

What is a database?

A collection of data, typically describing the activities of one or more related organizations.

What is a database management system?

Software designed to assist in maintaining and utilizing large collections of data.

Describe the evolution of database models from the 1960s through the 1980s.

network data model -> hierarchical data model -> relational data model + SQL

Which advances in database design date to the late '80s and '90s?

The ability to store new data types and to ask more complex queries.

What do enterprise/management resource planning packages provide?

A substantial layer of customizable application-oriented features on top of a DBMS.

What's wrong with storing data in the file system?

- The system won't be able to store (or address) all the data in memory, so programs have to bring data into main memory as needed.
- Each query type will require a special program.
- Applications are forced to deal with details of concurrent access and maintaining data consistency.
- Password security mechanisms are not flexible enough for complex permissions.

What are the advantages of a DBMS?

- Data independence.
- Efficient data access.
- Data integrity and security.
- Data administration.
- Concurrent access and crash recovery.
- Reduced application development time.

Describe the data independence advantage of DBMSs.

DBMSs hide the details of data representation and storage from the application.

Describe the data integrity and security advantage of DBMSs.

A DBMS can enforce integrity constraints on data inserted into the database. For example, making sure a person doesn't have a negative age. Also it can enforce access controls, giving each user a different levels of access to the data.

Describe the data administration benefits of DBMSs.

Centralizing data allows experienced professionals to control organization, access, and efficiency.

Describe the concurrent access advantage of DBMSs.

A DBMS schedules concurrent access to data in such a manner that users can think of data as being accessed by only one user at a time.

Why do DBMSs result in reduced application development time?

- The DBMS provides many common data-accessing functions.
- The DBMS provides a high-level interface.
- The DBMS is probably more robust and reliable than new code rolled out with the application for achieving the same functionality.

When should you *not* use a DBMS?

- If you need extremely high performance for a simple set of functions.
- If you need to manipulate data in ways not supported by the query language.

What is a data model?

How does a DBMS tie in?

A collection of high-level data description constructs that hide many low-level storage details.

A DBMS allows a user to define the data to be stored in terms of a data model.

What is a semantic data model?

Why is it needed?

A more abstract, high-level data model that makes it easier for a user to come up with a good initial description of the data in an enterprise.

It's needed because a DBMS's data model, though it hides many details, is still much lower level than how a user thinks about the underlying enterprise.

A relation can be thought of as ...

... a set of records.

Define *schema*.

A description of data in terms of a data model.

A field is also called ...

... an *attribute* or a *column*.

In a relational model, the schema for a relation specifies what?

Its name, the name of each field, and the type of each field.

What are the rows of a relation?

Records.

What are integrity constraints?

Conditions that the records in a relation must satisfy.

Name some data models.

- relational
- network
- hierarchical
- object-oriented
- object-relational

The data in a DBMS is described using what?

Schemas for the *conceptual* (AKA *logical*), *physical*, and *external* levels of abstraction.

What does the conceptual schema do?

It describes the stored data in terms of the data model of the DBMS.

In a relational DBMS, the conceptual schema describes all relations that are stored in the database.

What does the physical schema do?

It summarizes how the relations described in the conceptual schema are actually stored on secondary storage devices such as disks and tapes.

This includes determining file organization and indexes (auxiliary data structures for speeding up retrieval).

What is a *file* to a DBMS?

Either a collection of records or a collection of pages, **not** a string of characters as in an operating system.

What does an external schema do?

It allows data access to be customized (and authorized) at the level of individual users or groups of users.

Each external schema consists of what?

A collection of one or more views and relations from the conceptual schema.

What is a view?

Conceptually it is a relation, but the records in a view are not actually stored in the DBMS because they can be computed from relations already in the DBMS.

Why are relational databases so popular?

They have powerful query languages, allowing a rich class of questions to be posed easily.

Name to query languages.

- Relational calculus.
- Relational algebra.

What is a transaction?

Any *one execution* of a user program in a DBMS.

It is the basic unit of change as seen by the DBMS. Partial transactions are not allowed.

What is a locking protocol?

A set of rules to be followed by each transaction (and enforced by the DBMS) to ensure that, even though actions of several transactions might be interleaved, the net effect is identical to executing all transactions in the same serial order.

What is a lock?

Name two kinds of locks.

A lock is a mechanism used to control access to database objects.

- Shared lock - can be held by two different transactions simultaneously.
- Exclusive lock - ensures no other transaction holds *any* lock on the locked object.

What is a Write-Ahead Log?

What is its purpose?

A log of all writes to the database, written to *before* the action is taken on the database.

Since writes to the log occur first, the system can be brought back to consistency if the database crashes between the log write and the corresponding database change.

The log also guarantees that changes made by successfully completed transactions aren't lost either.

What is checkpointing?

Periodically forcing some information to the disk to reduce the time required to recover from a crash.

What are the three important points to remember about concurrency control and recovery?

- Objects read or written by a transaction must first be locked. This has a performance cost.
- For efficient log maintenance the DBMS must be able to selectively force a collection of pages in main memory to disk. OS support is not always satisfactory.
- Periodic checkpointing can reduce time needed to recover from a crash, but can decrease normal execution performance.

What do application programmers do?

- Access DBMS through an external schema to produce packages that facilitate data access for end users. They use the host/data languages and software tools the DBMS vendor provides.

What do database administrators do?

- Design and maintain a enterprise database. Responsible for conceptual/physical schemas, security, data availability and recovery, and database tuning.