Data Engineering :

Data engineers involve in preparing data for analytics or operational users. They also build data pipelines to pull all the information together from different sources.

A Data Engineer aims to make data secure and accessible for data scientists and analysts so that they can analyse it properly. Data engineers deal with raw data that often contains a lot of errors.

Data engineers use various tools and ways to improve the quality, reliability, and efficiency of data.

Roles and Responsibilities of Data Engineering

- Convert erroneous data into a usable form for further analysis.

- Create large data warehouses using ETL.

- Develop, test, and maintain architectures.

- Develop dataset processes.

- Deploy Machine Learning and statistical methods.

Let's what skills are required for Data Engineer-

1. Programming Language

Knowledge of programming language is mandatory for **data engineers**. There are various data engineering-specific programming languages like **Python, Java, and Scala**. But as you can see in the Jeff Hale analysis, the demand for Python is high as compared to Java and Scala.

That's why you should have a strong understanding of Python. Knowing Java and Scala is a plus.

2. In-Depth Database Knowledge

As a **Data Engineer**, you have to deal with data for a full day. That's why you should have in-depth knowledge of Database languages and tools. Knowledge of **SQL** is mandatory. The most demanding technology for data engineering is SQL.

3. Knowledge of Big Data Tools

Nowadays, data is increasing very fast. So to process a huge amount of data, you should be familiar with **Big Data** Tools. Most of the companies mention "Knowledge of Big Data tools" as compulsory for the Data Engineer post.

That's why you should know about these Big Data Tools-

1. **Hadoop and MapReduce.**

2. **Apache Spark**

3. **Apache Hive**

4. **Kafka**

5. **Apache Pig**

6. **Sqoop**

4. Data Warehousing and ETL Tools

As a **Data Engineer** most of the time, you need to perform ETL operations. Data warehousing is very important for managing huge amounts of data. So, knowledge of ETL tools like **Informatica** & **Talend** and Data warehousing solutions like **Redshift** or **Panoply** is highly valuable.

Informatica & **Talend** are the two well-known tools used in the industry. Informatica & Talend Open Studio are Data Integration tools with ETL architecture.

5. Data Engineering Cloud Platforms

There are various cloud or on-premise-based platforms available like- **Google Cloud Platform, AWS, Azure, and Apprenda.** You don't need to master all these tools. Even it is not mandatory to know all tools. But having a strong knowledge of at least one of them is required.

6. Familiar with Operating System

A **Data Engineer** should know the ins and outs of infrastructure components, such as virtual machines, networks, applications services, etc. That's why intimate knowledge of **UNIX, Linux,** and **Solaris** is very helpful for you.

7. Machine Learning

Knowledge of **Machine learning** is primarily considered the domain of a data scientist. But as a Data Engineer, you should have a basic understanding of machine learning algorithms.

8. Data Visualization Tools

Data Visualization is the representation of your finding with the help of **graphs, charts, or other visual formats. Tableau** and **PowerBI** are the two most popular Data Visualization tools. Knowledge of Tableau or PowerBI is a plus as a Data Engineer.

So, these are some must-have skills for **Data engineers**. Now let's see the Data Engineering Career Path.

**************************End***