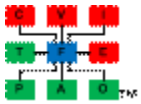

Definition of a Database

1. INTRODUCTION

Users deal with data as structures within forms, screens, reports, inquiries, etc. These data structures are not a database, they are components of the realization of Codd's Relational Model in which Codd describes a mathematical algebra of relations and operations that provides the specification for consistent data reporting.

When developing a database, a methodology (Database Development Life Cycle) guides the project through the processing required and its deliverables. The design phase of the database has to adhere to the definition of a database for the business data, the specification phase has to comply with business processing and regulatory requirements and the implementation phase has to ensure that the DBMS processing adheres to the application users' processing goals.

2. DATABASE DEFINITION

According to Codd, a database is a minimum cover set of the relational model. The database is a set of tables. A table is a set of attributes. And, the data within a table is a set of table rows.

To get to the database from Codd's relational model, replace the relation labels with business data terms (Kent) and domain labels with attribute names to produce a universe of discourse (Simsion) or data lake.

The start of the definition of a database, is Codd's Information Principle:

> "The entire information content of a logical data model is represented in one and only one way: namely, as attribute values within rows within tables."

This principle implies that there is no metadata information in attribute or table labels and requires using logical data types as part of the attribute definition for the logical data model. Attribute names that imply additional information, such as wages per hour or annual salary, mean that the information principle has been violated.

In the terms of Codd's definition, the properties of a database are:

2.1 No overlapping tables

There are no logical tables containing the same attributes. This is a requirement because the database is a minimum cover set. In the enterprise logical data model, this means that no tables have overlapping meanings. To have overlapping meaning, different logical tables would have to include the same attribute.

This property is restated by Date as the Principle of Orthogonal Design.

> "No two entities in a given logical data model should have overlapping meanings."

2.2 No overlapping attributes

There are no redundant attributes within a single logical table. This is a requirement because the logical table is a set. In the enterprise logical data model, this means that within a table there no duplicate attribute names or an attribute that can be derived algorithmically from other attributes in the table.

2.3 There are no duplicate rows in any table

For the rows in a logical table, there are no duplicate rows. This is a requirement because the contents of a logical table are a set. Table keys are the DBMS implementation of the set property of the relation as required by Codd's Relational Data Model To validate this in the enterprise logical data model, every table has a primary key or an alternate key that is the natural identifier of the data row. Using only a surrogate key on a table allows for data rows to be duplicated.

2.4 Each element of the database has a single meaning

Each element (table or attribute) of the database has a single definition. The ambiguity of multiple meanings would make it impossible to verify the requirement that tables and attributes are sets.

2.5 No redundant relationships

There are no redundant relationships between the logical tables within the database. This is a requirement because the database is a set. In the enterprise logical data model, redundancy of relationships means there is more than one path between two tables which leads to ambiguity of data reporting.

3. REFERENCES

Codd, Edgar, “A relational model for large shared databanks”, Communications of the ACM, vol 13, number 6, June 1970

Date, C.J., *The Relational Database Dictionary*, 1st Edition, O’Reilly Media, Inc., Sebastopol, CA, 2006.

Date C.J., McGoveran D, “The Principle of Orthogonal Design”, Database Programming & Design 7, No. 6 (June 1994)

Kent, William, *Data and Reality*, North Holland Publishing, New York, 1978

Simsion, Graeme C. and Graham C. Witt, *Data Modeling Essentials*, 3rd Ed., Morgan Kaufman Publishers, San Francisco, CA, 2005

Simsion, Graeme C., *Data Modeling Theory and Practice*, Technics Publications, LLC, Bradley Beach, New Jersey, 2007