بنام خدا

مبانی فناوری اطلاعات

Business Intelligence

هوش تجاری

بصیری
دانشکده برق و کامپیوتر
دانشگاه صنعتی اصفهان

مراجع

- Han, Jiawei, Micheline Kamber, and Data Mining. "Concepts and techniques." *Morgan Kaufmann* 340 (2006): 94104-3205.
- Kimball, Raiph. *The data warehouse toolkit*. John Wiley & Sons, 2006.
- Inmon, William H. *Building the data warehouse*. John wiley & sons, 2005.

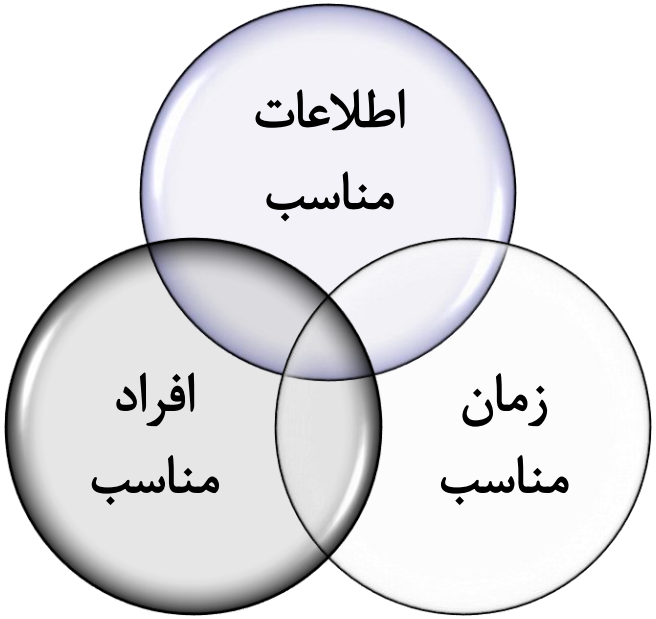
Outline

- Business Intelligence
  - Definition
  - Important
  - History
  - Architecture
- Data Warehouse
  - Definition
  - Characteristics
  - OLAP
  - Conceptual Modeling
  - Cube

هوش تجاری چیست؟

- 1989 Howard Dresner (Gartner Group)
  - مجموعه ای از مفاهیم و متدها جهت بهبود تصمیم سازی تجاری با بکارگیری سیستم های حقیقت-محور
    - ابزار ETL
    - ابزار OLAP
    - ابزار گزارشگیری
- **جمع‌آوری، نگهداری، تجزیه و تحلیل و انتشار داده/اطلاعات به منظور تسهیل در تصمیم‌گیری‌های مدیریتی.**

هوش تجاری (Business Intelligence)

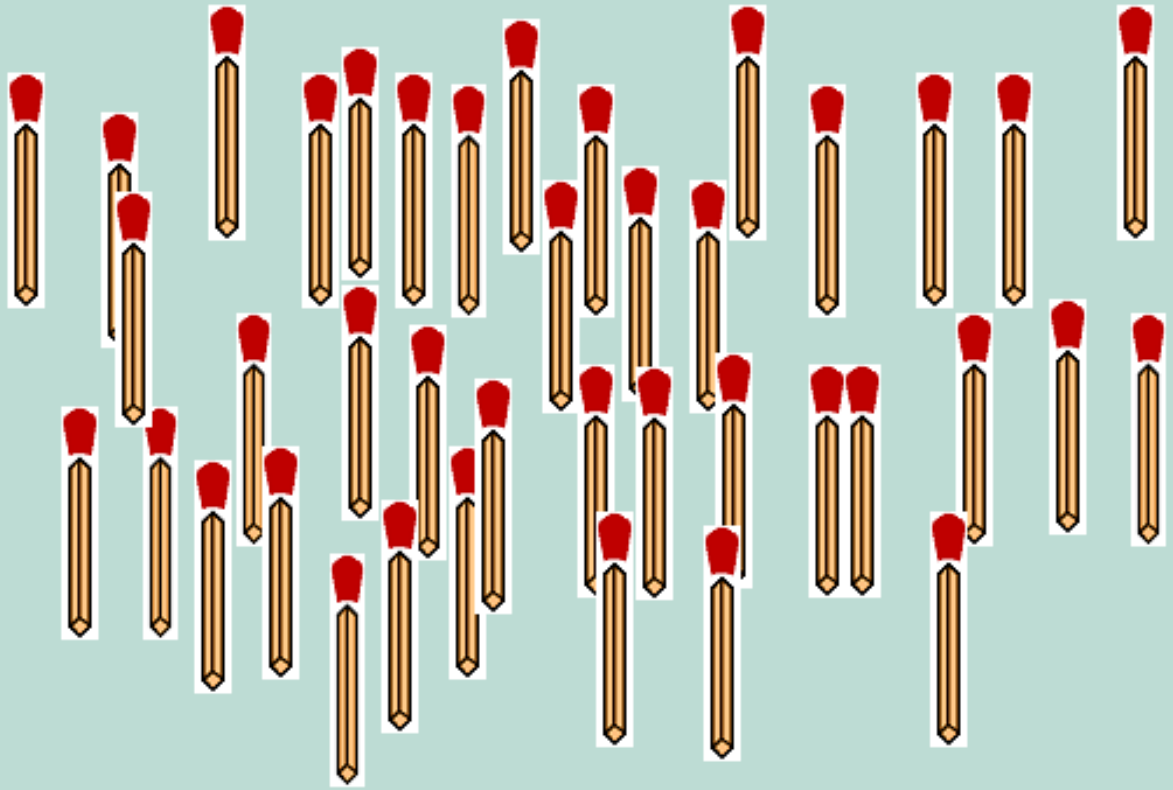


هوش تجاری یعنی دراختیار قراردادن **اطلاعات مناسب** به **افراد مناسب** در **زمان مناسب** برای اخذ **تصمیم مناسب**

