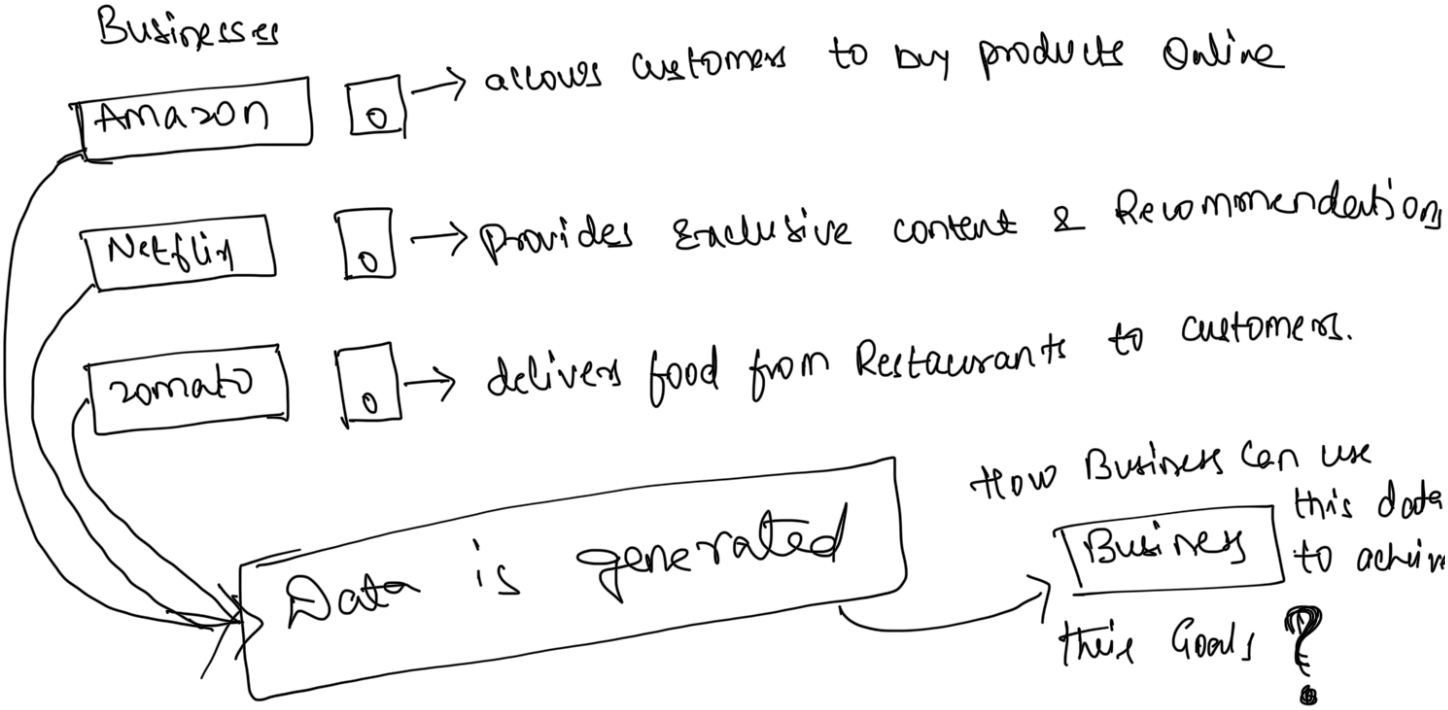
The Fundamentals of Data Engineering

Data Engineering ⟷ Internet

* Every thing Happens on the internet generates data



Business Goals

* understand customer Behaviour
* Increasing profits
* identifying & fixing problems

