

## Glossary

---

**Aggregate object set.** A relationship among object sets itself viewed as an object set.

**Anomalies.** Inconsistencies or errors resulting from manipulating data in random tables containing redundant data. Three types of anomalies are encountered: update, deletion, and addition.

**API (application programming interface).** A functional interface, usually supplied by the operating system, that allows one application program to communicate with another program. APIs are generally implemented through function calls.

**Application developer.** A DBMS component that enables the creation of an application with menus, toolbars, forms, and reports.

**ASCII (American Standard Code for Information Exchange).** A standard code, consisting of a set of seven-bit coded characters, used for information exchange between computers and communication systems.

**Atomic transaction.** A transaction in which either all database actions are executed completely, or no actions are executed at all.

**Attribute.** An intrinsic or inherent characteristic of an entity that is of interest to an organization.

**Authentication.** Determination and verification of the identity of a user.

**Authorization.** Granting of privileges to users for accessing the database based on the established security policy of an organization.

**Backward engineering.** The mechanism in CASE tools to obtain the logical data model from the physical data model of a certain database.

**Backward recovery.** Method used to back out or undo unwanted changes to the database. The database is restored to an earlier, consistent state.

**Binary relationship.** Relationship in which two object sets participate. This is the most common form of relationship between object sets.

**Binary search.** Widely used search technique where the data values are sorted. The search begins at the middle of the series of values and proceeds to the left or the right from the middle based on the value being searched for.

**Boyce-Codd normal form (BCNF).** A relation or table is in BCNF if it is already in the third normal form and no key attribute is functionally dependent on any nonkey attribute.

**Bitmapped indexing.** Compact, high-speed indexing method. A bitmap is an ordered series of bits, one bit for each distinct value of the indexed column.

**B-tree (balanced tree).** Arrangement of sorted data values in nodes forming an inverted tree with the root node at the top and an equal number of branch nodes on either side.

**B-tree indexing.** Common indexing technique used in most DBMSs based on a B-tree. The search begins at the root node and proceeds through the branch nodes on either side.

**Buffer.** Region of computer memory that holds data being transferred from one area to another. Data from database tables are fetched into memory buffers. Memory buffer management is crucial for system performance.

**Business rules.** Specifications based on business practices of an organization that need to be incorporated in the logical data model.

**Candidate key.** A single attribute or a set of attributes that uniquely identifies an instance of an object set or entity type and can be a candidate to be chosen as the primary key.

**Cardinality.** Cardinality of the relationship between two object sets indicates how many instances of the first object set may be related to how many of the second.

**CASE (computer-aided software engineering).** CASE tools or programs help develop software applications. A set of CASE tools many include code generators, data modeling tools, analysis and design tools, and tools for documenting and testing applications.

**Centralized database.** Database where all data is physically stored in one location.

**Checkpointing.** Procedure for automatically pausing database operations, committing completed transactions, and saving copies of the database at predetermined points. Database recovery may start at the latest checkpoint.

**Clustering.** Method for improving database performance. Related records that are likely to be accessed together are placed in the same data block.

**CODASYL (Conference on Data Systems and Languages).** Organization composed of hardware vendors, software vendors, and user groups. Among other efforts, famously known for the development of COBOL.

**Committed transaction.** Database transaction that has successfully completed all of its operations.

**Composite key.** Primary key made up of more than one attribute.

**Concatenated key.** Same as *Composite key*. Primary key consisting of more than one attribute.

**Conceptual data model.** Data model, independent of any specific DBMS, that represents the information requirements of an organization at a conceptual or logical level.

**Conceptual object set.** Set representing the types of an object, not the physical objects themselves.

**Conceptual schema.** Definition of a conceptual data model.

**Concurrent access.** Simultaneous access of the same data element in a database by two or more transactions.

**Concurrency control.** Method to maintain database integrity from potential problems that may result from concurrent access.

**Constraint.** A rule that is defined in the database. For example, the primary key or foreign key constraints on a data column imposes rules or restrictions on the data that can be entered into that column. Similarly, a business rule that  $\text{DailyWageRate} > 0$  can be defined as a constraint in the database.

**Conventional data model.** See *Logical data model*.

**Data administration.** Responsibility for the overall planning, managing, and control of the organization's data resource. Data administration is more managerial than technical.

