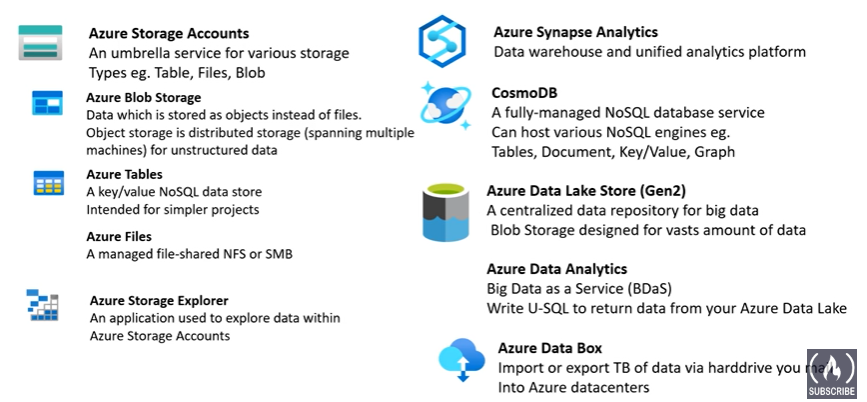
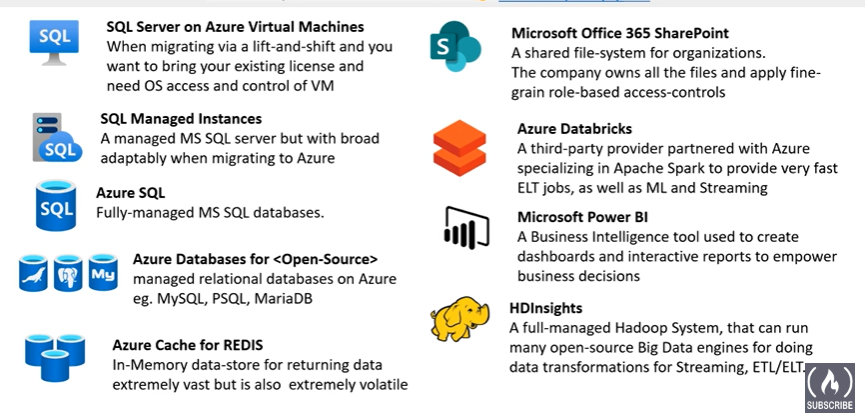
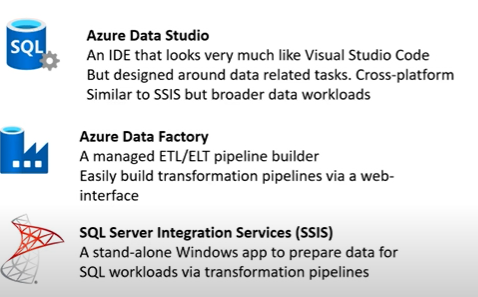
Azure DP 100

Core Data-related Azure Services:







**Types of Cloud Computing:**

**SAAS (software as a service):** it is specially designed for Customers

It is like a product that is run and managed by the Service provider

Note: Don’t need to worry about how the services are maintained. It just works and remains available

Eg: Power BI, Office 365

**PAAS (Platform as a Service):** For Developers

Focus on the deployment and management of your apps

Note: Don’t worry about provisioning, Configuring, or understanding the OS

eg: HDInsight’s, Managed SQL, Azure SQL, Cosmos DB

**IaaS (Infrastructure as a Service):** for Admins

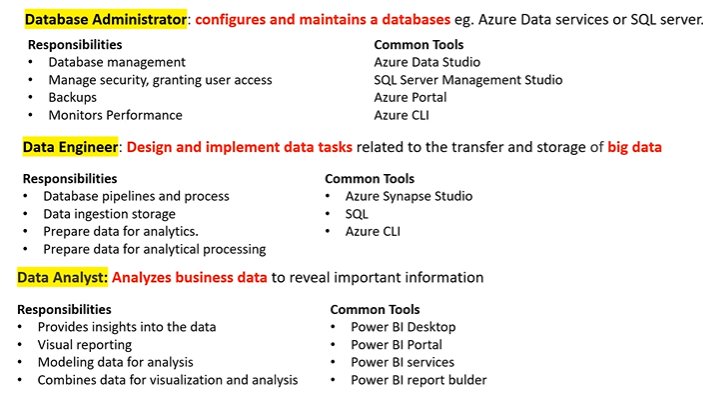
The basic building blocks for cloud IT.

