

Relational Databases

Why databases?

Without databases, we have a few options for persistent data storage

Unstructured Data

- General Textual Content
 - Movie reviews
 - Social media posts
 - Etc.
- Multimedia
 - Video
 - Images
 - Audio

Semi-Structured Data

- "Flat" files
- JSON
- XML

”Flat” files

- Delimited Files
 - CSV, TSV, etc.

Last name, First name, SSN, Test1, Test2, Test3, Test4, Final, Grade
Alfa, Aloysius, 123-45-6789, 40.0, 90.0, 100.0, 83.0, 49.0, D-
Alfred, University, 123-12-1234, 41.0, 97.0, 96.0, 97.0, 48.0, D+
Gerty, Gramma, 567-89-0123, 41.0, 80.0, 60.0, 40.0, 44.0, C

XML

```
<CATALOG>
  <PLANT>
    <COMMON>Bloodroot</COMMON>
    <BOTANICAL>Sanguinaria canadensis</BOTANICAL>
    <ZONE>4</ZONE>
    <LIGHT>Mostly Shady</LIGHT>
    <PRICE>$2.44</PRICE>
    <AVAILABILITY>031599</AVAILABILITY>
  </PLANT>
  ...
</CATALOG>
```

JSON

```
{  
  "PLANT":  
  [  
    {  
      "COMMON": "Bloodroot",  
      "BOTANICAL": "Sanguinaria canadensis",  
      "ZONE": "4",  
      "LIGHT": "Mostly Shady",  
      "PRICE": "$2.44",  
      "AVAILABILITY": "031599"  
    },  
    { ... }  
  ]  
}
```

Wait...those look pretty structured to me!

```
{
  "eBooks": [
    {
      "language": "Pascal",
      "edition": "third"
    },
    {
      "language": "Python",
      "edition": "four"
    },
    {
      "language": "SQL",
      "edition": 2
    }
  ]
}
```

```
import json

with open('books.json', 'r') as json_input:
    data = json.load(json_input)
    for book in data['eBooks']:
        # Works for every book except the one
        # I changed the edition to be a number (which is
        # valid in the JSON format
        print (
            f"{book['language']}: {book['edition'] + ' edition'}")
    )
```


Issues

- Ensuring data integrity
 - Validity, consistency, accuracy of data
- Identifying relationships between data
- Data redundancy
- Data access

Databases

- “A database is an organized collection of structured information, or data, typically stored electronically in a computer system.” (source: [Oracle](#))
- “A database is an organized collection of data used for the purpose of modeling some type of organization or organizational process [regardless of the medium].” (source: Database Design for Mere Mortals)
- “A database is [...] a set of related information.” (source: Learning SQL 3rd Edition)

Database Management System (DBMS)

- The actual software that
 - manages data storage on the physical medium (HDD, SSD, etc.)
 - provides interfaces for querying and modification of the database(s)
 - enforces constraints
 - controls access
- What we tend to think of when we talk about databases
 - MySQL
 - Oracle
 - MariaDB
 - Sqlite3
 - PostgreSQL

Relational Databases

- Data is stored in relations
 - We view these as **tables** (also called entities)
- Each relation is made up of tuples
 - These are the rows of our table which we also call **records**
- Each tuple is made up of data that represents the attributes of a record
 - These are the columns of our table which we also call **fields**
- The physical order of records in the table is unimportant, but each record is identified by a field that contains a unique value

Advantages of Relational Databases

- Data integrity at multiple levels
 - Fields, tables, relationships
- Data consistency and accuracy
- Data can be shared between tables to create a relationship which consolidates data to reduce redundancy
- Easy data retrieval from one or multiple (related) tables

SQL

- Is a language used to interact with most relational database management systems (RDBMS)
 - Can not only query information from tables in the databases, but metadata about a database and the RDBMS system itself
- Not all RDBMS have the same level of support for the SQL language standards
 - Can result in differences in features and query semantics (behavior)
- Declarative language
 - SQL tells the database system **what** to do, but not how do it

Schemas

- Describes the organization of a database including
 - table names
 - fields
 - data types
 - tables relationships
 - constraints
- For RDBMS, SQL statements are used to help the database define the database structure