CC5 - INFORMATION MANAGEMENT

Zyra Yell Fagyan

COURSE OUTLINE

Unit 1: Data and Information

- Definition
- Types of Databases
- Roles of a Database Administrator and Data Architect
- Database Manipulation

COURSE OUTLINE

Unit 2: Database Design

- Process
- Consideration
- Data Modelling
- ERD Notation

COURSE OUTLINE

Unit 3: Normalization

- Definition
- CRUD Analysis
- 1st-5th Normal Form
- Denormalization

COURSE OUTLINE

Unit 4: Relational Physical Modelling

- Introduction to SQL
- Data Type Design
- Index Design
- View Design
- Relational Operations

COURSE OUTLINE

Unit 5: Database Algorithms and Architectures

- Concurrency Controls
- Database Recovery
- Database query application

COURSE OUTLINE

Unit 6: Big Data and NoSQL

- Definition
- Contemporary Data Science
- NoSQL Attributes and Characteristics

COURSE OVERVIEW

Unit 1

CC5 Information Management

Zyra Yell Fagyan

LEARNING OUTCOMES:

- discuss the underlying importance of data in an organization;
- collect, organize, retrieve, and manage data efficiently;
- model, transform, and present data effectively;
- secure data and information; and
- integrate data and information systems to support the organization.

Unit 1: Data and Information

- Definition
- Types of Databases
- Roles of a Database Administrator and Data Architect
- Database Manipulation

“What is Data?”

1.1 Data

Data represents **raw elements or unprocessed facts**, including numbers and symbols to text and images.

Data comes in various forms:

- **Quantitative data**, like an item's weight, volume, or cost, is provided numerically.

- **Qualitative data** is descriptive but non-numerical, such as a person's name and sex.

Examples of Data:
- A list of numbers (e.g., 45, 78, 23)
- Survey responses (e.g., "Yes," "No," "Maybe")
- Dates (e.g., 2024-09-04)
- Sensor readings (e.g., temperature: 25°C)

“What is Information?”

1.1 Information

Information is **data that has been processed, organized, or structured** in a way that adds meaning or context, making it useful for decision-making. Information provides answers to questions such as who, what, where, when, and how.

Examples:

- The average score of a class based on raw test data.
- A weather forecast generated from temperature and humidity data.
- A report showing sales trends over the last quarter.

Data or Information?

1. An email summarizing the key takeaways from a recent meeting.

Data or Information?

2. A report showing the monthly sales growth of a company.

Data or Information?

3. Inventory levels in a warehouse on a specific date.

Data or Information?

4. Individual satisfaction scores on a customer service survey.

Data or Information?

5. Understanding that changes to a website have led to an increase or decrease in monthly site visitors.

Data or Information?

6. Time-stamped GPS coordinates collected from a mobile device.

Data or Information?

7. The price of a competitors' product.

Data or Information?

8. Finding areas for improvement with customer service based on a collection of survey responses.

Data or Information?

9. Determining if a competitor is charging more or less for a similar product.

Data or Information?

10. The average temperature recorded in July in a specific city.

INTRODUCTION TO

DATABASES

“Is an Excel file considered a database?”

DATABASE

- A database is an organized collection of structured information or data
- Typically stored electronically in a computer system
- Controlled by a database management system (DBMS)

“What’s the importance of a Database?”