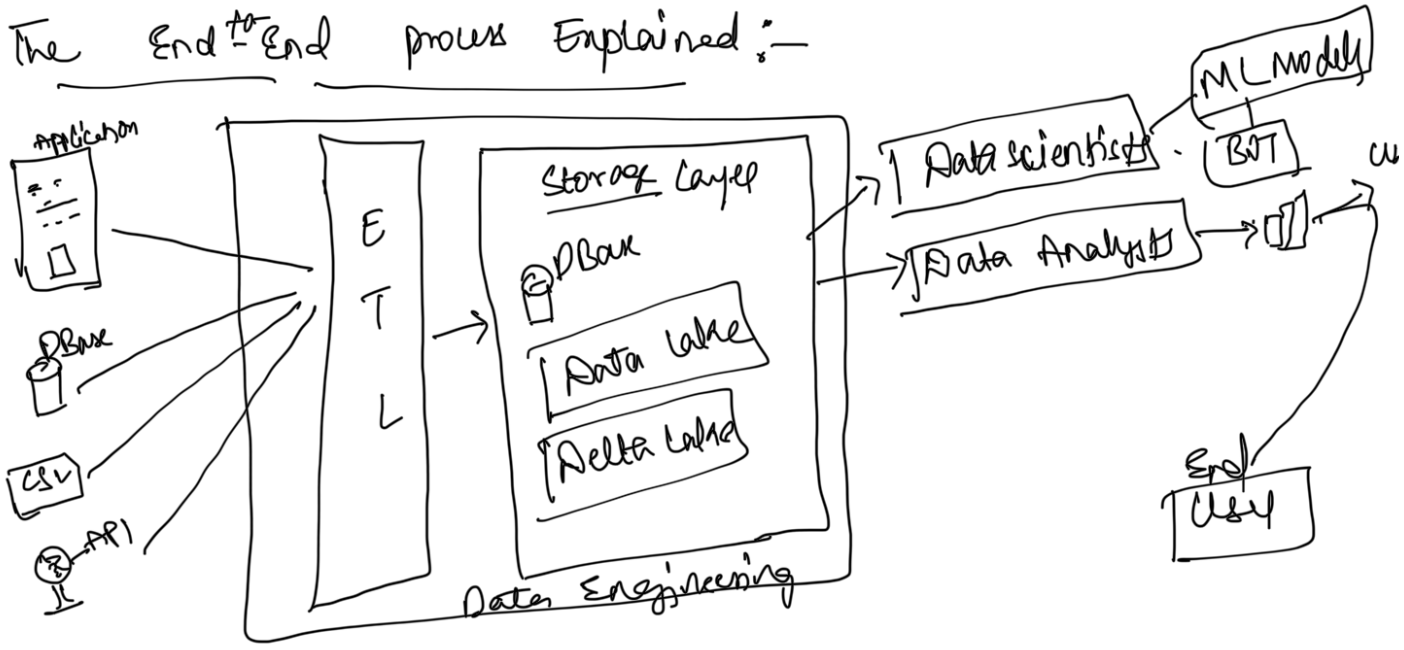
* Stake Holders Directly cannot relay on assumptions they Need Data Driven decision Making

↳ To achieve this data need to be collected, processed, Cleaned, Transformed and Stored

In Structured format.

↳ This structured data can be further used by the Business or Stake-holders in the form of Reports to get the insights about their Business.

The End to End process Explained :-



The Data Engineering Process

* Data is generated from different sources. (apps, website, transactions)
* Data is stored in DBMS
* Data Engineers process and transform the data to make it usable
* The cleaned data is then used by
  - ↳ Dashboards for Business Analysts
  - ↳ Machine learning Models for predictions
  - ↳ Data science teams for insights

Where does the Data Engineer

Where does the Data Engineering Fit?

* Bridge B/W Data collection and Data analysis.