Provides access to networking features, computers, and data storage spaces.

Note: Don’t need to worry about IT staff, hardware

Eg: SQL VM, Azure Disks, VM’s

**Azure Data-Related Roles:**



**Data Base Administrator 🡪 Common Tools:**

**Azure Data Studio:** Connect to Azur SQL, Azure SQL warehouse, Postgres SQL, SQL Server (big data clusters, on-premises)

Various lib and extensions along with automation tools

GUI for managing on-premise or cloud-based data services

Can run on MAC, Windows, Linus

Possibly a replacement for SSMS (lack of some features in SSMS)

**SQL Server Management Studio (SSMS):**

Automation tooling for running SQL commands pr common DB operations

GUI for managing on-premise or cloud-based data services

Runs on Windows only and has more features than Azure Data Studio

**Azure Portal and CLI:**

Manages SQL DB configurations eg create, delete, resize, no. of cores

Manage and provision other Azure Data Services

Automate creating, updating, or modifying resources via Azure Resource Manager templates (IaC)

**Data Engineer 🡪 Common Tools:**

Azure Synapse Studio:

Azure portal integrated to manage Azure Synapse, Data ingestion (Azure Data Factory), Management of Azure Synapse (SQL Pools/ spark Pool)

Knowledge SQL:

Create DB’s, Tables, Views, etc.,

Azure CLI:

Support operations SQL Commands to connect to Microsoft Server Azure SQL data and run talk queries and commands

HDInsights:

Streaming Data via Apache Kafka or Apache Spark

Applying for ELT jobs via HIVE, PIG, Apache Spark

Azure Databricks:

Using Apache Spark to create ELT or streaming jobs to data ware house or data lakes

Will also be working with Data lakes and Blob storage

**Data Analyst 🡪 Common Tools:**

Power BI Desktop: A stand-alone application for Data visualization

You can do modeling, connect to many data sources, create interactive reports

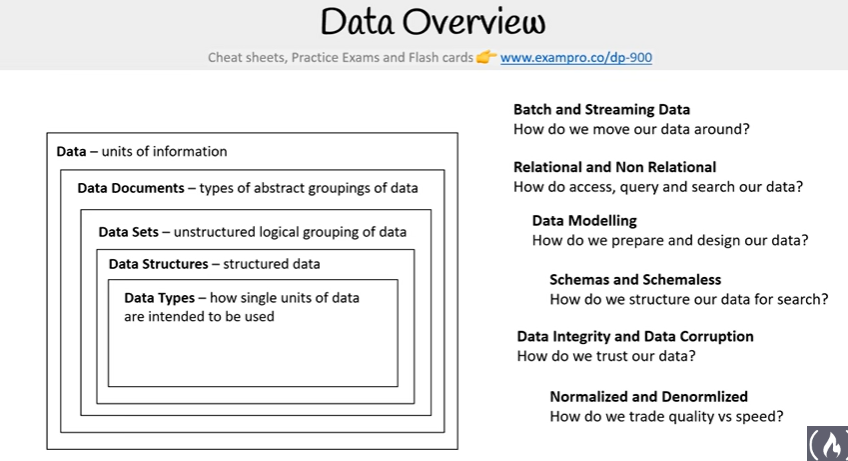
Power BI Portal/ Power BI service:

A web UI for creating interactive Dashboards

Power BI Report Builder:

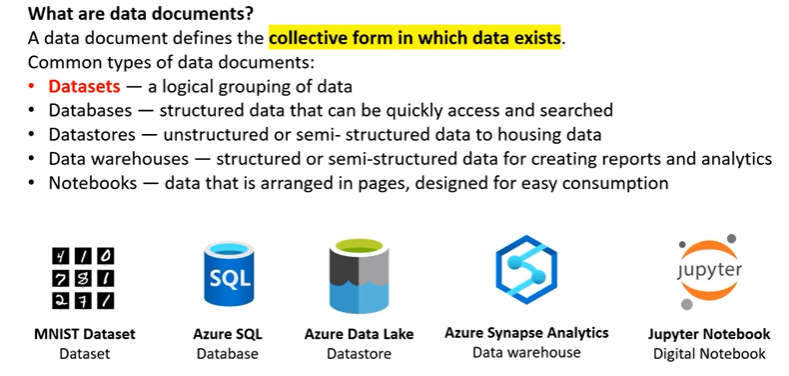
Create paginated reports (Printable reports)

