DATABASE TERMINOLOGY

RELATIONAL DATABASES TERMS GLOSSARY- BONUS RESOURCE

RARES ILEA

2018

ACID - The acronym standing for the properties maintained by standard database management systems, standing for Atomicity, Consistency, Isolation, and Durability. ACID refers to the properties on SQL Transactions.

Atomicity - The property of a transaction that guarantees that either all or none of the changes made by the transaction are written to the database.

Attribute - A database attribute is a characteristic of a database entity. Simply put, an attribute is the column in a database table, which itself is known as an entity.

Authentication - Databases use authentication to ensure that only authorized users can access the database or certain aspects of the database. For example, administrators might be authorized to insert or edit data, while regular employees might be able to only view data. Authentication is implemented with usernames and passwords.

BLOB - An abbreviation for Binary Large Object. In SQL, an example of a BLOB is a image file or a PDF file.

B-tree - An indexing method in which the values of the columns used in the index are efficiently maintained in sorted order that also provides fast access - logarithmic access complexity = O(log n) - to an individual index entry.

Candidate Key – is a synonym for key, in a relation schema with more than one key.

Cascade - A foreign key attribute that automatically migrates the changes made to a referenced (i.e., primary key) table to all of the referencing (foreign key) table rows.

Cloud - Cloud is a recently coined term used to describe an execution model for computing systems where functions and data are invoked by a name that refers to a remote system whose location is irrelevant (hence the concept of it being "out there somewhere." like a cloud). Cloud-based systems allow thin-client interfaces to access this functionality through the Internet, reducing the power requirements of the client computers.

Column - A single unit of named data that has a particular data type (e.g., number, text, or date). Columns only exist in tables.

Commit - The action that causes the all of the changes made by a particular SQL transaction to be reliably written to the database.

Consistency - The property of a SQL transaction that guarantees that the state of the database both before and after execution of the transaction remains consistent (i.e., free of any data integrity errors) whether or not the transaction commits or is rolled back.

Constraints - A database constraint is a set of rules that define valid data. Multiple types of constraints exist. The primary constraints are:

- Unique constraints: A field must contain a unique value in the table.
- CHECK constraints: A field can contain only specific data types and even certain allowable values.
- DEFAULT constraints: A field will contain a default value if it has no existing value; this eliminates a NULL value.
- PRIMARY KEY Constraints: The primary key must be unique.
- FOREIGN KEY Constraints: The foreign key must match an existing primary key in another table.

Data Entry - The process of getting information into a database, usually done by people typing it in by way of data-entry forms designed to simplify the process.

Data Redundancy - Data redundancy is a situation in a database in which copies of a given piece of data are housed in 2 different places. This redundancy can be achieved if data is held in multiple places in the same database, in multiple databases on the same computer, or in multiple databases across multiple computers, perhaps even using different database management server software.

Data Type - The basic kind of data that can be stored in a column. The basic data types that are: TEXT, NUMBERS and DATES.

DBMS - An acronym for Database Management System.

Durability - The property of a SQL transaction in which the DBMS guarantees that all committed transactions will survive any kind of system failure.

Entity - An entity is simply a table in a database. It is described using an Entity-Relationship Diagram, which is a type of graphic that shows the relationships between database tables.

Field - An alternative name for column or attribute. Fields describe a single aspect of each member of a table. A student record, for instance, might contain a last name field, a first name field, a date of birth field and so on. All records (rows) have exactly the same structure, so they contain the same fields. The values in each field vary from record to record, of course.

Flat File - A database that consists of a single table. Lightweight database programs such as the database component in Microsoft Works are sometimes called 'flat-file managers' (or list managers) because they can only handle single-table databases. More powerful programs, such as FileMaker Pro, Access, Approach and Paradox, can handle multi-table databases, and are called relational database managers, or RDBMSs.

Foreign Key - One or more columns in a table intended to contain only values that match the related primary/unique key column(s) in the referenced table. Foreign and primary keys explicitly define the direct relationships between tables. Referential Integrity is maintained when every foreign key refers to one and only one existing primary key.