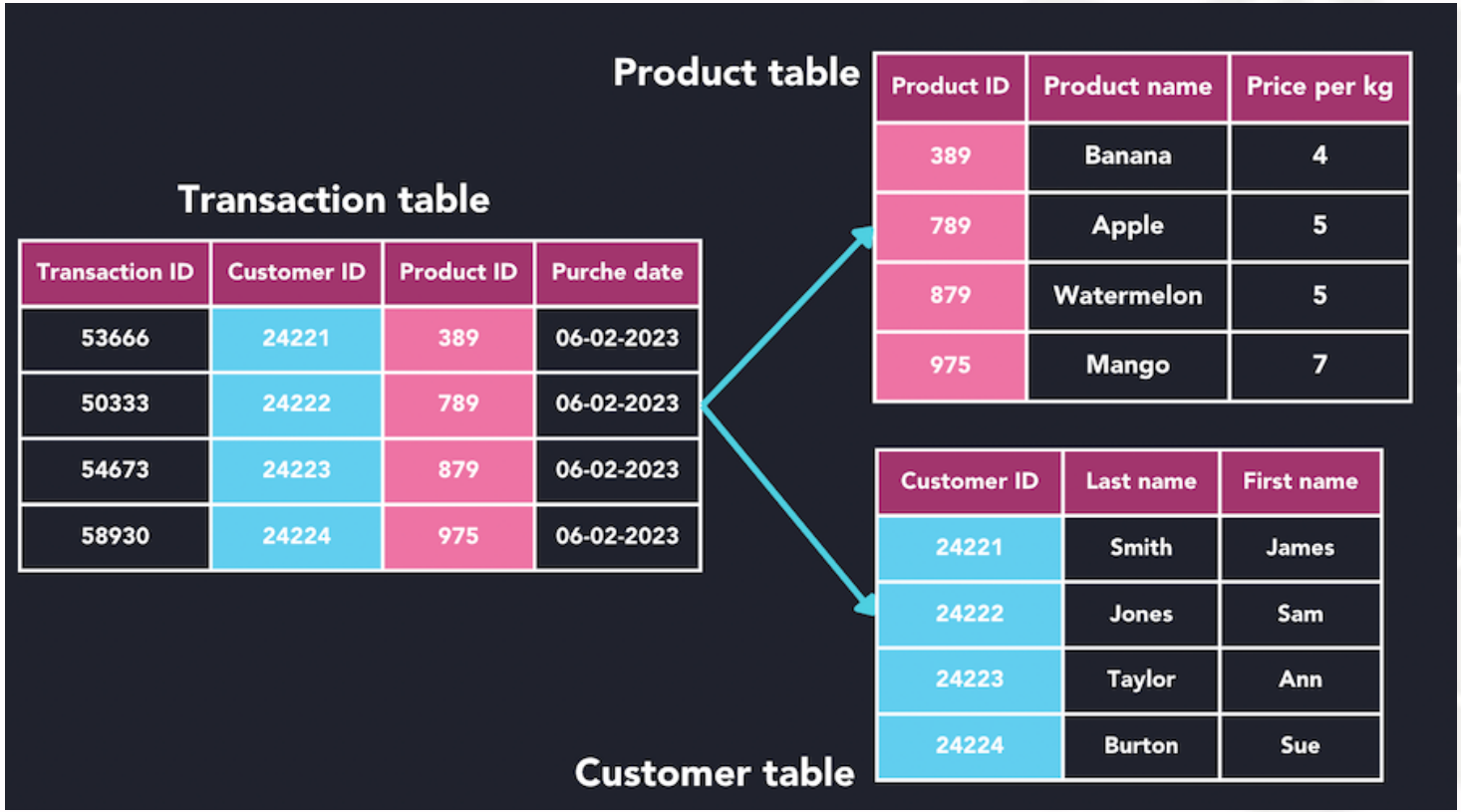
TYPES OF DATABASES:

1. Relational Databases

- Most common type of database
- Organizes data into tables with rows and columns
- Uses SQL (Structured Query Language) for querying
- Examples: MySQL, PostgreSQL, Oracle, Microsoft SQL Server

Structured query language (SQL)

- language for reading, creating, updating and deleting data.


Product table
Product ID
Product name
Price per kg
389
Banana
4
789
Apple
5
879
Watermelon
5
975
Mango
7
Transaction table
Transaction ID
Customer ID
Product ID
Purche date
53666
24221
389
06-02-2023
50333
24222
789
06-02-2023
54673
24223
879
06-02-2023
58930
24224
975
06-02-2023
Customer ID
Last name
First name
24221
Smith
James
24222
Jones
Sam
24223
Taylor
Ann
24224
Burton
Sue
Customer table

