**Database** – a logically coherent collection of related data.

Features of a database include:

A single source of information for user groups and multiple applications

Independence from the way information is actually stored in files on disk

Concurrency Control so that concurrent applications or users operations will happen with no interference.

Data integrity control – ensures that the data is not corrupted accidentally.

Data security – guards privacy and protects against deliberate corruption of data.

Data model – is the set of concepts characterising the way in which data is organised within a database.

Database management system (dbms) is a collection of programs that enable users to create and maintain a database.

A dbms is a general purpose software system that allows users to define, construct and manipulate databases.

A dbms usually supports one particular data model. Therefor most dbms are described by the data model they support.

**Types of databases:**

**Hierarchical** – represents data as a hierarchical tree structure.

**Network** – represents data as record types and also represents a limited kind of relationship between records. The networked data model (also known as CODASYL DBTG model), has an associative record-at-a-time language that must be embedded in an application programming language e.g COBOL

**Relational** – represents information as a collection of tables. Most relational DBMS’s support a high level query language called sql.

**Object Oriented** – represents information as objects which contain both data and operations which can process the data.

**Benefits of using a dbms:**

**Isolation of application programs and data** – by not embedding the storage and retrieval operations in an application, many users and applications may access the same information.

**Denial of unauthorised access** to information held within a database.

**Representation of complex data-types and relationships among data.**

**Enforcement of integrity constraints on data**. E.g. constraints on they type of information that a piece of data can have.

Provide **automatic backup and recovery** of information stored in the database.

**Relational Database:**

**Value** – is the smallest updatable portion of a table. A value is the intersection of a row and a column.

**Columns** – a set of values which may vary over time. All values of a column are of the same type.

A column specification includes its type and integrity constraints on the values it may hold. Column names are also called **attribute** names.

**Rows** – a non-empty set of values, also known as **tuples**.

The relational data model is based on the mathematics of set theory relational algebra.

A relation is a special type of set, where:

* A set is a collection of values of the same type.
* A set has no implied ordering.
* A set has no duplicate elements.

**Relational DBMS Architecture**

Simple data structures – since the user need only deal with tabular information, the data structures are both simple and intuitive.

Simple operators – all operations are based on tables. Therefore there is much greater uniformity in the operations and their syntax.

View mechanism – relational databases allow users to have different presentations of the database information which suit their different purposes for that information.

Standard language support – the relational database community have defined a standard language for database query and control. This language is therefore common across all implementations of relational databases.

Data independence – information is presented to users in tables. The underlying storage organisation and database file formats are completely hidden from the user. This allows physical data independence where the storage format may be changed without affecting the users view of the information. Changing the tables need not affect the users view of the database (logical data independence).

Dynamic data definition – relational databases automatically maintain information about all information contained in the database. For example it maintains dictionary (catalog) tables which h identify what attributes are stored in what tables, what data-types make up each attribute, the sizeof all tables etc.

Ease of application development – since the user only needs to keep track of what information to retrieve, application dev time is reduced as you don’t need to produce any data retrieval code.

Distribution support – most relational databases now provide distributed access to the database, e.g. the client may be a separate machine to the server, but it appears to the client that the database is local.

Concurrency control and security – relational databases provide mechanisms to enforce serialisation of queries and provide authorisation and authentication of the database.

Relational database management systems conform to the 3 – schema architecture:

1. The external view is the view presented to the user. Each view typically describes the part of the database that a particular user group is interesting in, and hides the rest of the database.
2. The conceptual level contains a description of the structure of the whole database for a community of users. The description is called the conceptual schema of a database. It hides the details of physical storage structures and concentrates on describing entities, data types, relationships and constraints.
3. The internal level contains the internal schema. This internal schema describes the physical storage structure of the database. The internal schema uses a physical data model and describes the complete details of data storage and access paths for the database.

Structured Query Language:

Sql is the standard relational query language and is based on relational algebra.

Sql supports the:

Creation and maintenance of database information.

Rapid prototyping and testing as it makes information retrieval and storage easy.

Porting of database application programs to different RDBMS implementations with only minor alterations, as sql is a standard language.

Control over security, concurrency and locking of tables, rows and values.

SQL has 3 sublanguages:

**Data definition language** – allows definition of relations (tables), their content type and integrity rules which govern attribute usage and values.

**Data manipulation language** – allows users to populate, manipulate, retrieve and delete information in the database.

**Data control language** – allows users and database administrators to specify security checks and controls on the database.

Benefits of sql:

Reduced training costs as the language is standard across all rdbms vendors.

Increases application longevity as the standard slowly evolves but maintains backwards compatibility.

Provides a basis for inter system communication.

Simplifies customer choice as comparison of different vendor offerings is easier.

Caveat – even though sql is standard some vendors have their own dialect to give users extra features.

Sql is designed to be simple

Supports English like statements

Statements are always terminated with a semicolon;

Other features of a dbms:

Report writers – allow automatic production of printed formatted reports based on current database values.

Screen generators forms – allow the rapid development of screens of forms to facilitate data entry or data querying.

Network connectivity tools: these allow applications which use the relational database to reside on a different computer to use the relation database without the user being aware of the distribution over the network.

Support for embedded 3rd generation languages – allows languages like c to communicate with the database. Usually comes in the form of a library.

Case dictionaries - ?

Data types – the datatype of a column specifies what kind of information the column will hold.