**Data block.** Records in a physical file are stored in data blocks, each block containing a specified number of records. A data block is the unit of data transfer between disk storage and main memory of a computer. Generally, one block of data is fetched or written in one I/O operation.

**Data dictionary.** Repository holding the definitions of the data structures in a database. In a relational database, the data dictionary contains the definitions of all the tables, columns, and so on.

**Data-driven.** Design and implementation of applications in a database approach are data-driven, not process-driven. The primary emphasis is on the business objects and the data about them, not on the processes.

**Data fragmentation.** Separation of data in the database into manageable fragments. In a relational database, a relation or table may be broken into fragments by separating the table into sets of columns or sets of rows.

**Data independence.** Ability in database systems to separate data descriptions from the application programs. As such, data definitions may be changed without the need to change the application programs.

**Data integration.** Combining of data as in a database for common use by various user groups in an organization.

**Data integrity.** Accuracy and consistency of the data stored in the organization's database system.

**Data manipulation.** Operations for altering data in the database. Data manipulation includes retrieval, addition, update, and deletion of data.

**Data mining.** A knowledge discovery process. Data mining algorithms uncover hidden relationships and patterns from a given set of data on which they operate. Knowledge discovery is automatic, not through deliberate search by analysts.

**Data model.** Representation of the real-world information requirements that gets implemented in a computer system. A data model provides a method and means for describing real-world information by using specific notations and conventions.

**Data record.** Data in physical files is stored as records. One record contains fields of information about one instance of an object set or entity type. A record in the customer file contains data about one customer.

**Data redundancy.** Duplication of data in a database.

**Data repository.** Storage of the organization's data in databases. Stores all data values that are part of the databases.

**Data security.** Protection of the database from unauthorized or fraudulent use.

**Data striping.** Distributing or striping data on separate physical devices for the purpose of improving performance by accessing the various segments on separate devices in parallel.

**Data view.** View of the database by a single user group. Therefore, a data view of a particular user group includes only those parts of the database that the group is concerned with. The collection of all data views of all the user groups constitutes the total data model.

**Data warehouse.** A specialized database having a collection of transformed and integrated data, stored for the purpose of providing strategic information to the organization.

**Database.** Repository where an ordered, integrated, and related collection of the organization's data is stored for the purpose of computer applications and information sharing.

**Database administration.** Responsibility for the technical aspects of the organization's database. Includes the physical design and handling of the technical details such as database security, performance, day-to-day maintenance, backup, and recovery. Database administration is more technical than managerial.

**Database administrator (DBA).** Specially trained technical person performing the database administration functions in an organization.

**Database engine.** Heart or kernel of a DBMS. Coordinates the tasks performed by the other components for storing, retrieving, and altering the data.

**Database monitoring.** Inspecting the performance and state of the database system. Database monitoring provides information for tuning the database for improved performance.

**Database plan.** Covers the activities to be performed in the various phases of the database development life cycle. Sets the tone for the database project, spells out the key activities, and provides a planning document to guide the project.

**Database practitioners.** Includes the set of IT professionals such as analysts, data modelers, designers, programmers, and database administrators who design, build, deploy, and maintain database systems.

**Database system.** Includes all the components that help define, design, build, and deploy databases. A database system not only includes hardware, systems software, DBMS, and the database, but also includes people and procedures.

**DBMS (database management system).** Software system to store, access, maintain, manage, and safeguard the data in databases.

**DBTG (Database Task Group).** A subgroup of CODASYL responsible for developing standards for DBMSs.

**DCL (data control language).** Language component in a DBMS to control the security of a database system. DCL is used to grant and revoke database access privileges to users.

**DDBMS (distributed database management system).** Software that manages a distributed database as a logical whole.

**DDL (data definition language).** Language component in a DBMS used for defining the data structures to the data dictionary as schema definitions.

**DDLC (database development life cycle).** A complete process from beginning to end, with distinct phases for defining information requirements, creating the data model, designing the database, implementing and maintaining the database.

**Deadlock.** A situation in which neither of two transactions can complete because each is holding exclusively a portion of data in the database that is needed by the other transaction. DBMSs contain provisions for deadlock detection and deadlock prevention.

**Decomposition of relations.** Splitting of relations or tables into smaller relations for the purpose of normalizing them.

**Degree.** The number of entity types or object sets that participate in a relationship. For a binary relationship the degree is 2.

