

# Database Management System and Big Data

[Struab] chap 2  
[powell] chap 1

**Dr. Yeffry Handoko Putra, M.T**  
**Magister Sistem Informasi**

UNIVERSITAS KOMPUTER INDONESIA

# Example Low Quality Information

ID	Nama Awal	Nama Keluarga	Jalan	Kota	Propinsi	Kodepos	Telp	Fax	Email
113	Nita		Mangga 2	Bandung	Jabar	40331	(022) 4208828	(022)4208829	<a href="mailto:nita@aol.com">nita@aol.com</a>
114	Dian	Nova	40	Bandung	Jabar	40221	(022) 5106679	(022) 5106679	(022) 5106679
115	Nurdin	Komara	Salak 3	Garut	Jabar	40851	4205531	4205534	<a href="mailto:nk@hotmail.com">nk@hotmail.com</a>
116	Nurddin	Komara	Salak 3	Garut	Jabar	40851	4212001	4212030	<a href="mailto:nk@hotmail.com">nk@hotmail.com</a>

3. Probable duplicate information (similar names, same address, phone number)

4. Potential wrong information (are the phone number and fax number the same or is this a error?)

6. Incomplete information (missing area codes)

# Five Common Characteristic of High Quality Information

Accurate

- Is There an incorrect value in the information?
- Example: Is the name spelled correctly ? Is the money amount recorded properly

Complete

- Is The value missing from the information
- Example : is the address complete including street, city, state, and zip code

Consistent

- Is aggregate or summary information In agreement with detailed

Timely

- Is the information current with respect to business needs

Unique

- Is each transaction and event represented only once in the information

### **Information level**

Individual, department, enterprise

Individual knowledge, goal  
and strategies

### **Information formats**

Document, presentations, spreadsheets,  
database

Letters, memo, faces, email,  
reports  
Product, Strategy, process,  
financial

### **Information Granularities**

Detail (fine), summary, Aggregate (coarse)

Reports for each  
salesperson, product, and  
part  
Report for sales personnel,  
all product and all parts

# Big Data Problem

- ❖ In 2000 25% of Information was digital, the rest were on paper, cassette, film
- ❖ Today
  - about 1300 exabytes of information exists in the world and only 2% is non-digital and 90% is unstructured.
  - Google processes > 24 petabytes of data every day
  - Face book 10 million photographs every hour
  - more than 400 million tweets per day

# Database Demand on Big Data

- ❖ Large user population
- ❖ Velocity → Very high Access )
- ❖ Volume → Large database size
- ❖ Granularity
- ❖ Venue → MapReduce
- ❖ Manageability (Hadoop, PIG)
- ❖ Veracity → Accuracy level

**Characteristics of Big Data:** Big data has the following characteristics:

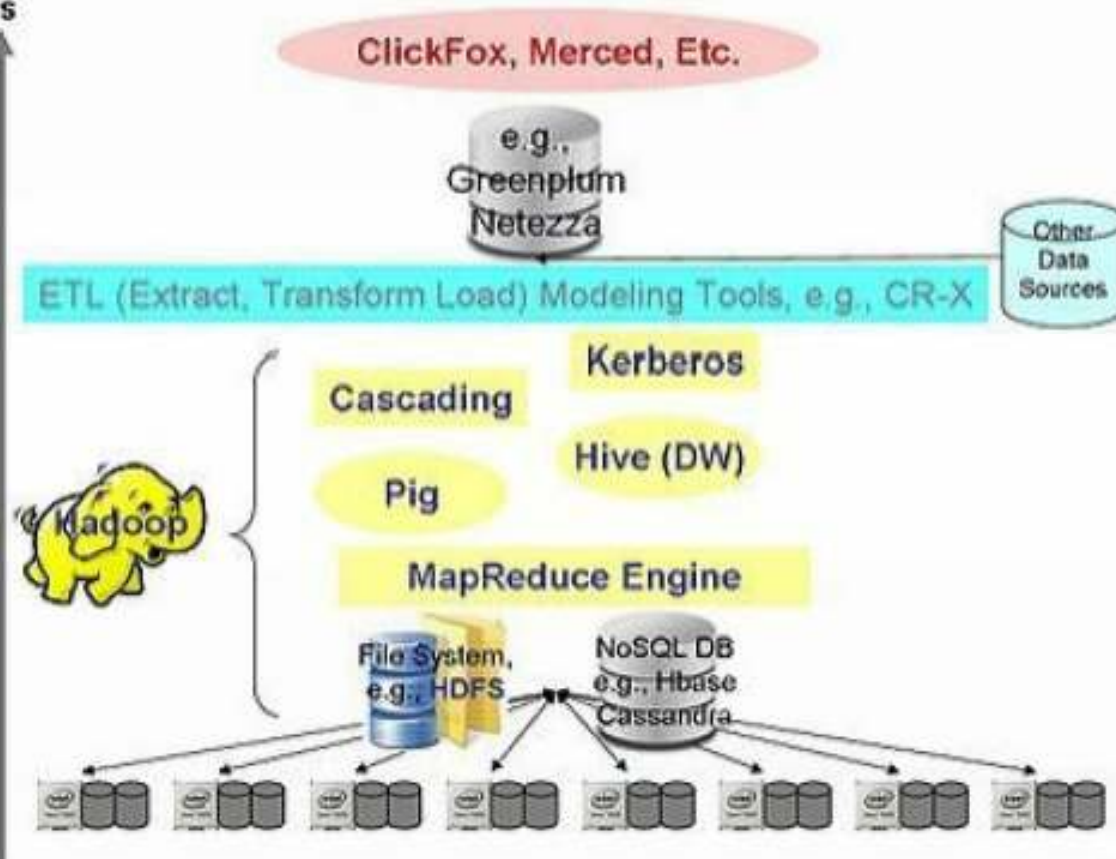
- Very large, distributed aggregations of loosely structured data – often incomplete and inaccessible:
- Petabytes/exabytes of data,
- Millions/billions of people,
- Billions/trillions of records,
- Loosely-structured and often distributed data,
- Flat schemas with few complex interrelationships,
- Often involving time-stamped events,
- Often made up of incomplete data,
- Often including connections between data elements that must be probabilistically inferred,



