CS 341
Database Systems

Welcome to the Course!

About the Instructor

Abeera Tariq

Lecturer – Computer Science

Email: abeeratariq@iba.edu.pk

Office location: F22/B, 1st Floor, Multi-purpose Building, Main Campus

Office hours:

M/W 11:30 am – 12:45 pm (Adamjee Faculty Lounge)

Tues 1:00 pm – 2:00 pm (Office)

Any changes in office hours will be communicated via LMS.

Introduce Yourself

```
SELECT
        Name,
        Batch,
        FunFactAboutYou,
FROM Class
```

Class Introductions

Course Outline

Introduction to Database Systems

One Minute on the Internet

2012 to 2022 – Data Scientist stays popular

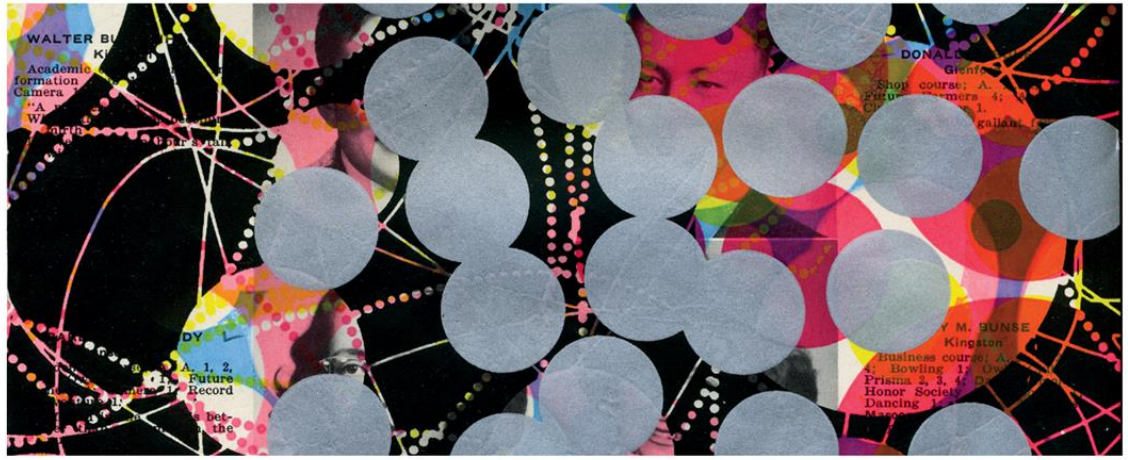
Harvard Business Review

Analytics And Data Science | Data Scientist: The Sexiest Job of the 21st Century

Data Scientist: The [redacted] Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)



https://hbr.org/2012/10/(

Is Data Scientist Still the [redacted] Job of the 21st Century?

by Thomas H. Davenport and DJ Patil

July 15, 2022

https://hbr.org/2022/07

Why 2023 Will Be the Year of Data Scientists

The hottest talent pool in 2023: Data scientists

Kareem Bakr Managing Director, Phaidon International

January 10, 2023

As we head into 2023, one of the most coveted talent pools will be data scientists as their skillset will become more valuable and vital for organizations across the board, says Kareem Bakr, managing director of Phaidon International.

Throughout 2022, hiring in the technology sector outperformed many other sectors that were impacted by the post-pandemic shift, the Great Resignation, and varied economic factors. In the coming year, expect tech hiring to continue to flourish, specifically within data science.

Tech companies will continue looking for secure candidates to fill their data scientist roles in 2023. The niche skills data scientists possess and the current skillset-driven market creates competition to find qualified candidates within the tech space. This means employees currently have the upper hand.

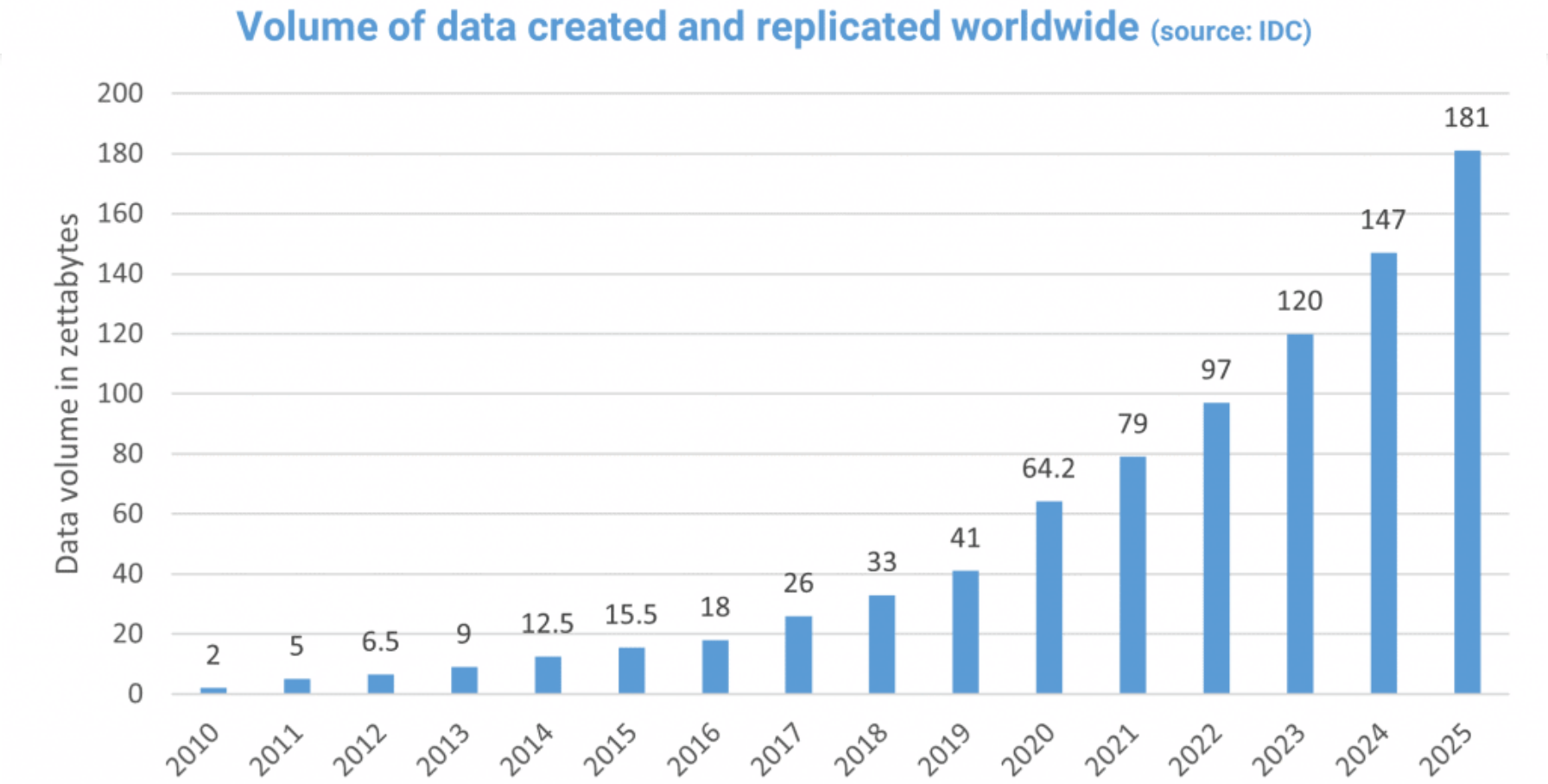
https://www.spiceworks.com/tech/big-data/guest-article/why-2023-will-be-the-year-of-data-scientists/

The Data Scientist Job Outlook Key Findings

- The employment rate for data scientists will grow by 36% from 2021 to 2031.
- 49% of the job ads on LinkedIn are in the IT & Tech industry.
- 6% of the data scientist job offers on LinkedIn are in California.
- Only 33% of job ads specifically require a data science degree.
- The most in-demand technical skills for data science experts are Python and SQL.
- Entry-level positions account for 55% of job offers.
- The average data scientist's salary in the US is $125,242/year.