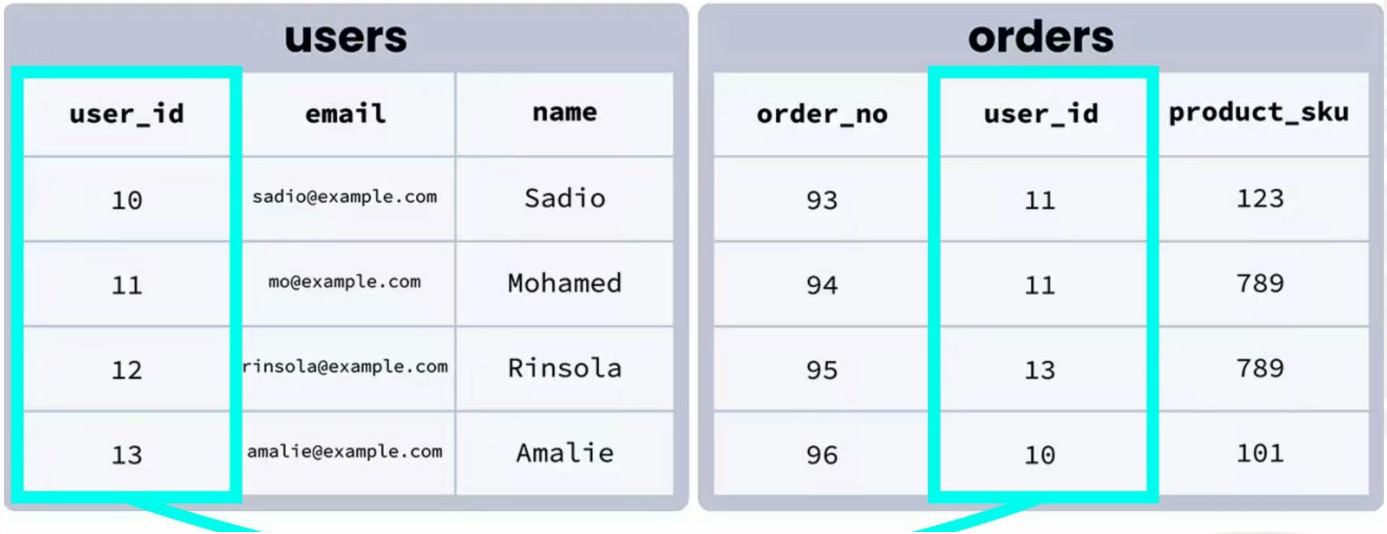
Popular Relational Databases

- MySQL: Open-source and known for its ease of use, speed, and reliability, often used in web applications.
- PostgreSQL: Open-source and highly extensible, offering advanced features and strong compliance with SQL standards.
- Oracle Database: A comprehensive, enterprise-grade solution known for its performance, scalability, and security.
- Microsoft SQL Server: Tightly integrated with the Microsoft ecosystem, offering a wide range of tools for business intelligence and analytics.