# Analytic Big Data Components

## Abstraction Layers

- Analytic Applications
- Fast-loading Analytic Database
- Modeling
- Management & Security
- Higher Level Languages
- Job & Task Trackers
- Location-aware File Systems
- Processing & Original Data





# Database Design

- ❖ Current System Analysis and Survey
- ❖ Logical Database Design
- ❖ Physical Database Design
- ❖ Implementation
- ❖ Testing and Debugging
- ❖ Training, Evaluation and Documentation

# Function of Database

- ❖ Transactional, e.g :
  - Client-server Database
  - OLTP Database
  - Transactional Database
- ❖ Decision Support System. :
  - Forecasting
  - Executive Information System
  - Office Automation System
- ❖ Hybrid

# Model Data

- ❖ Relational Database
- ❖ Hierarhical Database
- ❖ Network Database
- ❖ Object-oriented Database

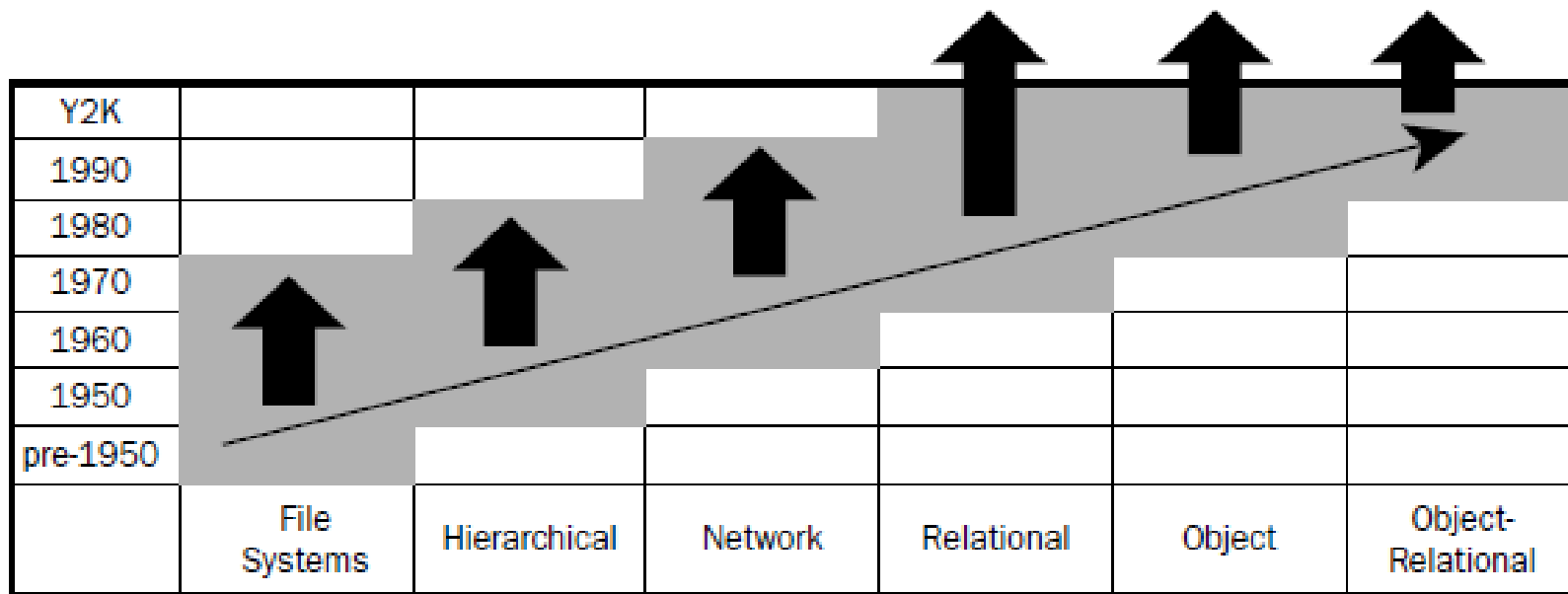


Figure 1-3: The evolution of database modeling techniques.

