Muqaddas Fatima

Dated: 16/5/23

**DATA ENGINEERING TRACK**

**Week 10, Task 1**

**Define the following terms:**

**S3**

Amazon S3 is a cloud storage service that enables you to store and retrieve various types of data, such as files, images, and videos. It offers scalability, data durability, high availability, and security features, and can be accessed programmatically via APIs.

**RDS**

Amazon RDS is a managed database service that simplifies the management of relational databases. It supports popular database engines and handles tasks like backups, software updates, and automatic scaling to ensure database availability.

**GLUE:**

AWS Glue is a managed service for data extraction, transformation, and loading (ETL). It facilitates data preparation and transformation tasks such as cataloging, cleansing, and transforming data, making it easier to build data pipelines and perform data analytics.

**Redshift:**

Amazon Redshift is a managed data warehousing service designed for analyzing large datasets. It efficiently stores and queries structured data using SQL and can handle complex analytical queries across massive amounts of data.

**IAM Role:**

IAM Roles are used to manage secure access to AWS services and resources. They define permissions and policies that determine what actions can be performed by users, applications, or services within an AWS environment.

**Crawler :**

In the context of AWS Glue, a Crawler is an automated component that scans and catalogs data from various sources, creating and updating a metadata catalog for easier data discovery and manipulation.

**Athena:**

Amazon Athena is an interactive query service that allows you to analyze data stored in Amazon S3 using SQL queries. It requires no infrastructure setup or data loading, making it convenient for ad-hoc analysis and exploration of large datasets.

**Quicksight**

Amazon QuickSight is a cloud-native business intelligence service that enables the creation of interactive dashboards and visual data analysis. It integrates with various data sources, including AWS services like S3 and Athena, to provide real-time insights.

**Lambda**

AWS Lambda is a serverless compute service that allows you to run code without the need to manage servers. It is well-suited for event-driven applications and allows code execution in response to events or on a predefined schedule.

**EC2**

Amazon EC2 is a service that provides scalable virtual servers in the cloud, allowing quick provisioning of compute capacity. It offers various instance types and configurations to cater to different workload requirements.