**Denormalization.** Introduction of controlled redundancy in a database, usually done for the purpose of improving data access performance.

**Distributed database.** A database where data is physically stored in more than one location.

**DKNF (domain-key normal form).** This is the ultimate goal in transforming a relation into the highest normal form. A relation is in DKNF if it represents one topic and all of its business rules are able to be expressed through domain constraints and key relationships.

**DML (data manipulation language).** Language component in a DBMS to perform data manipulation operations in the database.

**Domain.** The set of all permissible data values and data types for an attribute of an object set.

**DSS (decision-support system).** Application that enables users to make strategic decisions. DSSs are driven by specialized databases.

**EBCDIC (Extended Binary-Coded Decimal Interchange Code).** A coded character set of 256 eight-bit characters commonly used in mainframe computer systems.

**Encryption.** Encoding of data using key data strings so that the encrypted data becomes unintelligible to an intruder. Decryption restores the decoded data to its original form to be used by an authorized user.

**Entity.** A real-world “thing” of interest to an organization.

**Entity instance.** A single occurrence of an entity type. For example, a single invoice is an instance of the entity type called INVOICE.

**ERD (entity-relationship diagram).** A graphical representation of entities and their relationships.

**Entity set.** The collection of all entity instances of a particular type of entity.

**Entity type.** Refers to the type of entity occurrences in an entity set. For example, all customers of an organization form the CUSTOMER entity type.

**E-R data modeling.** Design technique for creating an ERD from the information requirements.

**External schema.** Definition of the data structures in a database that are of interest to various user groups in an organization. It is the way users view the database from outside.

**Feasibility study.** One of the earlier phases in the DDLIC conducting a study of the readiness of an organization and the technological, economic, and operational feasibility of a database system for the organization.

**Fifth normal form (5NF).** A relation that is already in the fourth normal form and without any join dependencies.

**File organization.** Method or technique of arranging the records of a file on physical storage.

**File-oriented data system.** Data systems earlier than database systems were file-oriented systems depending on sequential, ISAM, and VSAM files without the benefits of database technology.

**Firewall.** A set of defense mechanisms to protect data that is vulnerable over the Internet.

**First normal form (1NF).** A relation that has no repeating groups of values for a set of attributes in a single row.

**Foreign key.** An attribute in a relational table used for establishing a direct relationship with another table, known as the parent table. The values of the foreign key attribute are drawn from the primary key values of the parent table.

**Forms generator.** A component of the DBMS to enable the creation of online forms for data input into and data display from the database.

**Forward engineering.** The mechanism in CASE tools to obtain the physical data model for a target database from the logical data model.

**Forward Recovery.** Method used to start from an earlier state of the database and apply changes. The database is brought forward to a later, consistent state.

**Fourth normal form (4NF).** A relation that is already in the third normal and without any multivalued dependencies.

**Fragmentation.** See *Data fragmentation*.

**Functional dependency.** The value of an attribute B in a relation depending on the value of another attribute A. For every instance of attribute A, its value uniquely determines the value of attribute B in the relation.

**Generalization.** The concept that some objects are general cases of other objects. The objects in the general cases are known as supersets.

**Gerund.** Representation of a relationship between two entity types as an entity type by itself.

**Global information.** Information in a distributed database that is of interest to users in all locations.

**Global transaction.** A transaction in a distributed database system requiring data from multiple locations for completion.

**Hashing algorithm.** A software routine that converts the primary key value to a storage address to store that particular record.

**Hierarchical data model.** A data model where the data structures are arranged in a top-down manner as in an inverted tree.

**Homonyms.** Two or more data elements having the same name but containing different data.

**Horizontal partitioning.** In a relational data model, the division of a relation horizontally as a sets of rows.

**Identifier.** One or more attributes whose values can uniquely identify the instances of an object set or entity type.

**Identifying relationship.** A relationship between two entity types where one entity type depends on another entity type for its existence. For example, the entity type ORDER DETAIL cannot exist without the entity type ORDER.

**Index.** A table or data structure containing sorted values of one or more columns in a given relation and the storage addresses of the rows. An index speeds up data access.

**Inheritance.** The property of subsets inheriting the attributes and relationships of their superset.

**Internal schema.** Definition of the internal structure of a database.

**I/O (input/output).** Abbreviation used for indicating a database read/write operation. Excessive I/Os degrade system performance.

