# Introduction

# Why taking this course??

- "What is the use of all the courses I have taken so far?"
  - This course uses a lot of the basics introduced in the 200/300 level courses
- "I want to work in an interdisciplinary environment"
  - Be an application developer working with people from other areas (e-commerce, science, administration, law, etc. etc.)
- "I love the internals of how computers and systems work"
  - Be a database administrator or a DBS developer: a DBMS is an entire operating system and more
- "I am more a theoretical person"
  - Database systems have a very sound theoretical foundation and there are many exciting open problems
- "I want to work with computer languages, human-computer interaction, multimedia, logic, communication, distributed systems, knowledge management -- It's all there
- "I want to make a lot of money"

# Example Applications

- What data would you store in a database system (DBS, DBMS, RDBMS)?

Structured data/info

# University Data and its use

- What information is stored?
  - Provide at least 5 classes
- How is the data used?
  - Provide at least 3 queries (information that you want to retrieve from the system)
  - Provide at least 3 modifying actions (actions that insert data or modify data)
- Why cumbersome to do this with files?

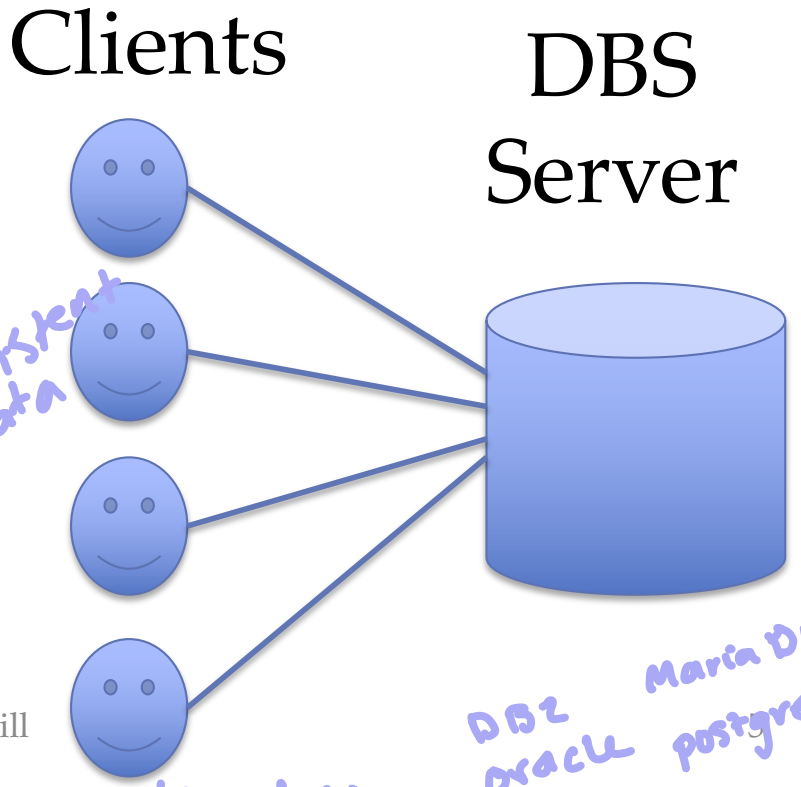
insert, update, delete, query

DBS DBMS rDBMS ↑relational

# Database Management System

Store data in a specific format + structure it

- Well-structured data model
- Powerful ***data-centric*** interface
  - Insert data
  - Change data
  - Advanced queries
- Efficiency
- Concurrency - concurrent users, consistent data
- Persistence - if machine fails data is still there




