

Databases

File System

- The file system is like a file format for your Operating System
- The OS needs to know how it should store, access, and organize data
 - mainly concerned with the hard disk

File System

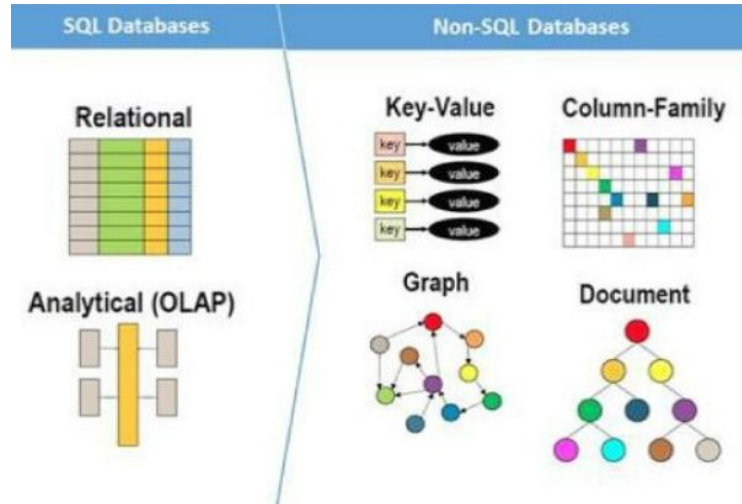
- 3 Main Layers:
 - *Logical File System*
 - What you, the user, and programs interact with
 - All files (and metadata) kept in a table, managed by OS, with their locations on the disk
 - Directories are a just files themselves with a list of filenames they contain (more lies!!!!)
 - Basic file operations: OPEN, CLOSE, READ, WRITE
 - *Virtual File System*
 - Helps with memory management and efficiency
 - Manages multiple file systems on the same disk (partitions)
 - *Physical File System*
 - Manages physical operations of storage devices
 - Divides physical storage device into different “sections” and “blocks”

Regular Files

- Uses a file format and an encoding
- Ends up as a very long string of “1s” and “0s” in a particular location on the hard drive (or in any kind of memory)
- Can browse the computer file system (or directory) to find the file (or data)
- This is for a single file (i.e., a small amount of data), what if we have a LOT of data to keep track of (especially a lot of SIMILAR data)?

Databases

- A collection of STRUCTURED data
- Two main types Relational (SQL) or Non-Relational (NoSQL)



Databases: Relational vs Non-Relational

SQL	NOSQL
Relational Database management system	Distributed Database management system
Vertically Scalable	Horizontally Scalable
Fixed or predefined Schema	Dynamic Schema
Not suitable for hierarchical data storage	Best suitable for hierarchical data storage
Can be used for complex queries	Not good for complex queries

Databases: Relational Example

Persons

Id	Name	SurName	Age
1	Jodie	Tucker	34
2	Jayden	Archer	56
3	Grace	Wheeler	18
4	Freddie	Humphries	56

Columns (points to the header row)

Rows (points to the data rows)

The diagram shows a table with four columns and four rows. The columns are labeled 'Id', 'Name', 'SurName', and 'Age'. The rows contain the following data: (1, Jodie, Tucker, 34), (2, Jayden, Archer, 56), (3, Grace, Wheeler, 18), and (4, Freddie, Humphries, 56). An arrow labeled 'Columns' points to the header row, and an arrow labeled 'Rows' points to the data rows.

Databases: Non-Relational Example

```
<bib name="IT">
  <book id="b001" year="1994">
    <title>TCP/IP Illustrated</title>
    <author><last>Stevens</last><first>W.</first></author>
    <publisher>Addison-Wesley</publisher>
  </book>
  <book id="b002" year="2000">
    <title>Data on the Web</title>
    <author><last>Abiteboul</last><first>Serge</first></author>
    <author><last>Buneman</last><first>Peter</first></author>
    <author><last>Suciu</last><first>Dan</first></author>
    <publisher>Morgan Kaufmann</publisher>
  </book>
  <journal id="j001" year="1998">
    <title>XML</title>
    <editor><last>Date</last><first>C.</first></editor>
    <editor><last>Gerbag</last><first>M.</first></editor>
  </journal>
</bib>
```


Databases: Relational

id (primary key)	first_name	last_name	age
1	Bob	Roberts	22
2	Tom	Jones	66
3	Sue	Williams	31

Databases: Relational

id (primary key)	first_name	last_name	age	course_id
1	Bob	Roberts	22	2
2	Tom	Jones	66	3
3	Sue	Williams	31	1

foreign key



id (primary key)	name	course_code
1	Software Engineering	3773
2	Secure Software	4683
3	Systems Programming	3423

Databases: Relational - In-Class Assignment

STORE		
Store_key	City	Region
1	New York	East
2	Chicago	Central
3	Atlanta	East
4	Los Angeles	West
5	San Francisco	West
6	Philadelphia	East

PRODUCT		
Product_key	Description	Brand
1	Beautiful Girls	MKF Studios
2	Toy Story	Wolf
3	Sense and Sensibility	Parabuster Inc.
4	Holiday of the Year	Wolf
5	Pulp Fiction	MKF Studios
6	The Juror	MKF Studios
7	From Dusk Till Dawn	Parabuster Inc.
8	Hellraiser: Bloodline	Big Studios

SALES_FACT				
Store_key	Product_key	Sales	Cost	Profit
1	6	2.39	1.15	1.24
1	2	16.7	6.91	9.79
2	7	7.16	2.75	4.40
3	2	4.77	1.84	2.93
5	3	11.93	4.59	7.34
5	1	14.31	5.51	8.80

Answer the following questions:

1. What was the total profit across all stores?
2. What is the name of the most popular product sold?
3. How much profit did the “East” Region make?
4. Which city had the highest cost?
5. What was the least popular product brand?
6. Which city (or cities) sold “Sense and Sensibility”?