Interpret the following information

Baker, Kenneth D.	324917628
Doyle, Joan E.	476193248
Finkle, Clive R.	548429344
Lewis, John C.	551742186
McFerran, Debra R.	409723145
Sisneros, Michael	392416582

Data in context

Class Roster

Course: MGT 500 Business Policy

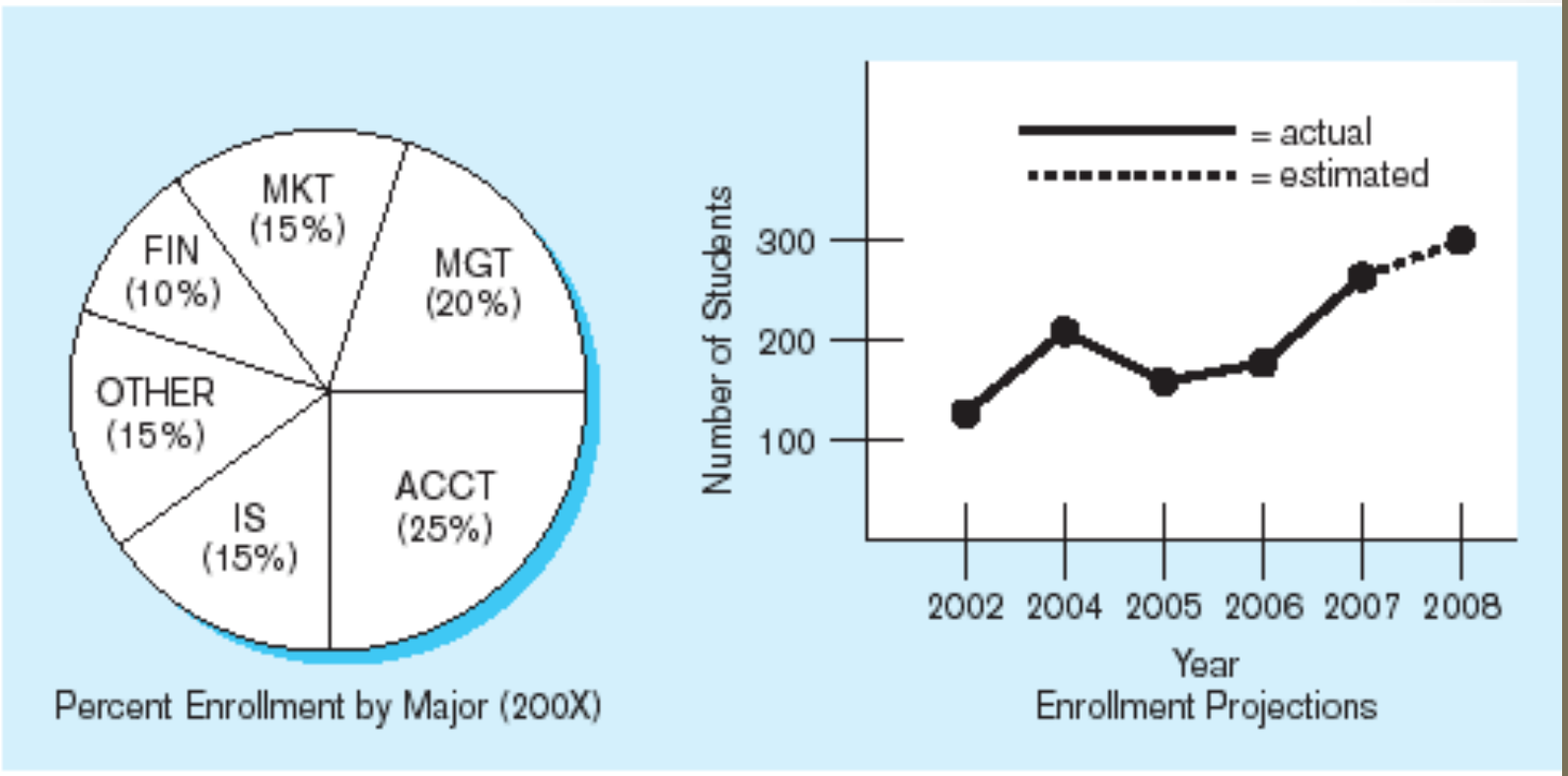
Semester: Spring 200X

Section: 2

Name	ID	Major	GPA
Baker, Kenneth D.	324917628	MGT	2.9
Doyle, Joan E.	476193248	MKT	3.4
Finkle, Clive R.	548429344	PRM	2.8
Lewis, John C.	551742186	MGT	3.7
McFerran, Debra R.	409723145	IS	2.9
Sisneros, Michael	392416582	ACCT	3.3