DB2 oracle MariaDB postgreSQL



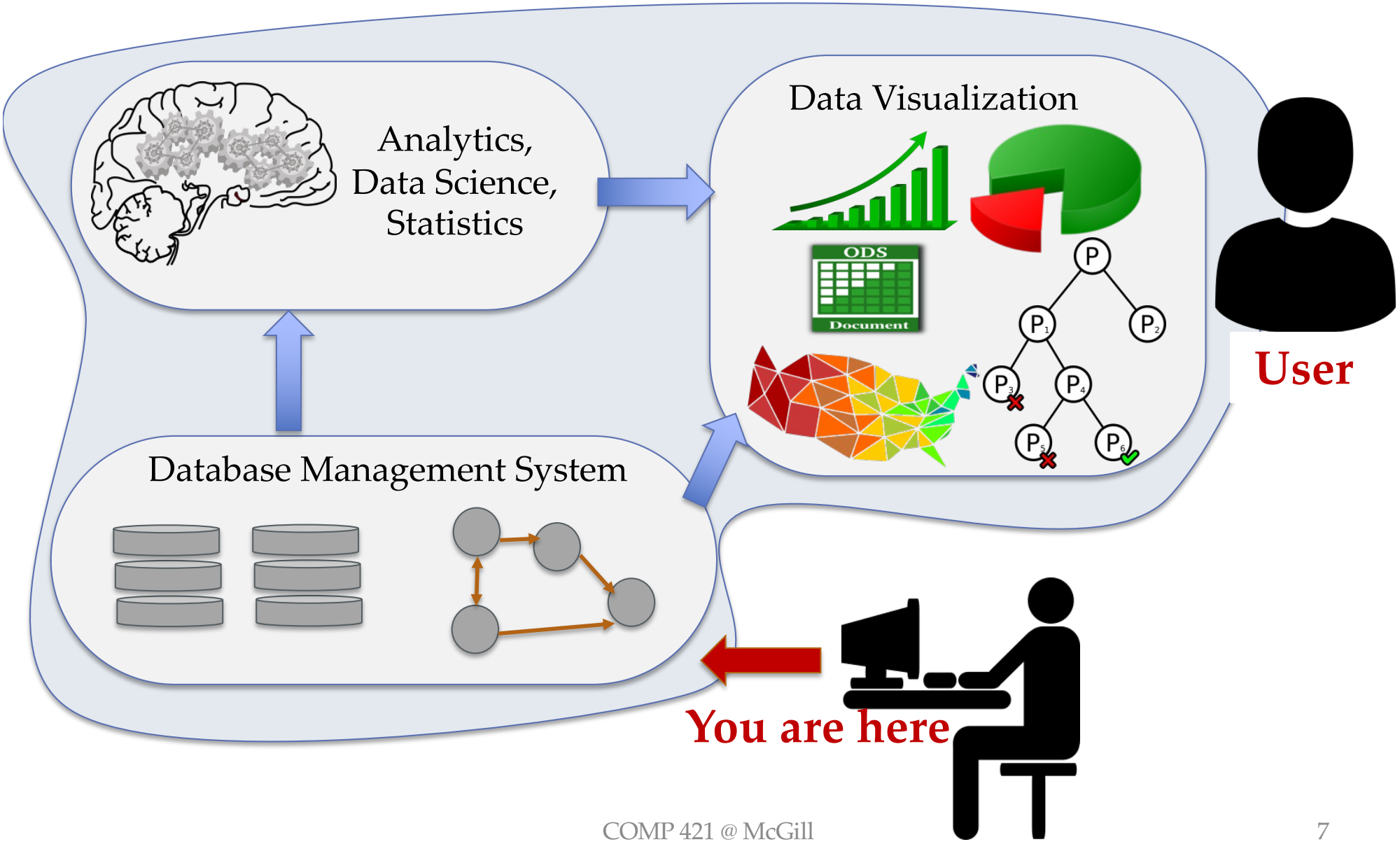
database: collection of data of the application

# Database Systems and Data Analytics

- Transform basic data into valuable knowledge
- In 2002 a statistician in Target came up with a list of 25 products in the shopping cart that could give an indication of someone being pregnant.
- Big Data Analytics is emerging as an important trend in analyzing and containing health issues like spread of epidemics. (Eg. COVID, Ebola in West Africa)

build data analytics on top of database systems

# Data Analytics



# Relational Data Model

Cartoon Characters

| Sid (int) | FirstName (string) | LastName (string) | … |
|---|---|---|---|
| 123 | Bugs | Bunny | |
| 124 | CardCapturer | Sakura | |
| 125 | Dora | The Explorer | |

Relation = Table

tuple = row

attribute = column header

Simple, powerful, mathematically sound

# Data Models covered

- Semantic data model
  - (Entity Relationship) ER
- Relational
- Key-value model
- Graph