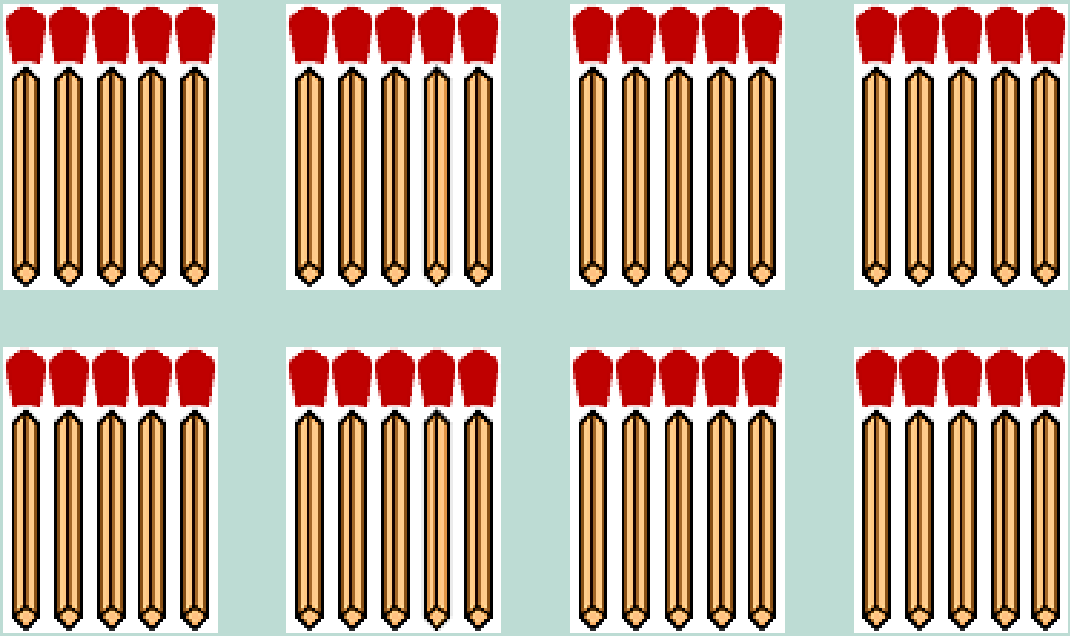
چرا به انبار داده و هوش تجاری نیاز داریم؟

- نیاز به یک چارچوب تصمیم‌گیری برای تسریع در اتخاذ تصمیمات موثر
- **دسترسی سریع به تمام اطلاعات موجود**
- **امکان ساخت هر گونه گزارش به صورت دینامیک**
- **امکان تحلیل اطلاعات به صورت Historical(مبتنی بر زمان) از کل به جزء(Drill Down) با استفاده از تجمیع ها(Aggregation)**
- **امکان انجام تحلیلهای آماری و مبتنی بر داده کاوی**
- **ایجاد مرجع واحد آمار و گزارشات**
- **امکان برنامه نویسی و تولید نرم افزارهای خاص منظوره**

تعداد چوب کبریت ها چند تاست؟



تعداد چوب کبریت ها اکنون چند تاست؟



ابزارهای هوش تجاری

- BizzScore Suite
- Board Management IntelligenceToolkit
- Business Objects Enterprise
- IBM Cognos
- JasperSoft
- Microsoft BI tools
- Microstrategy

- Oracle
- WebFocus
- Tableau Software
- Style Intelligence
- SAS
- SAP
- QlikView
- Pentaho BI Suite
- Actuate
- .....

History of Business Intelligence

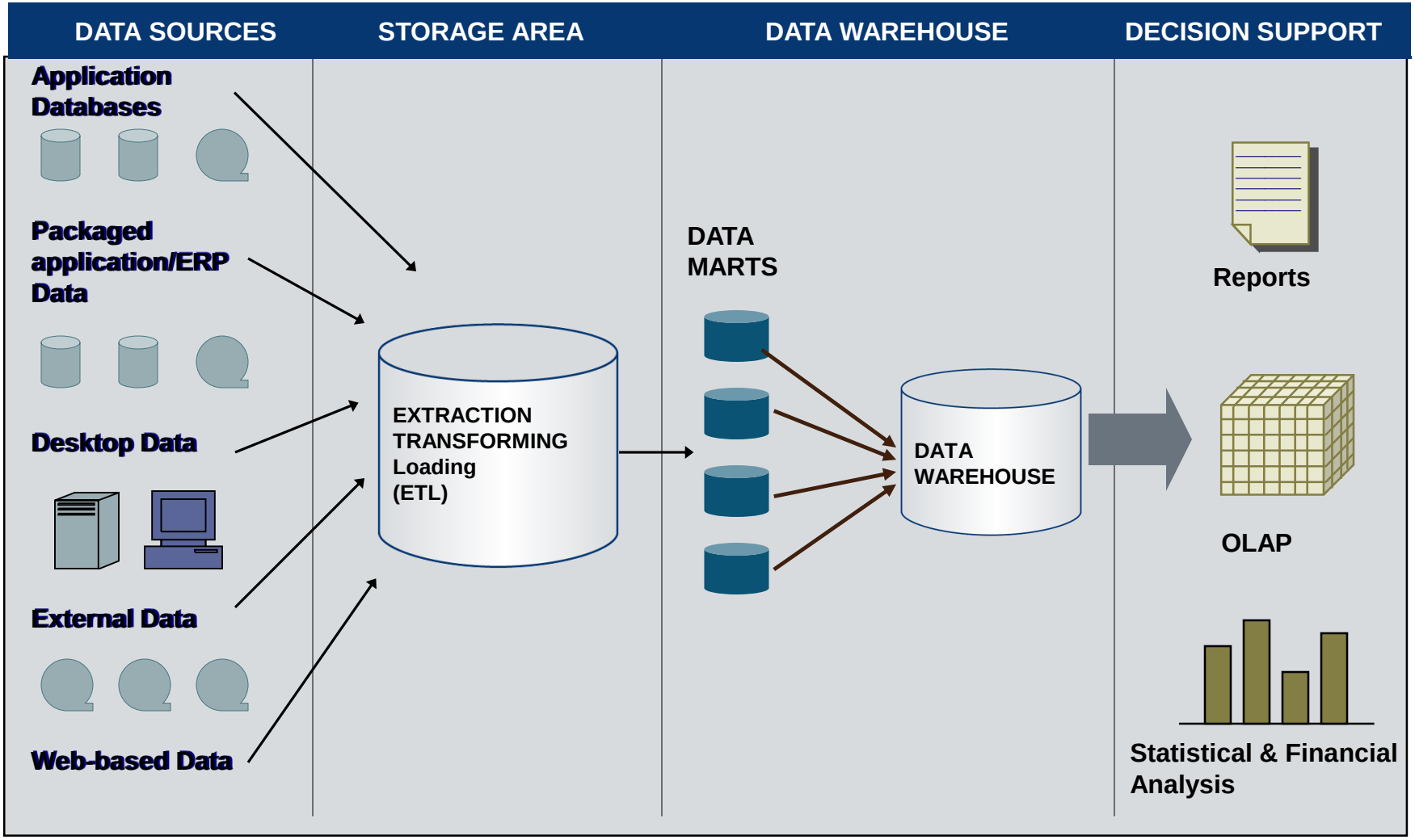
Previous Names

- Decision Support Systems
- Executive Information Systems
- Online Analytic Processing

Related Concepts

- Competitive Intelligence
- Market Intelligence
- Customer Intelligence
- Strategic Intelligence
- Technical Intelligence
- Knowledge Management
- Data Miming

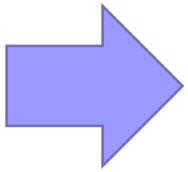
BI Architecture



What’s a Data Warehouse?