Context helps users understand data

Summarized data



Graphical displays turn data into useful information that managers can use for decision making and interpretation

Table 1-1 Example Metadata for Class Roster

Data Item		*Value*				
Name	**Type**	**Length**	**Min**	**Max**	**Description**	**Source**
Course	Alphanumeric	30			Course ID and name	Academic Unit
Section	Integer	1	1	9	Section number	Registrar
Semester	Alphanumeric	10			Semester and year	Registrar
Name	Alphanumeric	30			Student name	Student IS
ID	Integer	9			Student ID (SSN)	Student IS
Major	Alphanumeric	4			Student major	Student IS
GPA	Decimal	3	0.0	4.0	Student grade point average	Academic Unit

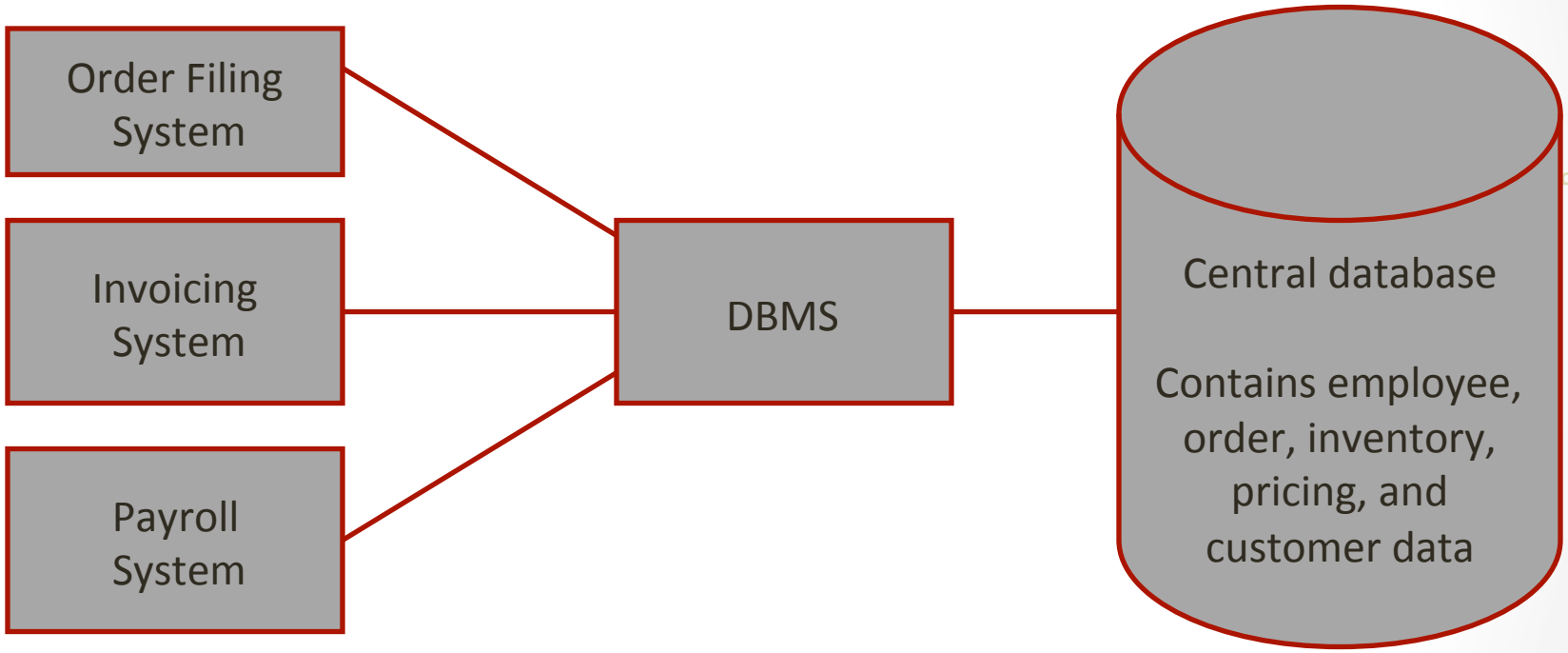
Descriptions of the properties or characteristics of the data, including data types, field sizes, allowable values, and data context

Database

- Database: organized collection of logically related data
- Data: stored representations of meaningful objects and events
  - Structured: numbers, text, dates
  - Unstructured: images, video, documents
- Information: data processed to increase knowledge in the person using the data
- Metadata(or Catalog): data that describes the properties and context of user data