https://365datascience.com/career-advice/data-scientist-job-outlook/

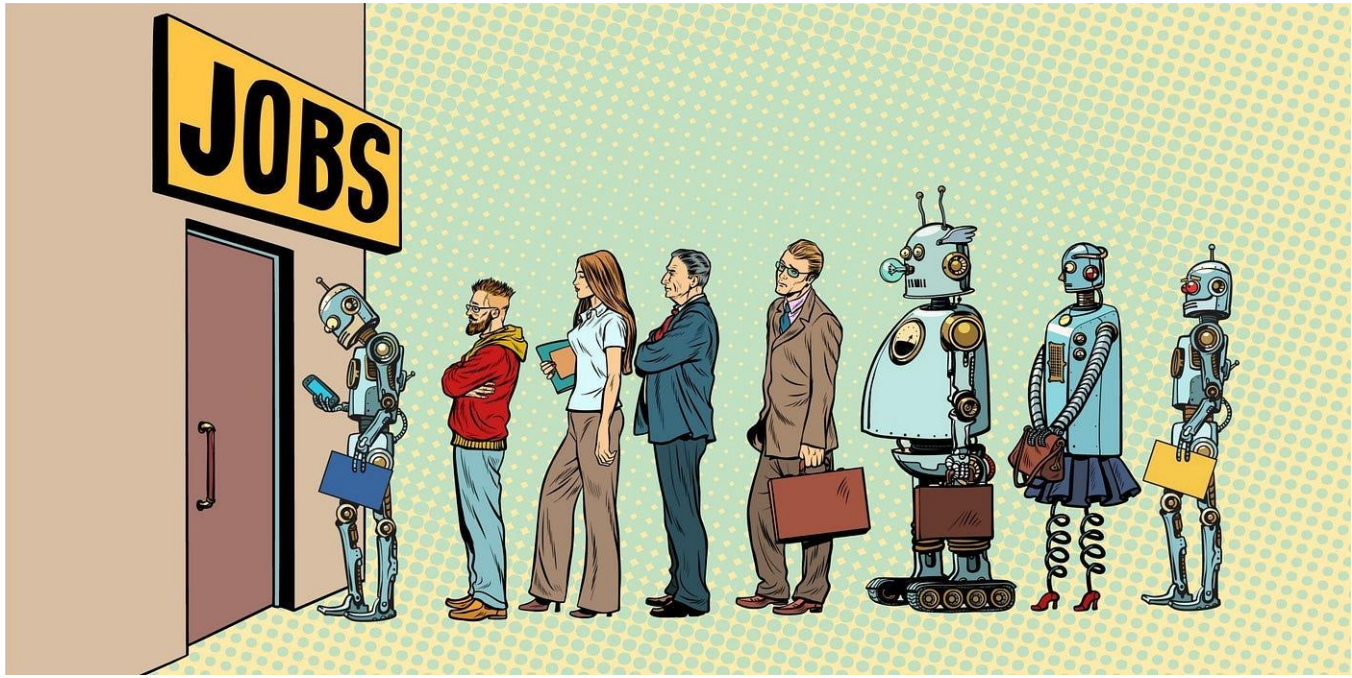
Data Trends


Volume of data created and replicated worldwide (source: IDC)
Data volume in zettabytes
200
180
160
140
120
100
80
60
40
20
0
2
5
6.5
9
12.5
15.5
18
26
33
41
64.2
79
97
120
147
181
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025

Data is the Present and the Future

- DBMS developers are in **demand** and still there are many **challenging unsolved problems in data management and processing.**
- Is the nature of the problem the same?

What about the threat of AI?



Will ChatGPT Put Data Analysts Out Of Work? (forbes.com)

Is ChatGPT a threat to jobs in data and analytics?

Forbes

Is ChatGPT a threat to jobs in data and analytics?

As we've seen, ChatGPT can easily automate some of the tasks that are traditionally carried out in analytical jobs – such as business, data, and financial analyst roles. Future iterations of the technology are likely to become even more effective and efficient at doing so.

But that doesn't mean that anyone who works in an analytical role is going to be out of a job right away. This is primarily because today's most sophisticated LLMs and NLP tools still lack abilities like critical thinking, strategic planning, and complex problem-solving. Most experts agree that it isn't likely that machine learning-based tools will be able to carry out these functions at the same level as humans any time soon.

It's likely that businesses and other organizations will still have a need for humans who are experts in this field for some time to come.

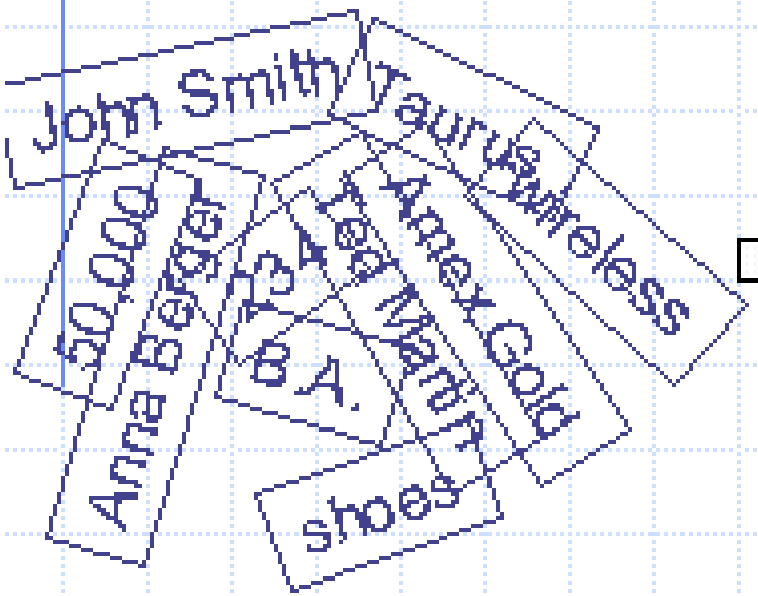
"The goal is to transform data into information, and information into insight"

Carly Fiorina (Executive and president of Hewlett-Packard Co. in 1999. Chairwoman in 2000)

Data → Info → Insights

EMPLOYEE_ID	LAST_NAME	SALARY	DEPARTMENT_ID	JOB_ID	HIRE_DATE
210	Hintz	25000		SH_CLERK	28-OCT-21
100	King	24000	90	AD_PRES	17-JUN-03
101	Kochhar	17000	90	AD_VP	21-SEP-05
102	De Haan	17000	90	AD_VP	13-JAN-01
103	Hunold	9000	60	IT_PROG	03-JAN-06
104	Ernst	6000	60	IT_PROG	21-MAY-07
105	Austin	4800	60	IT_PROG	25-JUN-05
106	Pataballa	4800	60	IT_PROG	05-FEB-06
107	Lorentz	4200	60	IT_PROG	07-FEB-07
108	Greenberg	12008	100	FI_MGR	17-AUG-02
109	Faviet	9000	100	FI_ACCOUNT	16-AUG-02
110	Chen	8200	100	FI_ACCOUNT	28-SEP-05
111	Sciarra	7700	100	FI_ACCOUNT	30-SEP-05
112	Urman	7800	100	FI_ACCOUNT	07-MAR-06
113	Popp	6900	100	FI_ACCOUNT	07-DEC-07
114	Raphaely	11000	30	PU_MAN	07-DEC-02
115	Khoo	3100	30	PU_CLERK	18-MAY-03
116	Baida	2900	30	PU_CLERK	24-DEC-05
117	Tobias	2800	30	PU_CLERK	24-JUL-05
118	Himuro	2600	30	PU_CLERK	15-NOV-06

Data, Information and Knowledge



Data

A. Berger M. Jones T. Martin J. Smith	Shoes Scarves Jewelry Groceries
50,000 46,800 29,200 75,500	MoU Qty Income Education