انبار داده یک منبع واحد از داده‌هاست که داده‌های منابع مختلف اطلاعاتی سازمان در آن جمع آوری، دسته‌بندی، خلاصه‌سازی و ذخیره شده تا تصمیم‌گیری را در سازمان تسهیل نماید.

- برای دسترسی آسان کاربران به حجم زیادی از داده‌ها طراحی شده است، و دسترسی به دیتا عموما به وسیله ابزارهای تحلیلی ویژه و اپلیکیشن‌ها پشتیبانی می‌شود.



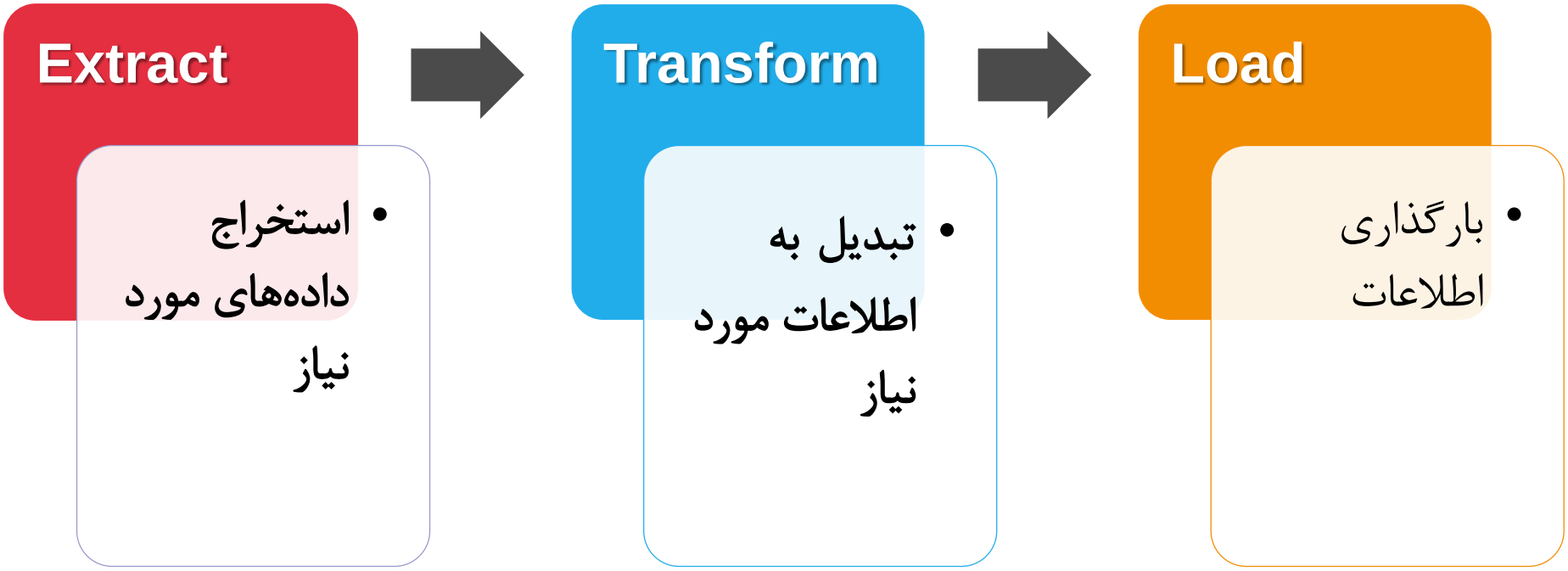
ابزارهای OLAP دسترسی به انبارداده را تسهیل نموده است.

Extraction, Transformation, and Loading (ETL)

- The process of data consolidation is often called Extraction, Transformation, and Loading (ETL)
  - The ETL process *extracts* data from the various source systems
  - Data is then *transformed* to make it consistent and improve data quality
  - The consolidated, consistent, and cleaned data is then *loaded* into a data repository
- Developing the ETL process often consumes 80% of the development time

Extraction, Transformation, and Loading (ETL)

- استخراج داده ها از منابع داده ای مختلف
- تحلیل و تبدیل داده‌ها مطابق با نیازمندی‌های کسب و کار
- بارگذاری اطلاعات



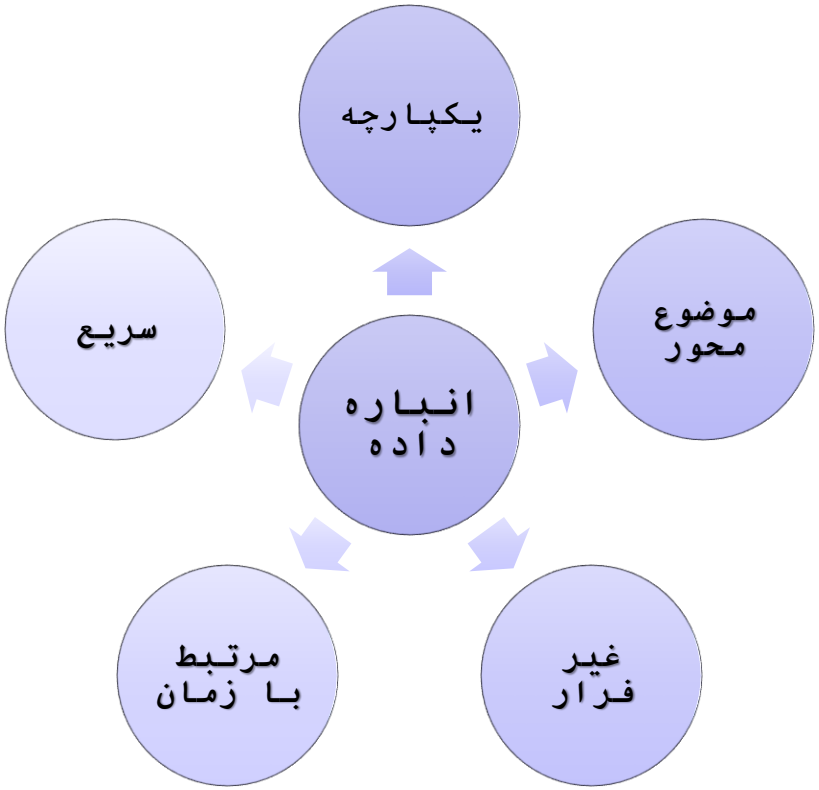
(ETL) Tools

- Some ETL Tools
  - Oracle Data Integrator (ODI)
  - Informatica
  - IBM Ascential
  - SSIS (Microsoft SQL SERVER)