**ISAM (indexed-sequential access method).** A data access method using indexes in file-oriented data systems.

**IT (information technology).** Covers all computing and data communications in an organization. Typically, the CIO is responsible for IT operations in an organization.

**JAD (joint application development).** A methodology for developing computer applications in which IT professionals and end-users cooperate and participate in the development effort.

**Key.** One or more attributes whose values can uniquely identify the rows of a relational table.

**Local information.** Information in a distributed database that is of interest to users in the local location.

**Local transaction.** A transaction in a distributed database system requiring data only from the local location for completion.

**Lock.** Mechanism to prevent access of a portion of the database by other transactions. Locks may be applied at different levels of database structures. Also, there are different types of locks such as shared or exclusive locks.

**Log.** A file used by the DBMS to record all database transactions. The log file is used for recovery of the database in case of failures. The log file is also known as the journal file.

**Logical data model.** Also referred as a conventional data model, consists of the logical data structure representing the information requirements of an organization. This data model conforms to the conventions of a class of database systems such as hierarchical, network, relational, and so on.



- Logical design.** Process of designing and creating a logical data model.
- Model transformation.** Process of mapping and transforming the components of a semantic data model to those of a logical or conventional data model.
- Network data model.** A data model where the data structures are arranged in a network of nodes.
- Nonprocedural language.** Language used for stating *what* result is required from a database transaction rather than stipulating the procedure on *how* to obtain the result.
- Normal form.** A state of a relation free from incorrect dependencies among the attributes. See *Boyce-Codd normal form*, *First normal form*, *Second normal form*, and *Third normal form*.
- Normalization.** The step-by-step method of transforming a random relation into a set of normalized relations free from incorrect dependencies and conforming to the rules of the relational data model.
- Null value.** A value of an attribute, different from zero or blank to indicate a missing, nonapplicable or unknown value.
- Object.** A physical or conceptual “thing” of interest to an organization, data about which is stored in the database.
- Object-based data modeling.** Data modeling method that combines the techniques of data modeling and object technology to create a semantic data model to represent the information requirements of an organization.
- Object instance.** A single occurrence in an object set.
- Object set.** Set of object instances of the same type. The CUSTOMER object set includes all customers.
- ODBC (Open Database Connectivity).** A programming interface from Microsoft that provides a common language interface for Windows applications to access databases on a network.
- Object-relational data model.** Combines the capabilities of object technology to handle complex data types and advanced relationship types with features of data integrity, reliability, and recovery found in the relational realm.
- OLAP (online analytical processing).** Powerful software systems providing extensive multidimensional analysis, complex calculations, and fast response times. Usually present in data warehousing systems.
- Partitioning.** See *Data fragmentation*.
- Performance tuning.** Actions usually taken by the DBA to improve the performance of the database on an ongoing basis.
- Physical data model.** Data model, consisting of the actual components such as data files, blocks, records, storage allocations, indexes, and so on, representing the information requirements of an organization at a physical level of hardware and system software.
- Physical design.** Process of designing the physical data model.
- Physical object set.** Set representing the physical objects themselves, not the types of objects.
- Pointer.** A logical or physical address of a data element in the database, usually embedded within a data record.



**Primary index.** An index file created with values of the primary key and corresponding addresses of the rows containing the primary key values.

**Primary key.** A single attribute or a set of attributes that uniquely identifies an instance of an object set or entity type and chosen as the primary key.

**Procedure.** Detailed steps of instructions for accomplishing a given task. In computer systems, procedures are written using programming languages.

**Procedural language.** Language used for stating the procedure on *how* to obtain a required result from a database transaction rather than just stating *what* result is desired.

**Process.** A set of tasks to accomplish a business function. In computer systems, a process produces outputs from inputs.

**Process-driven.** Design and implementation of applications where the primary emphasis is on the business processes rather than on the data relating to business objects.

**Query.** A computing function that requests data from the database, stating the parameters and constraints for the request.

**Query optimizer.** Part of the query processor in the DBMS responsible to enhance the query for faster processing.

**Query processor.** A component in the DBMS that executes queries.

**RAID (redundant array of inexpensive disks).** A system of disk storage where data is distributed across several drives for faster data access and improved fault tolerance.