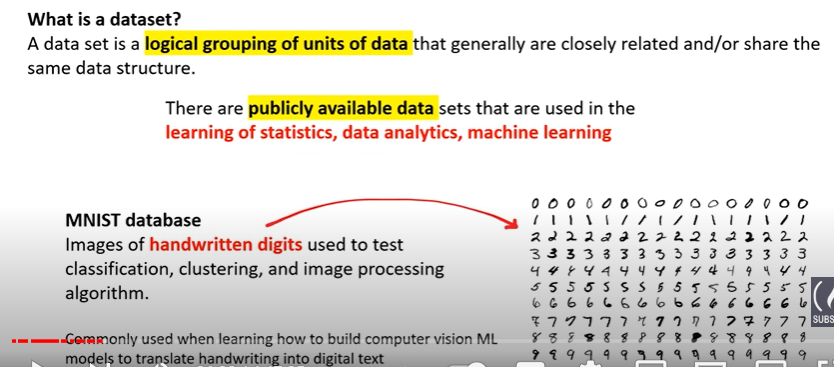
**Data Overview:**



**Data:**  it is units of Data it can be in form or image , text, number, etc.,

Data Documents: collective from which data exists





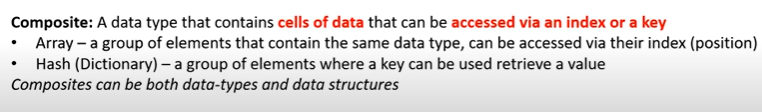
Some examples of Data Sets

* IMDB Reviews the Data set
* Free Music Archive
* Libri Speech

**Data Types:** A single unit of Data that tells a compiler or interpreter how the data is intended to be used

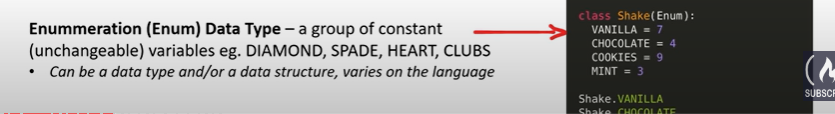
Numeric Data types: Integer, float

Text Data types: characters, String, Composite (Array, Hash(Dictionary))

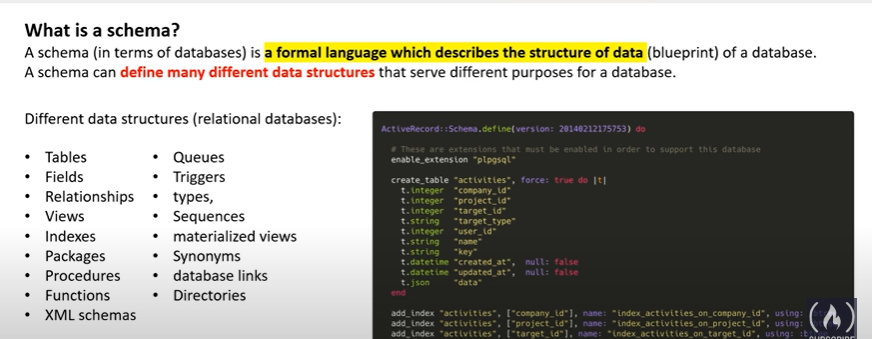


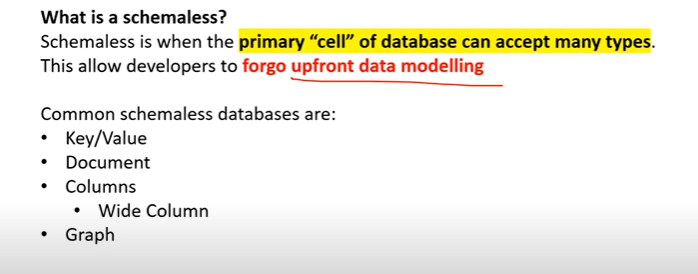
Binary Data types: bits or series os bits