Database Management System

A software system that is used to create, maintain, and provide controlled access to user databases

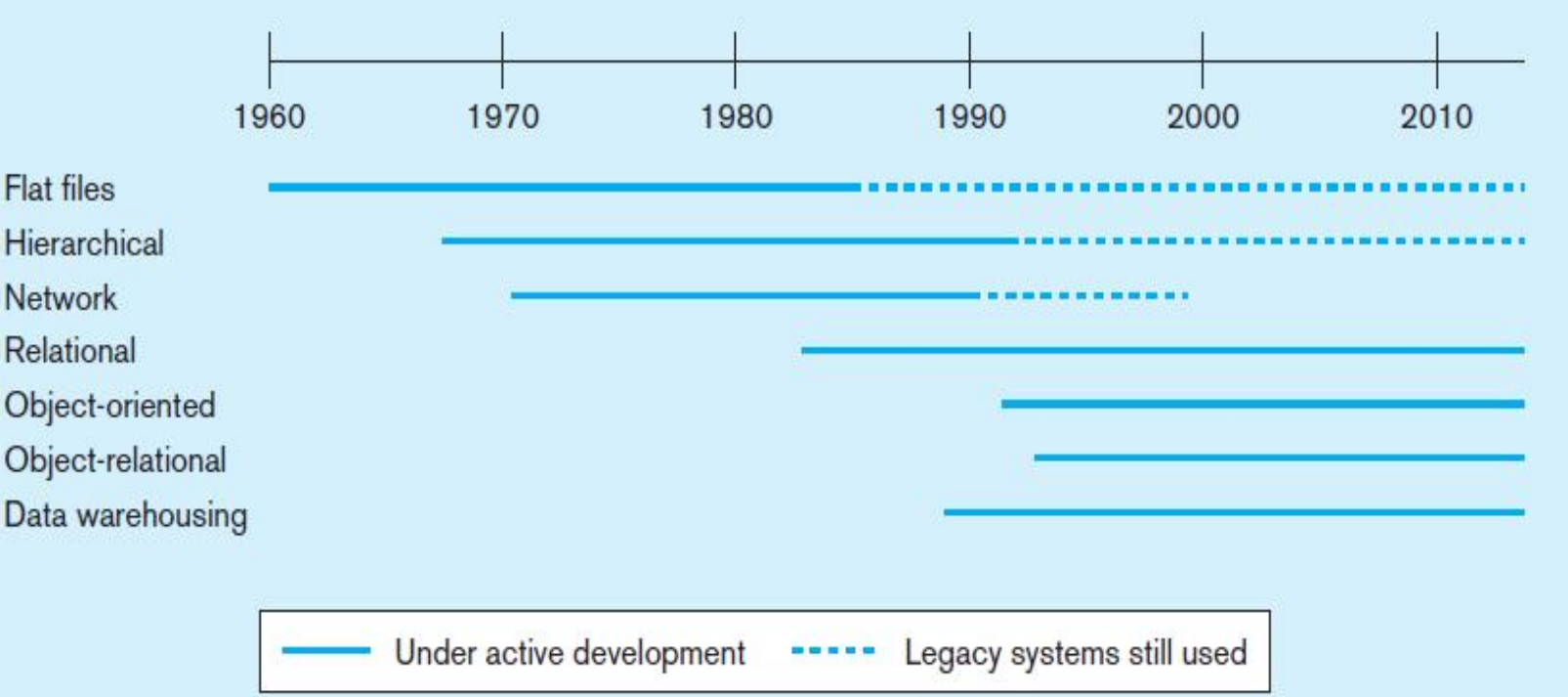


DBMS manages data resources like an operating system manages hardware resources

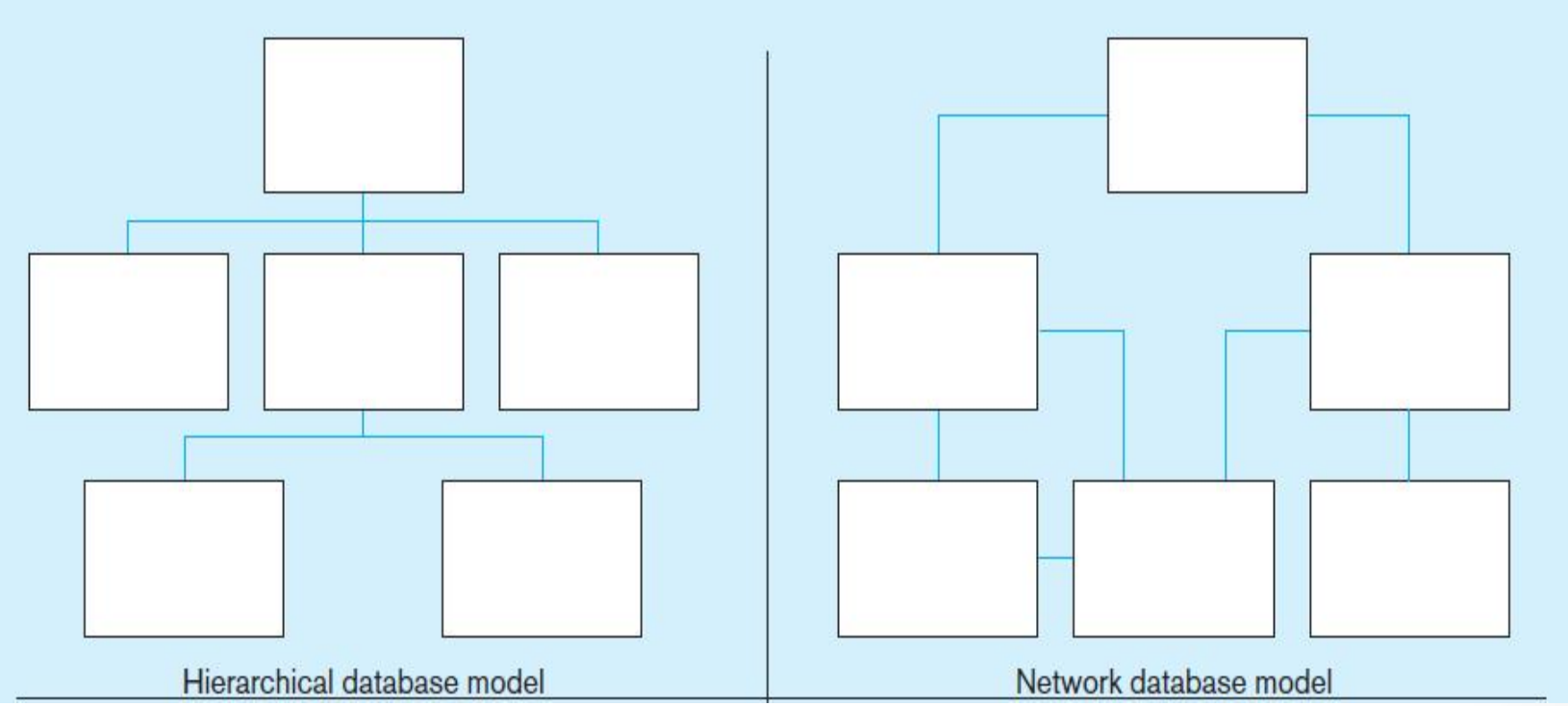
DBMS expectations

- Ability to create structures of the data using a data definition language (DDL)
- Ability to query the data using data manipulation language (DML)
- Storage and access to terabytes of data
- Recovery of the data in case of failures
- Data control (DCL) and integrity

Evolution of DB Systems



Database Models(1)



Hierarchical Database

- A hierarchical database consists of a collection of *records* that are connected to each other through *links*.
- Each record is a collection of fields (attributes), each of which contains only one data value.
- A link is an association between precisely two records.
- Tree like structure
- http://en.wikipedia.org/wiki/Hierarchical_database_model