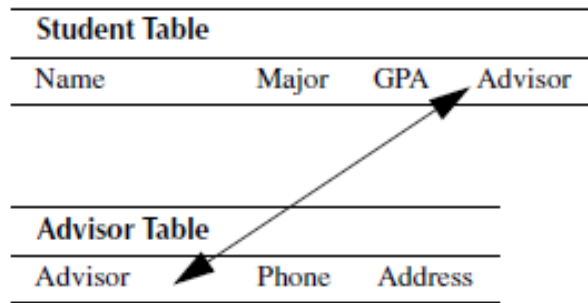


FIGURE 6.2 A relational database.

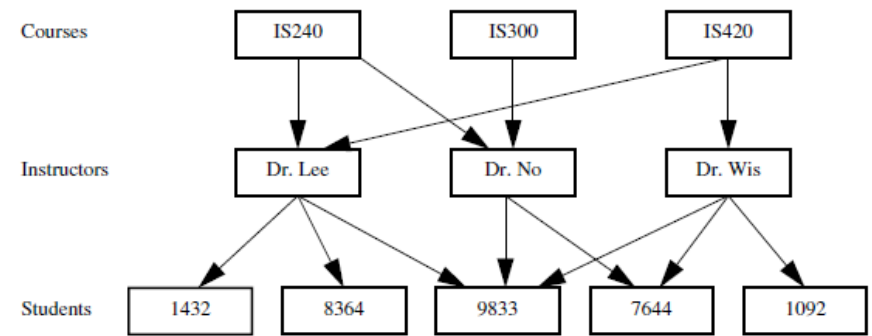


FIGURE 6.4 A network database.

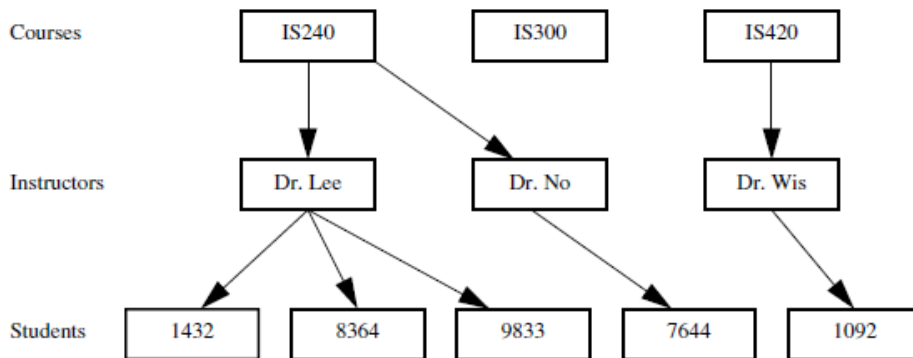


FIGURE 6.3 A hierarchical database.

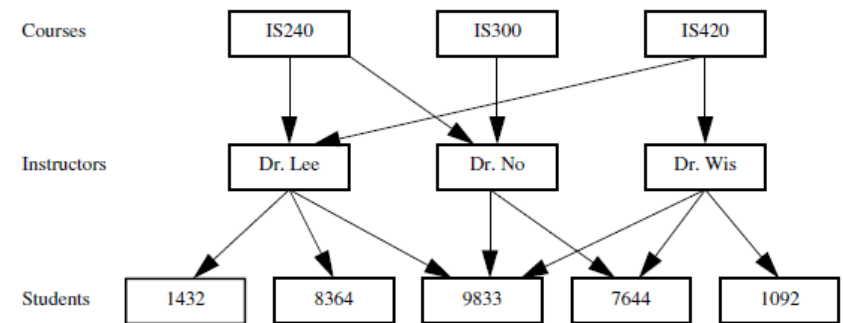


FIGURE 6.4 A network database.

# Benefit Database Management System

- ❖ Mengurangi Redudansi Data
- ❖ Meningkatkan Integritas Data
- ❖ Meningkatkan keamanan Data
- ❖ Mengurangi waktu pengembangan

## Penentuan Integritas Data

- ❖ Penentuan Dependency yang jelas:  
Primary Key, Secondary Key, Foreign Key
- ❖ Keterkaitan antara Desain Konseptual (ERD), dengan desain logika
- ❖ Tuning Phase: Evaluasi database seperti pemberian index, normalisasi lanjutan, denormalisasi, sistem keamanan data,



# CIA's Principle in Data Security

- ❖ **Confidentiality**
- ❖ **Integrity**
- ❖ **Availability**

# STRIDE in Data Security

- ❖ **S**poofing : Penipuan identitas  
Contoh: pada http authentication secara basic atau digest (RFC2617)
- ❖ **T**ampering: Modifikasi Data
- ❖ **R**epudiation: Transaksi yang benar disangkal oleh vendor
- ❖ **I**nformation Disclosure: Informasi rahasia terbuka
- ❖ **D**enial of Service: membuat server tidak berfungsi
- ❖ **E**levation of Privilege: user biasa berubah hak menjadi admin