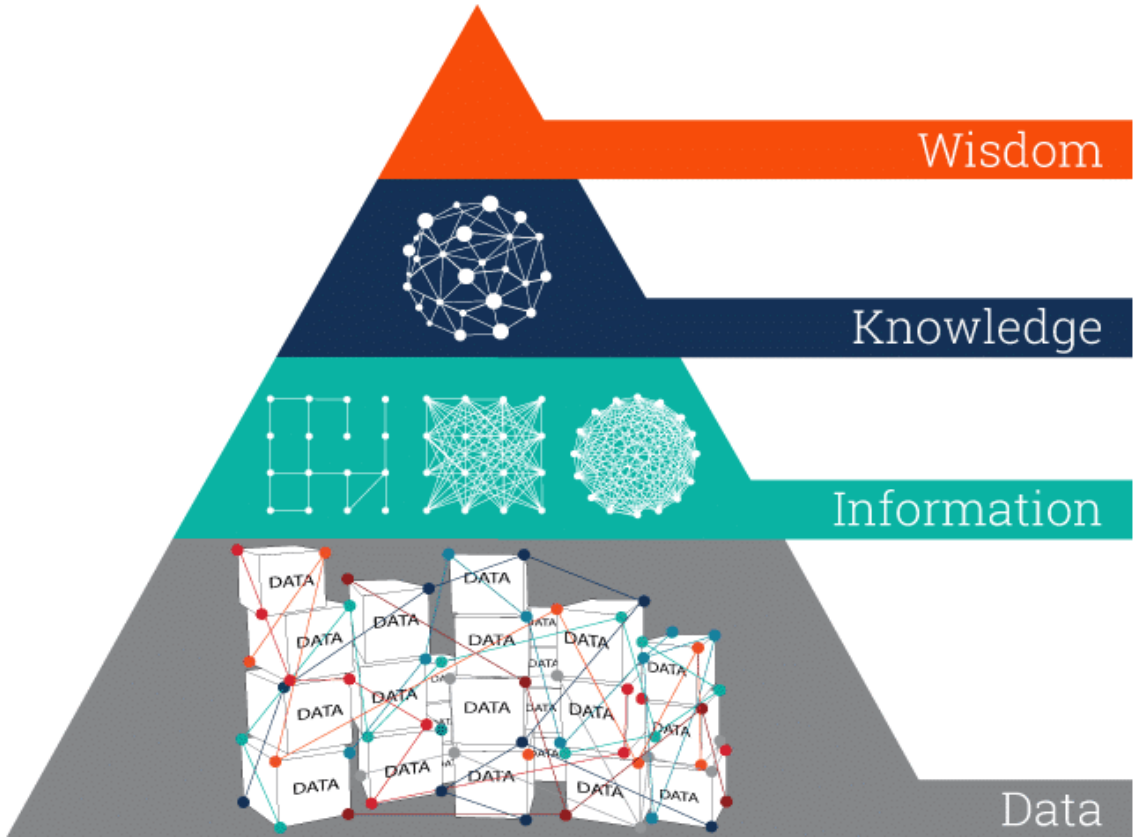
Information

A. Berger is most likely to buy new product

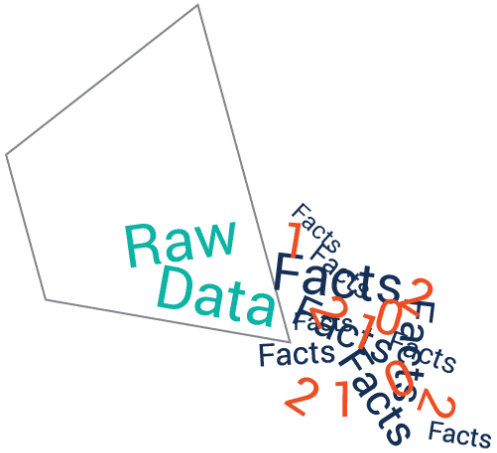
T. Martin is profitable customer and is likely to switch carriers

Knowledge & Customer Insight

D-I-K-W Pyramid

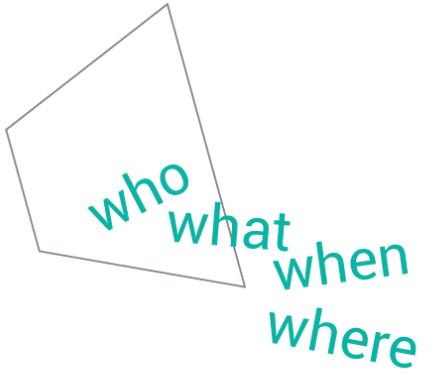


Each step up the pyramid answers questions about and adds value to the initial data.



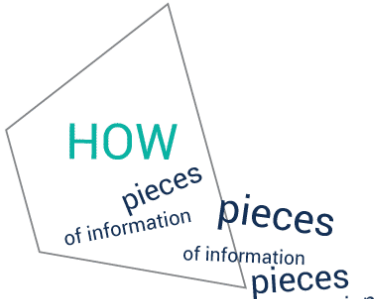
= a collection of facts in a raw or unorganized form

Base building block - Raw Data



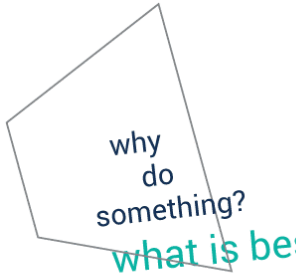
= easier to measure, visualize and analyze data for a specific purpose

Second building block - Derived Information



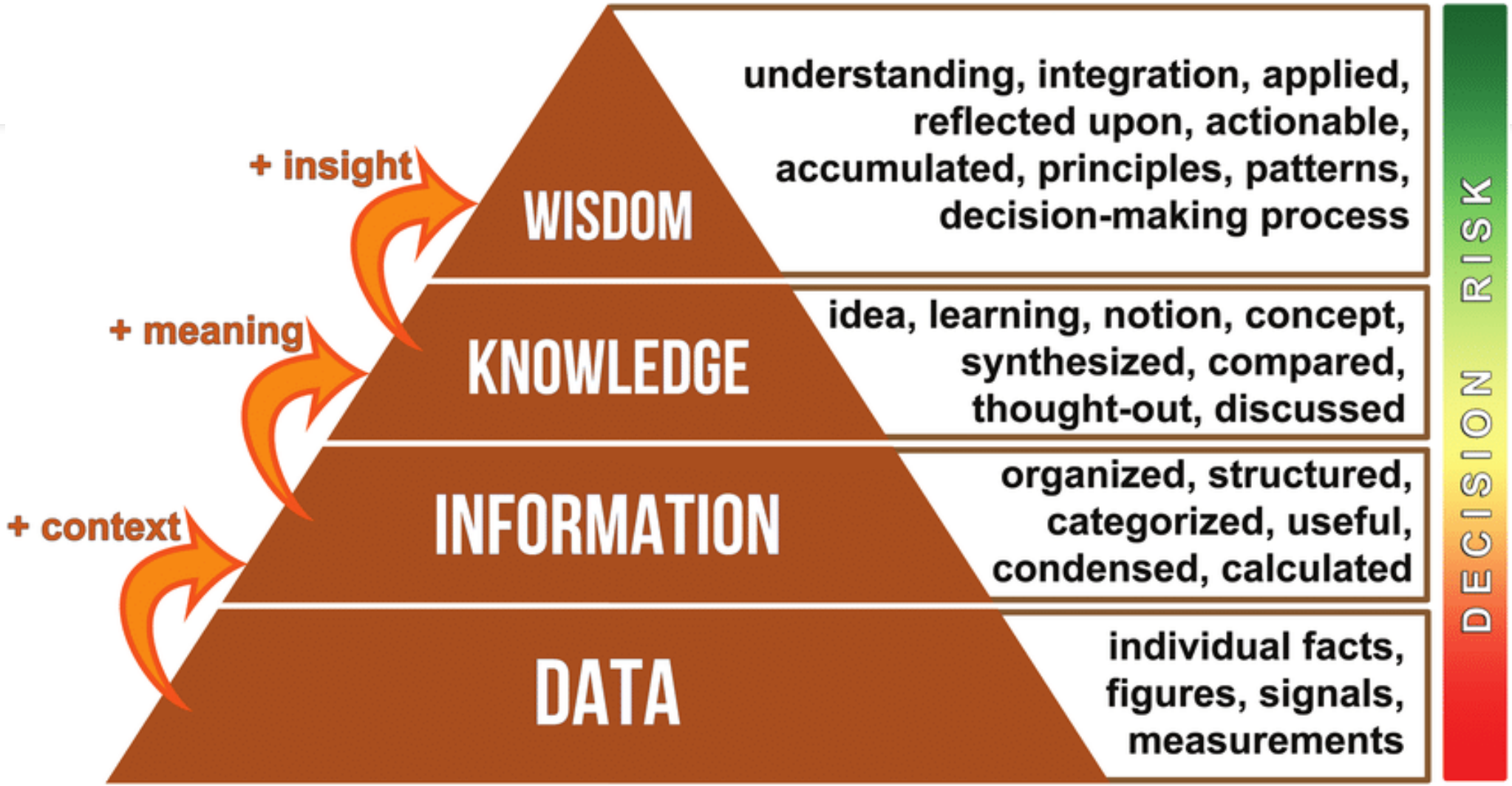
connected to other pieces help us understand how to apply information to achieve our goal

Third building block - Relevant Knowledge



Wisdom is knowledge applied in action

The top of the DIKW hierarchy - Guiding Wisdom


+ insight
+ meaning
+ context
WISDOM
KNOWLEDGE
INFORMATION
DATA
understanding, integration, applied, reflected upon, actionable, accumulated, principles, patterns, decision-making process
idea, learning, notion, concept, synthesized, compared, thought-out, discussed
organized, structured, categorized, useful, condensed, calculated
individual facts, figures, signals, measurements
DECISION RISK

Databases are at the heart of many systems we interact with daily.