Boolean Data type True or False, 0 or 1, T or F

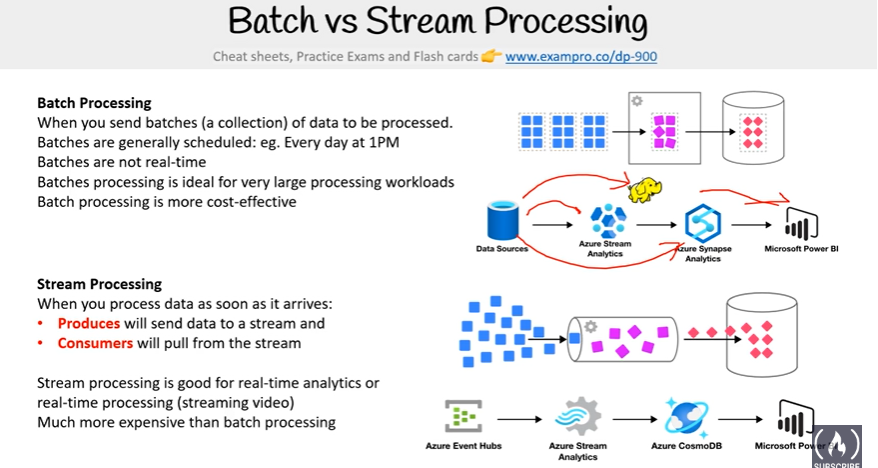


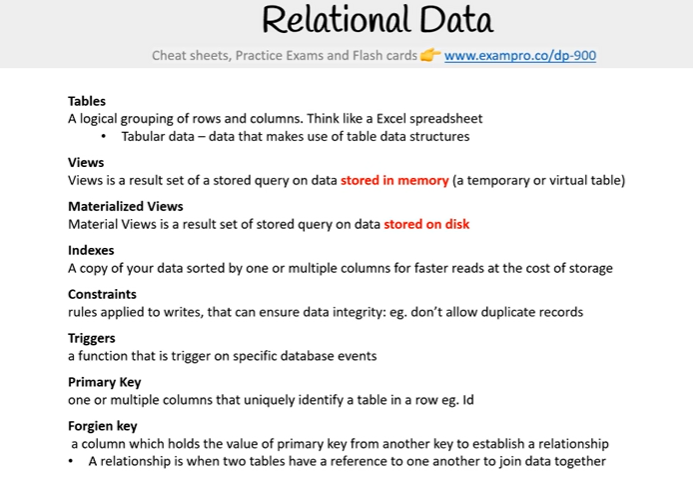
Schema vs Schemeless:



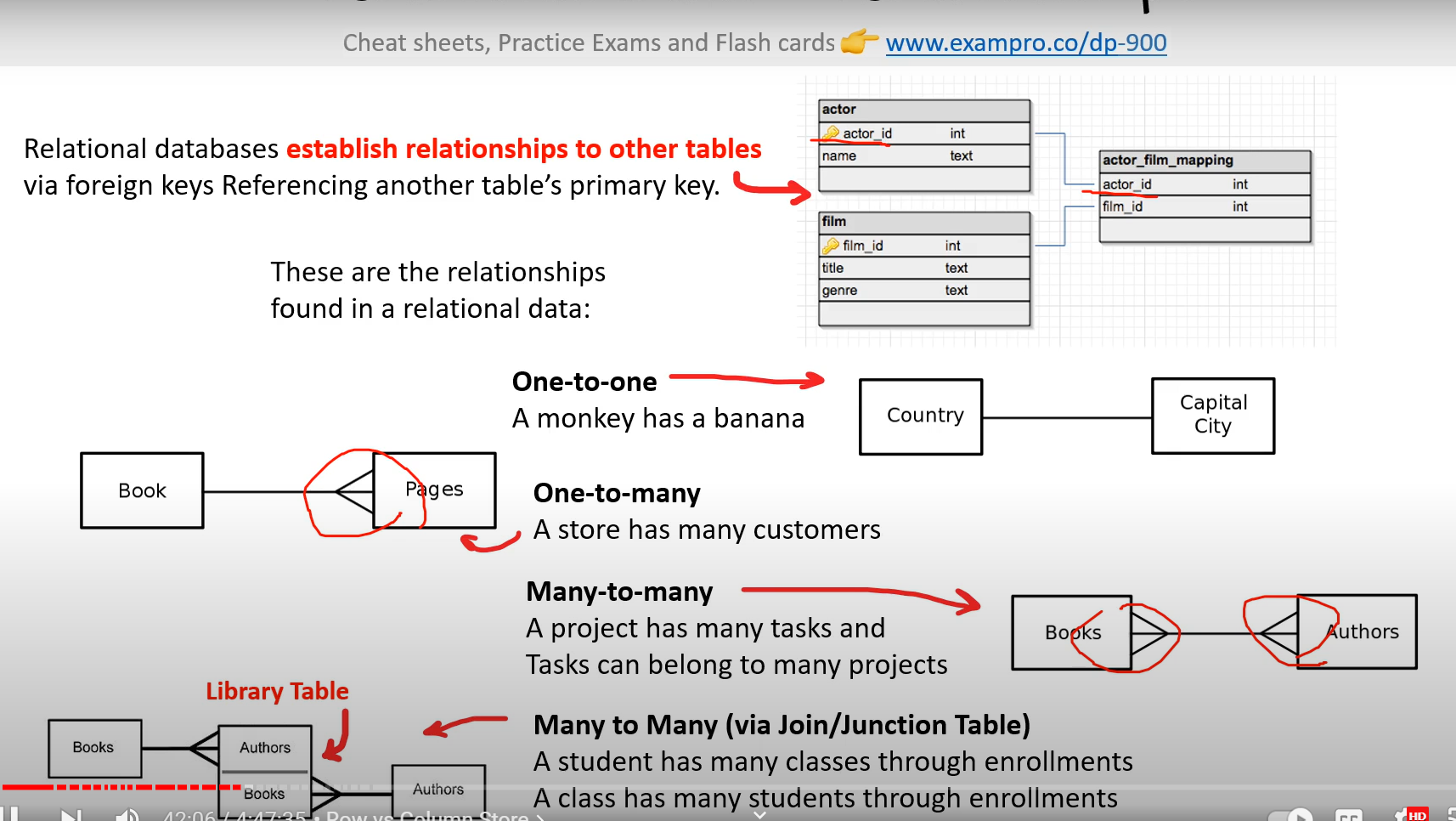


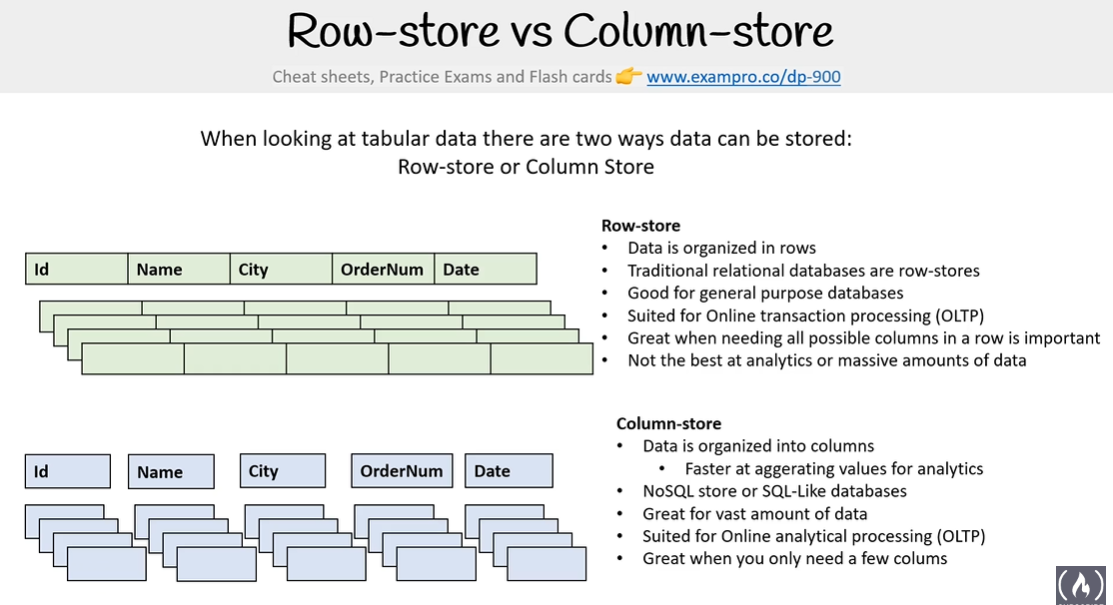
Batch Processing vs Stream Processing:

