

SQL Data Science for Beginners

What is a Database?

A database is an organized collection of structured data stored electronically in a computer system. Databases allow for efficient data storage, retrieval, updating, and management. They serve as the backbone for most modern applications, from simple websites to complex enterprise systems.

Real-world examples:

- Bank systems storing customer accounts and transactions
- E-commerce platforms managing product catalogs and orders
- Hospital systems maintaining patient records
- Social media networks storing user profiles and connections

Database Structure :



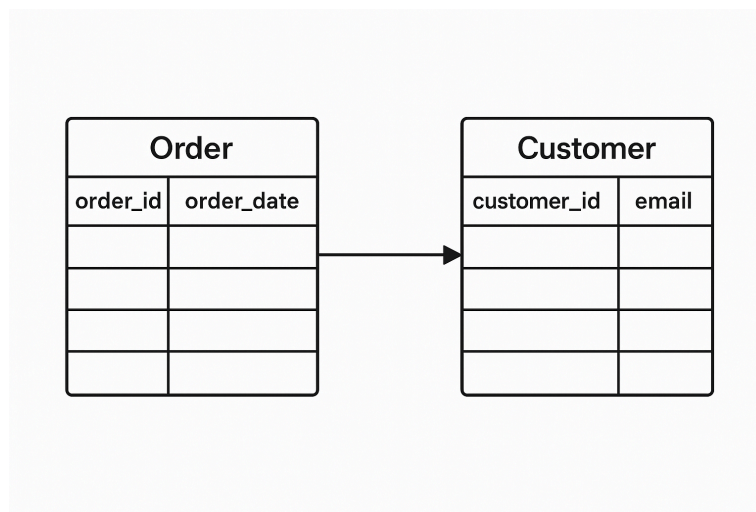
SQL Data Science for Beginners

Types of Databases

Relational Databases (SQL Databases)

Relational databases organize data into tables (relations) with rows and columns, where relationships between tables are established through keys.

Basic Structure of a Relational Database :



Characteristics:

- Use Structured Query Language (SQL) for operations
- Data is stored in tables with strict schemas
- Support ACID properties (Atomicity, Consistency, Isolation, Durability)
- Excellent for complex queries and transactions

Popular relational databases:

1. MySQL (open-source, widely used for web applications)

SQL Data Science for Beginners

2. PostgreSQL (advanced open-source RDBMS)
3. Oracle Database (enterprise-grade commercial database)
4. Microsoft SQL Server (Microsoft's enterprise RDBMS)
5. SQLite (lightweight, embedded database)

Example use cases:

- Financial systems requiring strict data integrity
- Inventory management systems
- Customer relationship management (CRM) systems

NoSQL Databases

NoSQL (Not Only SQL) databases provide flexible schemas and scale horizontally to handle large volumes of unstructured or semi-structured data.

Structure of a Non-Relational (NoSQL) Database : Document Database

```
{
  "_id": "customer1",
  "name": "John Doe",
  "age": 30
  orders: [
    "order_id": "order1",
    amount: 100.0
  ],
  "order_id": "order2",
  amount: 200.0
}
```

SQL Data Science for Beginners

Types of NoSQL databases:

1. **Document databases** (MongoDB, CouchDB) - Store data in JSON-like documents
2. **Key-value stores** (Redis, DynamoDB) - Simple pairs of keys and values
3. **Wide-column stores** (Cassandra, HBase) - Use tables, rows, and dynamic columns
4. **Graph databases** (Neo4j, Amazon Neptune) - Store data in nodes and relationships

Characteristics:

- Flexible schema design
- Horizontal scalability
- Optimized for specific data models and access patterns
- Often sacrifice some ACID properties for performance and scalability

Example use cases:

- Big data applications
- Real-time web applications
- Content management systems
- Social networks with complex relationships

Introduction to MySQL

MySQL

SQL Data Science for Beginners

MySQL is the world's most popular open-source relational database management system (RDBMS), originally developed by MySQL AB and now owned by Oracle Corporation.

Key features:

- Free and open-source (under GNU GPL)
- Cross-platform (Windows, Linux, macOS)
- High performance and reliability
- Works well with web applications (often used with PHP)
- Supports large databases (up to 50 million rows or more in a table)

Database Basics:

Creating a Database: `CREATE DATABASE`

The `CREATE DATABASE` statement is used to create a new database in your database management system.

Syntax:

```
CREATE DATABASE database_name;
```

Real-world Example:

Let's create a database for an e-commerce application:

```
CREATE DATABASE ecommerce_store;
```

Important Notes:

SQL Data Science for Beginners

- Database names are typically case-sensitive in Linux/Unix but not in Windows
- Avoid using spaces or special characters in database names
- In MySQL, you need appropriate privileges to create databases

Using a Database: `USE database_name`

The `USE` statement selects a specific database to work with for all subsequent operations.

Syntax:

```
USE database_name;
```

Real-world Example:

After creating our e-commerce database, we need to select it before creating tables:

```
USE ecommerce_store;
```

Why this is important:

- Most database systems allow multiple databases on a single server
- You must specify which database you're working with before creating tables or querying data
- Without `USE`, you would need to qualify every table reference with the database name (e.g., `ecommerce_store.customers`)

Practical Scenario:

1. Connect to MySQL server

SQL Data Science for Beginners

2. Create database: `CREATE DATABASE ecommerce_store;`
3. Select database: `USE ecommerce_store;`
4. Now you can create tables within this database

Dropping a Database: `DROP DATABASE`

The `DROP DATABASE` statement permanently deletes an entire database, including all its tables, data, and associated files.

Syntax:

`DROP DATABASE database_name;`

Real-world Example:

To remove a test database that's no longer needed:

`DROP DATABASE old_test_store;`

Safety Considerations:

1. Use **IF EXISTS** to avoid errors if database doesn't exist:

`DROP DATABASE IF EXISTS old_test_store;`