Data Warehouse Characteristics

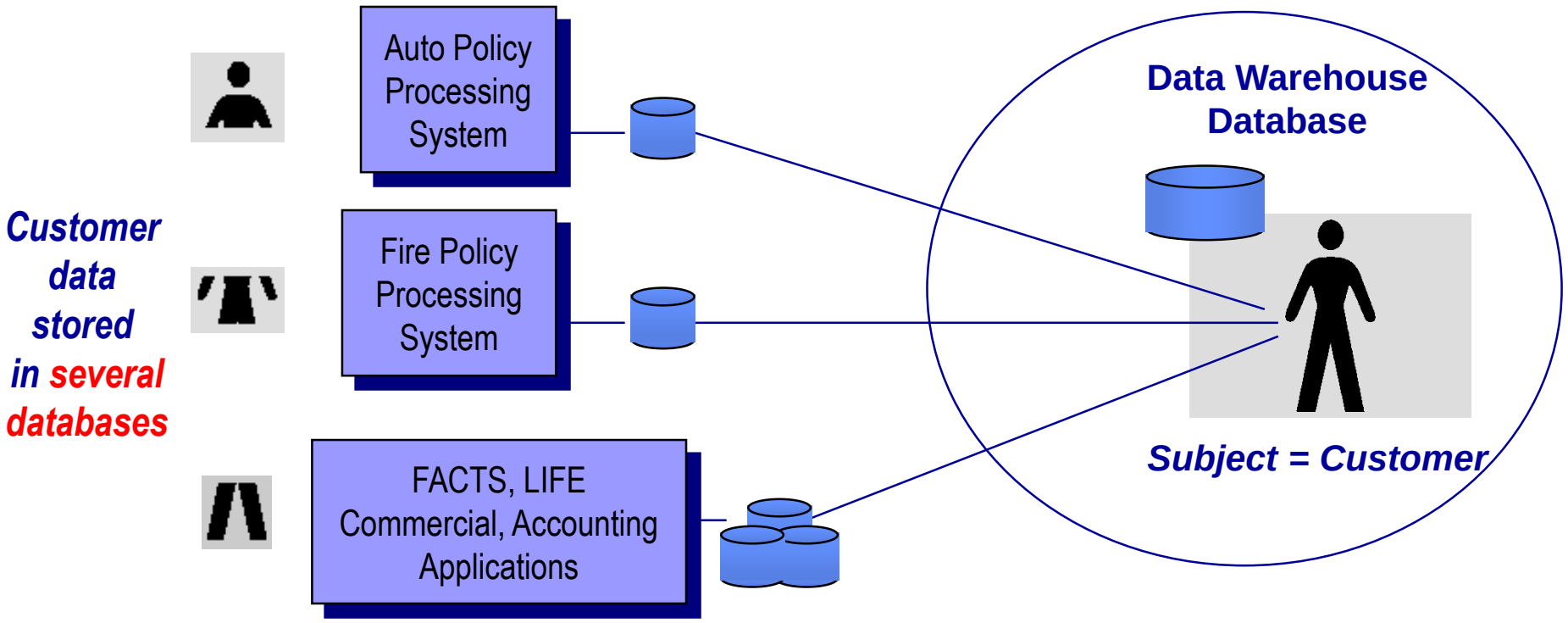
- Key Characteristics of a Data Warehouse

- Integrated
- Time-variant
- Non-volatile
- Subject-Oriented
- Fast



Integrated

- **Data is stored once in a single integrated location (e.g. insurance company)**