Think of some Database applications

10 minutes

A Day in Susan's Life

See how many databases she interacts with each day

In the morning, Susan goes grocery shopping

Later, she picks up her prescription at the pharmacy

In the afternoon, she orders some items online

At night, she plans for a trip and buys airline tickets and hotel reservations online

Before going to bed, Susan checks her social media accounts

A Day in Susan's Life

See how many databases she interacts with each day

In the morning, Susan goes grocery shopping

Where are the product data stored?

Is the product quantity in stock updated at checkout?

Does she pay with a credit card?

Later, she picks up her prescription at the pharmacy

Where is the pharmacy inventory data stored?

What data about each product will be in the inventory data?

What data is kept about each customer and where is it stored?

In the afternoon, she orders some items online

Where are the product and stock data stored?

Where does the system get the data to generate product "recommendations" to the customer?

At night, she plans for a trip and buys airline tickets and hotel reservations online

Where does the online travel website get the airline and hotel data from?

What customer data would be kept by the website?

Where would the customer data be stored?

Before going to bed, Susan checks her social media accounts

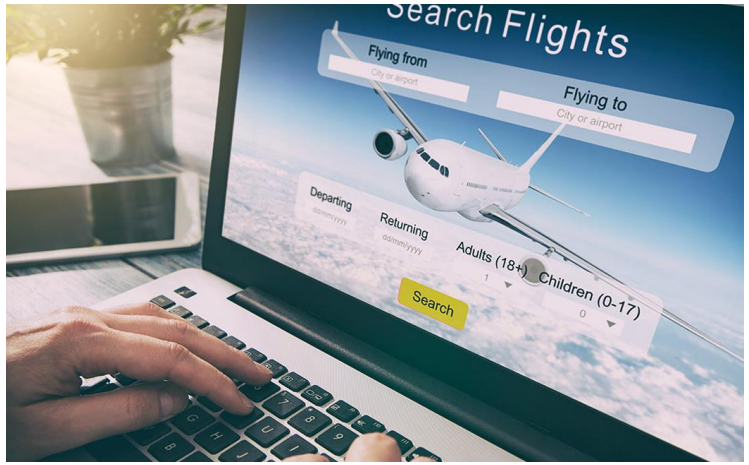
What particular customer data is kept by each website?

Where is the data about the friends and groups stored?

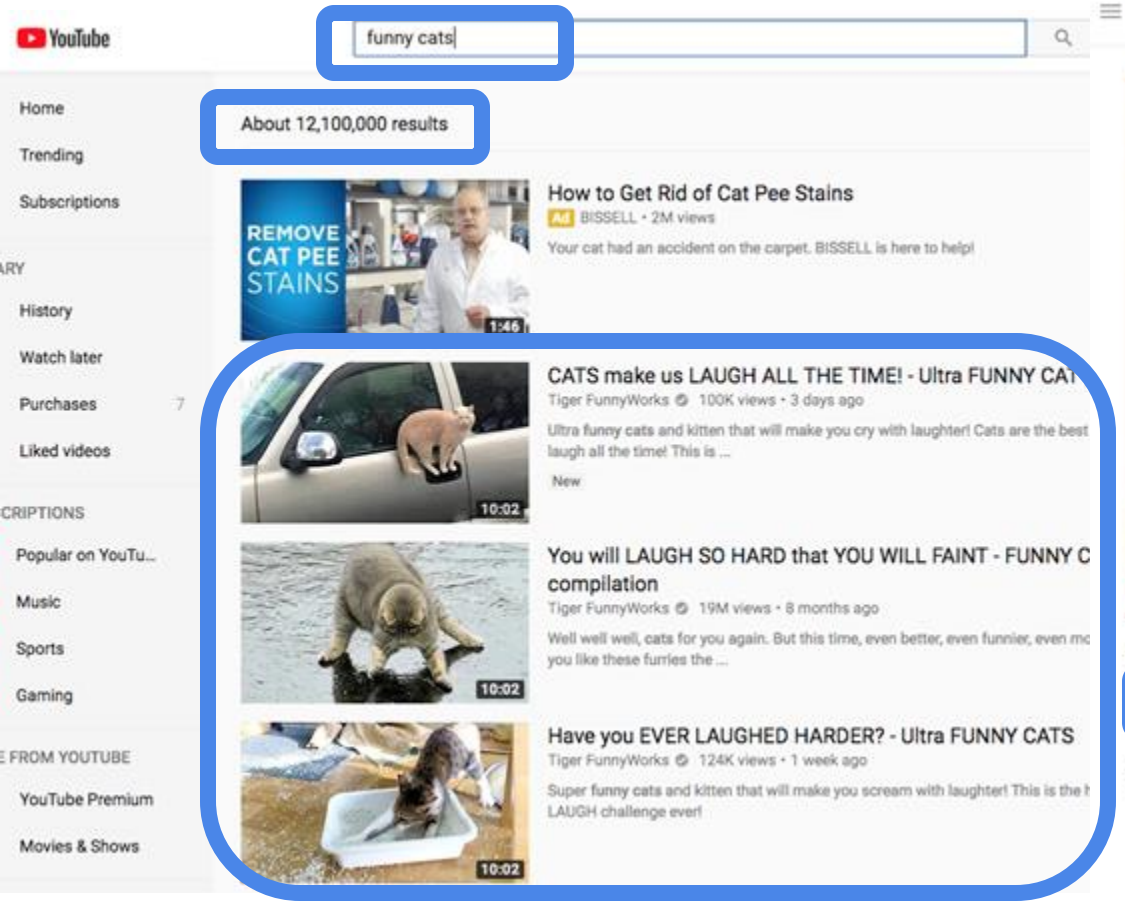
Where are the "likes" stored and what would they be used for?

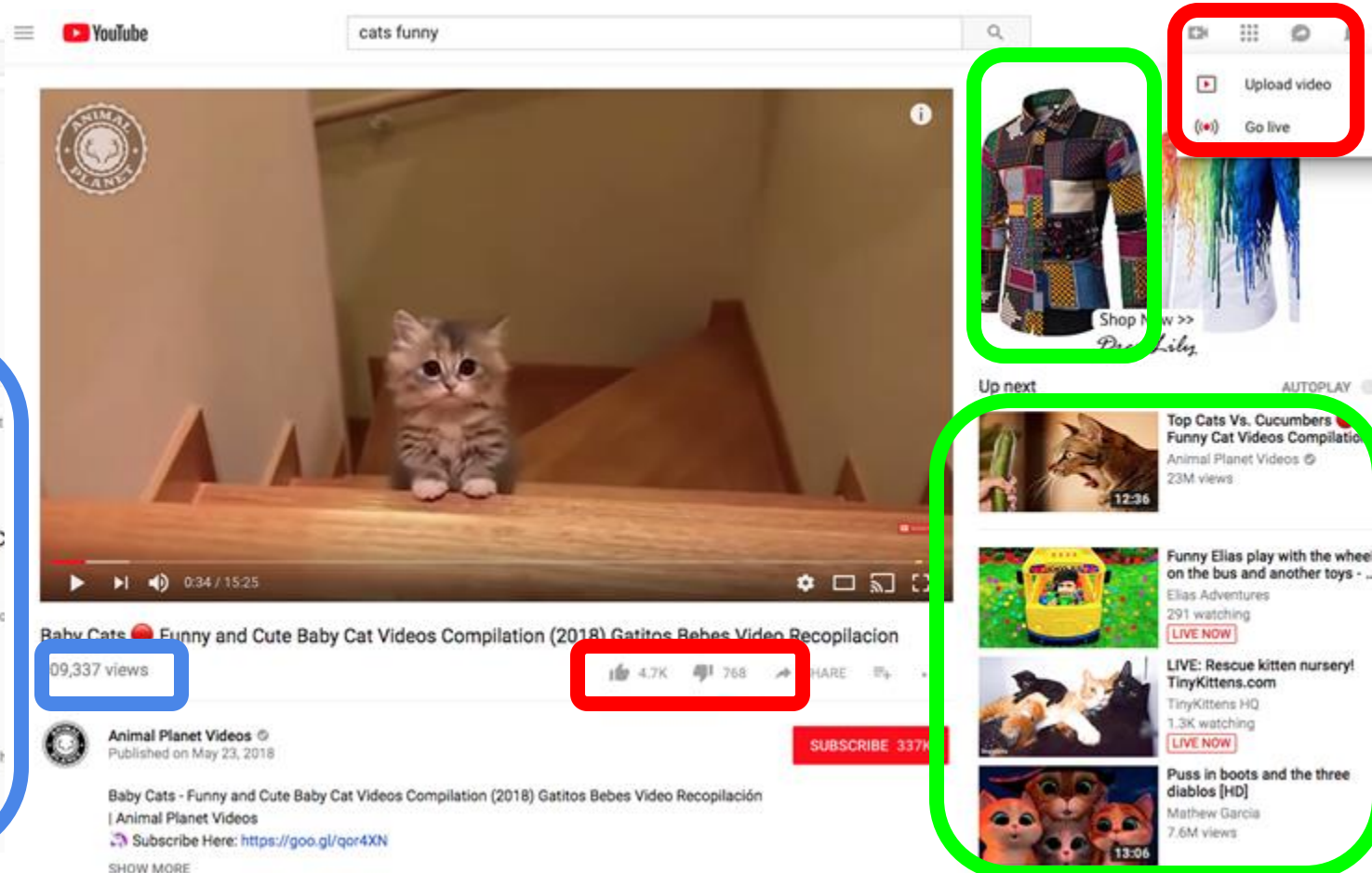
Examples of Database Applications

- Purchases from the supermarket
- Purchases using your credit card
- Booking a holiday at the travel agents
- Using the local library
- Taking out insurance
- Using the Internet
- Studying at university



