

Research and write about - Big Data service in AWS

Like other Cloud service provider Amazon is also one of the most popular Cloud service providers and it also provide end to end services for Big Data Application. Amazon Web Service (AWS) provides all the necessary infrastructure and tools to run the big data projects on cloud. Some of these services are

Amazon Kinesis

It is the data streaming platform provided by AWS. Data streaming is the process of transferring and collecting the data is almost real time. With the kinesis application data and logs can be written into databases, data warehouse or data lakes. It can be easily integrated with the other AWS services and real time streaming application can be built relying on this service thus offering the possibilities of real time analytics.

AWS Lambda

AWS Lambda helps in running different types of functions or services without worrying about the server management. Users just need to upload the code and lambda will take care of the rest of the process; it is useful in developing backend services without any user involvement. The pricing model is quite fascinating too, users may for the used compute time only. The compute time for every request is calculated and users have to pay for that. It is specially used for those Streaming of AWS events and real time processing.

Amazon EMR

Amazon EMR is distributed computing framework which deal with the storage and processing of the data. It leverages opensource Apache Hadoop framework for the storage of data and its processing. User can easily create and modify the cluster attributes. EMR supports the general Hadoop tools such as Spark and hive. Some of the possible use cases with amazon is log processing, threat analysis and predictive analytics but any kind of the function can be carried out on it which is possible via Apache Hadoop.

Amazon Glue

It is Extraction, Transformation and Load (ETL) tool which is used for classification of data. It helps in migration of data from different sources and cleaning the data so that it can be used for further processes. It reduces the time and resource complexity involved with managing the ETL Jobs. Glue provides automatic data crawling services, so user do not have to worry about loading processes, generating execution code. It can be easily integrated with other AWS services such as redshift and EMR for big data applications.

There are other AWS services which are used for big data

- Amazon Machine Learning, Amazon Elastic Search Service, Amazon Redshift
- Amazon Dynamo DB, Amazon Quick search and many more..