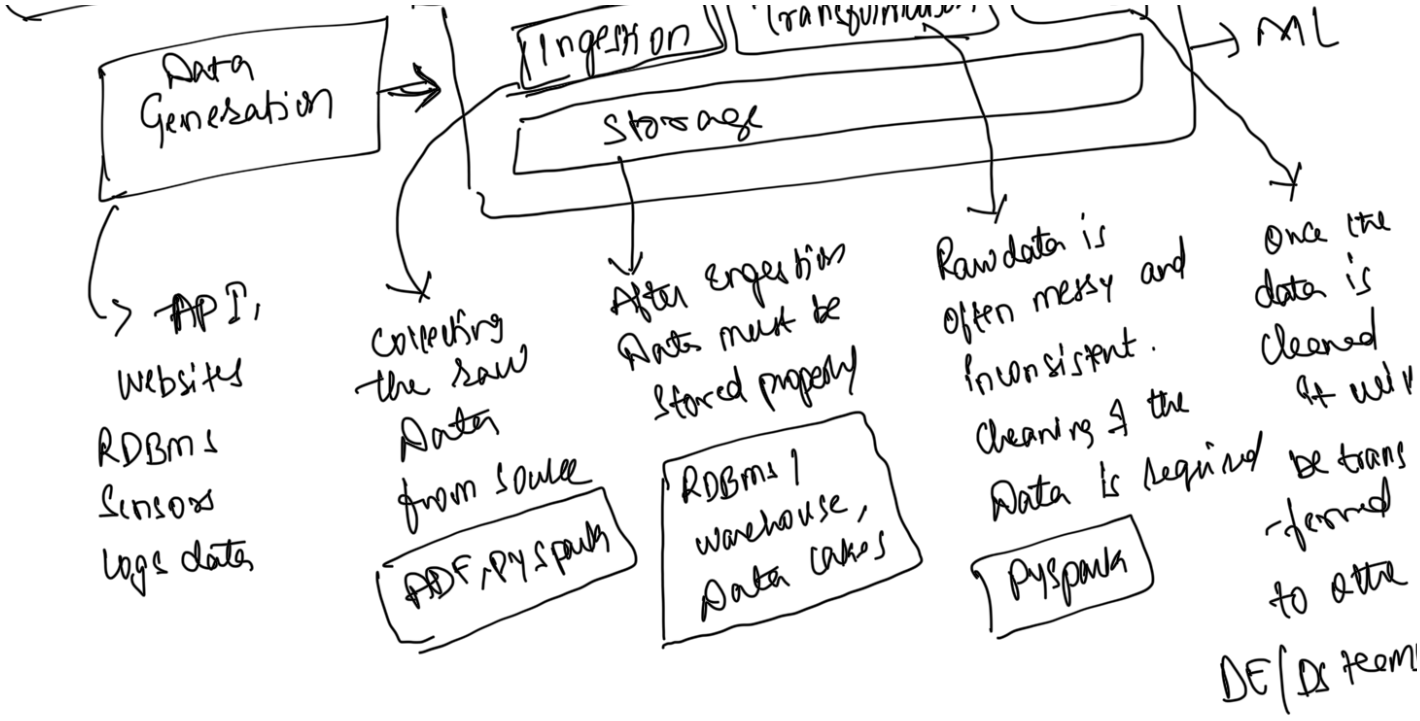
The Key Role in Engineering?

* Software Engineer → App development
* Data Base Administrators (DBA) → Data Base Management & Design
* Data Engineers → End to end pipelines to collect & Transform data.
* Data Analysts & Scientists → Build predictions & trends based on the cleaned data
* ML Engineers → Build and deploy AI models that automate decision-making.

Life cycle & Data Engineering ??

Why → Data cannot simply be taken and given to data science team
→ There must be a step by step approach to ensure data pipeline serves a meaningful purpose
→ This structured approach is called Data Engineering lifecycle

The stages of DE life cycle:-



Major work area of Data Engineering will be in

Transformation part

↳ Formatting

↳ Cleaning

↳ Aggregation

↳ Joining Data

Why Data Engineering is important ?

* without clean and structured data, business cannot make informed decisions.

* Data Engineers Ensuring that data is :-

  * Accurate

- well organized
- useful