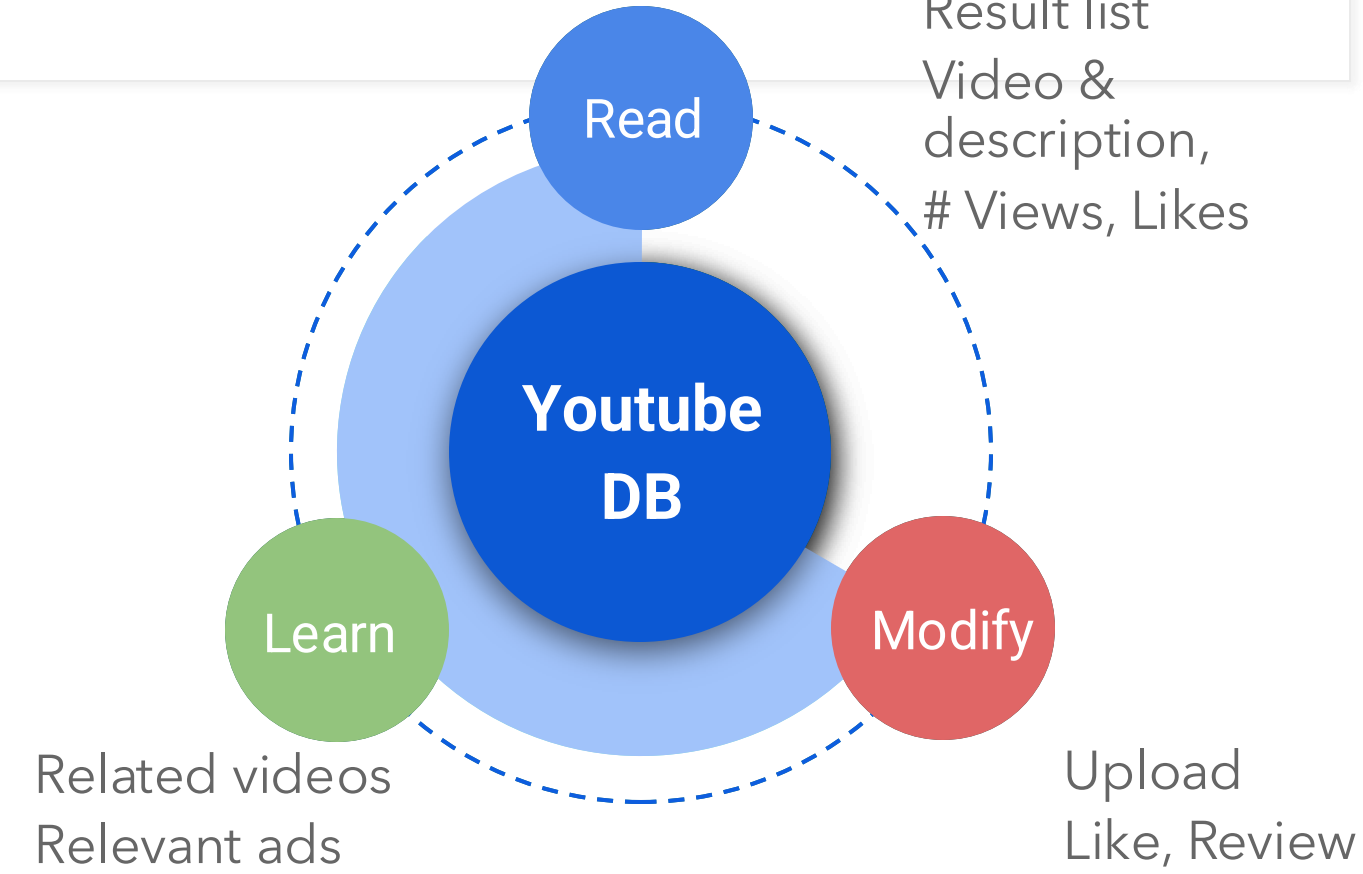
Example: Youtube DB





YouTube Database

How do you interact with the YouTube database?

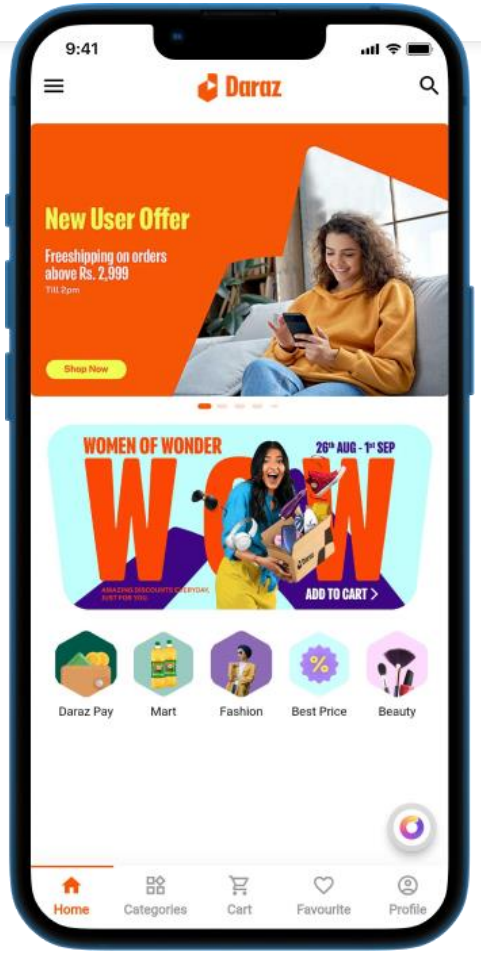


Netflix


WITCHER
THE CROWN
STRANGER THINGS
Available Now
Strange sightings. Government secrets. Fearless kids.
And a dark force that turns a small town upside down.
Play
More info
Trending Now
NARCOS

eCommerce


BUY
BUY


9:41
Daraz
New User Offer
Freeshipping on orders
above Rs. 2,999
Shop Now
WOMEN OF WONDER
26th AUG - 1st SEP
WOW
ADD TO CART
Daraz Pay
Mart
Fashion
Best Price
Beauty
Home
Categories
Cart
Favourite
Profile

Database

Key Definitions

- A **database** can be defined as a collection of *related data items* within a specific business process or problem setting
  - has a target group of users and applications
- A **Database Management System (DBMS)**, is the software package used to define, create, use and maintain a database
  - consists of several software modules
- The combination of a **DBMS** and a **database** is then often called a **database system**

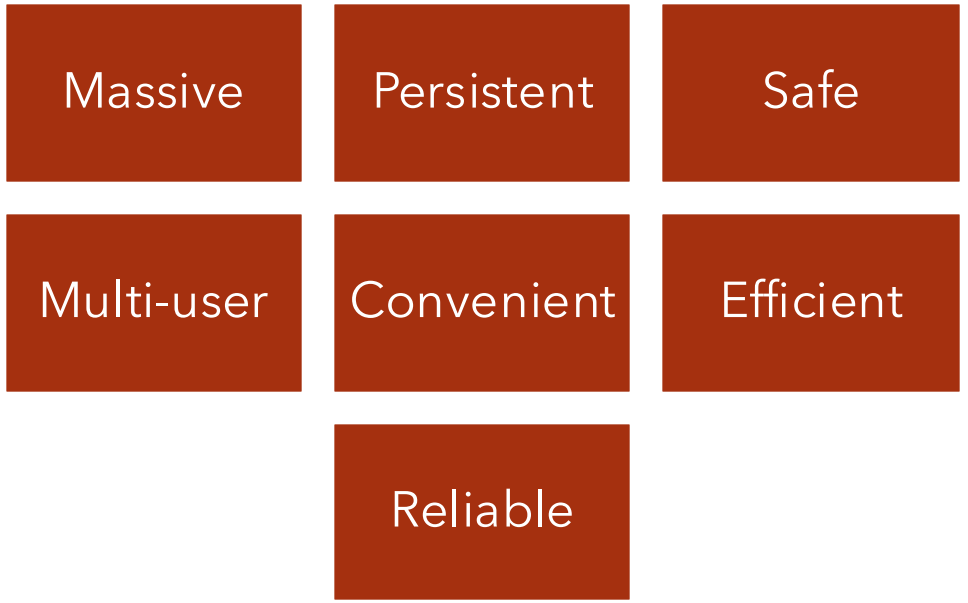
DATABASE – definitions

- **Shared** collection of **logically related data** (and a description of this data), designed to meet the information needs of an organization.
  - **System catalog (metadata)** provides description of data to enable program–data independence.
  - Logically related data comprises **entities**, **attributes**, and **relationships** of an organization's information.