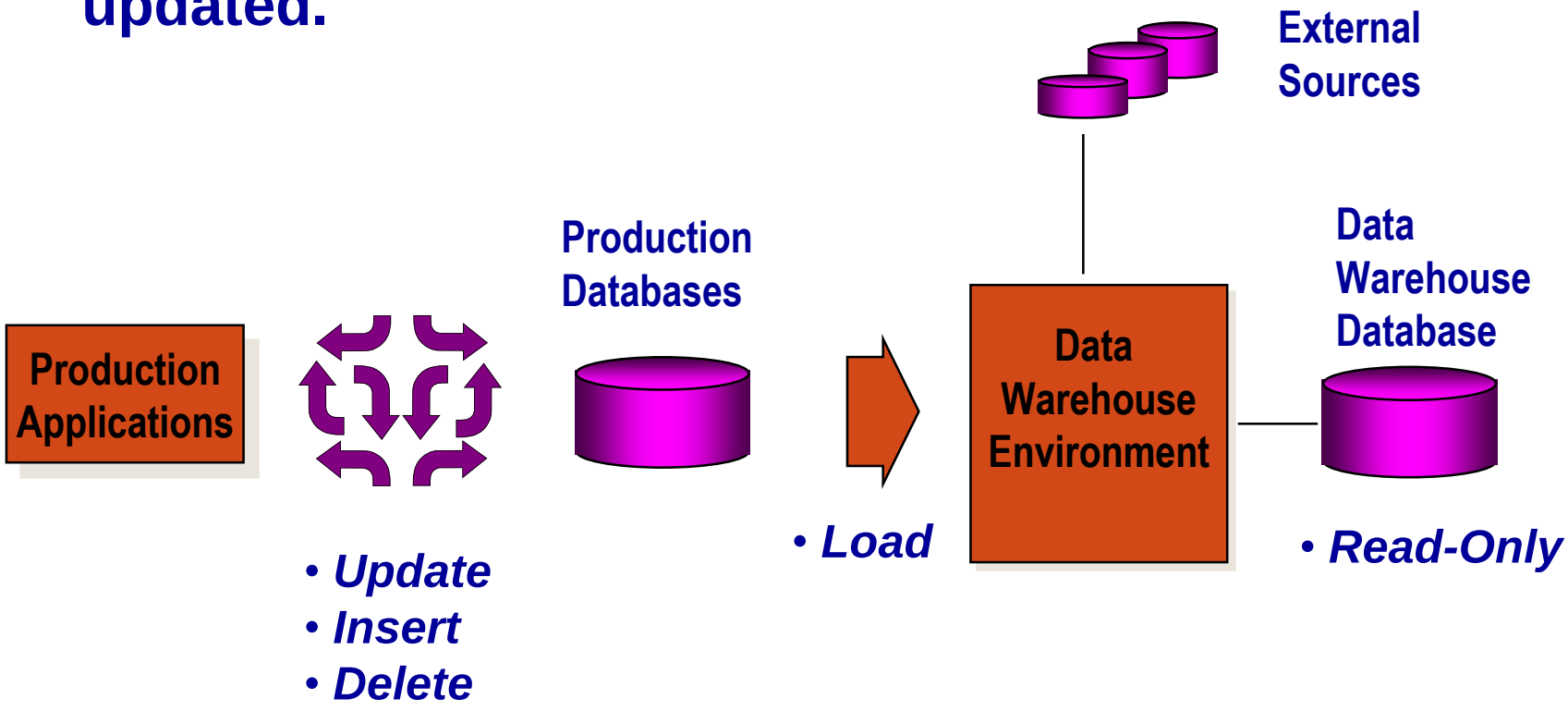


Data Warehouse—Time Variant

- The time horizon for the data warehouse is significantly longer than that of operational systems
  - Operational database: current value data
  - Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)
- Every key structure in the data warehouse
  - **Contains an element of time**, explicitly or implicitly
  - But the key of operational data may or may not contain "time element"

Non-Volatile

- **Existing data in the warehouse is not overwritten or updated.**



Data Warehouse—Subject-Oriented

- Organized around major subjects, such as customer, product, sales
- Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing
- Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process

Data Warehouse vs. Operational DBMS

- OLTP (on-line transaction processing)
  - Major task of traditional relational DBMS
  - Day-to-day operations: purchasing, inventory, banking, manufacturing, payroll, registration, accounting, etc.
- OLAP (on-line analytical processing)
  - Major task of data warehouse system
  - Data analysis and decision making
- Distinct features (OLTP vs. OLAP):
  - User and system orientation: customer vs. market
  - Data contents: current, detailed vs. historical, consolidated
  - Database design: ER + application vs. star + subject
  - View: current, local vs. evolutionary, integrated
  - Access patterns: update vs. read-only but complex queries

OLAP

- ✓ **OLAP (On-Line Analytical Processing)** با استفاده از تجمیع مقیاس‌ها و شکستن آنها بر روی ابعاد مختلف کسب‌وکار، یک شمای متناسب با نیازهای گوناگون بخش‌های مختلف در اختیار بانک قرار می‌دهد.
- ✓ **Cube** قالب نگهداری داده ها در مدل چند بعدی است.
- ✓ با استفاده از **Cube** داده ها به آسانی و به سرعت می تواند در اختیار کاربران غیر فنی قرار گیرد.
- ✓ در **Cube** اطلاعات به صورت تجمیع شده محاسبه و نگهداری می شوند و در زمان واکشی داده از **Cube** محاسبات انجام نمی گیرد.

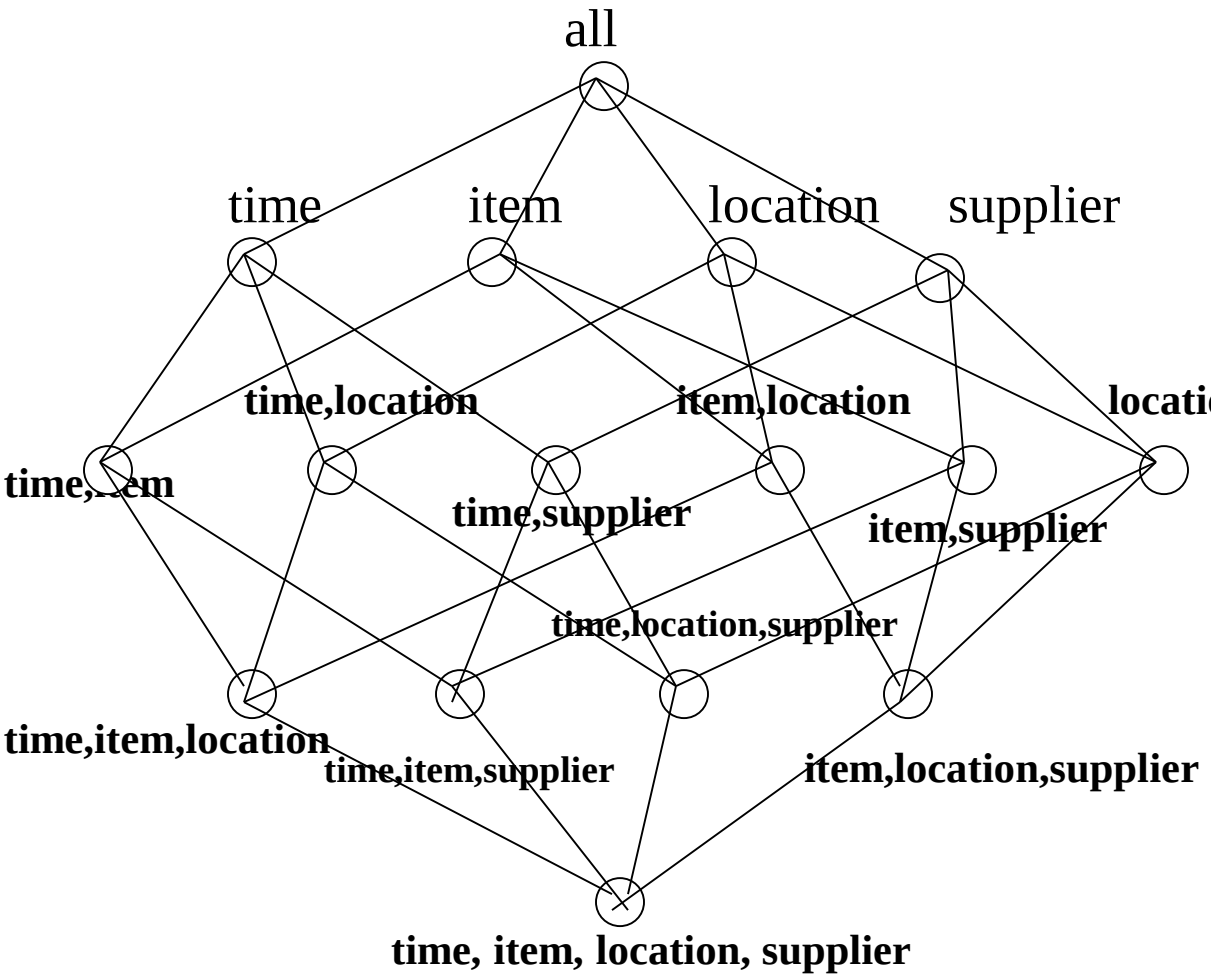
OLTP vs. OLAP

	OLTP	**OLAP**
users	clerk, IT professional	knowledge worker
function	day to day operations	decision support
DB design	application-oriented	subject-oriented
data	current, up-to-date detailed, flat relational isolated	historical, summarized, multidimensional integrated, consolidated
usage	repetitive	ad-hoc
access	read/write index/hash on prim. key	lots of scans
unit of work	short, simple transaction	complex query
# records accessed	tens	millions
#users	thousands	hundreds
DB size	GB to high-order GB	>=TB
metric	transaction throughput	query throughput, response

