DATA ANALYTICS AND AI

Department Industrial Vertical (DIV)
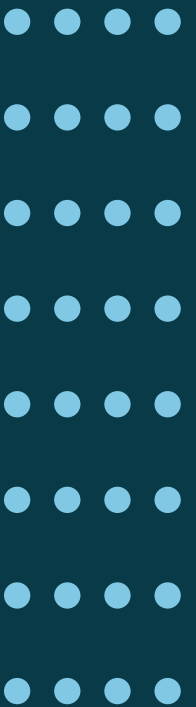
Foundation of Data Engineering

Overview

What is ?

- Data,Schema,Table

- Types of Data

  ▶ Structured Data
  ▶ Un-Structured Data
  ▶ Semi-Structured Data

- Data Life Cycle **(ETL)**

- Business Strateyg
  Key Performance Indication **(KPI)**

- Definig Data Model with
  realtime examples

- Granular level

- Types of Loads

Extract

Load

Transform
Compute

No.	Name	Age	Department	DOA	Fee	Gender
1	Piyush	24	Computer	10/01/97	120	Male
2	Shivam	21	History	24/03/98	200	Male
3	Sudha	22	Hindi	12/12/96	300	Female
4	Surya	25	History	01/07/99	400	Male
5	Rohit	22	Hindi	05/09/97	250	Male
6	Mohit	30	History	27/06/98	300	Male
7	Sneha	34	Computer	25/02/97	210	Female
8	Yash	23	Hindi	31/07/97	200	Male

**Raw Data from
different sources**

**Data are stored in
Mass storage
database like cloud**

Data Pipeline Process
(ETA)

Extract ⟶ **Extract Raw Data from Different Data source**

Transform ⟶ **Clean, format, and organize the data.**
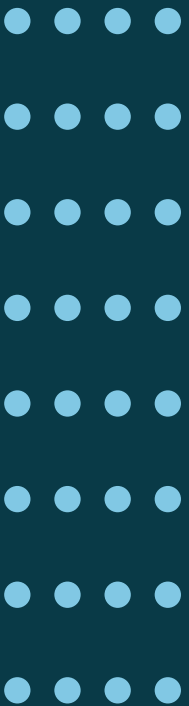
Load ⟶ **Extract Raw Data from Different Data source**

Basics of
Tools

- **ETL TOOLS**
  - ➡ SSRS, SSIS, SSAS,T-SQL

- **Spark**

- **Airflow**

- **Apache Kafka**

Additional

Information about

- **Online Transaction Process** **(OLTP)**

- **Why data no extraction during Business ours**

- **Onprem Setups**

- **Data Quality and Data Vault**

- **Medallian Achitecture**

  ▶ **(from Bronze ⇒ Silver ⇒ Gold layer tables)**

- **Dimension,facts in data cleaning**

  ▶ **Star Schema**
  Snow Schema

- **Data warehouse,Data lake,Data lake house**