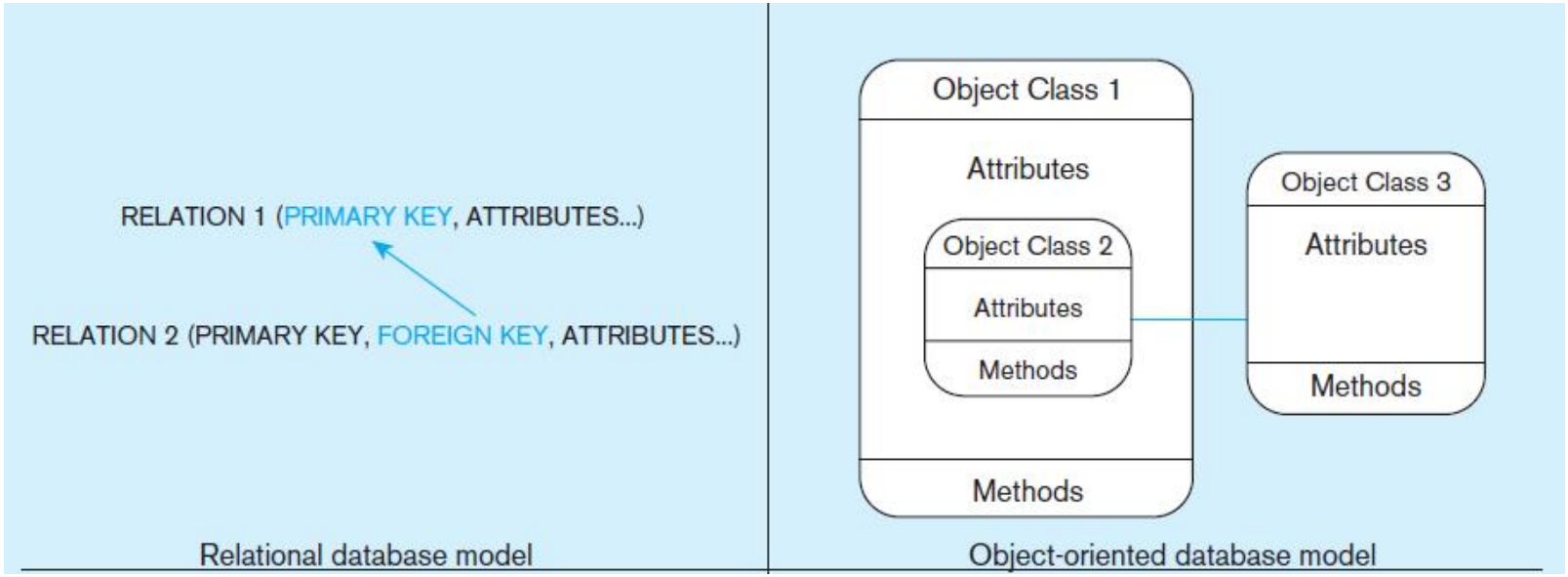
EmpNo	First Name	Last Name	Dept. Num
100	Sally	Baker	10-L
101	Jack	Douglas	10-L
102	Sarah	Schultz	20-B
103	David	Drachmeier	20-B

Serial Num	Type	User EmpNo
3009734-4	Computer	100
3-23-283742	Monitor	100
2-22-723423	Monitor	100
232342	Printer	100

