

RDS Overview

- Managed PostgreSQL / MySQL / Oracle / SQL Server
 - Must provision an EC2 instance & EBS Volume type and size
 - Support for Read Replicas and Multi AZ
 - Security through IAM, Security Groups, KMS , SSL in transit
 - Backup / Snapshot / Point in time restore feature
 - Managed and Scheduled maintenance
 - Monitoring through CloudWatch
-
- **Use case:** Store relational datasets (RDBMS / OLTP), perform SQL queries
transactional inserts / update / delete is available

RDS for Solutions Architect

- **Operations:** small downtime when failover happens, when maintenance happens, scaling in read replicas / ec2 instance / restore EBS implies manual intervention, application changes
- **Security:** AWS responsible for OS security, we are responsible for setting up KMS, security groups, IAM policies, authorizing users in DB, using S
- **Reliability:** Multi AZ feature, failover in case of failures
- **Performance:** depends on EC2 instance type, EBS volume type, ability to add Read Replicas. Storage auto-scaling & manual scaling of instances
- **Cost:** Pay per hour based on provisioned EC2 and EBS

Aurora Overview



- Compatible API for PostgreSQL / MySQL
- Data is held in 6 replicas, across 3 AZ
- Auto healing capability
- Multi AZ, Auto Scaling Read Replicas
- Read Replicas can be Global
- Aurora database can be Global for DR or latency purposes
- Auto scaling of storage from 10GB to 128 TB
- Define EC2 instance type for aurora instances
- Same security / monitoring / maintenance features as RDS
- **Aurora Serverless** – for unpredictable / intermittent workloads
- **Aurora Multi-Master** – for continuous writes failover
- **Use case:** same as RDS, but with less maintenance / more flexibility / more performance

Aurora for Solutions Architect

- **Operations:** less operations, auto scaling storage
- **Security:** AWS responsible for OS security, we are responsible for setting up KMS, security groups, IAM policies, authorizing users in DB, using S
- **Reliability:** Multi AZ, highly available, possibly more than RDS, Aurora Serverless option, Aurora Multi-Master option
- **Performance:** 5x performance (according to AWS) due to architectural optimizations. Up to 15 Read Replicas (only 5 for RDS)
- **Cost:** Pay per hour based on EC2 and storage usage. Possibly lower costs compared to Enterprise grade databases such as Oracle

ElastiCache Overview

- Managed Redis / Memcached (similar offering as RDS, but for caches)
 - In-memory data store, sub-millisecond latency
 - Must provision an EC2 instance type
 - Support for Clustering (Redis) and Multi AZ, Read Replicas (sharding)
 - Security through IAM, Security Groups, KMS, Redis Auth
 - Backup / Snapshot / Point in time restore feature
 - Managed and Scheduled maintenance
 - Monitoring through CloudWatch
-
- **Use Case:** Key/Value store, Frequent reads, less writes, cache results for D queries, store session data for websites, cannot use SQL.

ElastiCache for Solutions Architect

- **Operations:** same as RDS
- **Security:** AWS responsible for OS security, we are responsible for setting up KMS, security groups, IAM policies, users (Redis Auth), using SSL
- **Reliability:** Clustering, Multi AZ
- **Performance:** Sub-millisecond performance, in memory, read replicas, sharding, very popular cache option
- **Cost:** Pay per hour based on EC2 and storage usage

DynamoDB Overview

- AWS proprietary technology, managed NoSQL database
- Serverless, provisioned capacity, auto scaling, on demand capacity (Nov 2018)
- Can replace ElastiCache as a key/value store (storing session data for example)
- Highly Available, Multi AZ by default, Read and Writes are decoupled, DAX for read cache
- Reads can be eventually consistent or strongly consistent
- Security, authentication and authorization is done through IAM
- DynamoDB Streams to integrate with AWS Lambda
- Backup / Restore feature, Global Table feature
- Monitoring through CloudWatch
- Can only query on primary key, sort key, or indexes
- **Use Case:** Serverless applications development (small documents 100s KB), distributed serverless cache, doesn't have SQL query language available, has transactions capability from Nov 2018

DynamoDB for Solutions Architect

- **Operations:** no operations needed, auto scaling capability, serverless
- **Security:** full security through IAM policies, KMS encryption, SSL in flight
- **Reliability:** Multi AZ, Backups
- **Performance:** single digit millisecond performance, DAX for caching reads, performance doesn't degrade if your application scales
- **Cost:** Pay per provisioned capacity and storage usage (no need to guess in advance any capacity – can use auto scaling)