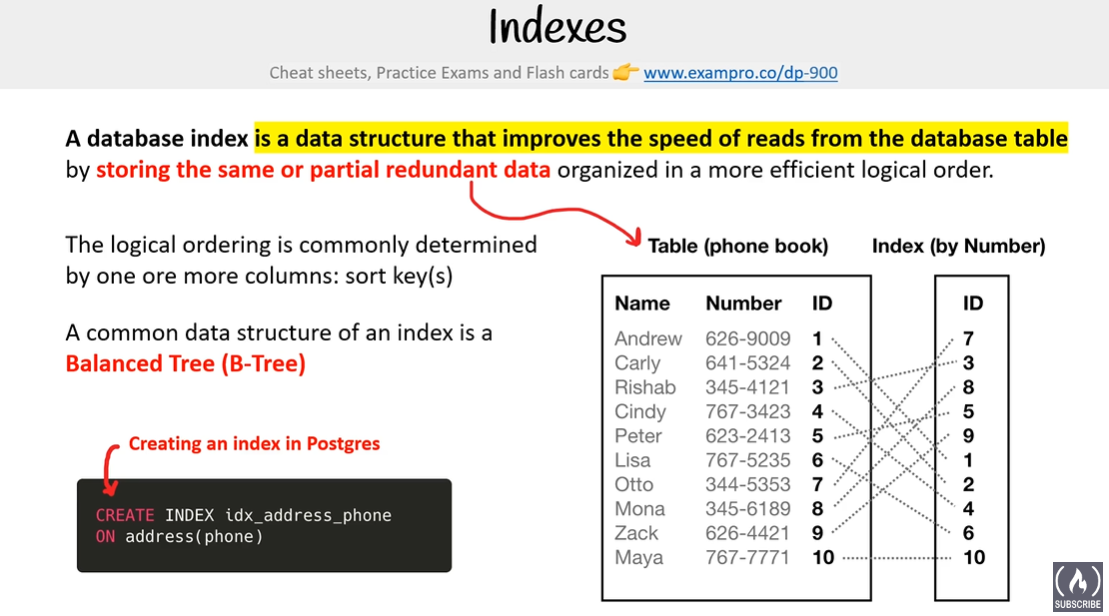




Relations between tables:





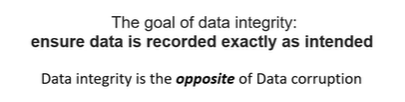


**Data Integrity and Data Corruption:**

Data integrity is the maintenance and Assurance of data accuracy and consistency over its entire life cycle

Proxy term for Data Quality

Data validation is a pre-requisite for data integrity



**Data Corruption:** it is an act of data not being in the intended state

Which results in data loss or misinformation

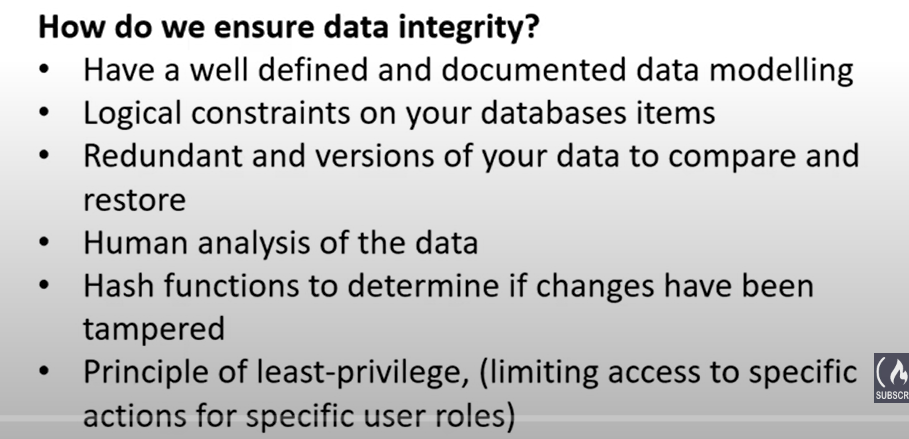
Data corruption occurs when unattended changes result when reading and **writing:**

Unexpected hardware failure

Human error when inputting or modifying data

Unforeseen side effects for automated operations via computer code

Malicious actors with the intent of corrupting your data



**Normalized VS DeNormalized:**