	Table	Action	Rows	Type	Collation	Size	Overhead
☐	wpep_cfs_contacts	Browse Structure Search Insert Empty Drop	4	MyISAM	utf8_general_ci	4.6 KiB	-
☐	wpep_cfs_countries	Browse Structure Search Insert Empty Drop	240	MyISAM	utf8_general_ci	11.9 KiB	-
☐	wpep_cfs_forms	Browse Structure Search Insert Empty Drop	11	MyISAM	utf8_general_ci	108.6 KiB	-
☐	wpep_cfs_modules	Browse Structure Search Insert Empty Drop	11	MyISAM	utf8_general_ci	3.3 KiB	-
☐	wpep_cfs_modules_type	Browse Structure Search Insert Empty Drop	2	MyISAM	utf8_general_ci	2 KiB	-
☐	wpep_cfs_statistics	Browse Structure Search Insert Empty Drop	130	MyISAM	utf8_general_ci	5.9 KiB	-
☐	wpep_cfs_usage_stat	Browse Structure Search Insert Empty Drop	0	MyISAM	utf8_general_ci	1 KiB	-
☐	wpep_commentmeta	Browse Structure Search Insert Empty Drop	0	MyISAM	utf8mb4_unicode_ci	4 KiB	-
☐	wpep_comments	Browse Structure Search Insert Empty Drop	1	MyISAM	utf8mb4_unicode_ci	7.3 KiB	-
☐	wpep_links	Browse Structure Search Insert Empty Drop	0	MyISAM	utf8mb4_unicode_ci	1 KiB	-
☐	wpep_options	Browse Structure Search Insert Empty Drop	502	MyISAM	utf8mb4_unicode_ci	140.5 KiB	7.4 KiB
☐	wpep_photo_gallery_wp_gallerys	Browse Structure Search Insert Empty Drop	2	MyISAM	utf8_general_ci	3.2 KiB	-
☐	wpep_photo_gallery_wp_images	Browse Structure Search Insert Empty Drop	9	MyISAM	utf8_general_ci	9 KiB	-
☐	wpep_photo_gallery_wp_like_dislike	Browse Structure Search Insert Empty Drop	0	MyISAM	utf8_general_ci	1 KiB	-
☐	wpep_postmeta	Browse Structure Search Insert Empty Drop	238	MyISAM	utf8mb4_unicode_ci	73.3 KiB	-
☐	wpep_posts	Browse Structure Search Insert Empty Drop	213	MyISAM	utf8mb4_unicode_ci	196.5 KiB	-
☐	wpep_termmeta	Browse Structure Search Insert Empty Drop	0	MyISAM	utf8mb4_unicode_ci	4 KiB	-
☐	wpep_terms	Browse Structure Search Insert Empty Drop	3	MyISAM	utf8mb4_unicode_ci	13.1 KiB	-
☐	wpep_term_relationships	Browse Structure Search Insert Empty Drop	8	MyISAM	utf8mb4_unicode_ci	3.2 KiB	-
☐	wpep_term_taxonomy	Browse Structure Search Insert Empty Drop	3	MyISAM	utf8mb4_unicode_ci	4.1 KiB	-
☐	wpep_usermeta	Browse Structure Search Insert Empty Drop	26	MyISAM	utf8mb4_unicode_ci	11.9 KiB	-
☐	wpep_users	Browse Structure Search Insert Empty Drop	1	MyISAM	utf8mb4_unicode_ci	8.1 KiB	-
	22 tables	Sum	1,404	InnoDB	utf8mb4_unicode_ci	617.4 KiB	7.4 KiB

When to use Relational Database?

- **Strong consistency**: Ensuring all users see the same data simultaneously.
- **Complex queries**: Joining data from multiple tables to gain insights.



users

user_id	email	name
10	sadio@example.com	Sadio
11	mo@example.com	Mohamed
12	rinsola@example.com	Rinsola
13	amalie@example.com	Amalie

orders

order_no	user_id	product_sku
93	11	123
94	11	789
95	13	789
96	10	101

When to use Relational Database?

- **Strong consistency**: Ensuring all users see the same data simultaneously.
- **Complex queries**: Joining data from multiple tables to gain insights.
- **ACID compliance**: Guaranteeing reliable transaction processing for critical applications.

Not best fit for:

- **Unstructured data**: Handling data that doesn't fit neatly into a tabular format (e.g., social media posts, sensor data).
- **Massive scalability**: When your application needs to scale horizontally across numerous servers.

TYPES OF DATABASES:

2. NoSQL Databases

- more flexible because the data on the object isn't limited to the same table.

- can handle unstructured or semi-structured data without the constraints of a fixed schema. This means we can store data in various formats, such as **JSON documents, key-value pairs, or graph structures**, without having to define a rigid structure upfront.

Key	Document
1001	{ "CustomerID": 99, "OrderItems": [{ "ProductID": 2010, "Quantity": 2, "Cost": 520 }, { "ProductID": 4365, "Quantity": 1, "Cost": 18 }], "OrderDate": "04/01/2017" }
1002	{ "CustomerID": 220, "OrderItems": [{ "ProductID": 1285, "Quantity": 1, "Cost": 120 }], "OrderDate": "05/08/2017" }

When to use NoSQL Database?

- **Agility is key**: Rapid development cycles and evolving data models.
- **Scale is a priority**: Applications with exponential data growth or high traffic.
- **Performance matters**: Real-time applications requiring fast read/write operations.
- **Variety is the norm**: Diverse data types (e.g., social media posts, sensor data).