S3 Overview

- S3 is a... key / value store for objects
- Great for big objects, not so great for small objects
- Serverless, scales infinitely, max object size is 5 TB
- Strong consistency
- Tiers: S3 Standard, S3 IA, S3 One Zone IA, Glacier for backups
- Features: Versioning, Encryption, Cross Region Replication, etc...
- Security: IAM, Bucket Policies, ACL
- Encryption: SSE-S3, SSE-KMS, SSE-C, client side encryption, SSL in transit
- **Use Case:** static files, key value store for big files, website hosting

S3 for Solutions Architect

- **Operations:** no operations needed
- **Security:** IAM, Bucket Policies, ACL, Encryption (Server/Client), SSL
- **Reliability:** 99.999999999% durability / 99.99% availability, Multi AZ, C
- **Performance:** scales to thousands of read / writes per second, transfer acceleration / multi-part for big files
- **Cost:** pay per storage usage, network cost, requests number

Athena Overview



- Fully Serverless database with SQL capabilities
 - Used to query data in S3
 - Pay per query
 - Output results back to S3
 - Secured through IAM
-
- **Use Case:** one time SQL queries, serverless queries on S3, log analytics

Athena for Solutions Architect

- **Operations:** no operations needed, serverless
- **Security:** IAM + S3 security
- **Reliability:** managed service, uses Presto engine, highly available
- **Performance:** queries scale based on data size
- **Cost:** pay per query / per TB of data scanned, serverless

Redshift Overview

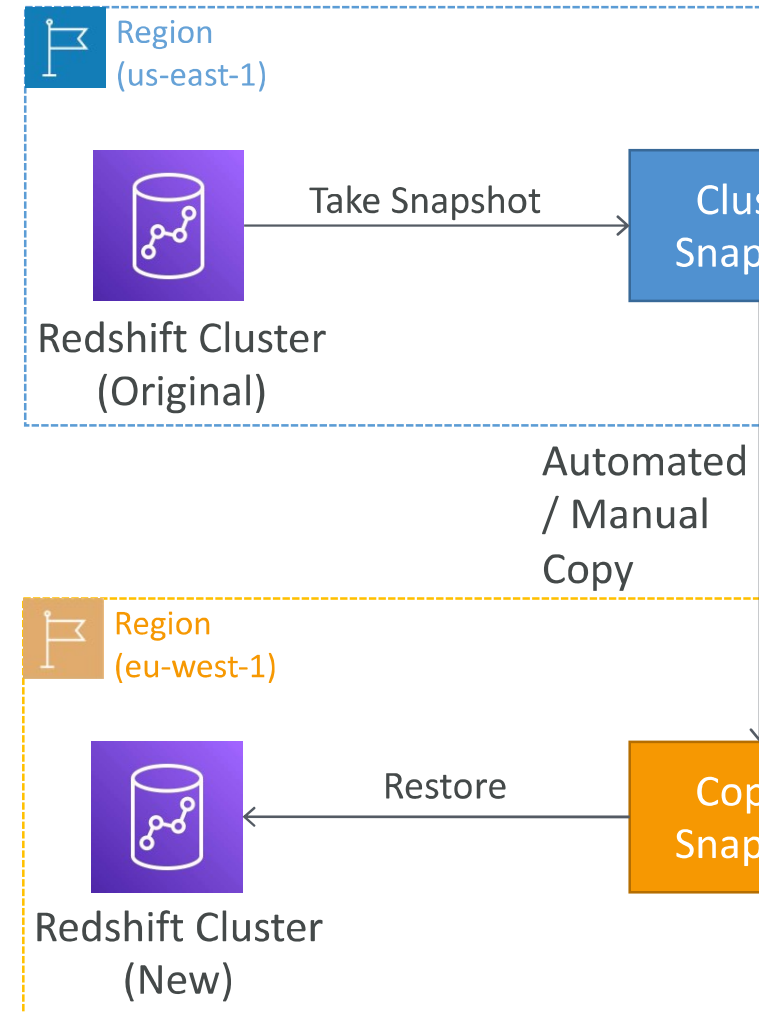
- Redshift is based on PostgreSQL, but it's not used for OLTP
- It's OLAP – online analytical processing (analytics and data warehouse)
- 10x better performance than other data warehouses, scale to PBs of data
- **Columnar** storage of data (instead of row based)
- Massively Parallel Query Execution (MPP)
- Pay as you go based on the instances provisioned
- Has a SQL interface for performing the queries
- BI tools such as AWS Quicksight or Tableau integrate with it

Redshift Continued...