# Information Security Management

Theory	Activities	Characteristics
Security policy theory	Policy establishment Policy implementation	<ul style="list-style-type: none"> <li>• Policy is the main focus</li> <li>• Emphasize sequential, structured procedures</li> </ul>
Risk management theory	Policy maintenance Risk assessment  Risk control	<ul style="list-style-type: none"> <li>• Understand and cope with insecure environments</li> <li>• Ignore security policy and information audit mechanisms</li> </ul>
Control and auditing theory	Review and modification Establish control systems  Implement control systems	<ul style="list-style-type: none"> <li>• Overemphasize structures</li> <li>• Internal control and information audit is the main focus; ignore security policy and risk management</li> <li>• Lack of requirements planning and contingency for the unexpected</li> </ul>
Management system theory	Information auditing Establish security policy  Define security scope	<ul style="list-style-type: none"> <li>• Information auditing is ignored and the implementation is affected</li> <li>• Lack of periodic check</li> <li>• Lack of feedback</li> </ul>
Contingency theory	Risk management Implementation Policy strategy  Risk management strategy Control and audit strategy Management system strategy	<ul style="list-style-type: none"> <li>• Consider environments both outside and inside of an organization, and choose appropriate security strategies</li> <li>• Lack of integration and structures</li> </ul>



# Framework for Information Security Management System ISO/IEC 17799:2005

Consist of 10 Control Clause:

1. Security Policy
2. Security Organisation
3. Asset classification and control
4. Personnel security
5. Physical and environmental security
6. Communication and operation Management
7. Access Control
8. System development and maintenance
9. Business continuity management
10. Compliance

## Other Security IS Standard

- ❖ BS7799:2002

- ❖ ISO 27001:2005

For Mobile

- ❖ OMA (open mobile Alliance) DRM 2.0

- ❖ CSS (Content Scrambling System)

- ❖ Fairplay from Apple

- ❖ Window Media DRM



# Masalah dalam DBMS

- ❖ **Redudansi Data** : Kerancuan value, Anomali Penghapusan
- ❖ **Concurrency, real-time query database**
- ❖ **Skema database berdasarkan level management**
- ❖ **Kesalahan Normalisasi:**
  - Lossy Join Decomposition
  - over normalisation
- ❖ **Bentuk struktur Data yang berbeda pada sistem sharing Database**
- ❖ **Efektivitas SQL**
- ❖ **Ukuran Data**



# Efektivitas SQL : Penggunaan Select

- ❖ Hindari Query semua field
  - SELECT NAME FROM AUTHOR; lebih cepat dari
  - SELECT \* FROM AUTHOR;
- ❖ Gunakan index  
SELECT ISBN FROM EDITION;
- ❖ Alias yang sederhana  
SELECT A.NAME, P.TITLE, E.ISBN  
FROM AUTHOR A JOIN PUBLICATION P USING (AUTHOR\_ID)  
JOIN EDITION E USING (PUBLICATION\_ID);

This is much easier to deal with than this query:

```
SELECT AUTHOR.NAME, PUBLICATION.TITLE, EDITION.ISBN  
FROM AUTHOR JOIN PUBLICATION USING (AUTHOR_ID)  
JOIN EDITION USING (PUBLICATION_ID);
```



# Penggunaan Where

- ❖ *Single record searches*

```
SELECT * FROM AUTHOR WHERE  
AUTHOR_ID = 10;
```

- ❖ *Record range searches—Using the >, >=, <, and <= operators executes range searching*

```
SELECT * FROM AUTHOR WHERE  
AUTHOR_ID >= 5 AND AUTHOR_ID <= 10;
```



Bentuk relasi semula

Contoh :