From Tables and Spreadsheets to Data Cubes

- A data warehouse is based on a multidimensional data model which views data in the form of a data cube
- A data cube, such as sales, allows data to be modeled and viewed in multiple dimensions
  - Dimension tables, such as item (item_name, brand, type), or time(day, week, month, quarter, year)
  - Fact table contains measures (such as dollars_sold) and keys to each of the related dimension tables
- In data warehousing literature, an n-D base cube is called a base cuboid. The top most 0-D cuboid, which holds the highest-level of summarization, is called the apex cuboid. The lattice of cuboids forms a data cube.

Cube: A Lattice of Cuboids



0-D(apex) cuboid

1-D cuboids

2-D cuboids

3-D cuboids

4-D(base) cuboid

Conceptual Modeling of Data Warehouses

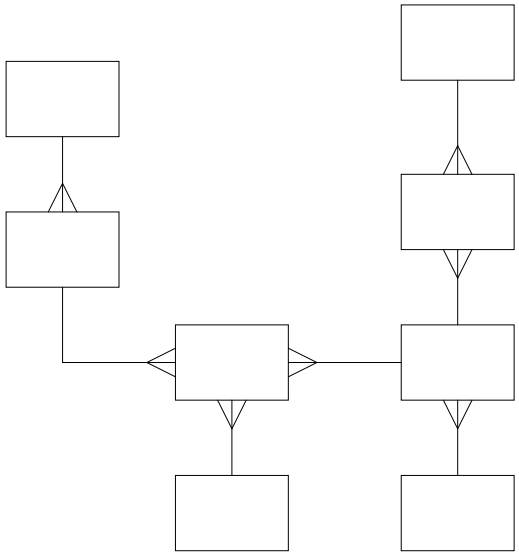
- Modeling data warehouses: dimensions & measures
  - Star schema: A fact table in the middle connected to a set of dimension tables
  - Snowflake schema: A refinement of star schema where some dimensional hierarchy is normalized into a set of smaller dimension tables, forming a shape similar to snowflake
  - Fact constellations: Multiple fact tables share dimension tables, viewed as a collection of stars, therefore called galaxy schema or fact constellation

Dimensional Modeling

- Dimensional modeling = data warehouse modeling technique
- 2 types of tables: facts and dimensions.
- A fact table contains one or more measures (usually numerical) of a subject that is being modeled for analysis.
- Dimension tables contain various descriptive attributes (usually textual) that are related to the subject depicted by the fact table.
- The intent of the dimensional model is to represent relevant questions whose answers enable appropriate decision making in a specific business area