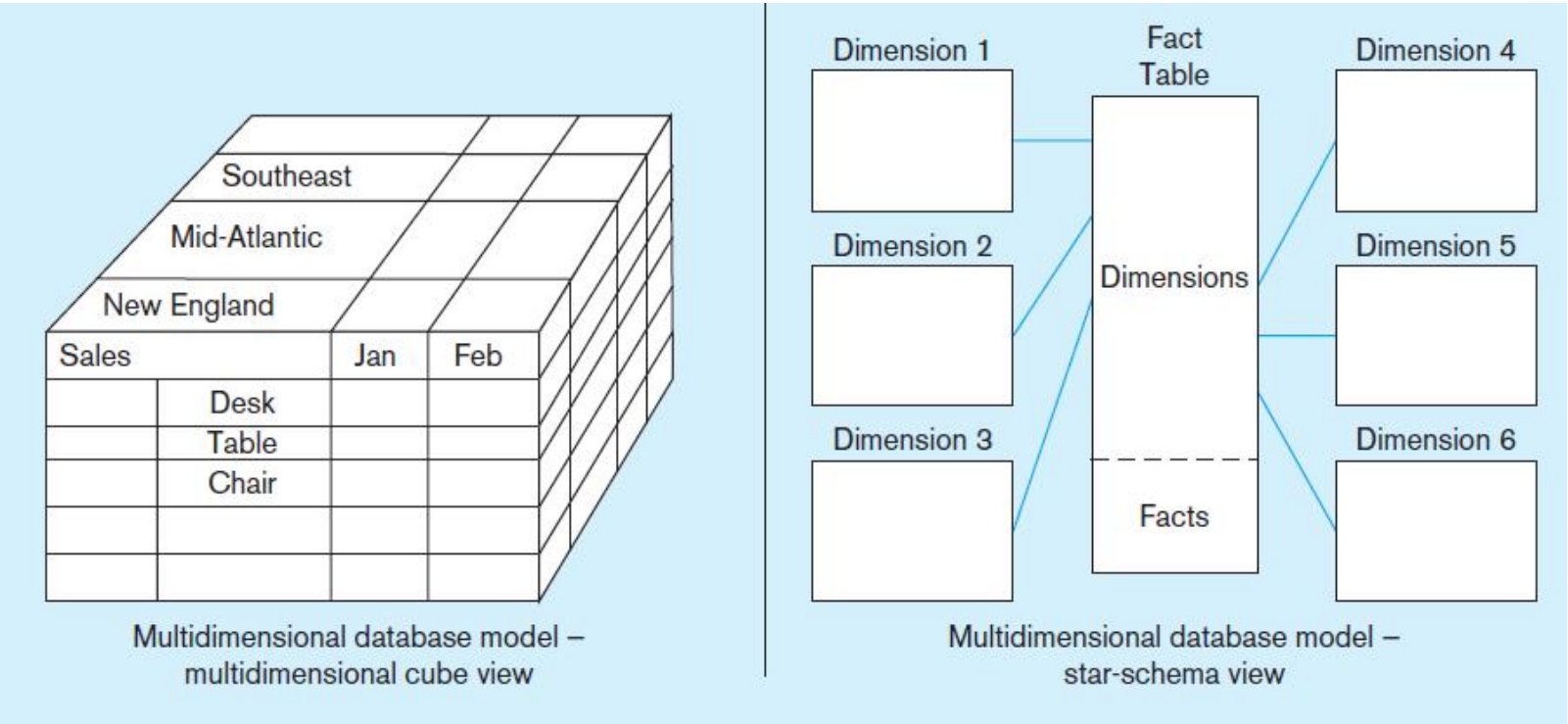
Network Databases

- A network database consists of a collection of records connected to one another through links.
- A record is in many respects similar to an entity (table) in the E-R model.
- Each record is a collection of fields (attributes), each of which contains only one data value.
- A link is an association between precisely two records.
- Thus, a link can be viewed as a restricted (binary) form of relationship in the sense of the E-R model.
- http://en.wikipedia.org/wiki/Network_model

Database Models(2)



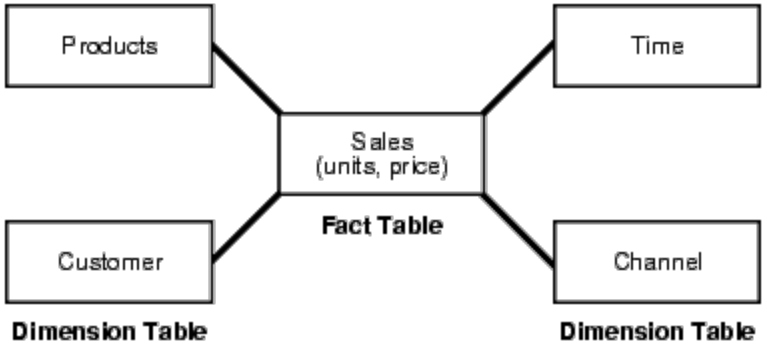
Database Models(3)



http://www.stanford.edu/dept/itss/docs/oracle/10g/olap.101/b10333/multimodel.htm

Star Schema

The star schema is the simplest data warehouse schema. It is called a star schema because the diagram of a star schema resembles a star, with points radiating from a center. The center of the star consists of one or more fact tables and the points of the star are the dimension tables.



http://docs.oracle.com/cd/A87860_01/doc/server.817/a76994/schemas.htm

The Range of Database Applications

- Personal databases
- Two-tier and N-tier Client/Server databases
- Enterprise applications
  - Enterprise resource planning (ERP) systems
  - Data warehousing implementations

