy directly by normalizing the dimension tables
- Easy to maintain
- Saves storage, but it is alleged that it reduces effectiveness of browsing (Kimball)

References

- W.H. Inmon: Building the Data Warehouse(2nd Edition). John Wiley, 1996
- R.Kimball : The Data Warehouse Toolkit, John Wiley, 1996
- E.F. Codd, S.B. Codd, C.T. Salley : Providing OLAP(On-line Analytical Processing) to User Analysts : An IT Mandate. Available from Arbor Software's web site
- Articles in trade journals : Datamation, Databased Advisor, Database Programming and Design, DBMS Magazine.
- Web sites of all product vendors.
- Web site of the OLAP Council.

  Http://pwp.starnetic.com/larryg/articles.html has many references.

  Http://www.olapcouncil.org/research/whitepapco.htm