DBMS – Database Management system

DBMS provides efficient, reliable, convenient, and safe, multi-user storage of and access to massive amounts of persistent data.

A software system that enables users to define, create, and maintain the database which provides controlled access to this database.



Database Approach

- **Controlled access to database may include:**
  - A security system.
  - An integrity system.
  - A concurrency control system.
  - A recovery control system.
  - A user-accessible catalog.

- **A view mechanism.**
  - Provides users with only the data they want or need to use.

Relational DBMS (RDBMS)

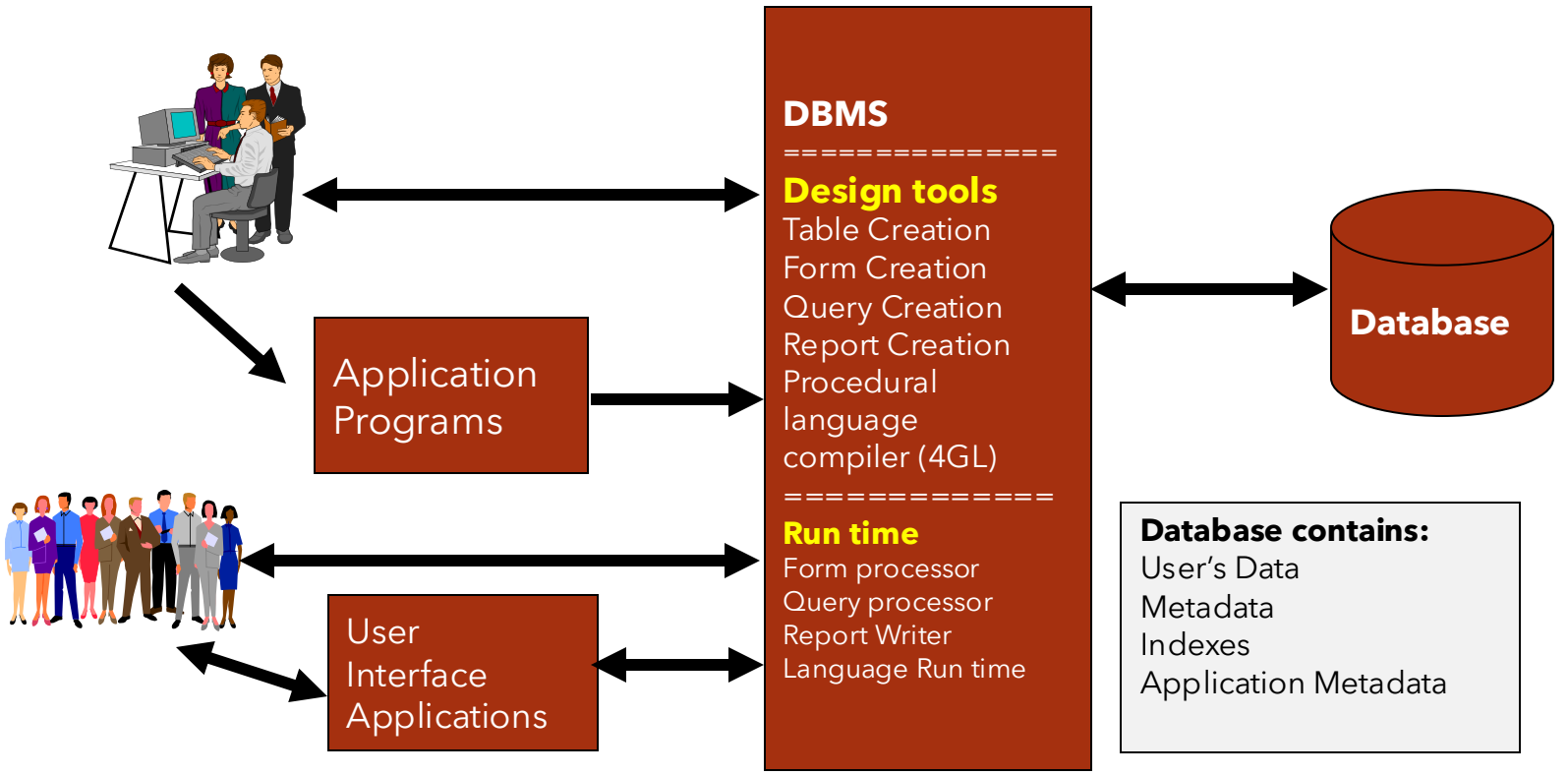
Relational database management system (RDBMS)

- a DBMS based on the relational data model.

Relational model

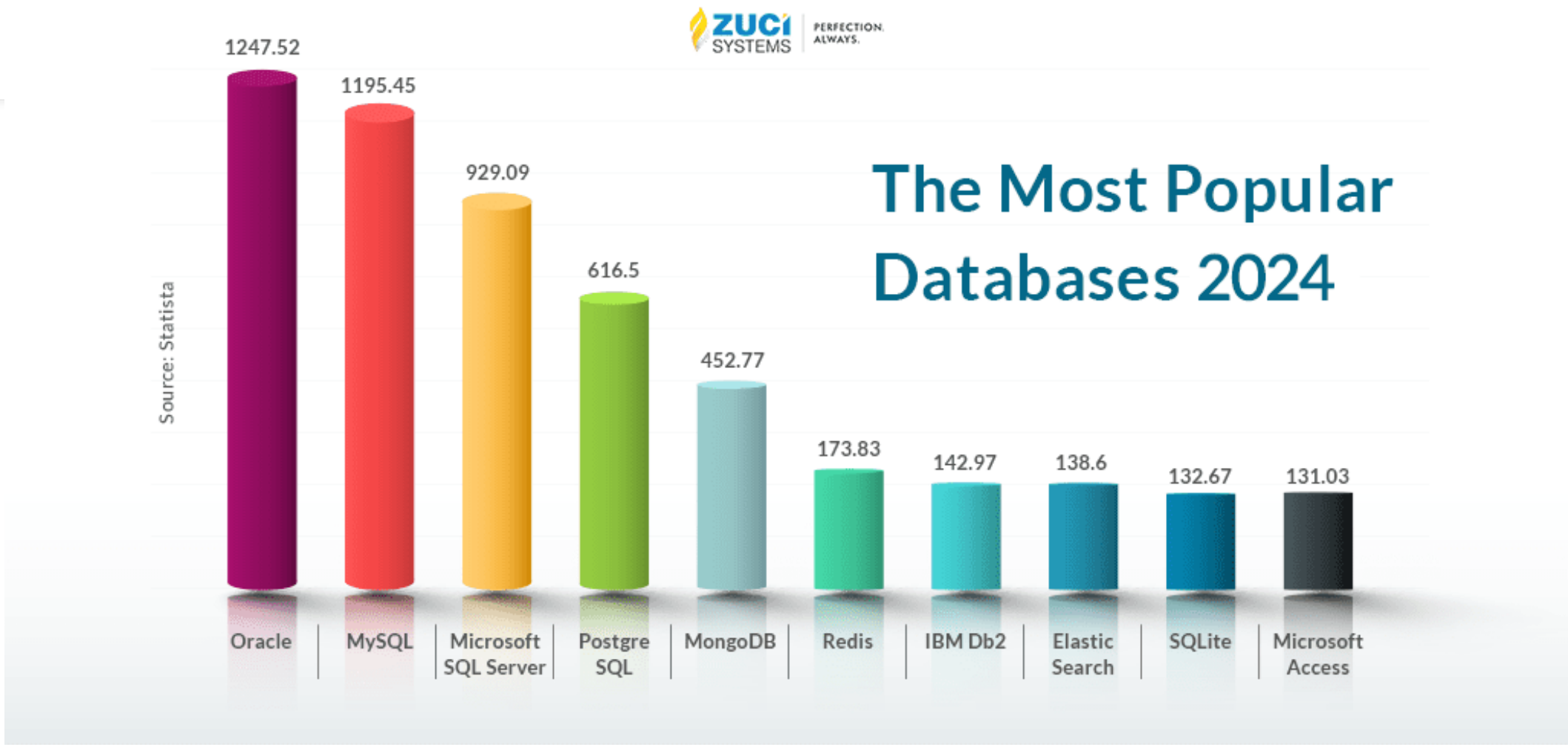
- a formal data model with a sound mathematical foundation, based on set theory and first order predicate logic in which a database is represented as a collection of **relations**.