} move cloud-y

– tables

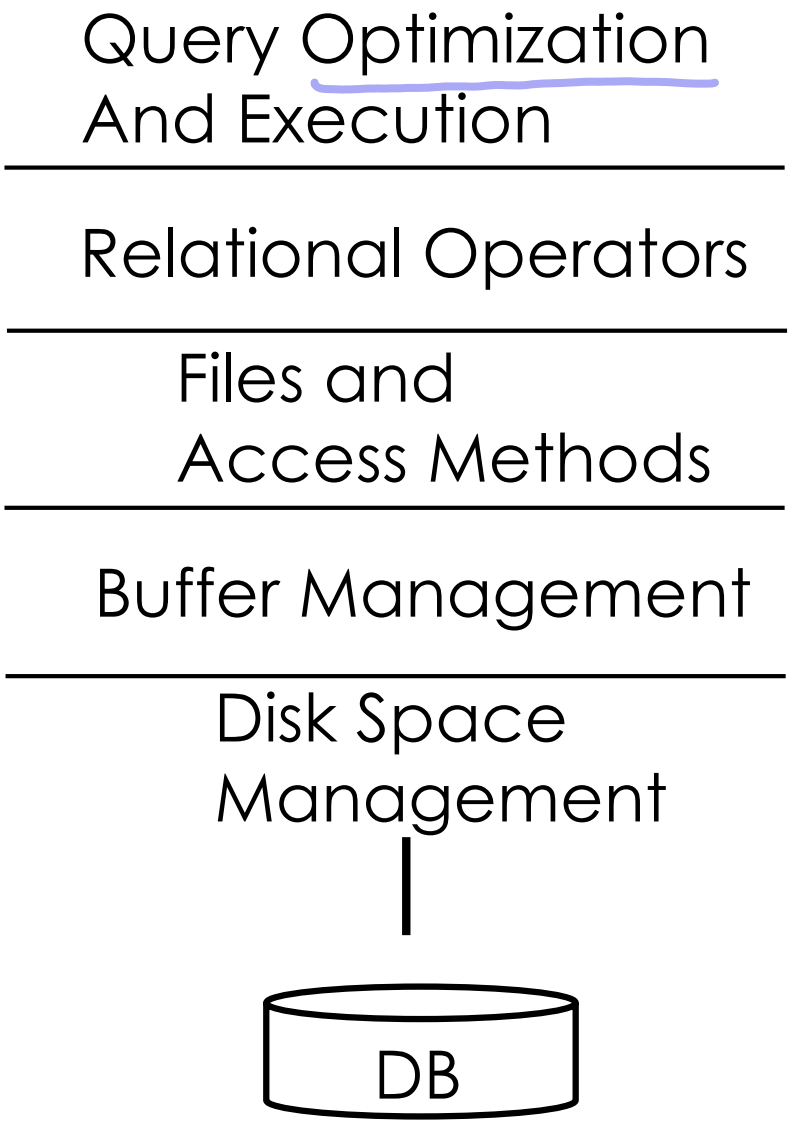
types { relational, Semi-structured, object-oriented, graph

# Working with Data

- Defining a Schema
- Inserting Data/Updating Data
- Querying Data
- Application Programming

- Focus: Relational data and **SQL**
- Other data models:
  - Differences and similarities with SQL