Functional Dependency - A functional dependency constraint helps to ensure data validity, and exists when one attribute determines the value of another, described as A -> B which means that the value

of A determines the value of B, or that B is "functionally dependent" on A. For example, a table in a university that includes records of all students might have a functional dependency between the student ID and the student name, i.e. the unique student ID will determine the value of the name.

Hierarchical Model - A special case of a network model database in which each record type can participate only as the member of one set.

IEC - International Electrotechnical Commission. Along with the ISO, the IEC controls the SQL standard (ISO/IEC 9075) and many others as well.

Index - An index is a data structure that helps speed database queries for large datasets. Database developers create an index on particular columns in a table. The index holds the column values but just pointers to the data in the rest of the table, and can be searched efficiently and quickly.

In-memory - A feature in which the DBMS keeps the entire contents of a database or table available in computer memory at all times while the database is opened. Frequently, in-memory databases are volatile, meaning that they have little or no durability if the computer malfunctions.

ISO - International Organization for Standardization. Along with the IEC, the ISO controls the SQL standard (ISO/IEC 9075) and many others as well.

Isolation - The property of a SQL transaction that guarantees that the changes made by a transaction are isolated from the rest of the system until after the transaction has committed.

Java - A multi-platform, object-oriented programming language, which is freely available to any and all software developers. It is particularly important in the development of internet/web and mobile applications.

JDBC - Java Database Connectivity API. JDBC provides a standard database access and manipulations API for Java programs. MySQL supports JDBC.

Join - An operation in which the rows of one table are related to the rows of another through common column values.

Key - A key is a database field whose purpose is to uniquely identify a record. Keys help enforce data integrity and avoid duplication. The main types of keys used in a database are candidate keys, primary keys foreign keys.

Candidate keys: The set of columns that can each uniquely identify a record and from which the primary key is chosen.

Primary keys: The key chosen to uniquely identify a record in a table. This key cannot be NULL.

Foreign keys: The key linking a record to a record in another table. A table's foreign key must exist as the primary key of another table.

Locking - A method for safely protecting objects from being changed by two or more users (processes/threads) at the same time. A write (exclusive) lock allows access from only one user (process/thread) at a time. A read (shared) lock allows read-only access from multiple users (processes/threads).

Metadata - "Data about data." In a DBMS context, data stored in columns of a table have certain attributes, such as the *type*, *length*, *description* or other characteristics that allow the DBMS to process the data meaningfully, or allow the users to understand it better.

Network Model - A database in which inter-record type relationships are organized using one-to-many sets. This differs from a Hierarchical Model in that it allows a record type to be a member of more than one set. Individual rows can be retrieved using API functions that allow an application to navigate through individual set instances.

Normalization - To normalize a database is to design its tables (relations) and columns (attributes) in a way to ensure data integrity and to avoid duplication. The primary levels of normalization are First Normal Form (1NF), Second Normal Form (2NF), and Third Normal Form (3NF). There are other 3 normal forms, rarely used in practical applications: Boyce-Codd Normal Form (BCNF), 4NF and 5NF.

NoSQL - NoSQL is an umbrella term, one which encompasses a number of different technologies. These different technologies aren't even necessarily related in any way beyond the single defining characteristic of NoSQL: they are not relational in nature. This lack of relational structure results in unstructured or semi-structured data in storage; there may be structure, but it is loose in nature, lax in enforcement.

Often, NoSQL is used to mean "not only SQL," meaning that these solutions are more flexible and less rigid in nature. I'm sure there are die-hards in this argument of terminology ownership, but just be aware of the potential difference in definition.

Null - The value NULL is frequently confused to mean "none" or zero; however, it actually means "unknown." If a field has a value of NULL, it is a placeholder for an unknown value. Structured Query Language (SQL) uses the IS NULL and IS NOT NULL operators to test for null values.

Object-oriented - A computing programming paradigm that defines the computing problem to be solved as a set of objects that are members of various object classes each with its own set of data manipulation methods. Individual objects which have been instantiated (created) can be manipulated only using those prescribed methods.

