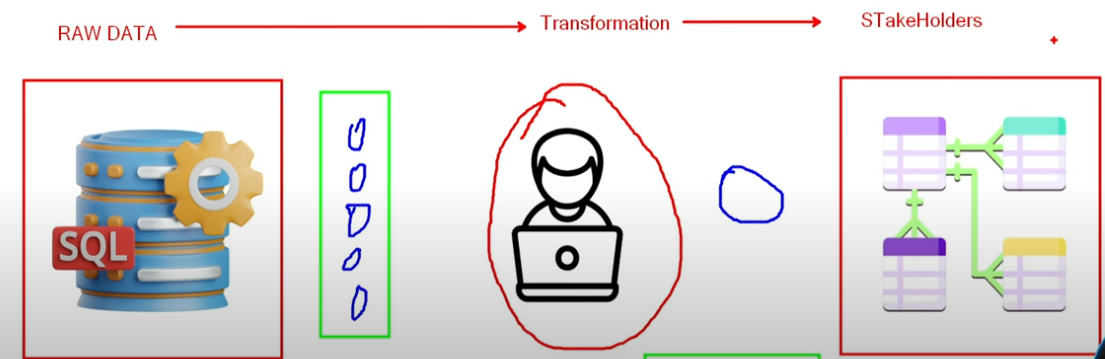
**Data dictionary**

**Data engineering**

Data Engineering ek process hai jisme raw data ko collect, store, process aur optimize karke analysis ke liye tayar kiya jata hai.

**ex**

Swiggy jaise apps data engineers ka use karte hain taaki order tracking aur delivery ko fast aur smooth banaya ja sake. 🚀



**Data Engineering Lifecycle**

ek process hai jisme data ko ikattha karna, save karna, badalna aur analyze ke liye tayar karna hota hai.

**Ex**

Amazon jaise websites par, logon ke orders aur clicks ka data collect karke, process karke, store kiya jata hai aur fir use analysis ke liye use kiya jata hai.

**Raw data**

woh hota hai jo bilkul asli form mein hota hai, bina kisi badlav ya safai ke.

**Ex.**

Data engineering mein Raw data ka real world example kya ek sentence mein

**Microsoft Fabric**

ek unified data analytics platform hai jo data engineering, data science, real-time analytics, aur business intelligence ko ek hi ecosystem me integrate karta hai.

Microsoft Fabric ek platform hai jo data ko store, process, aur analyze karne ke liye alag-alag tools ek jagah par lata hai.

**Data Generation**

Data Engineering mein Data Generation ka matlab hai naye data ka banana, jo manually ya automatic tareeke se create kiya jata h

**Ex**

E-commerce websites jaise Amazon har second lakhon transactions generate karti hain, jo real-time data engineering pipelines ke through store aur process hoti hain.

**Data Transformation**

Data engineering mein Data Transformation ka matlab hai raw data ko clean, format, aur structure karke usable format mein convert karna, taaki analysis aur processing aasaan ho sake.

**Ex**

E-commerce platforms ETL (Extract, Transform, Load) pipelines ka use karke raw sales data ko transform karti hain, jaise alag-alag sources se aaye order details ko clean karke aur customer ID ke saath merge karke analytics ke liye ready karna.

**Data Serving**

Data Engineering mein Data Serving ka matlab hai processed aur optimized data ko end-users ya applications ke liye efficiently accessibile banana, taki woh analytics, reporting, ya machine learning models ke liye use ho sake.

**EX.**

Data Serving ka ek real-world example **Netflix** hai, jo users ko unki pasand ke hisaab se movies aur shows recommend karta hai.

**Data Upstream**

un sources ya systems ko kehte hain jo data generate ya provide karte hain, jaise databases, APIs, ya IoT devices., aur fir processing ya storage ke liye downstream systems tak flow karta hai.

**Ex.**

Data Engineering mein Data Upstream ka real world example kya hai ek sentence mein

**Data downstream**

ka matlab hai jab data process hone ke baad users, apps ya analytics tools tak pahunchta h.

**Ex.**

E-commerce platforms par customer orders ka data warehouse mein store hone ke baad analytics dashboard tak pahunchna ek data downstream ka example hai.

**OLTP**

(Online Transaction Processing) database ek high-performance database hota hai jo real-time transactions ko efficiently handle karta hai, jaise banking, e-commerce, aur inventory management systems mein.

**Ex.**

Banking transactions system, jisme har deposit, withdrawal aur fund transfer turant process hota hai, ek real-world example hai OLTP (Online Transaction Processing) ka.

**OLAP**

(Online Analytical Processing) database ek type ka database hota hai jo complex queries aur multidimensional analysis ke liye optimized hota hai, jise business intelligence aur reporting ke liye use kiya jata hai.

We use relational db - mysql, postresql

**Ex.**

OLAP ka ek real-world example hai Amazon ka sales analysis system, jo purane aur naye data ko milaakar alag-alag dimensions (jaise product category, region, aur time) ke basis par insights generate karta hai.

**Diff**

OLAP data analyze karne aur reports banane ke liye hota hai, jabki OLTP real-time transactions handle karta hai, jaise banking ya shopping apps.

**ETL**

(Extract, Transform, Load) ek data engineering process hai jisme data ko alag-alag sources se extract kiya jata hai, uska transform karke safai aur format change kiya jata hai, aur fir use ek database ya data warehouse me load kiya jata hai analysis ke liye.