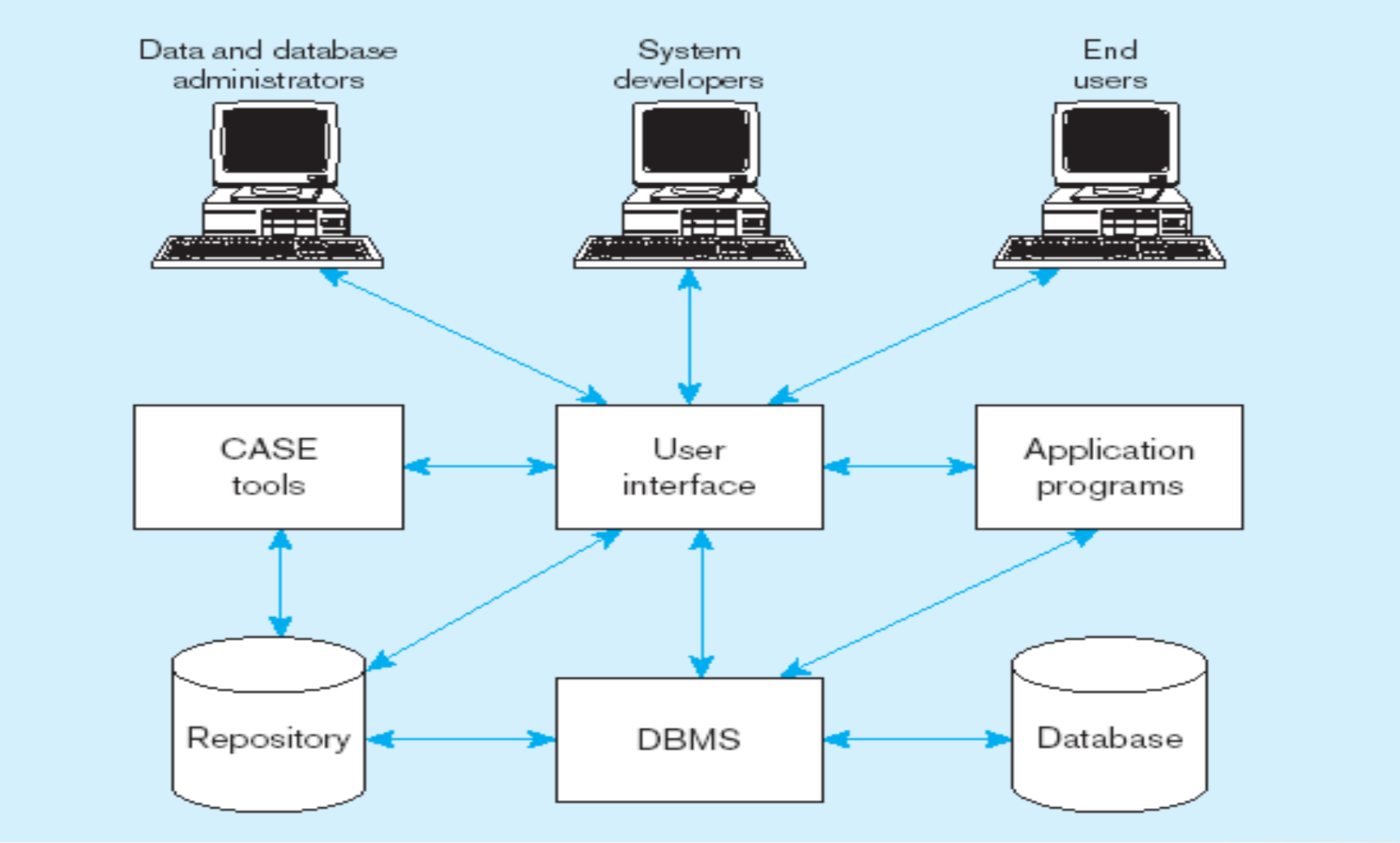
TABLE 1-5 Summary of Database Applications

Type of Database / Application	Typical Number of Users	Typical Size of Database
Personal	1	Megabytes
Two-tier	5–100	Megabytes–gigabytes
Three-tier	100–1000	Gigabytes
Enterprise resource planning	>100	Gigabytes–terabytes
Data warehousing	>100	Terabytes–petabytes

Enterprise Database Applications

- Enterprise Resource Planning (ERP)
  - Integrate all enterprise functions (manufacturing, finance, sales, marketing, inventory, accounting, human resources)
- Data Warehouse
  - Integrated decision support system derived from various operational databases

Components of the Database Environment



DBMS Functions

- Query Processing
- Storage and Buffer Management
- Transaction Processing

- Two sources of commands to the DBMS:
  - Conventional users and application programs (ask for data or modify)
  - DBAs (responsible for structure)

