AWS Services

Amazon Elastic MapReduce (EMR) is a web service that provides a managed framework to run data processing frameworks such as Apache Hadoop, Apache Spark, and Presto in an easy, cost-effective, and secure manner. It is used for data analysis, web indexing, data warehousing, financial analysis, scientific simulation, etc. Amazon EMR allows to compute large number of instances to process data at any scale.

Amazon Redshift is a fully managed data warehouse service in the cloud.

The initial process to create a data warehouse is to launch a set of compute resources called nodes, which are organized into groups called cluster. After that you can process your queries.

Amazon S3 - Simple Storage Service is a scalable, high-speed, low-cost web-based service designed for online backup and archiving of data and application programs. It allows to upload, store, and download any type of files up to 5 GB in size.

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL. Athena is serverless, so there is no infrastructure to manage, and you pay only for the queries that you run.

Amazon RDS (Relational Database Service) is a fully-managed SQL database cloud service that allows to create and operate relational databases. Amazon RDS is a relational database for primary data, running software like SQL, MySQL, Aurora, MariaDB, Oracle and PostgreSQL. Can be used as your main, supporting or transactional database.

Redshift is Amazon's analytic database with ParAccel technology this is designed for heavy lifting, crunching big data queries against large datasets.

Redshift is OLAP (Analytics) while RDS is OLTP (Transactions).

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud.

Amazon IAM - https://www.slideshare.net/EdurekaIN/aws-iam-tutorial-identity-and-access-management-iam-aws-training-videos-edureka

Amazon Kinesis is a managed, scalable, cloud-based service that allows real-time processing of streaming large amount of data per second. It is designed for real-time applications and allows developers to take in any amount of data from several sources, scaling up and down that can be run on EC2 instances. For real-time graphs, real-time data analytics.

Amazon Lambda inspects actions within the application and responds by deploying the user-defined codes called functions. It automatically manages the compute resources across multiple availability zones and scales them when new actions are triggered.

Amazon Virtual Private Cloud (VPC) allows the user to use AWS resources in a virtual network.

AWS Route 53 is a highly available and scalable DNS web service.

Designed for developers and corporates to route the end users to internet applications like www.mydomain.com to some numeric IP address like 192.0.2.1 that computers use to connect to each other.

CloudFront is a CDN. It retrieves data from Amazon S3 bucket and distributed it to multiple datacenter locations.

NoSQL solutions are great for large data sets with high volumes of writes that need to be distributed over a wide geography (multiple regions or data centers),

where there is some tolerance for marginal loss of data in the event of a disaster recovery scenario.

NoSQL (Not Only SQL Database)

NoSQL is an approach to database design that can accommodate a wide variety of data models, including key-value, document, columnar and graph formats.

NoSQL, which stand for "not only SQL," is an alternative to traditional relational databases in which data is placed in tables and data schema is carefully designed before the database is built.

NoSQL databases are especially useful for working with large sets of distributed data.

Amazon DynamoDB is a fully managed NoSQL database service that allows to create database tables that can store and retrieve any amount of data.

AWS Data Pipeline is a web service, designed to make it easier for users to integrate data spread across multiple AWS services and analyze it from a single location.

Amazon machine learning (ML) is a service that allows to develop predictive applications by using algo, mathematical models based on user's data.

It reads data through Amazon S3, Redshift, RDS then visualizes the data through the AWS mgmt console and the Amazon ML API.

A workflow is a set of activities that carry out some objective, including logic that coordinates the activities to achieve the desired output.