NIM	NAMA	PROGRAM STUDI
1104001	ALI	EKONOMI
1104002	EDI	EKONOMI
1104003	ALI	FISIKA

**(a) Contoh dekomposisi aman**

Relasi : NIM\_NAMA

NIM	NAMA
1104001	ALI
1104002	EDI
1104003	ALI

Relasi : NIM\_PROG

NIM	PROGRAM STUDI
1104001	EKONOMI
1104002	EKONOMI
1104003	FISIKA

**(b) Contoh dekomposisi tidak aman**

Relasi : NIM\_NAMA

NIM	NAMA
1104001	ALI
1104002	EDI
1104003	ALI

Relasi : NAMA\_PROG

NAMA	PROGRAM STUDI
ALI	EKONOMI
EDI	EKONOMI
ALI	FISIKA



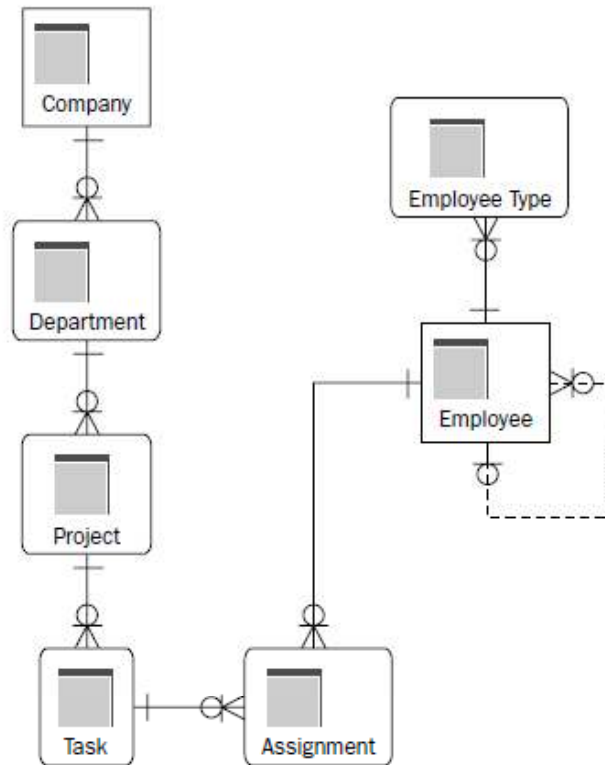


Figure 1-6: The relational database model.

PROJECT_ID	DEPARTMENT_ID	PROJECT	COMPLETION	BUDGET
1	1	Software sales data mart	4-Apr-05	35,000
2	1	Software development costing application	24-Apr-05	50,000
3	2	Easy Street construction project	15-Dec-08	25,000,000
4	1	Company data warehouse	31-Dec-06	250,000

TASK_ID	PROJECT_ID	TASK
1	1	Acquire data from outside vendors
2	1	Build transformation code
3	1	Test all ETL process
4	2	Assess vendor costing applications
5	3	Hire an architect
6	3	Hire an engineer
7	3	Buy lots of bricks
8	3	Buy lots of concrete
9	3	Find someone to do this because we don't know how

Figure 1-7: The relational database model—a picture of the data.





## Other Feature

- ❖ Data Dictionary
- ❖ Database Utilities
- ❖ Database Recovery
  - Recovery via reprocessing
  - Recovery via rollback/rollforward
  - Transaction logging
  - Write-ahead log



# Query Language

## ❖ Structured Query Language

### ❖ Contoh :

- Query by Example (QBE) helps the manager construct a query by displaying a list of fields that are available in the files from which the query will be made.
  - SELECT name, gpa
  - FROM student
  - WHERE gpa >=3.0
  - ORDER BY name



## Entity Relationship Diagram

- ❖ Kardinalitas : one to many, one to one, many to many
- ❖ Jenis2 atribut: null, mandatory, simple, composite, derived, single /multivalue
- ❖ Key: primary key, guest/foreign key
- ❖ Variasi entitas: subentitas, weak entitas
- ❖ Variasi relasi: redudansi, multi relasi, spesialisasi/generalisasi, agregasi
- ❖ Implementasi ERD pada tabel
- ❖ Struktur penyimpanan

# Manajemen Sistem Informasi Database

## ❖ Analyzing Database Model

- Identify the operations of the company.
- Draw up a picture of basic tables.
- Establish simple relationships.
- Create basic fields within each table.

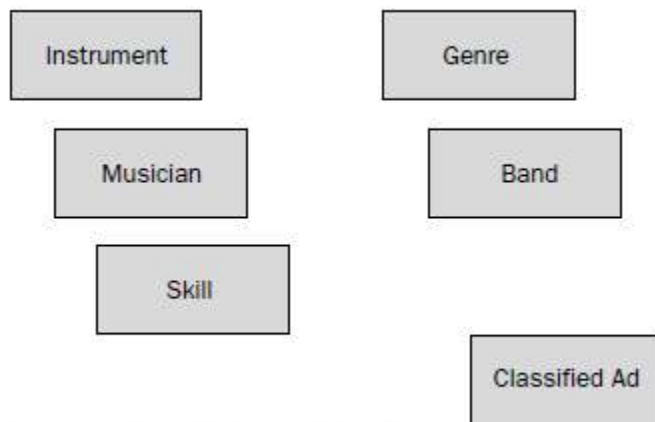


Figure 9-17: Identifying basic operations.

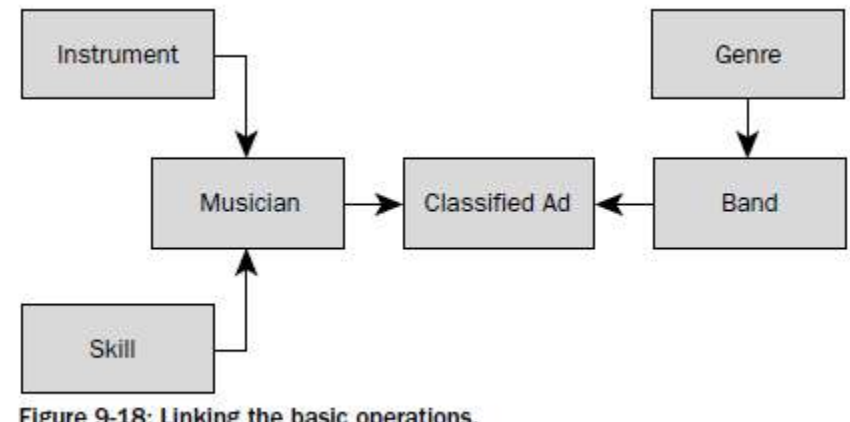


Figure 9-18: Linking the basic operations.



