**Data Engineering**

**By: Shaharyar Sabir**

**Week 1**

**Task#1**

Familiarize yourself with the basics of Data Engineering which include the following topics:

-Big Data

- Data Lake

- Database

- Data Warehouse

**Big data** is a term used to depicts extremely large amount of data set that cannot be easily managed, processed by traditional tools. As data engineering is defined; to design, build and maintain the systems for managing and processing large group of data. So, traditional tolls are not enough to manage, process and store big data. So data engineers design and built distributed processing systems that allow parallel processing of large datasets across multiple nodes in a cluster and enable processing efficiently. Data engineers also look for security and quality of big data so, they develop pipelines that ensure data quality and security by validate data and detecting errors.

**Data Lake** An important tool in data engineering that allows to store large volumes of data that may be structured, semi structured or unstructured at any scale. Used different tools to store and process data. Data is organized using metadata that facilitate user to easily search and discover data they need. So, it is scalable and flexible platform for ingesting and processing large volume of data.

**Database** a software system to store, organize and manage data. There are many types of databases used in data engineering.

**Relational Database**: Data stored in tables in rows and columns.

**NoSQL Database:** No table based data models, instead the use document base, key value, graph or column family data model. Used in large scales etc.

**Data Warehouses** stores large amount of historical data for business and analytical purposes. Support complex queries. Uses schema based approach to organize data where data is structured into tables, columns and rows. Organized around specific subject areas. Such as sales, marketing and design. Store historical data so possible to analyze trends and make changes. Once data is loaded into warehouses it cannot be updated or deleted.