- Data is loaded from S3, DynamoDB, DMS, other DBs...
- From 1 node to 128 nodes, up to 128 TB of space per node
- Leader node: for query planning, results aggregation
- Compute node: for performing the queries, send results to leader
- Redshift Spectrum: perform queries directly against S3 (no need to load data)
- Backup & Restore, Security VPC / IAM / KMS, Monitoring
- Redshift Enhanced VPC Routing: COPY / UNLOAD goes through VPC

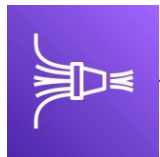
Redshift – Snapshots & DR

- Redshift has no “Multi-AZ” mode
- Snapshots are point-in-time backups of a cluster, stored internally in S3
- Snapshots are incremental (only what has changed is saved)
- You can restore a snapshot into a **new cluster**
- Automated: every 8 hours, every 5 GB, or on a schedule. Set retention
- Manual: snapshot is retained until you delete it
- You can configure Amazon Redshift to automatically copy snapshots (automated or manual) of a cluster to another AWS Region

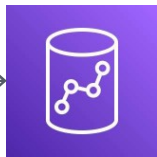


Loading data into Redshift

Amazon Kinesis Data Firehose



Amazon Kinesis
Data Firehose



Amazon Redshift
Cluster
(through S3 copy)

S3 using COPY command



S3 Bucket
(mybucket)