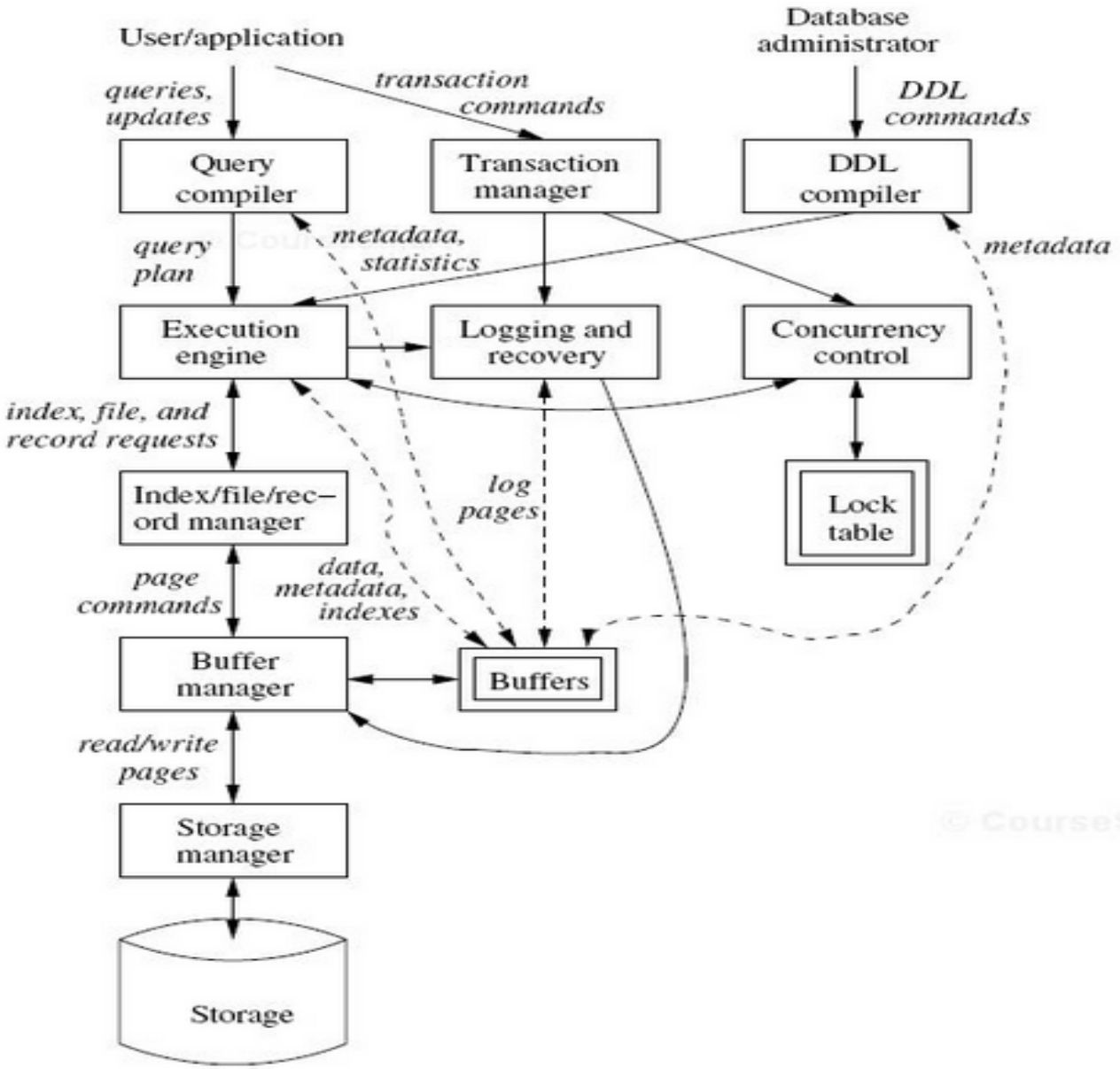


Figure 1.1: Database management system components

Query Processing

- DBAs follow the right path of the Figure 1.1
  - A DDL command is issued by an authenticated DBA.
  - The command is processed by the DDL processor and passed to the execution engine.
  - Execution engine goes through index/file/record manager to alter the metadata or schema.
- Normal query follows the left path of Figure 1.1
  - A DML command is issued by a user or program.
  - It is parsed and optimized by a query compiler that creates a query plan.
  - The query plan is executed by the execution engine to request records of a relation using data files or index files to find the data quickly.
  - Results are passed to a Buffer manager to bring data from disk to memory buffers using a disk controller.

Transaction Processing

- Queries and DML statements can be grouped into transactions that must be executed atomically.
- Must follow the ACID rules.

- Transactions go through a concurrency-control manager to make sure that atomicity and isolation happen.
- Logging and recovery manager is responsible for the durability of the transactions.

Storage and Buffer management

- Data must be brought from disk to main memory. This is the job of Storage manager.
- Buffer manager is responsible of partitioning the available main memory into buffers.
- All DBMS components need information that may include:
  - Data
  - Metadata
  - Log records
  - Statistics
  - Indexes

Query processing

- Query processing follows the following steps:
  - Query compiler to translate to a query plan.
    - A query parser – Creates a tree structure that we will discuss in relational algebra from the query text.
    - A query preprocessor – Performs semantic checks.
    - A query optimizer – Transforms the query plan into the best available sequence of operations.
  - Use of metadata and statistics to speed up the query using indexes when present.

References

- Jeffrey A. Hoffer, V. Ramesh, Heikki Topi . “The Database Environment" Modern Database Management 11.0.
- Ullman, Widom. “Database Systems: The Complete Book, Second Edition.