# Creating ERD

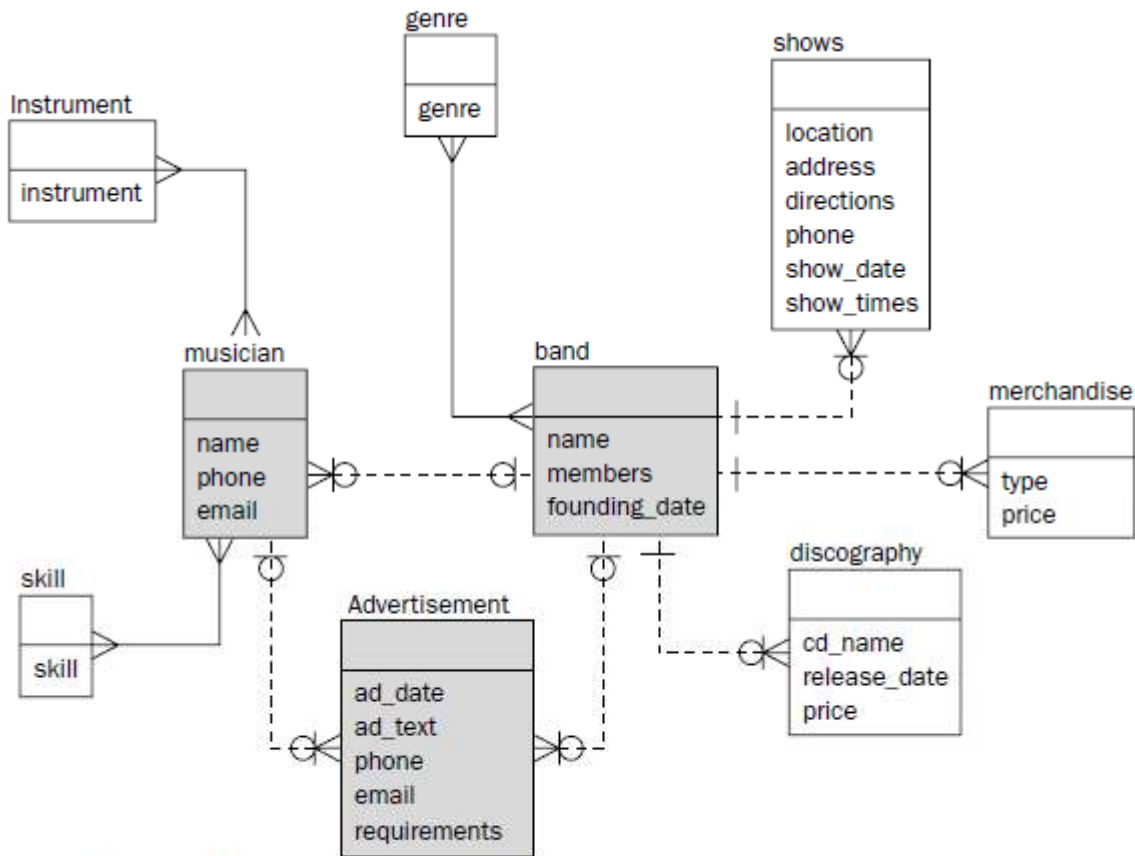


Figure 9-19: Creating a basic analytical ERD of business rules

# Establishing Company Operation

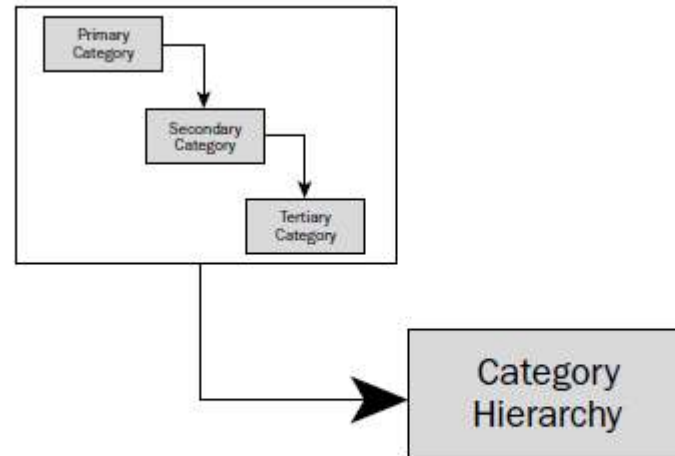


Figure 9-20: Data warehouse data modeling denormalizes multiple hierarchical static tables into single static structures.

