AWS cloud services

EC2

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud computing easier for developers.

Amazon EC2 is integrated with most AWS services such as Amazon Simple Storage Service (Amazon S3), Amazon Relational Database Service (Amazon RDS), and Amazon Virtual Private Cloud (Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

EMR

Amazon EMR provides a managed Hadoop framework that makes it easy, fast, and cost-effective to process vast amounts of data across dynamically scalable Amazon EC2 instances. You can also run other popular distributed frameworks such as [Apache Spark](https://aws.amazon.com/emr/details/spark/), [HBase](https://aws.amazon.com/emr/details/hbase/), [Presto](https://aws.amazon.com/emr/details/presto/), and [Flink](https://aws.amazon.com/blogs/big-data/use-apache-flink-on-amazon-emr/) in EMR, and interact with data in other AWS data stores such as Amazon S3 and Amazon DynamoDB. EMR Notebooks, based on the popular Jupyter Notebook, provide a development and collaboration environment for ad hoc querying and exploratory analysis.

RDS

Amazon Relational Database Service (Amazon RDS). [PostgreSQL](https://aws.amazon.com/rds/postgresql/), [MySQL](https://aws.amazon.com/rds/mysql/), [MariaDB](https://aws.amazon.com/rds/mariadb/), [Oracle Database](https://aws.amazon.com/rds/oracle/), and [SQL Server](https://aws.amazon.com/rds/sqlserver/). You can use the [AWS Database Migration Service](https://aws.amazon.com/dms/) to easily migrate or replicate your existing databases to Amazon RDS.

Redshift

Massively parallel(MPP): Amazon Redshift delivers fast query performance on datasets ranging in size from gigabytes to exabytes. Redshift uses columnar storage, data compression, and zone maps to reduce the amount of I/O needed to perform queries. It uses a massively parallel processing (MPP) data warehouse architecture to parallelize and distribute SQL operations to take advantage of all available resources.

Machine Learning: For example, queries such as dashboards and reports with high concurrency requirements are routed to an express queue for immediate processing.

Result Caching: When a query executes, Redshift searches the cache to see if there is a cached result from a prior run.