**Data Engineer**

1. Linux
   * Command Line
2. Programming
   * Python
   * SQL
   * NOSQL
3. Databases RDBMS
   * SQL Server
   * Oracle
   * PostgreSQL
   * MySQL
4. Databases NOSQL any one
   * MongoDB
   * Cassandra
5. Big Data Tools
   * HDFS
   * Hive
   * Spark
6. ETL Pipeline
   * Kafka
   * Airflow
   * Spark Streaming
7. Data Processes
   * Data Collecting
   * Data Cleaning
   * Data Exploration
   * Data transformation
8. Cloud any one
   * Azure
   * AWS
   * GCP
9. Clod Platform any one
   * Debricks
   * Snowflake
10. Data Store Concepts
    * Data Wearhouse
    * Data Mart
    * Data Lake
    * Delta Lake
    * Lake houses
    * Data Mech
    * Data Catalog
11. analytics platform
    * Databricks
12. Data Visualizations any one
    * Microsoft Power BI
    * Tubule
13. **Databricks** is an analytics platform
14. **Spark** is an analytics engine for large-scale data processing
15. **Airflow** is a workflow management platform
16. **Kafka** is a distributed messaging system
17. **Hadoop** Distributed File System (HDFS) is a distributed file system designed to store large amounts of data reliably across a cluster of nodes.
18. **YARN** is a resource management framework in Hadoop ecosystem
19. **Hive** is a data warehouse software built on top of Hadoop
20. **Data lake** is a centralized repository that stores all of an organization's data in its raw format.
21. **Google Cloud Platform (GCP**) is a suite of cloud computing services
22. **MongoDB** is NoSQL databases, which means they do not use the traditional relational database model
23. **Big Query** is a fully managed, serverless, petabyte-scale analytics data warehouse that enables businesses to analyze all their data very quickly
    1. **Data warehousing**: used to store and analyze large amounts of data from a variety of sources, including transactional data, log data, and sensor data.
    2. **Business intelligence**: used to build business intelligence applications that help you to make informed decisions based on your data.
    3. **Machine learning**: used to build machine learning models that can be used to make predictions or identify patterns in your data.