**Architecture Details**

**AWS Kinesis Data Firehose**

* Great for real-time big data, application logs, metrics, IoT, clickstreams
* Great for streaming processing frameworks (Spark, NiFi, etc…)
* Data is automatically replicated to 3 AZ
* Fully Managed Service, no administration, automatic scaling, serverless
* Load data into Redshift / Amazon S3 / ElasticSearch / Splunk
* Serverless data transformations with Lambda
* Near real time (lowest buffer time is 1 minute)
* Automated Scaling
* No data storage
* Supports many data formats, conversions, transformations, compression
* Pay for the amount of data going through Firehose
* Control access / authorization using IAM policies
* Encryption at rest using KMS

**S3 Storage and Data Management**

* Infinitely scaling storage
* Object values are the content of the body (Max Object Size is 5TB)
* You can version your files in Amazon S3
* Encryption in transit (SSL/TLS)

**AWS Glue**

* Fully-managed ETL (Extract, Transform & Load) service
* Automating time consuming steps of data preparation for analytics
* Serverless, pay as you go, fully managed, provisions Apache Spark and ML models
* Crawls data sources and identifies data formats (schema inference)
* Sources: Amazon Aurora, RDS, Redshift & S3
* Sinks: S3, Redshift, etc…
* Glue Data Catalog: Metadata (definition & schema) of the Source Tables

**AWS Athena**

* Serverless service to perform analytics directly against S3 files
* Uses SQL language to query the files
* Has a JDBC / ODBC driver
* Charged per query and amount of data scanned
* Supports CSV, JSON, ORC, Avro, and Parquet (built on Presto)
* Use cases: Business intelligence / analytics / reporting

**Amazon Aurora Serverless**

* Aurora is a proprietary technology from AWS (not open sourced)
* Aurora is AWS cloud optimized and claims 5x performance improvement
* over MySQL on RDS, over 3x the performance of Postgres on RDS
* Aurora storage automatically grows in increments of 10GB, up to 64 TB.
* Aurora can have 15 replicas while MySQL has 5, and the replication process

is faster (sub 10 ms replica lag)

* Failover in Aurora is instantaneous. It’s HA (High Availability) native.
* Support for Cross Region Replication
* instantiation and autoscaling based on actual usage
* Good for infrequent, intermittent or unpredictable workloads
* No capacity planning needed
* Pay per second, can be more cost-effective
* Automatic fail-over
* Backup and Recovery
* Isolation and security
* Encryption at rest using KMS
* Automated backups, snapshots and replicas are also encrypted
* Encryption in flight using SSL (same process as MySQL or Postgres)
* Possibility to authenticate using IAM token

**Amazon QuickSight**

* Pay only for what you use
* Scale to tens of thousands of users
* Leverage Amazon SageMaker models
* Build end-to-end BI solutions
* Quicksight mobile