Database Components

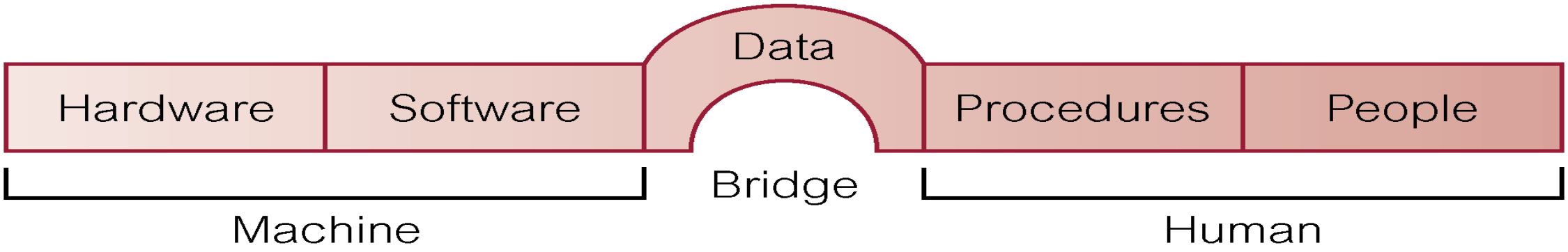


Modern Database Programming

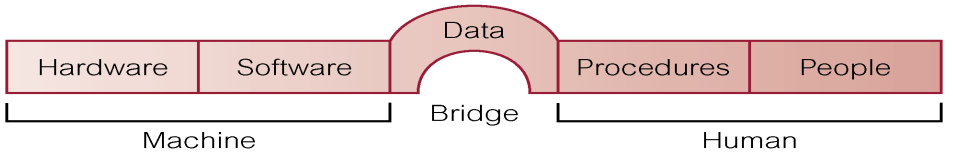
- **Frameworks**: Popular tools like Django, Flask, Laravel, React, Angular, Spring Boot, and Swift are used to build modern database-driven applications.
- **Middleware**: Middleware like Node.js, Express, and API gateways connect applications with databases, ensuring smooth data flow.
- **Beyond DBMS**: Some modern, data-heavy applications may use alternatives like cloud storage, NoSQL databases, or data lakes instead of traditional DBMS.


ZUCi SYSTEMS
PERFECTION. ALWAYS.
The Most Popular Databases 2024
Source: Statista
1247.52
1195.45
929.09
616.5
452.77
173.83
142.97
138.6
132.67
131.03
Oracle
MySQL
Microsoft SQL Server
Postgre SQL
MongoDB
Redis
IBM Db2
Elastic Search
SQLite
Microsoft Access

Components of DBMS Environment

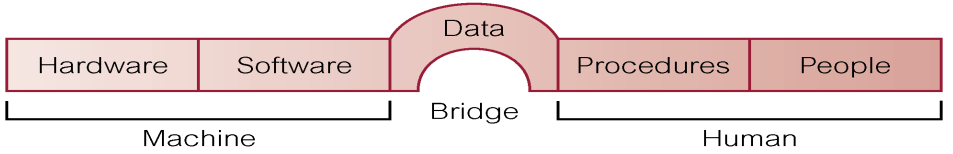


Components of DBMS Environment



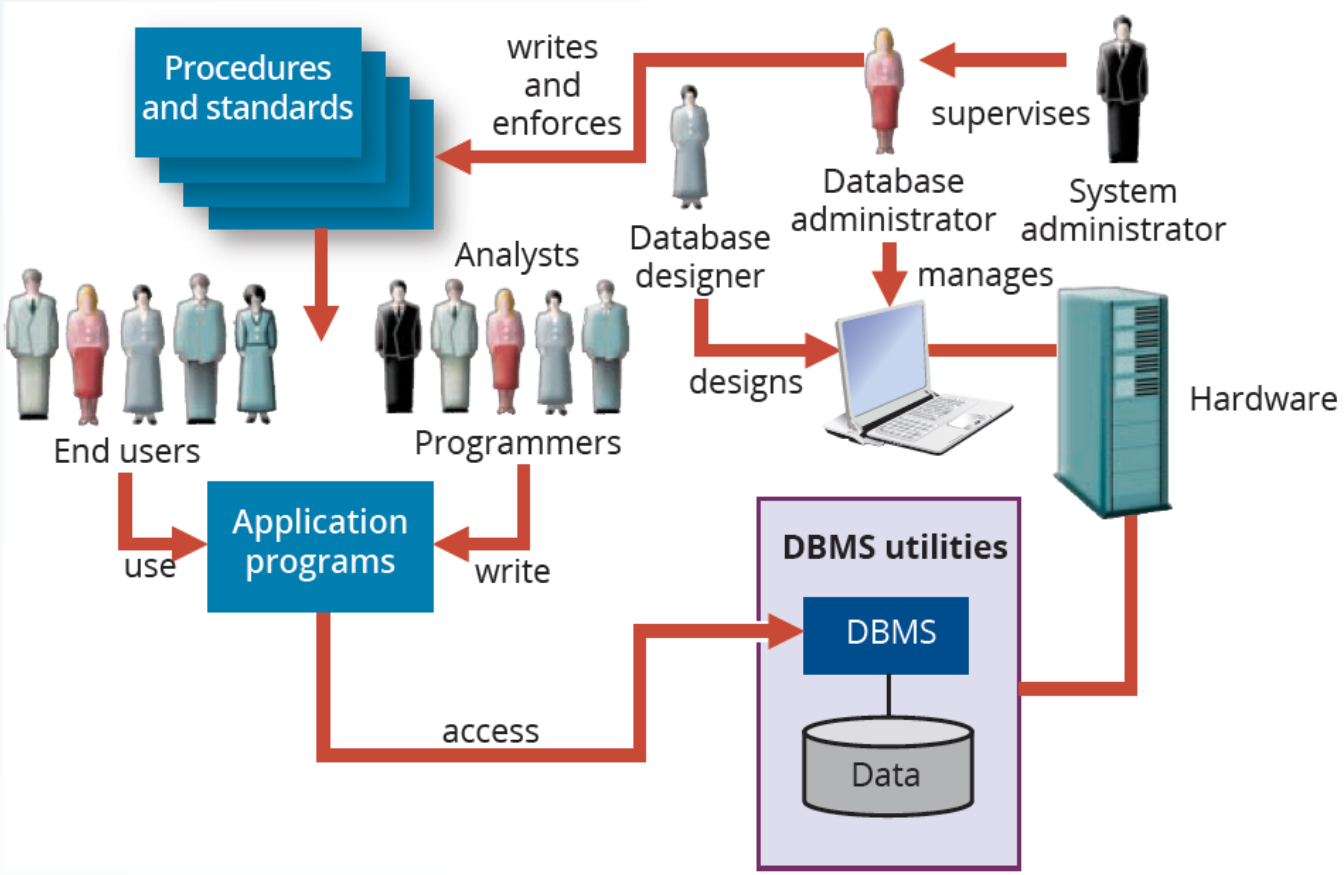
- **Hardware**
  - Can range from a PC to a network of computers.
- **Software**
  - DBMS, operating system, network software (if necessary) and also the application programs.
- **Data**
  - Used by the organization and a description of this data called the schema.