Internet

Without Enhanced VPC Routing

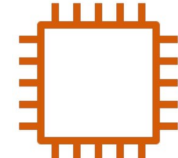
Through VPC

With Enhanced VPC Routing



Amazon Redshift
Cluster

EC2 Instance JDBC driver



EC2 Instance

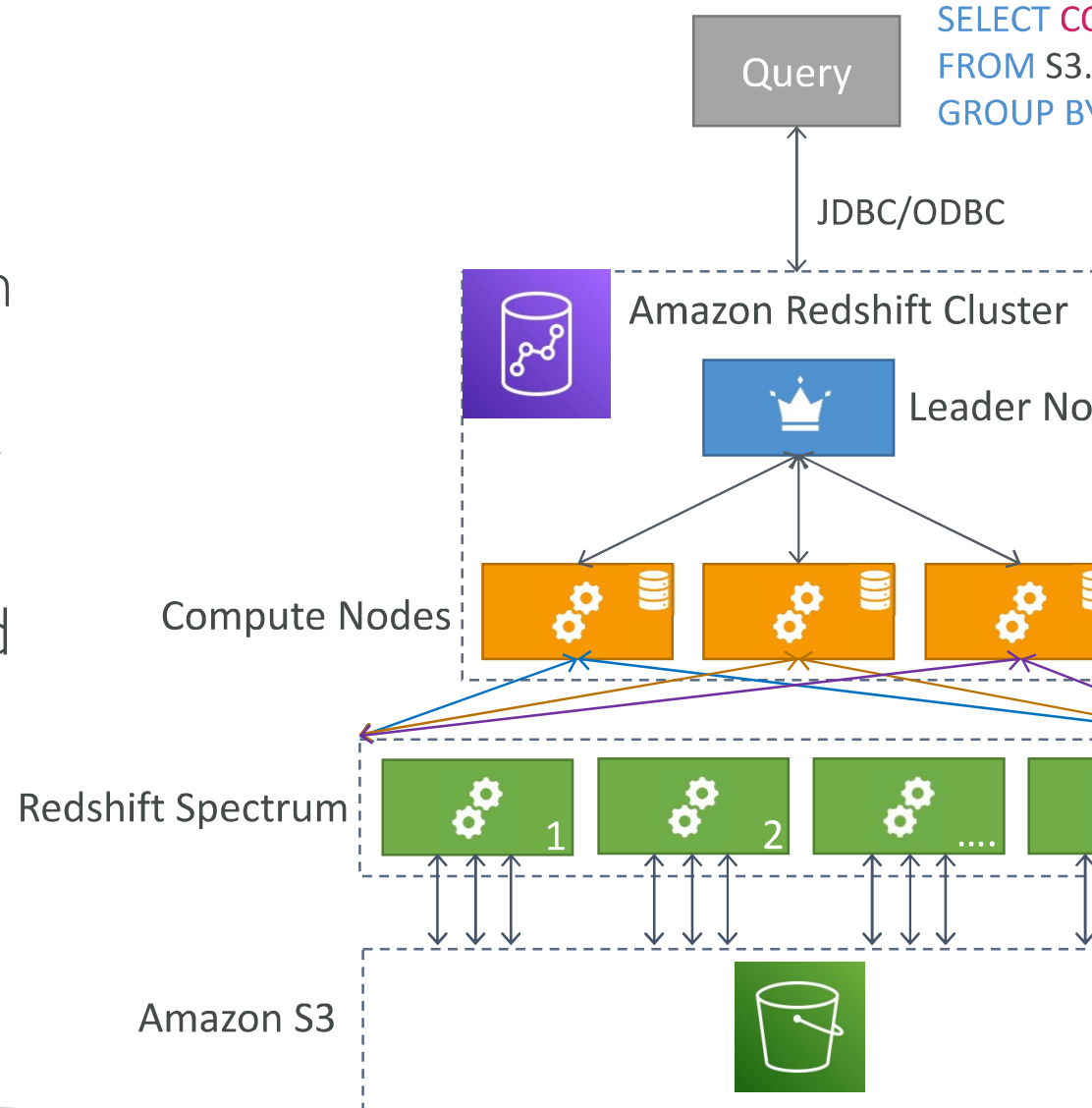
Amazon Redshift

*Better to write
Data in batches*

```
copy customer
from 's3://mybucket/mydata'
iam_role 'arn:aws:iam::0123456789012:role/MyRedshiftRole';
```


Redshift Spectrum

- Query data that is already in S3 without loading it
- Must have a Redshift cluster available to start the query
- The query is then submitted to thousands of Redshift Spectrum nodes

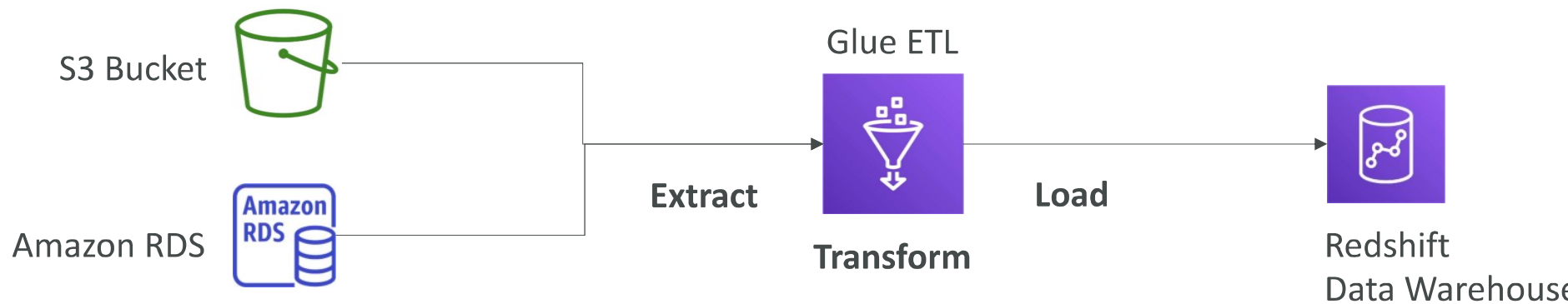


Redshift for Solutions Architect

- **Operations:** like RDS
- **Security:** IAM, VPC, KMS, SSL (like RDS)
- **Reliability:** auto healing features, cross-region snapshot copy
- **Performance:** 10x performance vs other data warehousing, compression
- **Cost:** pay per node provisioned, 1/10th of the cost vs other warehouses
- **vs Athena:** faster queries / joins / aggregations thanks to indexes
- **Remember:** Redshift = Analytics / BI / Data Warehouse