# Internals of a DBS




Selections, projections, etc

how to store on disk/ stable storage

# Transaction Management

- Managing updates

- Handle concurrent access
- Handle failures

- consistency of updates

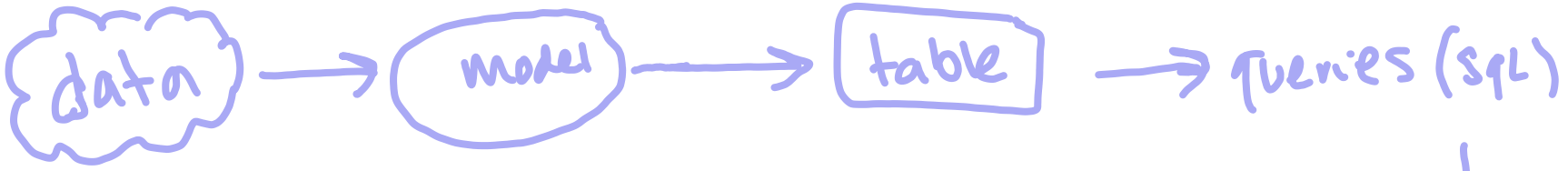
# Large-Scale Management

- Distribution of data
- Distributed data analysis
  - Map-reduce

graph-based DBMS

↳ differences and similarities to relational model + SQL

↳ particularities of the query language



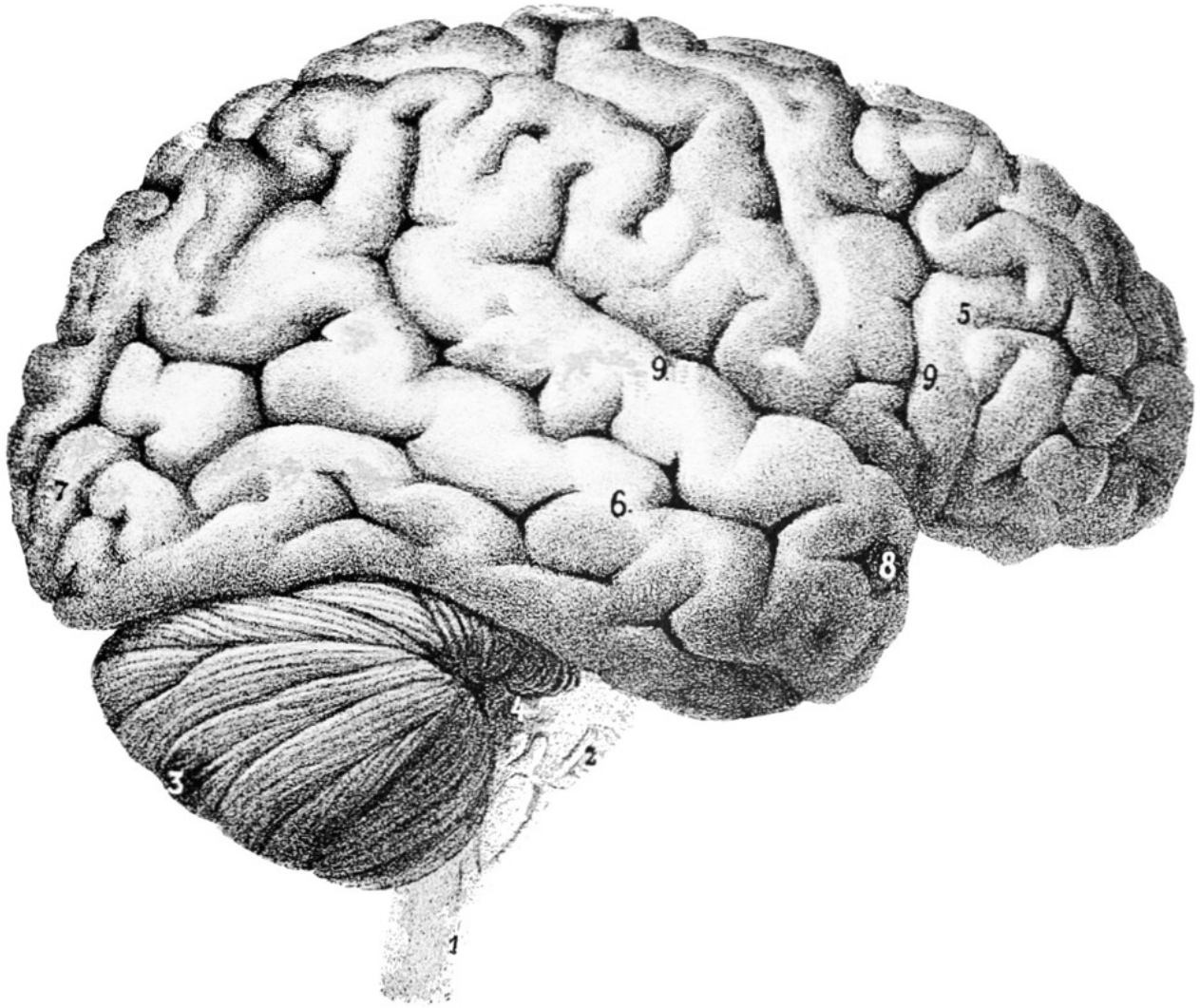

# Emphasis of the Course

- **Design** of databases
- **Use** of a database
- **Internals**

- Database technology is continuously evolving
  - Understand what are the fundamental concepts of any database system

# The Importance of Being Data

Natural, Built in



# The Importance of Being Data

Ancient History


# The Importance of Being Data

# The Importance of Being Data

- Capture the (right) data.
- Extract (meaningful) information out of it.
- Turn information into (profitable) actions.

# The Importance of Being Data

Walmart and 9/11 US Flag Sales.

- Sept 11, 2000 – 6,400
- Sept 11, 2001 – 116,000

The ability to detect changes immediately and then take quick decisions are key.

# The Importance of Being Data

- In 2002 a statistician in Target came up with a list of 25 products in the shopping cart that could give an indication of someone being pregnant.

# The Importance of Being Data

- Big Data Analytics is emerging as an important trend in analyzing and containing health issues like spread of epidemics. (Eg. Ebola in West Africa)

# Dominant/Promising/Emerging

- In – Database Analytics
- Columnar Databases
- In-memory Databases
- Graph Databases

# Some Interesting Reads/Activities

(These are not required for class)

1. How Companies Learn your secrets.
http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1&hp=&pagewanted=all

2. Walmart and the American flag.
http://www.pipelinepub.com/0207/pdf/Pipelinev3i8Ariticle9.pdf

3. Ebola and Big Data
http://www.bbc.com/news/business-29617831

4. Attracting students from near and far to McGill
http://mcgillnews.mcgill.ca/s/1762/news/interior.aspx?sid=1762&gid=2&pgid=1843

5. Look around for how pandemic has affected the economic landscape. Can you find stories about businesses that turned a profit out of the current situation? (That are not so obvious)

# Applications

Who needs a data management system?

- ?