Common use cases:

- **Big data analytics**: Processing massive datasets.
- **Real-time applications**: Delivering up-to-the-minute information.
- **Content management systems**: Storing and managing diverse content.
- **Internet of Things (IoT)**: Handling continuous data streams.
- **Personalization engines**: Tailoring user experience.

Types of NoSQL Databases

- **MongoDB**: A document-oriented database that is great for storing JSON-like documents with dynamic schemas.
- **Redis**: A key-value store often used for caching and as a fast in-memory datastore.
- **Cassandra**: A column-family store known for its scalability and fault tolerance.
- **Neo4j**: A graph database that excels in managing and querying highly connected data.

TYPES OF DATABASES:

3. Cloud Databases

- These databases reside on **remote servers** and are accessed over **the internet**, eliminating the need for organizations to invest in and maintain their own hardware and infrastructure.
- Cloud databases operate on a *pay-as-you-go* model

Querying cloud databases

- Querying cloud databases typically involves using the same tools and languages we'd use with on-premises databases
- For relational databases in the cloud, we'd use SQL to interact with the data
- NoSQL databases in the cloud typically have their own query languages or APIs, similar to their on-premises counterparts

When to use Cloud databases

- **Scalability is crucial**: Easily adapt to changing demands.
- **Flexibility is a priority**: Wide range of database options available.
- **Global accessibility is important**: Low-latency access for users worldwide.
- **Cost-effectiveness is a concern**: Pay-as-you-go model and scalable resources.

Popular Cloud Databases

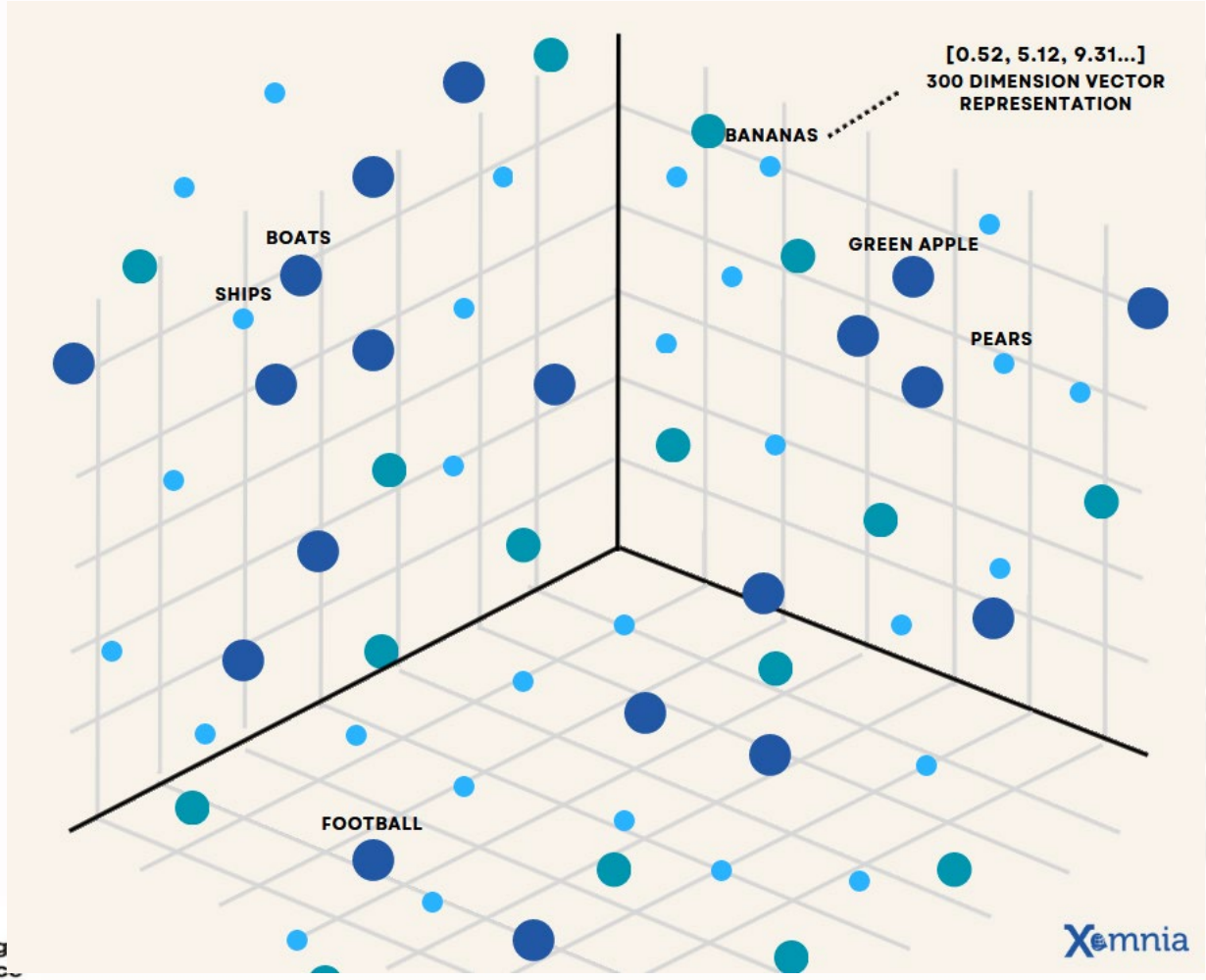
- **Amazon RDS**: Supports multiple database engines like MySQL, PostgreSQL, and Oracle, offering managed relational database services.
- **Google Cloud SQL**: A fully-managed service that allows running MySQL, PostgreSQL, and SQL Server databases in the cloud.
- **Azure SQL Database**: Provides scalable, intelligent, and fully-managed database services in the Microsoft Azure cloud.