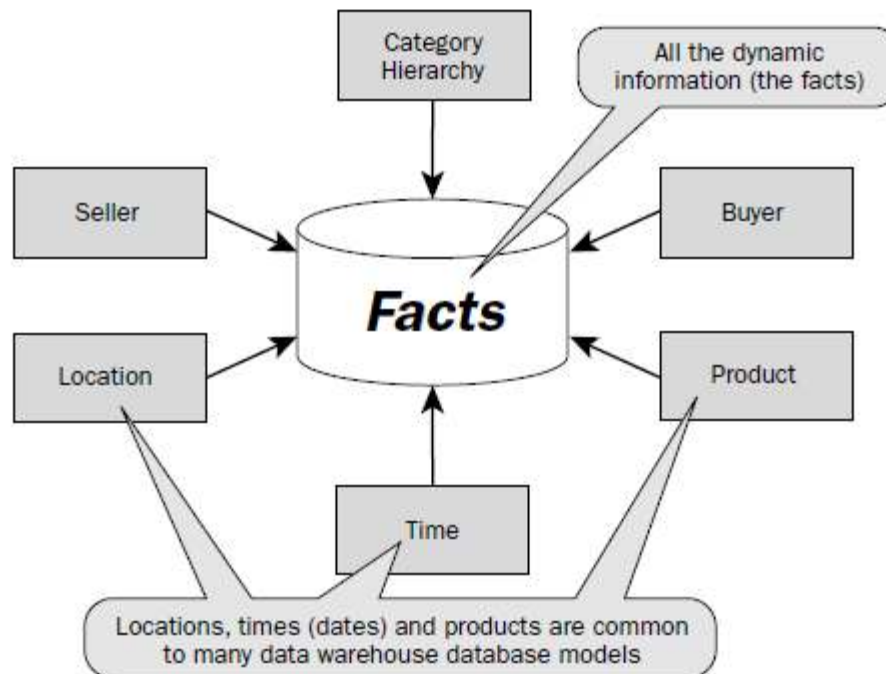


Figure 9-21: A data warehouse star schema database model for the online auction house.



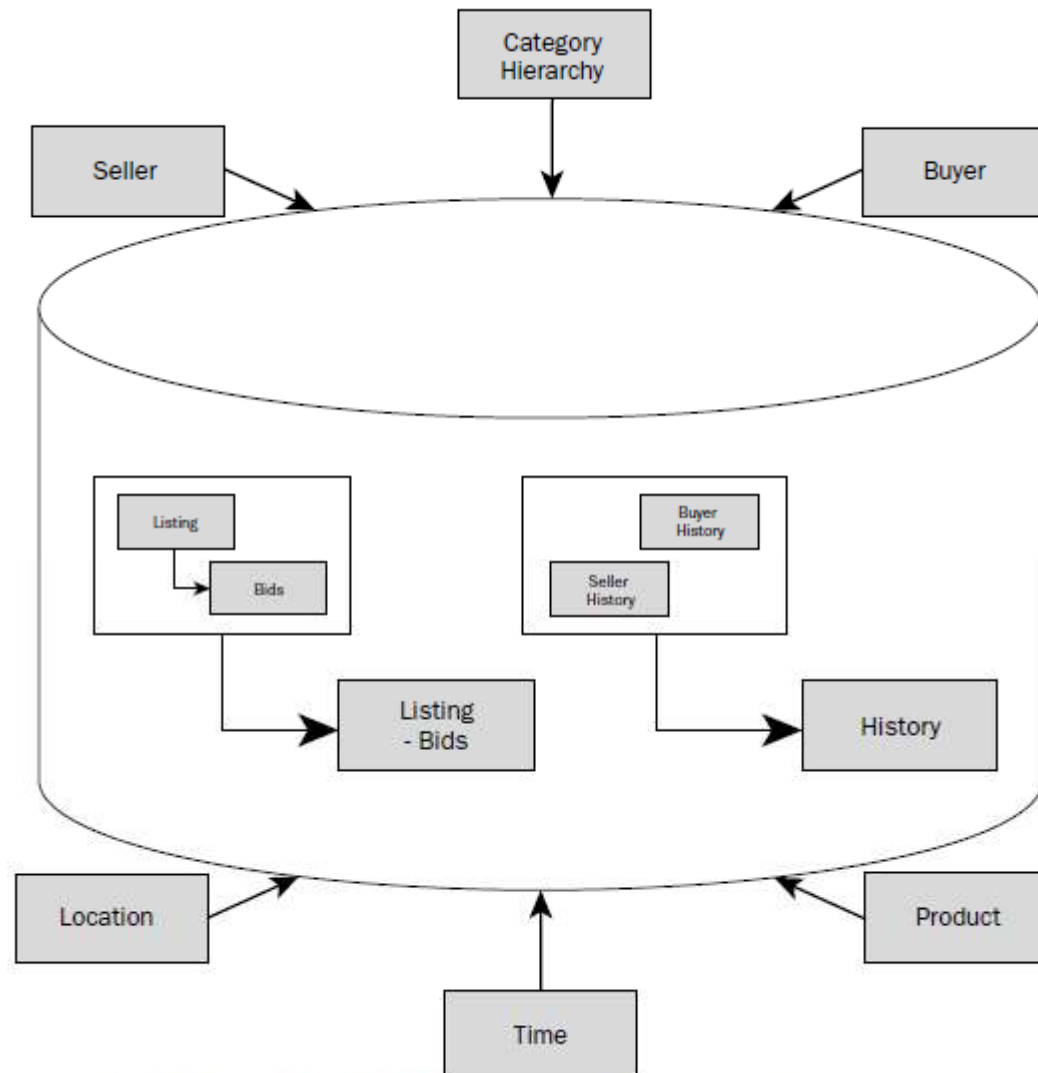


Figure 9-22: Analyzing the facts in a data warehouse database model.