AWS Glue

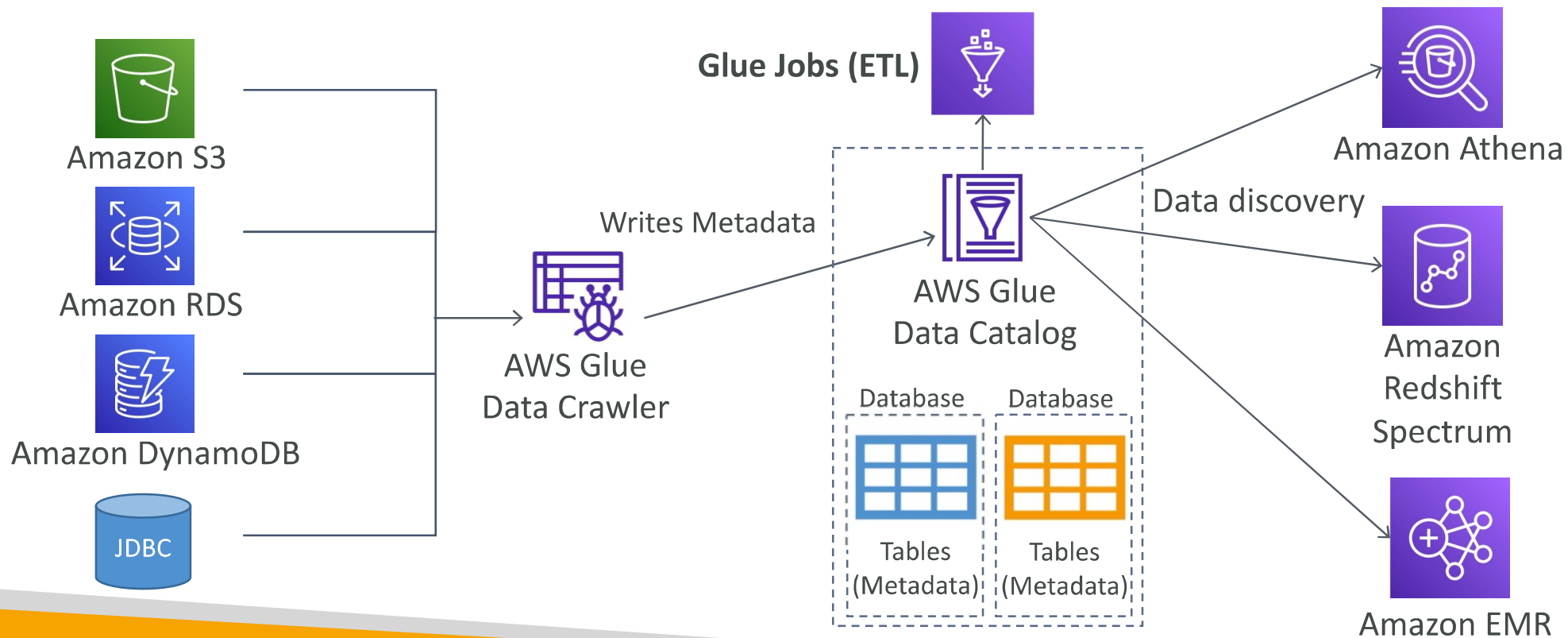
- Managed **extract, transform, and load (ETL)** service
- Useful to prepare and transform data for analytics
- Fully **serverless** service



Glue Data Catalog

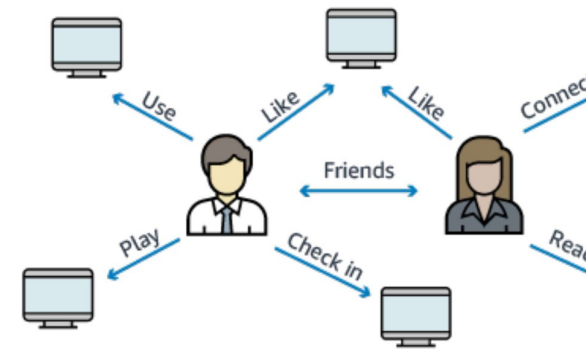


- Glue Data Catalog: catalog of datasets



Neptune

- Fully managed **graph** database
- **When do we use Graphs?**
 - High relationship data
 - Social Networking: Users friends with Users, replied to comment on post of user and likes other comments.
 - Knowledge graphs (Wikipedia)
- Highly available across 3 AZ, with up to 15 read replicas
- Point-in-time recovery, continuous backup to Amazon S3
- Support for KMS encryption at rest + HTTPS



Neptune for Solutions Architect

- **Operations:** similar to RDS
 - **Security:** IAM, VPC, KMS, SSL (similar to RDS) + IAM Authentication
 - **Reliability:** Multi-AZ, clustering
 - **Performance:** best suited for graphs, clustering to improve performance
 - **Cost:** pay per node provisioned (similar to RDS)
-
- Remember: Neptune = Graphs

ElasticSearch

- Example: In DynamoDB, you can only find by primary key or indexes
- With ElasticSearch, you can **search any field**, even partially matches
- It's common to use ElasticSearch as a complement to another database
- ElasticSearch also has some usage for Big Data applications
- You can provision a cluster of instances
- Built-in integrations: Amazon Kinesis Data Firehose, AWS IoT, and Amazon CloudWatch Logs for data ingestion
- Security through Cognito & IAM, KMS encryption, SSL & VPC
- Comes with Kibana (visualization) & Logstash (log ingestion) – ELK stack

ElasticSearch for Solutions Architect

- **Operations:** similar to RDS
 - **Security:** Cognito, IAM, VPC, KMS, SSL
 - **Reliability:** Multi-AZ, clustering
 - **Performance:** based on ElasticSearch project (open source), petabyte
 - **Cost:** pay per node provisioned (similar to RDS)
-
- **Remember:** ElasticSearch = Search / Indexing

AWS Monitoring, Audit and Performance

CloudWatch, CloudTrail & AWS Config