

: *What is the theory of relational databases?*

by Dan Power

Editor, DSSResources.com

Relational databases organize data elements in clusters that are meaningful to users and useful for performing tasks. The goal is to create a logical, well-structured data model. Rules help a database designer make sure a collection of data is well-defined and meaningful. Designers usually implement data relations as a table that is organized in rows and columns. In data-driven DSS, relational online analytical processing (ROLAP) tools access data in a relational database and generate SQL queries to create information for users.

The theory of relational databases was proposed by Edgar Codd at IBM Almaden Research Center in 1970. Codd published a paper entitled "A Relational Model of Data for Large Shared Data Banks" (June 1970 issue of Communications of the ACM). In that paper, Codd introduced a set of rules intended to eliminate the need to store redundant data. These rules are the foundation of relational database theory.

In the relational model, a primary key uniquely defines a relationship within a database. A foreign key is a reference to a primary key in another relation, i.e., the referencing table includes as a field the values of a primary key in the referenced table.

Normalization was first proposed by Codd as an integral part of the relational model. The process of normalizing eliminates the duplication of data, which in turn prevents data manipulation anomalies and loss of data integrity. The rules of normalization applied to databases define the normal forms.

The relational model is a data model that represents data in the form of relations. A data model is a conceptual collection of the data, relationships, and constraints on the data. Therefore, the relational data model is a conceptual representation of the objects, events, and associations in a relational database system. The relational data model requires that data be stored in relations.

Why are relational database management systems (RDBMS) and data warehouses needed?

The flat file processing approaches that preceded RDBMS and DW created data redundancy and inconsistency, difficulty in accessing data, data isolation, data integrity problems and security problems.

: *What is the theory of relational databases?*

Often the same information was duplicated in several files and all copies may not be updated. Because of the many files it was often necessary to write a new application program to satisfy an ad hoc data request. Because data was in many files often with many file formats, it was difficult to write new application programs. In general it was difficult to enforce or to change integrity constraints with the file-processing approach. Finally, security is difficult to manage with many files and many isolated application programs.

The major purpose of a database system is to provide users with an understandable view of the data. The system hides most details of how data is stored and created and maintained from users. The relational data model describes the organization of the database. It is often created as an entity relationship diagram.

What are the rules? The following rules are modified from Rettig.

1. Eliminate repeating groups of data. Make a separate table for each set of related attributes, and give each table a primary key.
2. Eliminate redundant data. If an attribute depends on only part of a multi-valued key, move it to a separate table.
3. All fields in a relation must be dependent on the primary key. If attributes do not contribute to a description of the key, move them to a separate table.
4. A table should use foreign keys to create relationships. Isolate independent multiple relationships. No table may contain two or more 1:n (one-to-many) or n:m (many-to-many) relationships that are not directly related.
5. Separate logically related many-to-many relationships

Wiorowski and Kull in the DB2 Design & Development Guide summarize the relational theory and the rules leading to and including the third normal form for creating relations: **"Each attribute must**

: *What is the theory of relational databases?*

be a fact about the key, the whole key, and nothing but the key."

References

Chapple, M. About.com Guide, "Database Normalization Basics," URL <http://databases.about.com/od/specificproducts/a/normalization.htm>, retrieved May 11, 2010.

Codd, E.F. (1970). "A Relational Model of Data for Large Shared Data Banks". *Communications of the ACM* 13 (6): 377–387.

Edgar F. Codd, from Wikipedia, the free encyclopedia at URL http://en.wikipedia.org/wiki/Edgar_F._Codd

Rettig, M. 5 Rules of Data Normalization, Database Programming & Design poster, San Francisco, Miller Freeman.

ROLAP, from Wikipedia, the free encyclopedia at URL <http://en.wikipedia.org/wiki/ROLAP> .

Author: Daniel Power

Last update: 2010-04-11 02:29