

Foundations of Computer Science - SQL

1 Introduction to Relational DataBases

The relational DB are organized in tables that can be linked together. A table in the relational model represents a relation. For example in the following database each row describes a single book:

ISBN	Title
978-1-449-30321-1	Scaling MongoDB
978-1-491-93200-1	Graph Databases
978-1-449-39041-9	Cassandra: The Definitive Guide
007-709500-6	Database Systems

Table 1: Some sample books

The data in a relational model is organized into columns and each entry(cell) contains a **SINGLE** piece of data, so the main problem of relational model is to handle data with multiple elements in a single column for example a book with two authors.

The solution proposed by the relational model is to link two tables and to do so we need an identifier ID for each row. Here is an example of two linked tables.

book_id	ISBN	Title
1	978-1-449-30321-1	Scaling MongoDB
2	978-1-491-93200-1	Graph Databases
3	978-1-449-39041-9	Cassandra: The Definitive Guide
4	007-709500-6	Database Systems

Table 2: Books Table with ID

author_id	Name	Surname
1	Ian	Robinson
2	Kristina	Chodorow
3	Riccardo	Torlone
4	Paolo	Atzeni
5	Stefano	Ceri
6	Stefano	Paraboschi
7	Eben	Hewitt

Table 3: Authors Table with ID

To connect these two tables we use a “relation” or “join” table. In the example the “join” table is the following:

book_id	author_id
2	1
3	7
1	2
4	4
4	5
4	6
4	3

Table 4: BookAuthors Table

In the Table 4 we can see the need of using an ID to identify each row, so it can be referenced uniquely. In some cases we might not need a third table (join table): one of these cases is the “One-to-Many” relationship. The One-to-Many relationship also requires that rows in tables have unique IDs, but unlike the join table used in Many-to-Many relationship, the table with the many side of data has a column reserved for the IDs of the one side of data.

The IDs used to uniquely identify the rows described in the tables are called “Primary Keys” (PK). If this primary key is used in another table it is called “Foreign Key” (FK) and it is usually not unique in the new table. The purpose of the foreign key is to link the two tables in one-to-many relationship.

Usually the primary key generation process is left to the RDBMS (it’s safer this way), which automatically generates the key and usually it is an ordinary integer. For the join tables, the primary key is a combination of the foreign keys. A primary key comprised of more than one attribute is called a “Composite Primary key” (CPK).

2 Introduction To SQL

SQL stands for Structured Query Language and it is a standard language for querying and manipulating data. It is a very high-level programming language and it is very well optimized.

SQL is a:

- Data Definition Language (DDL):
It defines relation *schema* and creates/alters/deletes tables and their attributes.
- Data Manipulation Language (DML):
It Inserts/deletes/modifies tuples in tables and queries one or more tables.

A **relation** (or **table**) in SQL is a multiset of tuples having the attributes specified by the schema. Where a **multiset** is an unordered list (so multiple duplicates instances are allowed) and an **attribute** is a typed data entry present in each tuples in the relation and it must have an atomic type in standard SQL. The atomic types are:

- Characters: CHAR(20), VARCHAR(50)
- Numbers: INT, BIGINT, SMALLINT, FLOAT
- Others: MONEY, DATETIME, etc.

A **tuple** (**row**) is a single entry in the table having the attributes specified by the schema.

The **schema** of a table is the table name, it’s attributes and they types, a key is an attribute whose values are unique. A **key** is a minimal subset of attributes that acts as a unique identifier for tuples in a relation. A key is an implicit constraint on which tuples can be in the relation: so if two tuples agree on the same value of the key, then they must be the same tuples.

If some information is missing or is not known SQL uses the NULL. To check if a value is NULL we cannot say value=NULL since it’s unknown we need to use “IS NULL” command of SQL. SQL offers the possibility to constrain a column to be NOT NULL or supports other constraints such as maximum number of values

per attributes. It's thanks to the schema and constraints that the databases understand the semantics of the data.