

Introduction to Databases and DBMS

DSC 301

January 22, 2021

Objectives

- Why databases are important?
- What is a database?
- What are the common Databases
- What role associated with database environments
- Types of databases
- Features of databases
- Database Architecture

1 Why are databases important?

Databases have become ubiquitous in society. Annual sales from the database market is approximately \$50 billion (and growing).

- Most important development in software engineering
- Underlying framework of information systems
- Fundamentally changed the way organizations operate
- Backbone of the web
 - Nearly all website have backend databases driving content
 - * WordPress (CMS), Joomla, Blackboard
 - * Amazon (online store), Facebook, IRS,
 - * Online banking, travel (airline reservations)
- Every business leverages databases (employees, tax records, etc.)

Data is important

- Data is new oil.
 - Oil and energy companies dominated the top of the valuable firms in the world until around 2016.
 - * Alphabet, Apple, Facebook, Amazon, and Microsoft
 - If the product is free, then YOU are the product.

The 5V's of data

1. **Volume** - In the digital universe there is ≈ 45 zettabytes (10^{21})
2. **Velocity** - high rate of data collection.
 - a. 95 million photos and videos shared daily on Instagram
 - b. 305 billion emails per day
 - c. 5 million Tweets per day
 - d. 3.5 billion Google searches per day
3. **Variety** - data can be structured data, semi-structured, unstructured. Also data comes from a variety of sources and can be highly heterogeneous.
4. **Veracity** - data is inaccurate, inconsistent, and difficult to maintain. High dimensionality with data. Missing data????
5. **Value** - data is meaningless. Only if information can be extracted makes it valuable. For example, clickstream are sequences of user interactions with a website. A data scientist wants to convert these clickstreams into actionable information.

2 What is a database?

Definition 1. A *database* is any collection of related data organized for systematic management.

Examples of database

- Notebook containing a list of friends and family.
- Box of recipes
- Filing cabinet
- Spreadsheet
- Expense ledger

However, we think of databases as “electronic database”

Definition 2. A *database* refers to an electronic database managed by software. **Database Management System (DBMS)** is software used to create, store, retrieve, modify, and delete data in conjunction with various data-processing operations such as generating reports, printing invoices, etc.

There are many DBMS. Here are a few:

- Microsoft SQL Server (MS SQL Server)
 - American Airlines
- PostgreSQL
- Amazon RDS
- Oracle RDBMS
- IBM Db2 (and Informix)
- SAP Sybase
- Apache (open-source foundation) Cassandra (Big Data)
 - Apple, Constant Contact, Comcast, eBay, Hulu, Macy's Microsoft, McDonalds, Netflix, New York Times, Uber, Walmart
- Google Cloud BigTable (NoSQL)
- MongoDB
- MySQL
 - LinkedIn, Etsy, Twitter, Uber, Tesla, Toyota, Netflix, Verizon, Bank of America, PayPal, Pinterest, Yahoo, YouTube, Google, Facebook, Yelp, Dropbox, Github, Airbnb, ...
 - LAMP - Linux (OS), Apache - Webserver, M-MySQL, P=php or perl or Python.
 - WordPress, Online Shopping, etc.

Key Features of a DBMS

1. Allows users to create new and multiple databases
2. Provide ability to access, retrieve, and update data
3. Supports storage of large amounts of data
4. Control who access and different levels of access
5. ACID Transaction management

- a. A = Atomicity - transaction is “all or nothing”
- b. C = Consistency - only valid data will be written to the database.
Sum of balances must be equal (Credit = Debit)
- c. I = Isolation - multiple transactions cannot interfere
- d. D = Durability - a transaction will endure (will not be lost). Also
provides a mechanism for back-up and recovery.

Google to get additional perspectives on these concepts.

Database Architecture

Components of the DBMS:

- Storage Manager
- Concurrency controller (manager)
- Query language compilers
 - DDL = Data definition language (create databases, etc.)
 - DQL = Data query language (retrieve data)
 - DML = Data Manipulation language (update/modify the data)
- Recovery engine
- Buffer manager
- Authorization manager
- Transaction manager

Client-server Model

- Two-tier
 - Client → Database (Server)
- Three-tier
 - Client → Application Server → Database server
 - Client → Web Server → Database server

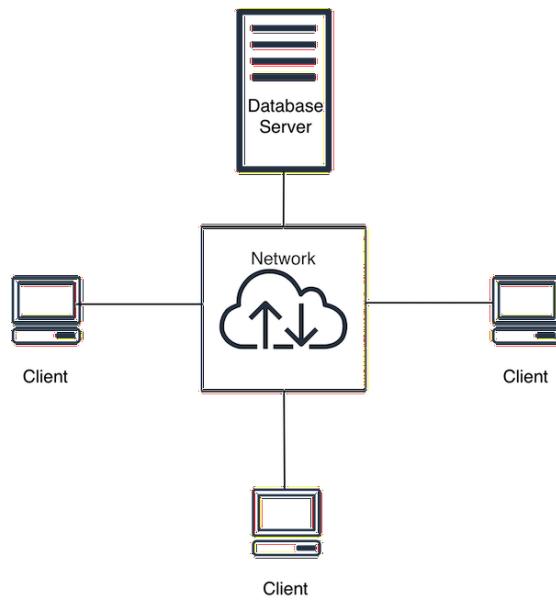


Figure 1: Two-tier client-server model.