TYPES OF DATABASES:

4. Vector Databases

- Vector databases have emerged as a specialized tool for handling the unique demands of **artificial intelligence and machine learning applications.**

- Vector databases are **designed to store, index, and manage vector embeddings.** This enables efficient similarity search, where the database can **quickly identify vectors** that are "close" to a given query vector based on distance metrics like cosine similarity or Euclidean distance.


[0.52, 5.12, 9.31...]
300 DIMENSION VECTOR
REPRESENTATION
BANANAS
GREEN APPLE
PEARS
BOATS
SHIPS
FOOTBALL
Xemnia

Querying Vector Databases

- **Embedding the query**: The input query (e.g., an image, a piece of text) is converted into a vector embedding using an appropriate embedding model.
- **Similarity search**: The vector database performs a similarity search to find the nearest neighbors of the query embedding in the vector space. This is often done using approximate nearest neighbor (ANN) algorithms to ensure efficiency at scale.
- **Returning results**: The database returns the identified nearest neighbors along with their associated metadata or original data objects.

Popular Vector Databases

- **Faiss**: Developed by Facebook AI Research, it provides efficient similarity search and clustering of dense vectors.
- **Milvus**: An open-source vector database that supports scalable similarity search and AI applications.
- **Pinecone**: A vector database service that simplifies the deployment and scaling of similarity search in production environments.

OTHER TYPES OF DATABASES

5. Time Series Database

- Time-series databases are optimized for **storing and analyzing time-stamped data**, such as sensor readings, stock prices, or server logs. They **excel at handling high-volume data ingestion** and efficiently querying data points based on time ranges.
- Popular options include InfluxDB, TimescaleDB, and Prometheus.