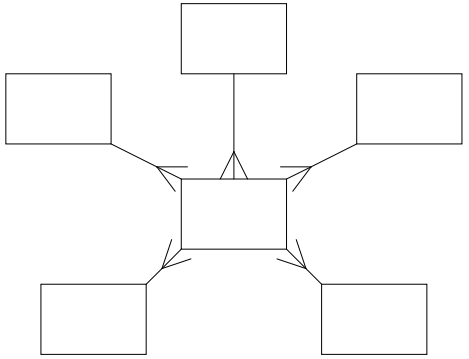
مدل ذخیره داده

Operational System



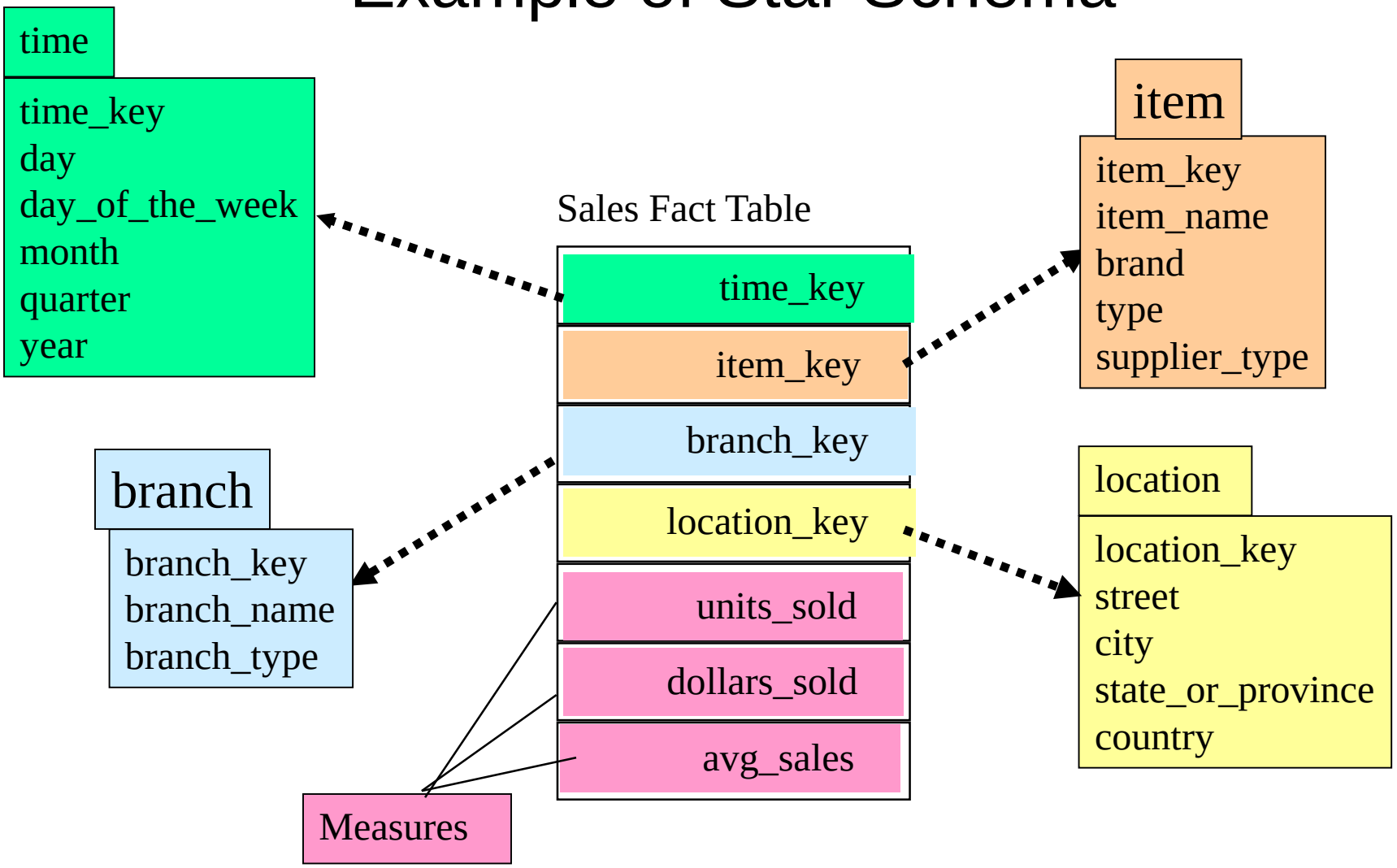
ER Diagram

Data Warehouse

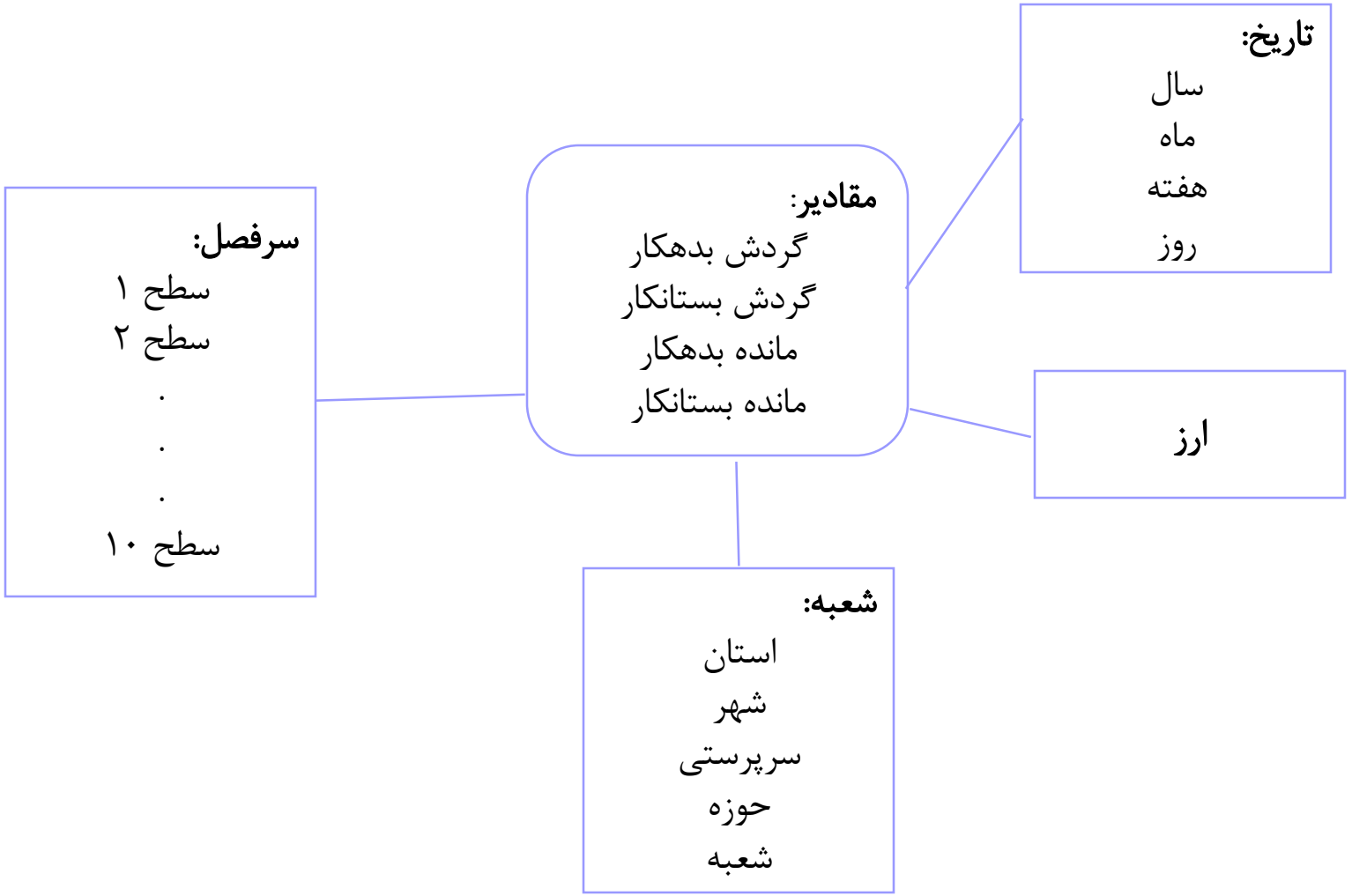


Star Schema

Example of Star Schema



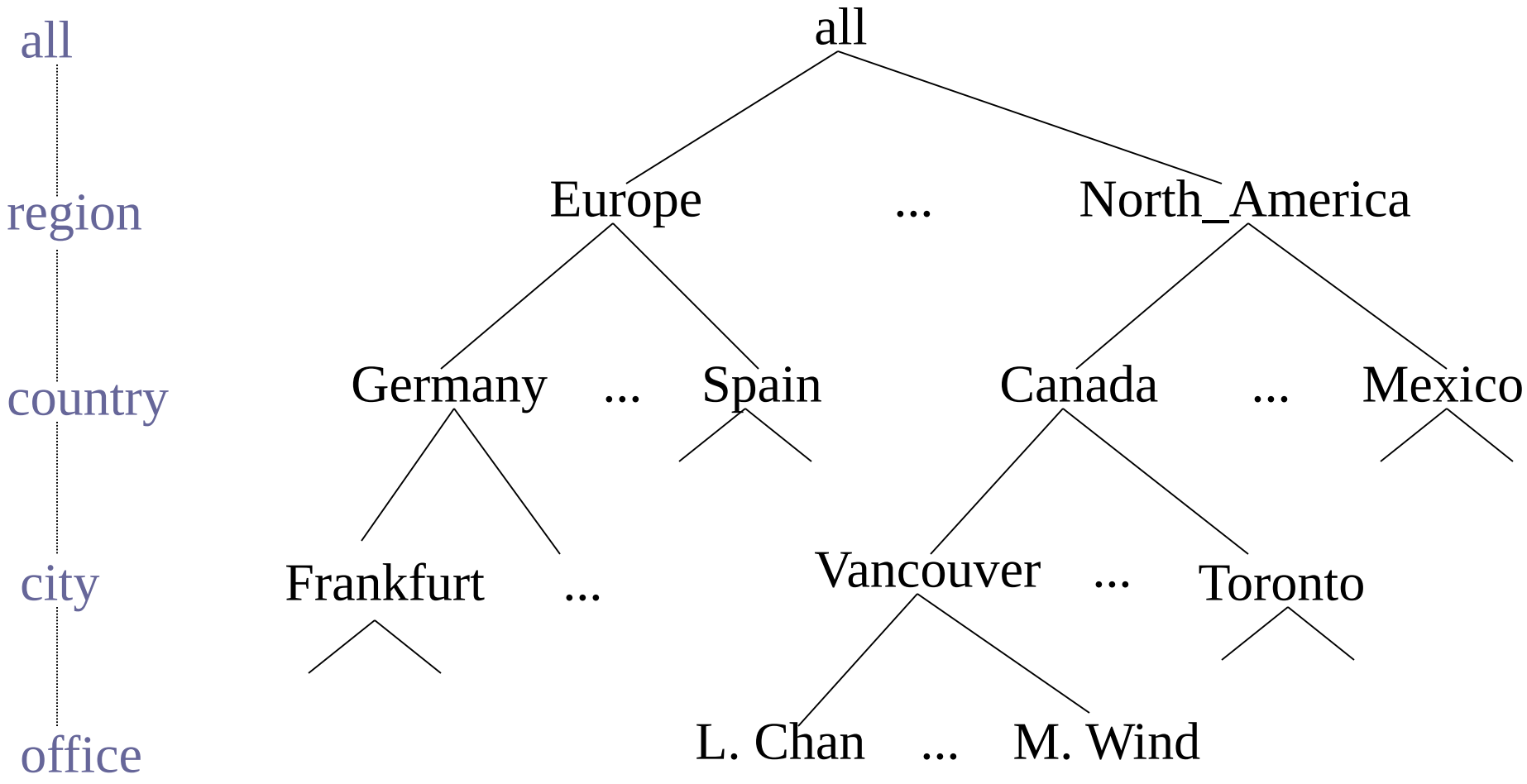
مثال مدل STAR



Measures of Data Cube: Three Categories

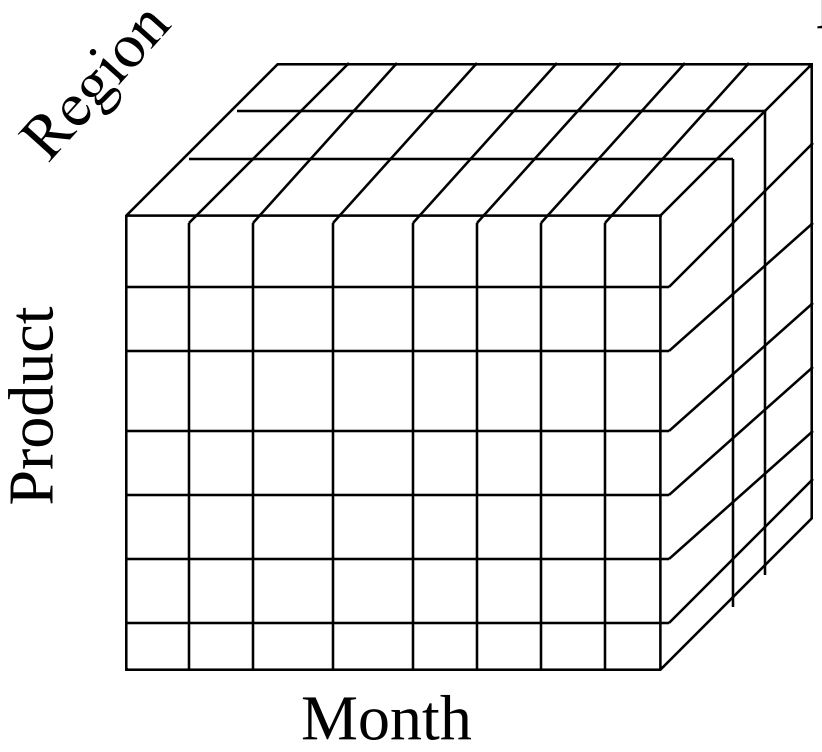
- Distributive: if the result derived by applying the function to *n* aggregate values is the same as that derived by applying the function on all the data without partitioning
  - E.g., count(), sum(), min(), max()
- Algebraic: if it can be computed by an algebraic function with *M* arguments (where *M* is a bounded integer), each of which is obtained by applying a distributive aggregate function
  - E.g., avg(), min_N(), standard_deviation()
