INTRODUCTION TO DATABASES

Acknowledgements: Sharanya Jayaraman

DATABASE

- A database is a collection of data stored electronically, usually organized and structured in some form.
- Databases can be very simple – a text file or a CSV file, for example.
- When they are complex, we use formal methods of design and engineering principles, to build, update and maintain these systems.
- A DataBase Management System (DBMS) is the software used to interact between the database and an application that uses the database.
- Data in databases is usually stored as tables.

DIFFERENCES BETWEEN TABLES AND DATABASES

- When we think of databases, we often think of tables of information.
- However, we encounter several types of tables in software, and they focus on different aspects.
- HTML tables
  - Focus on display – tables have no basis of "identity" for data
- Spreadsheets
  - Focus on positional data – fields are arranged and calculated based on their position in the spreadsheet
- Databases
  - Focus on the data itself – metadata is used to identify data fields.
- Data stored in transience during program execution (classes, structs, etc) do not count, since they do not offer permanence.

DATA IN 3 LEVELS

- Level 1 – The actual database
  - A file that stores all the information. The format depends on the type of the Database (text, relational, graph,etc.)
- Level 2 – The DataBase Management System (DBMS) – interaction between data and an application
  - A software package that helps to update, maintain and secure the data.
- Level 3 – Application
  - Any software package that is built on top of the other 2 layers that provides interaction with the user. A web app, a mobile app, BI/ERP, etc.

TYPES OF DATABASES

- Flat Model
- Navigational databases
- ▫ Hierarchical (tree) database model
- ▫ Network/Graph model
- Relational Model
- Object model
- Document model
- Entity–attribute–value model
- Star schema

DISADVANTAGES OF FILE STORAGE

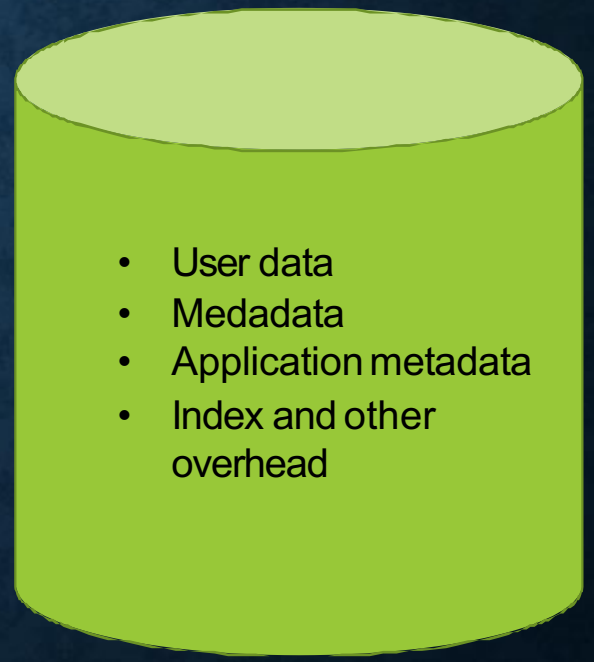
- Redundancy – We might have multiple copies of the same data.
- Inconsistency – Data on the same business entity may appear in different forms, for example, state name, phone number, etc. This makes it hard to modify data and track changes.
- Mixture of data – No clear logical relationships between the data columns, making it hard to understand and manage complex structures.
- Hard to maintain and manage
- No concurrency support – limiting scalability

HOW DO DATABASES HELP?

- Self-describing data collection of related records (meta data, data about data)
- Databases contains not just data, but also include the structure of data
- At the application level, databases can also store other application related meta data.
- This facilitates the personalization and customization of the application according to a user's profile.
- Two of the most important roles in defining metadata
  - Identify the type of data with a unique tag
  - Define the affinity of the data (tags enclose all data that is logically related)

CONTENTS OF A DATABASE

- User Data: The actual data used by the application
- Meta data: The structure (schema) of the data, including table names, column names and types, and constraints over the column(s)
- Application meta data: Meta data on user settings or functions of the application
- Index and other overhead data: used for performance improvement and maintenance. Logs, tracks, security, etc.



3 ACRONYMS

- ACID - Atomicity, Consistency, Isolation and Durability
  - 4 essential properties that a DBMS has to implement to effectively interact with a database
- SQL – Structured Query Language
  - A query is a result of a request to the database for information or for modification.
  - SQL is the language most relational databases (the most common for of databases) use for interaction with the data.
- CRUD – Create, Read, Update, Delete
  - Self-explanatory
  - The 4 major operations performed on the database.

ACID

- Atomicity – transactions are either all or none (commit/rollback)
- Consistency – only valid data is saved
- Isolation – transactions should not affect each other
- Durability – written data will not be lost

- Good example : Bank transactions

- Most of the challenges related to ACID compliance come from multiple users/concurrent use of the database

RELATIONAL DATABASES - RELATIONS

- A relational database describes the relationships among different kinds of data
  - Captures ideas like those defined in the Affinity and Collection rules
  - Allows software to answer queries about them
- Any relational DB can be described in XML
  - But it is not the case that every XML description defines a relational DB