Components of DBMS Environment



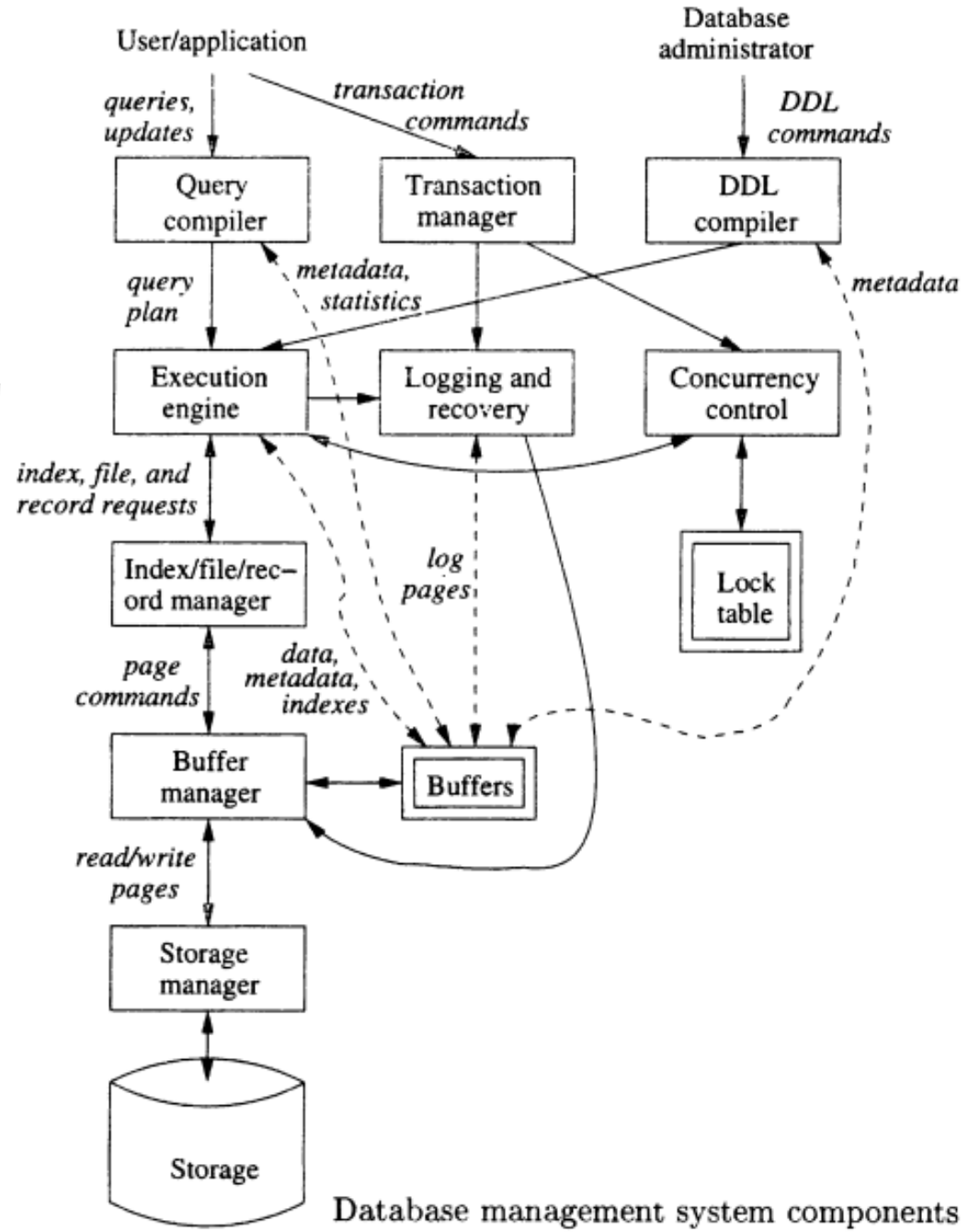
- **Procedures**
  - Instructions and rules that should be applied to the design and use of the database and DBMS.
- **People**
  - DBMS implementer
  - Database designer (Logical and Physical)
  - Database application developer
  - Database administrator
  - End Users (naive and sophisticated)

Figure 1.10 The Database System Environment



Functions of a DBMS

Data Storage, Retrieval, and Update.	A User-Accessible Catalog.	Transaction Support.	Concurrency Control Services.	Recovery Services.
Authorization Services.	Support for Data Communication.	Integrity Services.	Services to Promote Data Independence.	Utility Services.



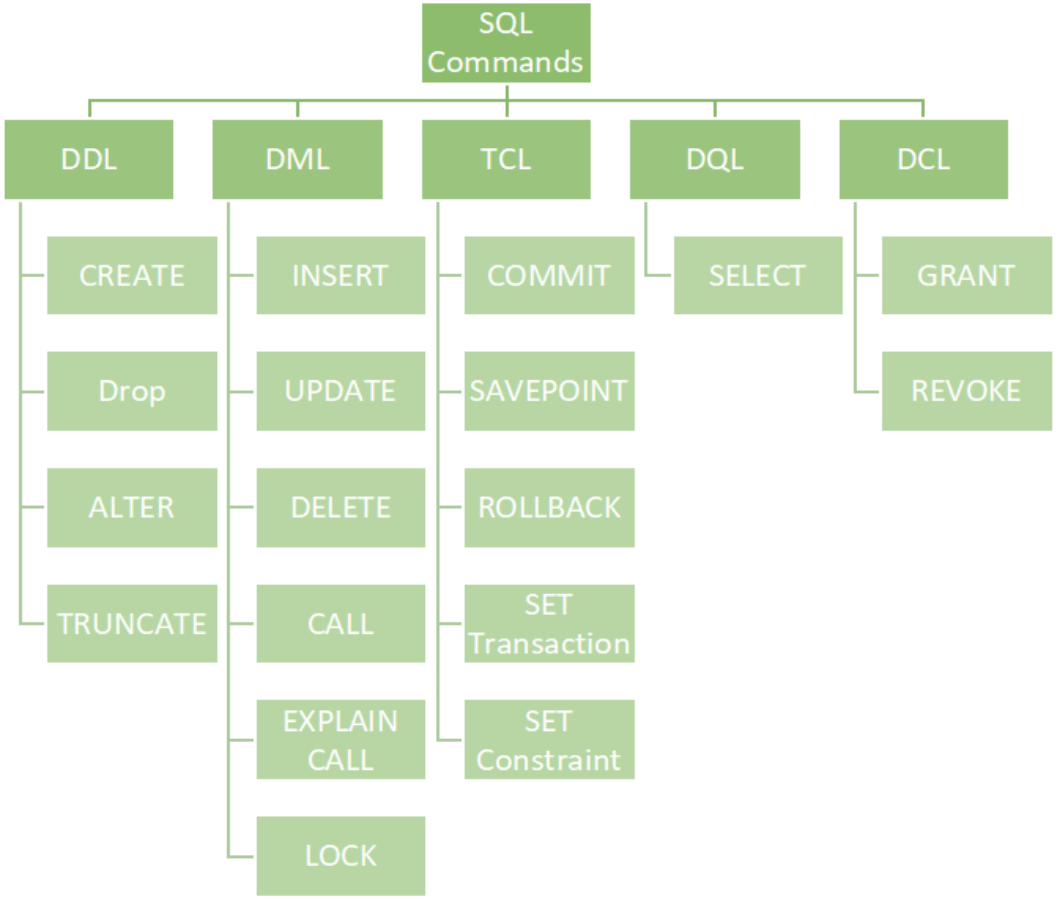
Database management system components

Key Database Activities (CRUD)

- **Create** - Add new data to the database
- **Read** - Read current data from the database
- **Update** - Update or modify current database data
- **Delete** - Remove current data from the database

Database Communication

- **DDL – Data Definition Language**
- **DML – Data Manipulation Language**
  - **DQL – Data Query Language**
- DCL – Data Control Language
- TCL – Transaction Control Language



Model Vs Instance

Database Model

The **database model** or **database schema** provides the description of the database data at different levels of detail and specifies the various data items, their characteristics, and relationships, constraints, storage details, etc.

Database Instance

- The database state then represents the data in the database at a particular moment. It is sometimes also called the current set of instances. Depending upon data manipulations, such as adding, updating, or removing data, it typically changes on an ongoing basis.

Database Model Vs Instance

Database model

- Student (number, name, address, email)
- Course (number, name)
- Building (number, address)

STUDENT

Number	Name	Address	Email
0165854	Bart Baesens	1040 Market Street, SF	Bart.Baesens@kuleuven.be
0168975	Seppe vanden Broucke	520, Fifth Avenue, NY	Seppe.vandenbroucke@kuleuven.be
0157895	Wilfried Lemahieu	644, Wacker Drive, Chicago	Wilfried.Lemahieu@kuleuven.be

COURSE

Number	Name
D0I69A	Principles of Database Management
D0R04A	Basic Programming
D0T21A	Big Data & Analytics

BUILDING

Number	Address
0600	Naamsestraat 69, Leuven
0365	Naamsestraat 78, Leuven
0589	Tiensestraat 115, Leuven

Figure 1.3 Example database state.