

Relational Data Models

Wednesday, April 29, 2020 12:17 AM

Importance of Relational Databases:

- **Standardization of data model:** Once your data is transformed into the rows and columns format, your data is standardized and you can query it with SQL
- **Flexibility in adding and altering tables:** Relational databases gives you flexibility to add tables, alter tables, add and remove data.
- **Data Integrity:** Data Integrity is the backbone of using a relational database.
- **Structured Query Language (SQL):** A standard language can be used to access the data with a predefined language.
- **Simplicity :** Data is systematically stored and modeled in tabular format.
- **Intuitive Organization:** The spreadsheet format is intuitive but intuitive to data modeling in relational databases.

OLAP VS OLTP

Online Analytical Processing (OLAP):

Databases optimized for these workloads allow for complex analytical and ad hoc queries, including aggregations. These type of databases are optimized for reads.

Online Transactional Processing (OLTP):

Databases optimized for these workloads allow for less complex queries in large volume. The types of queries for these databases are read, insert, update, and delete.

Structuring the Database:

Objectives of Normal Form:

1. To free the database from unwanted insertions, updates, & deletion dependencies
2. To reduce the need for refactoring the database as new types of data are introduced
3. To make the relational model more informative to users
4. To make the database neutral to the query statistics

Normal Forms

1. How to reach First Normal Form (1NF):

- Atomic values: each cell contains unique and single values

- Be able to add data without altering tables
- Separate different relations into different tables
- Keep relationships between tables together with foreign keys

2. Second Normal Form (2NF):

- Have reached 1NF
- All columns in the table must rely on the Primary Key

3. Third Normal Form (3NF):

- Must be in 2nd Normal Form
- No transitive dependencies
- Remember, transitive dependencies you are trying to maintain is that to get from A \rightarrow C, you want to avoid going through B.

When to use 3NF:

- When you want to update data, we want to be able to do in just 1 place. We want to avoid updating the table in the Customers Detail table (in the example in the lecture slide).

Normalization

- The process of trying to improve the read performance of a database at the expense of losing some write performance by adding redundant copies of data

Logical Design Change

- The Designer is in charge of keeping data consistent
- Reads will be faster(select)
- Writes will be slower(insert, update, delete)

Key Concepts:

Normalization is about trying to increase data integrity by reducing the number of copies of the data. Data that needs to be added or updated will be done in as few places as possible.

Denormalization is trying to increase performance by reducing the number of joins between tables (as joins can be slow). Data integrity will take a bit of a potential hit, as there will be more copies of the data (to reduce JOINS).

Fact and Dimension Tables

Fact Tables

- Fact table consists of the measurements, metrics or facts of a business process