- Holistic: if there is no constant bound on the storage size needed to describe a subaggregate.
  - E.g., median(), mode(), rank()

A Concept Hierarchy: Dimension (location)

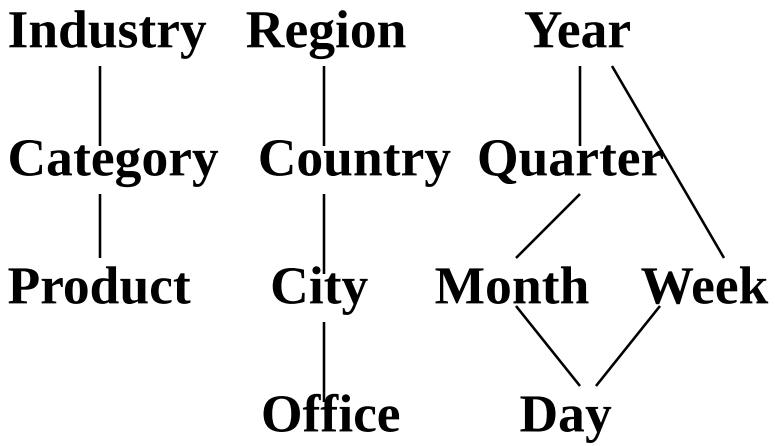


Multidimensional Data

- Sales volume as a function of product, month, and region



Dimensions: Product, Location, Time
Hierarchical summarization paths



Typical OLAP Operations

- Roll up (drill-up): summarize data
  - by climbing up hierarchy or by dimension reduction
- Drill down (roll down): reverse of roll-up
  - from higher level summary to lower level summary or detailed data, or introducing new dimensions
- Slice and dice: project and select
- Pivot (rotate):
  - reorient the cube, visualization, 3D to series of 2D planes

Typical OLAP Operations