

AN INTRODUCTION TO DATA MODELS

INTRODUCTION

A data model organizes data elements and standardizes their relation to each other and the properties of real-world entities. A database model is a type of data model that determines a database's logical structure and how data can be stored, organized and manipulated in that database. A normalized relational database and unnormalized star schema are two of the most common database models, each of which has its advantages.



RELATIONAL DATABASE DATA

E.F. Codd introduced the relational database model in 1970, which uses tables to make databases independent of other software applications. Database normalization is the process of structuring a relational database to reduce data redundancy and improve data integrity. It generally consists of organizing the columns, or attributes in tables to enforce various data integrity constraints, either by creating a new database design or improving the existing design. Normalization is a progressive process, meaning that the requirements for one level of normalization must be met before the next level can be achieved.

FIRST NORMAL FORM

Unnormalized data may have a primary key, meaning the table doesn't have duplicate records. First normal form (1NF) requires a primary key and also requires all cells in the table to have single values. The most common solution to this problem is to move attributes that could have multiple values to a separate table, along with a primary key to connect the new table to the old table. Assume for this example that a table's attributes include the title and subject of books such that one book may have more than one subject. One way to place this table into 1NF would be to move the subject attribute to another table, which would also include the book's title.



SECOND NORMAL FORM

Once a table is in 1NF, placing it in second normal form (2NF) requires removing its partial dependencies, meaning that all attributes must depend on the entire primary key. Consider a table that contains the attributes of title, format and price such that the same book can have a different price depending on its format, which could be a hardback, paperback or ebook. Furthermore, the only candidate for the primary key is the combination of the title and format attributes. Placing this table in 2NF requires moving the format and price attributes to a second table that also contains the title attribute. Note that the title attribute now becomes the only candidate for the primary key in the first table.

THIRD NORMAL FORM

Third normal form (3NF) requires the removal of all transitive dependencies, meaning that attributes may only depend on the primary key. Assume for this example that the book table contains the genre id and genre name for the books. While both of these attributes depend on the primary key of title, the genre name also depends on the genre id. Placing this table in 3NF requires the genre name to be moved to another table that also contains the genre id.

STAR SCHEMA

Some data models are more generalized than the relational model and are sometimes known as post-relational models. A star schema is one such model and is the most common data model used in data warehouses. Star schemas are unnormalized and consist of at least one fact table that references any number of dimension tables.

FACT TABLES

Fact tables record measurements for specific events, so they usually consist of only numeric values and a foreign key to a dimensional table containing descriptive information for those values. Fact tables should record measurements at a low level of detail, or granularity, which typically results in the accumulation of many records over time. Fact tables may be classified into three types, including transaction fact tables, snapshot fact tables and accumulating snapshot tables.

DIMENSION TABLES

Dimension tables usually have a smaller number of records compared to fact tables, although the number of attributes in their records can be very large. These tables can define many types of dimensions, but the most common ones include time dimensions that describe the time at which events are recorded in the fact table. Range dimensions describe a range of measurable quantities to simplify reporting. Other common dimension tables include tables for employees, geography, and products.



DESIGN CONSIDERATIONS

The primary advantages of a normalized relational database as compared to a star schema are that a normalized database strictly reinforces data integrity, which prevents anomalies like one-off inserts and updates. Normalized databases also have greater flexibility in performing analytical tasks, provided they follow the database model's business logic. Furthermore, normalized database models support many-to-many relationships between business entities more easily than star schemas, which must simplify these relationships to conform to the dimensional data model.

The benefits of star schemas include simpler reporting logic compared to a normalized database, especially as-of and period-over-period reporting. Star schemas also have simpler join-logic, which improves performance on read-only operations like reporting and queries. This advantage also improves the performance of aggregation operations in star schemas. In addition, all online analytical processing (OLAP) systems use star schemas to build proprietary cubes efficiently. An OLAP cube is a three-dimensional dataset, as opposed to a traditional two-dimensional table. For example, an OLAP user can represent financial data by product, time period, and location with a cube.



IDERA'S SOLUTION

ER/Studio Data Architect

Document, design, and communicate data assets faster and easier.

Round-trip database support gives ER/Studio Data Architect users the power to easily reverse- and forward-engineer, compare and merge, and visually document data assets across multiple relational, NoSQL, and ETL platforms and data sources. Built-in facilities automate routine data modeling tasks so users can analyze and optimize database and data warehouse designs faster than ever.

Start a free, fully-functional, 14-day trial today!

Start for FREE