**RDBMS.** A relational database management system.

**Recovery manager.** Software component in the DBMS that enables database backup and recovery from failures.

**Referential integrity.** Refers to two relational tables that are directly related. Referential integrity between related tables is established if non-null values in the foreign key attribute of the child table are primary key values in the parent table.

**Relation.** In relational database systems, a relation is a two-dimensional table with columns and rows, conforming to relational rules.

**Relational algebra.** A generic, procedural language containing well-defined operations for data manipulation in the relational data model.

**Relational calculus.** A generic, nonprocedural language for data manipulation in the relational data model.

**Relational data model.** A conventional or logical data model where data is perceived as two-dimensional tables with rows and columns. Each table represents a business object; each column represents an attribute of the object; each row represents an instance of the object.

**Relational database.** A database system built based on the relational data model.

**Relationship.** A relationship between two objects sets or entity types represents the associations of the instances of one object set with the instances of the other object set. Unary, binary, or ternary relationships are the common types depending on the number of object sets participating in the relationship. A unary rela-

tionship is recursive—instances of an object set associated with instances of the same object set. Relationships may be mandatory or optional based on whether some instances may or may not participate in the relationship.

**Repeating group.** A group of attributes in a relation that has multiple sets of values for the attributes.

**Replication.** The method of storing full or partial copies of relations at different locations in a distributed database system.

**Report writer.** A component in the DBMS that enables creation and running of reports.

**Scalability.** In a database system, the ability to support increasing numbers of users and transactions.

**Schema.** Definition in the data dictionary of the entire logical structure of a database.

**Second normal form (2NF).** A relation that is already in the first normal form and without partial key dependencies.

**Secondary index.** An index file created with values of the secondary key and corresponding addresses of the rows containing the secondary key values.

**Secondary key.** One or more attributes in a relation whose values identify a set of rows. These attributes do not constitute the primary key.

**Security manager.** Software component in the DBMS that controls database security and enables the granting and revoking of data access privileges to users.

**Semantic data model.** A generic data model that captures the true meaning of the information requirements of an organization. Does not conform to the conventions of any class of database systems such as hierarchical, network, relational, and so on. The two common data modeling techniques for creating a semantic data model are: object-based and entity-relationship.

**Sequential file.** In a sequential file, data records are stored in the order in which they are entered into the file. Data systems earlier than database systems used sequential files extensively.

**Serializability.** A set of concurrent database transactions are said to be serializable if they produce the same result as if the transactions were executed serially one after the other. Serializability is a significant requirement for concurrency control.

**Space management.** Administration and control of storage space for a database, usually performed by the DBA.

**Specialization.** The concept that some objects are special cases of other objects. The objects in the special cases are known as subsets.

**SQL (Structured Query Language).** Has become the standard language interface for relational databases.

**Stored procedure.** A software program stored in the database itself to be executed on the server based on stipulated conditions.

**Subquery.** A second query within a main query.

**Subschema.** Subsets of the full schema defining specific user views of the database.

**Subset.** An object that is a special case of another object known as the superset.

**Superset.** An object that is a general case of another object known as the subset.

**Surrogate key.** A unique value generated by the computer system used as a key for a relation. A surrogate key has no business meaning apart from the computer system.

**Synonyms.** Two or more data elements containing the same data but having different names.

**Third normal form (3NF).** A relation that is already in the second normal form and without any transitive dependencies—that is, the dependency of a nonkey attribute on the primary key through another nonkey attribute, not directly.

**Transaction.** Database operations that are executed as a unified set on the data in the database. For example, the transfer of money from the savings to the checking account involving database changes to two relations is considered as a transaction.

**Transaction manager.** Software component in the DBMS that enables the proper execution of transactions.

**Transitive dependency.** In a relation, the dependency of a nonkey attribute on the primary key through another nonkey attribute, not directly.

**Tuple.** A row in a relational table.

**Two-phase commit.** A technique for concurrency control in a distributed database environment. Commit of a transaction happens in two phases—first, the coordinating DBMS sends updates to other databases asking them to execute; next, coordinating the DBMS sends messages to commit after participating databases agree on the updates.

**Vertical partitioning.** In a relational data model, the division of a relation vertically as sets of columns.

**VSAM (virtual sequential access method).** An improved data access method in file-oriented data systems.

**Weak entity.** An entity that depends on another entity for its existence. For example, the entity type ORDER DETAIL cannot exist without the entity type ORDER.