Optimizer - A component of the SQL system that estimates the optimum (i.e., fastest) method to access the database data requested is by particular SQL SELECT, UPDATE, or DELETE statement.

Page Size - The size in bytes of a database page.

Page - The basic unit of database file input/output. Database files may be organized into a set of fixed-sized pages containing data associated with one or more record occurrences (table rows).

PL/SQL - A SQL based programming language. This allows for a SQL programmer to use programming constructs like variables, conditionals and loops purely through the use of SQL statements.

Primary Key - A column or group of columns in a given table that uniquely identifies each row of the table. The primary key is used in conjunction with a foreign key in another (or even the same) table to relate the two tables together. For example, the primary key in an author table would match the foreign key in a book table in order to relate a particular author to that author's books.

Query - A database query is how users interact with a database. It is usually written in SQL and can be either a select query or an action query. A select query requests data from a database; an action query changes, updates or adds data. Some databases provide forms that hide the semantics of the query, allowing users to easily request information without having to understand SQL.

RDBMS - Relational database management system. A program which lets you manage structured information stored in tables and which can handle databases consisting of multiple tables linked in a meaningful way using relationships.

Record - One set of related data field values associated with a specific record type—equivalent to an table row.

Referential Integrity - A condition in which the foreign key column values in all of the rows in one table have matching rows in the referenced primary key table. Referential integrity is maintained by SQL during the processing of an INSERT and DELETE statement and any UPDATE statement that modifies a foreign or primary key value.

Relational Databases - Relational Databases are the most common database systems used today. Some examples are: SQL Server, Oracle Database, IBM Informix, MySQL and many others.. The relational database management systems (RDMS) feature very good performance for managing data, they allow multiple users (even thousands!) to work with the data at the same time, and are providing advanced security for accessing the data. As we already saw, relational databases store data in columns and rows, which in turn make up tables. These tables are linked to each other using relationships, thus the name relational. A set of tables makes up a schema. A number of schemas create a database. Many databases can be created on a single server.

Replication - A process where selected modifications in a master database is replicated (re-played) into another database.

Result Set - The complete set of rows that is returned by a particular SELECT statement.

Rollback - An operation, usually performed by the SQL ROLLBACK statement, that discards all of the changes made by all INSERT, UPDATE and DELETE statements that have been executed since the most recently started transaction (e.g., START TRANSACTION statement).

Row - One set of related values for all of the columns declared in a given table; also known as a record.

Runtime - A portion of a DBMS that is included within the process space of an application program.

Scalability - A software system is scalable when its performance and overall system throughput continues to improve as more computing resources are made available for its use. This usually comes in the form of the number of CPUs and cores available in the computer on which the software system is run.

Schema - A database schema is the design of tables, columns, relations, and constraints that make up a database.

Server (Software) - A Seat that resides on a single Server machine and is capable of accepting connections from one or more Client machines.

SQL - The standardized and commonly accepted language used for defining, querying and manipulating a relational database. The etymology of "SQL" is unclear, possibly a progression from "QueL" (Query Language) to "SeQuel" to "SQL." The Data Manipulation Language (DML) contains the subset of SQL commands used most frequently and includes SELECT, INSERT, UPDATE and DELETE.

Stored Procedure - A stored procedure is a pre-compiled query, or SQL statement that can be shared across multiple programs and users in a Database Management System. Stored procedures improve efficiency, help enforce data integrity and boost productivity.

Table - A collection of closely related columns. A table consists of rows each of which shares the same columns but vary in the column values.

Transaction - A set of logically related database modifications that is written to the database as a unit. The database changes associated with a given transaction are guaranteed by the DBMS to be

written completely to the database; in the event of a system failure, none are written. The state of the database both before and after a transaction will be consistent with its design.

Trigger - A trigger is a stored procedure set to execute given a particular event, usually a change to a table's data. For example, a trigger might be designed to write to a log, gather statistics or compute a value.

View - A database view is a filtered set of data displayed to the end user in order to hide data complexity and streamline the user experience. A view can join data from two or more tables and contains a subset of information.