**Ex.**

E-commerce platforms jaise Amazon apne databases se raw sales data extract karte hain, usse clean aur transform karke structured format mein convert karte hain, aur phir analytics aur reporting ke liye data warehouse (jaise Redshift ya Snowflake) mein load karte hain.

**Data Warehouse**

ek centralized repository hota hai jo multiple sources se structured data ko store, process, aur analyze karne ke liye design kiya jata hai, taaki business intelligence aur reporting ko support kiya ja sake.

**Ex.**

Amazon Redshift ek real-world example hai, jo Amazon ka cloud-based data warehouse hai aur large-scale data analytics ke liye use hota hai.

**Data pipeline**

ek automated process hai jo raw data ko extract, transform, aur load (ETL/ELT) karke usable insights ya storage systems tak pahunchata hai.

**Data warehousing**

ek process hai jisme alag-alag sources se data ko extract, transform aur load (ETL) karke ek centralized repository me store kiya jata hai, taki analysis aur reporting aasaan ho sake.

**Data layers**

alag-alag stages hain jismein raw data ko process, transform aur store kiya jata hai, jaise ki ingestion layer, processing layer, storage layer, aur analytics layer.

**Data eng workflow or 3 piller of data eng**



**Data production**

Data production ka matlab hai naye data ka banana ya existing data ko badalna taaki usey aage process kiya ja sake.

**Data transformation**

Data engineering mein data transformation ka matlab raw data ko clean, restructure, aur enrich karke ek usable format mein convert karna hota hai, taaki analysis ya downstream systems ke liye useful bane.

**Data serving**

Data serving ka matlab hai users ya applications ke liye processed data ko fast aur easily available banana.

**Staging layer**

Staging layer ek temporary storage area hoti hai jahan raw data ko transform ya process karne se pehle load kiya jata hai.

**Ex**

Staging layer ek temporary storage hota hai jahan raw data ko multiple sources se collect karke clean aur transform kiya jata hai, jaise ek warehouse mein naye maal ka inspection aur sorting hota hai before final storage.

**Core Layer**

wo central layer hoti hai jo raw data ko process, transform, aur optimize karke downstream systems ya analytics ke liye ready karti hai.

**Ex**

Core Layer ka real-world example: E-commerce platform (e.g., Amazon) mein Core Layer raw transactional data (orders, payments, user clicks) ko store karta hai, jo aage transformation aur analytics ke liye use hota hai.

**Incremental loading**

ek data integration technique hai jisme sirf naye ya updated records ko source se destination par load kiya jata hai, taaki performance aur efficiency improve ho.

**Ex.**

Incremental loading ka real-world example ho sakta hai: "E-commerce website pe, har din sirf naye orders aur updates ko database mein load karna, taaki purane data ko dobara se process na karna pade."

**Fact Table**

ek main table hoti Main Table ek central table hoti hai jo data ka core part hoti

**Ex.**

Data engineering mein, ek real-world example of Fact Table ho sakta hai ek sales fact table, jisme transactions ke details jaise total sales, quantity sold, aur transaction date store kiye jate hain, jo ki analysis ke liye use kiye jaate hain.

**Dimension Tables**

woh tables hoti hain jo Fact Table ke data ko describe karti hain, jaise date, location, product, customer, etc., taki data ko alag-alag angles se analyze kiya ja sake.

**Ex.**

Dimension tables ka real-world example ek retail store ka product catalog ho sakta hai, jisme products ki details jaise ki product name, category, brand, aur size stored hoti hain, jise sales data ke saath join karke analysis kiya jata hai.

**Dimensional Data Modeling**

ek technique hai jo data ko fact tables (measurable events) aur dimension tables (contextual attributes) mein organize karti hai, taaki analytical queries fast aur efficient ho sakein.

**Ex.**

Dimensional Data Modeling ka real-world example retail industry mein hota hai, jahan sales data ko product, time, aur store ke dimensions ke saath model kiya jata hai taaki analysis aur reporting aasani se ki ja sake.

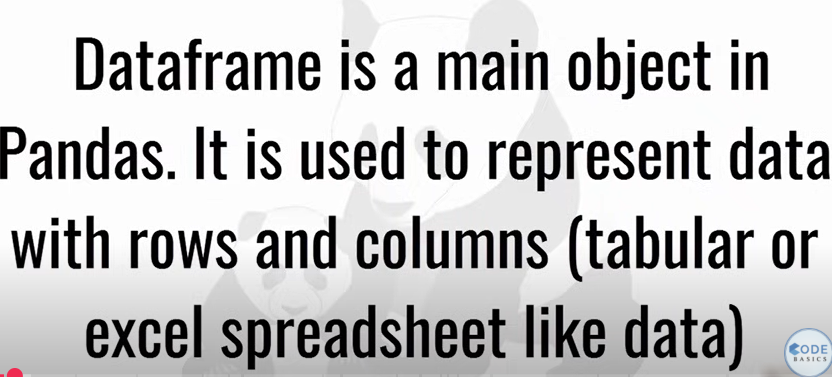
Star Schema

ek data modeling technique hai jo ek central fact table ko multiple dimension tables se connect karti hai, jisse queries fast execute hoti hain aur data analysis optimized hota hai.

Snowflake Schema

ek normalized database design hai jo star schema ka extended version hota hai, jisme dimension tables further normalize ki jati hain redundancy kam karne ke liye.





**api**