5. Time Series Database

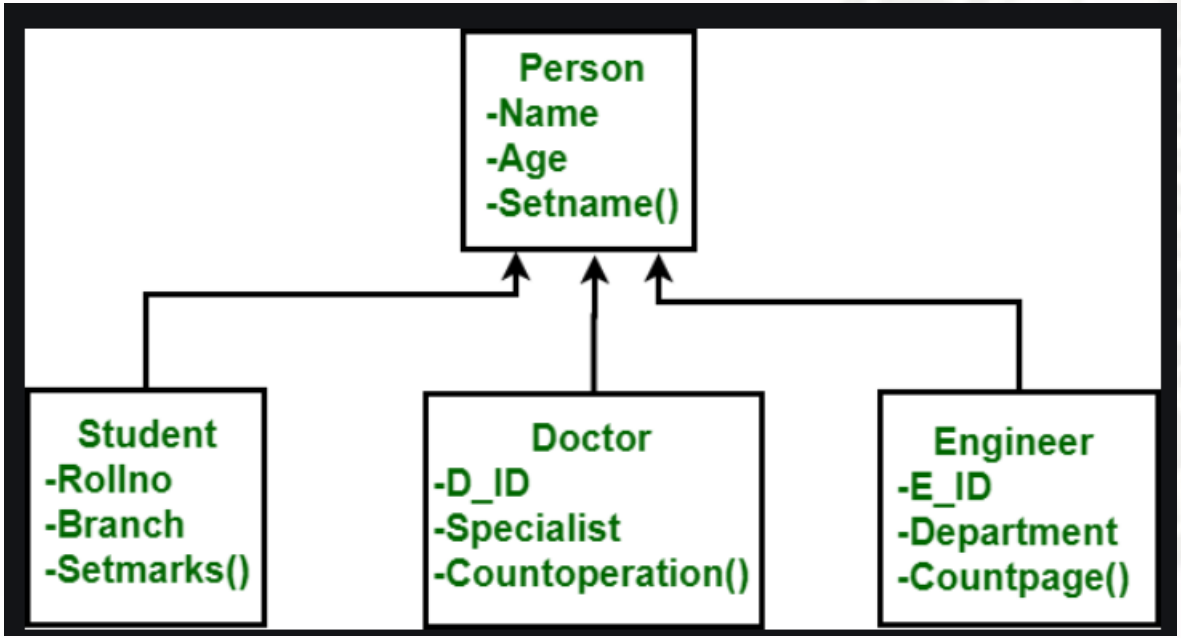
ID	Timestamp	Air quality	Temperature
1	1640120331	Poor	15
2	1640120332	Poor	16
3	1640120333	Poor	13

OTHER TYPES OF DATABASES

6. Object Oriented Database

- Object-oriented databases (OODBs) store data as objects, similar to **object-oriented programming**. This can simplify modeling complex data structures and relationships.
- However, OODBs have not gained widespread adoption due to **challenges with standardization and query optimization**.
- Popular options include ObjectDB and Versant Object Database.

6. Object Oriented Database

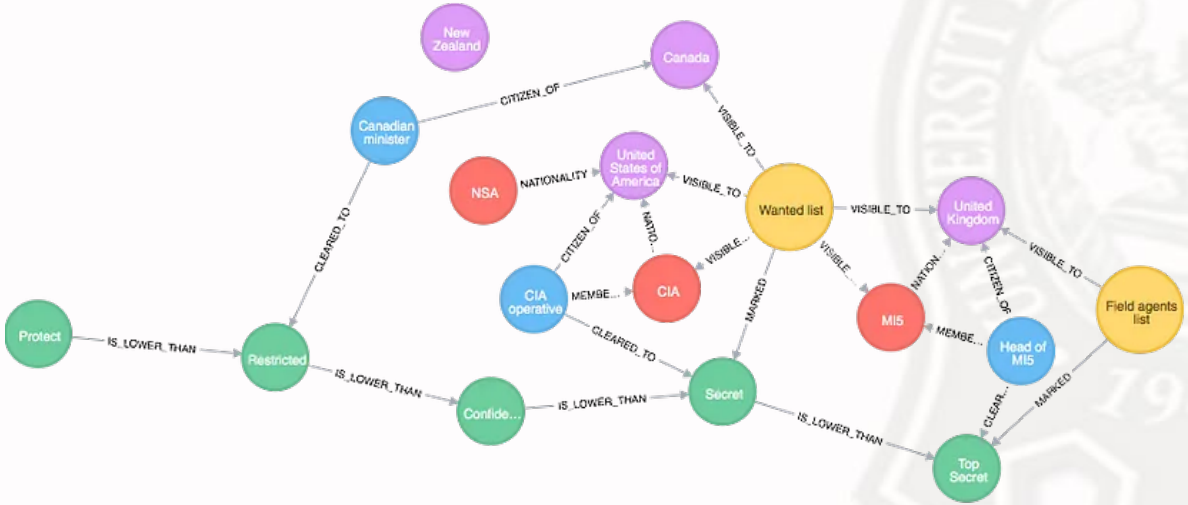


OTHER TYPES OF DATABASES

7. Graph Database

- Graph databases excel at representing and querying relationships between entities. **They store data as nodes (entities) and edges (relationships),** making them well-suited for social networks, recommendation engines, fraud detection systems, and knowledge graphs. Popular options include Neo4j, Amazon Neptune, and JanusGraph.