# Discovering Business Rule

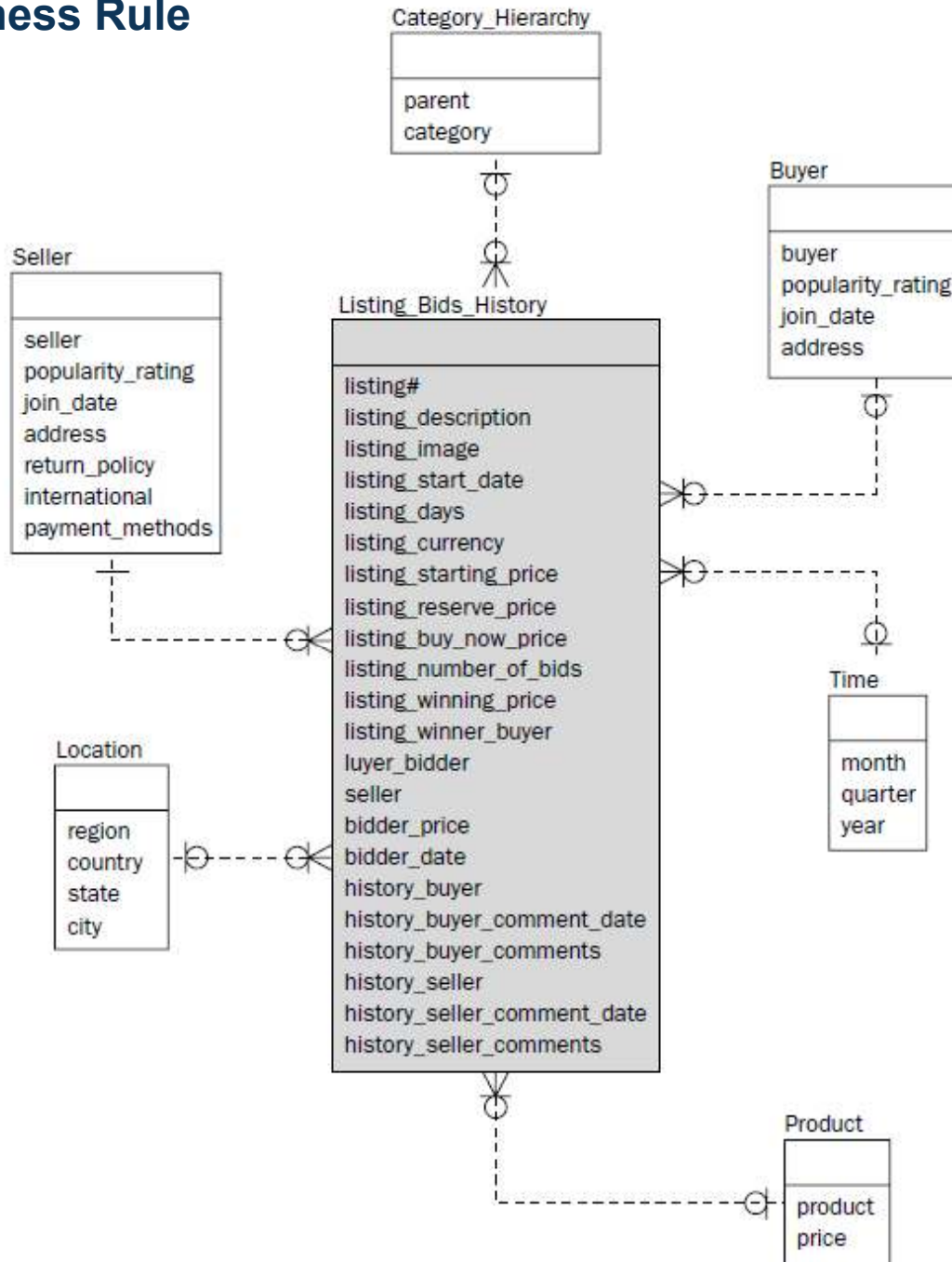


Figure 9-25: A data warehouse database model ERD for an online auction house.



# DSS Practical Application

1. Hewlett-Packard developed *Quality Decision Management* to perform production and quality-control functions. It can help with raw material inspection, product testing, and statistical analysis.
2. *Manufacturing Decision Support System (MDSS)*, developed at Purdue University to support decisions in automated manufacturing facilities, is especially useful for CAD/CAM operations.
3. RCA has developed a DSS to deal with personnel problems and issues. The system, called *Industrial Relations Information Systems (IRIS)*, can handle problems that may not be anticipated or that may occur once, and it can assist in difficult labor negotiations.
4. The Great Eastern Bank Trust Division developed a DSS called *On-line Portfolio Management (OPM)* that can be used for portfolio and investment management. The DSS permits display and analysis of various investments and securities.
5. *RealPlan*, a DSS to assist with commercial real estate decisions, is useful for various decision aspects of purchasing, renovating, and selling property.
6. *EPLAN (Energy Plan)* is a DSS being developed by the National Audubon Society to analyze the impact of U.S. energy policy on the environment.