**S3 buckets:** This globally named S3 buckets are easy way of storing & accessing data within cloud service from across the globe. It may be used to store input/output files, directories and standard/user metadata. Generally S3 buckets are universal persistent storage for all types of objects at low latency and hugh throughput performance. It is recommended that customers choose regions physically close to them in order to avoid network latency and cost. S3 usage cost varies depending upon the underlying storage class. The storage classes may be: a) S3 Standard, b) S3 Intelligent-tiering, c) S3 Standard-Infrequent Access (IA), d) S3 OneZone-IA, e) S3 Glacier, f) S3 Glacier Deep Archive.

From AWS Management Console, user can create S3 buckets within his account. User can create folders, can upload files in the buckets. User can assign various access privileges to the objects or to the folders using either IAM service or bucket policies and access control lists. An AWS customer can interact with Amazon S3 bucket using AWS Management Console or AWS Command Line Interface or AWS APIs. S3 bucket offers various features. The owner can easily enable versioning for S3 buckets (which will preserve every version of the object operated through copy or delete operation), setup server access logs or object-level API logs or setup tags and encryption.

Also there is AWS S3 Transfer Acceleration helps execute fast & secure transfers from a client to an S3 bucket via AWS Edge locations.

**Athena:** Amazon Athena service is a serverless interactive SQL-based query service which integrates easily with various types of big data repositories including S3 buckets in different formats. It does not require any infra-structure of its own. Athena helps to analyze unstructured, semi-structured or fully structured data stored in AWS S3 buckets in formats like CSV (or TSV), JSON or columnar formats such as Apache Parquet or Apache ORC. To use Athena, one does not need to write complex ETL jobs to import data anywhere because it runs directly on S3 buckets. In addition, Athena integrates well with AWS Glue Data Catalog which offers a persistent central meta-data store for the data in AWS S3. Also AWS Glue possesses various ETL options and data discovery options which indirectly may be exploited by Athena for meta-data access. Also Athena may be used to directly access service logs like AWS CloudFront, CloudTrail, CloudFormation, VPC, ELB logs etc. Inside, Athena uses Presto, a distributed SQL engine to run queries. It also uses Apache Hive to create, drop or alter tables or partitions. One can write Hive-compliant DDL statements and ANSI-SQL statements in Athena. Athena query language (or Preston) supports different library functions and various options (like aggregate functions, window functions, regular expression processing functions, lateral joins, with table selects etc.). It supports various advanced data structures like string, array, maps etc. Athena uses an approach as schema-on-read, which allows system to project the schema on data during query runtime. Using common table expression (cte) is a very common practice in Athena queries.

Because Athena is serverless and scalable, it can run queries returning data very fast. Its versatile library with array or map processing functions add more power to this tool. Using this, one can pick up text data and parse them as array of words and process them easily. Converting a string into array or array into string becomes very easy with library functions. Regular expression pattern matching is a great poAlso from any SQL client that supports JDBC driver or ODBC driver, it is possible to connect to Athena SQL engine and run queries in a session. Athena integrates very well with AWS QuickSight visual presentation service.

**QuickSight:** This AWS service provides a way to perform BI analysis by discovering any datasource and generating visual presentations on any columns. One can select the columns, filter data and choose type of visual from a very broad range of visual types (like scatterplots, hitcharts etc.). Various types of dashboards or graphs or plots can be built on connecting even instream data. There is another option **InSights** within QuickSight that is based on Machine Learning. You can use this tool to analyze this imported data using machine intelligence and report its findings. It uses random cut forest algorithm within learning methodologies and tells us which data is the anomaly part or it can give us predictions etc.  It also has auto-narrative based dashboards where it tells us its findings in plain English. It is an initial predictive tool.

**SageMaker:** This AI-based service can be used to build, train and deploy machine learning models at any scale. In addition SageMaker includes built-in modelues that may be mixed with developer’s machine learning models. It may be used to develop and train, test, run and deploy models in other systems also.