7. Graph Database

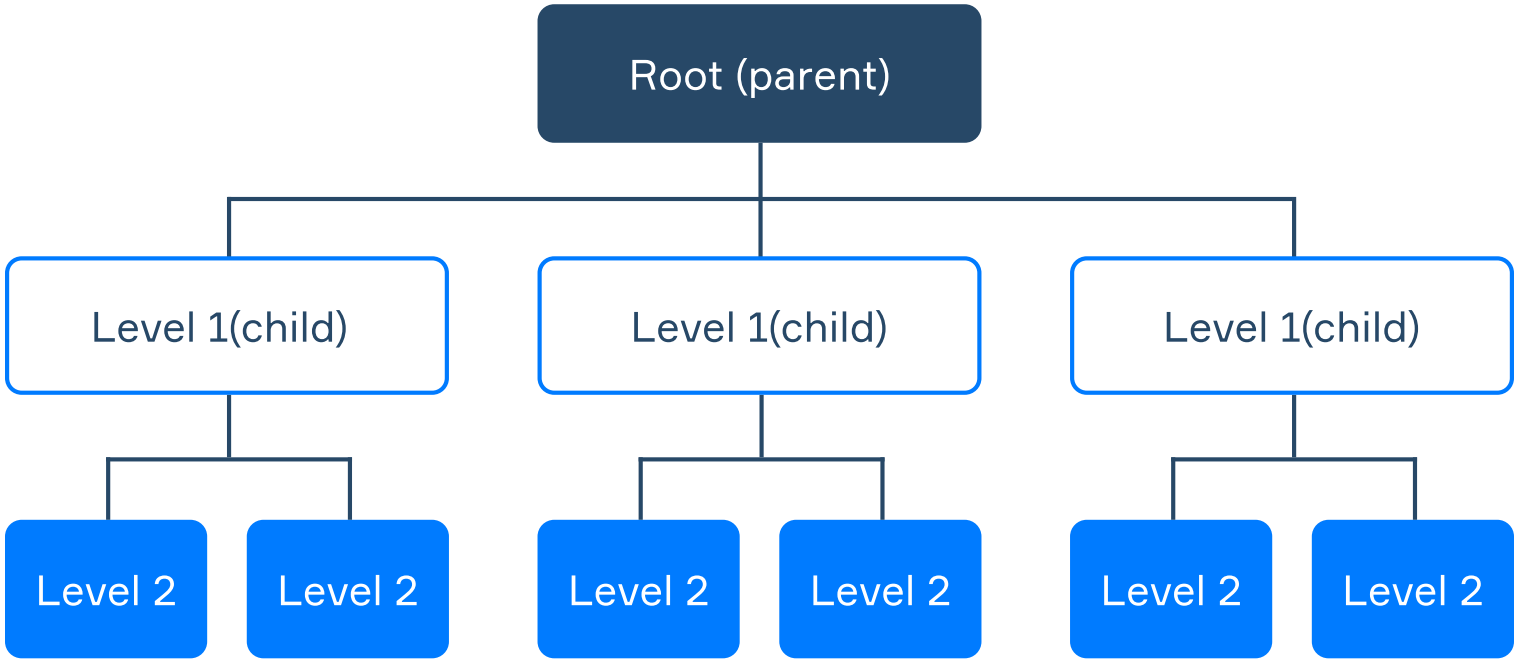


OTHER TYPES OF DATABASES

8. Hierarchal Database

- Hierarchal databases organize data in a tree-like structure, with parent-child relationships between records. This structure is suitable for some specialized applications but can be inflexible for complex data models. While historically significant, hierarchical databases are less common in modern applications.

The Hierarchical Database Model



OTHER TYPES OF DATABASES

9. Network Database

- Network databases are similar to **hierarchical databases** but allow for more complex relationships between records. While they offer flexibility, they can also be more challenging to manage and query. Network databases have largely been replaced by relational and graph databases in most applications.

9. Network Database