AWS Database Services

Thursday, February 8, 2018 7:14 PM

Relational Databases:

- Looks like Excel spreadsheets with links between them
- Can use SQL language to perform queries / lookups

Non-SQL Databases:

- NoSQL → Non Relational database
- Built for specific data models and have flexible schemas
- Benefits are flexibility, scalability, high performance, highly functional, key-value, JSON doc, graph etc.

AWS RDS: Relational Database

- Managed DB service by AWS
- Allows to run various database like Postgres, MySQL, Oracle etc.

Advantages:

- Automated provisioning, OS patching
- Continuous backups
- Monitoring dashboards
- Multi AZ setup for disaster recovery
- Maintenance windows for upgrades
- Scaling capability
- Storage backed by EBS

- But you can't SSH into your instance

Amazon Aurora:

- DB service from AWS (not open sourced)
- Postgres & MySQL - supported as Aurora DB.
- 3X & 5X AWS Cloud Optimized - 5X perf over MySQL on RDS and 3X over Postgres on RDS.
- Aurora storage grows in increment of 10GB, upto 128GB
- Costs 20% more than RDS but worth on the free tier

Amazon Aurora Serverless:

- Automated database instantiation & auto scaling based on actual usage
- Compatible DBs → Postgres & MySQL
- No capacity planning needed
- Cost management overhead
- Pay per second, more cost effective

RDS Deployments:

Read Replicas:

- Create upto 15 read replicas of RDS
- As it can distribute the workload of incoming queries across DB
- Writes can be done only to main DB

Multi AZ:

- Failover in case of AZ outage.
- Data is only read/written to main DB.
- Only failover AZ DB can be created

Multi Region (Read Replicas)

- Read Replicas created in different regions for disaster recovery

AMAZON ELASTICACHE:

- Elasticache is to get managed Redis or Memcached
- Caches are in-memory DBs with high performance, low latency
- Helps in lightening the load off DB for read intensive workloads
- AWS managed DB.

- It stores cache and reduces the load off the main RDS

DYNAMO DB:

- Fully managed highly available NoSQL database. - Serverless
- Fast & Consistent in performance
- Low latency Retrieval
- Integrated with IAM for security, auditing & administration
- Low cost and auto scaling capabilities
- Standard & IA Table Class

DYNAMO DB - ACCELERATOR (DAX)

- Fully managed in-memory cache for Dynamo DB
- 10x performance improvement
- DAX is only used for & integrated with Dynamo DB but Elastic Cache can be used for other databases

DYNAMODB - GLOBAL TABLES:

- Make a dynamo DB table accessible with low latency in multiple regions
- Active-Active application / read/write to any AWS Region.

REDSHIFT:

- Based on Postgres - but SQL not used for OLTP (Online Transaction Processing)
- OLAP (Online Analytical Processing) - Analytics & Data Warehousing
- Petabyte-scale storage of data.
- Massively parallel query execution
- ISO 27001 such as SOC1 Type2 & PCI
- Tableau integrated with Redshift

Redshift Spectrum:

- Automatically scales & provisions data warehouse.
- Run analytics workloads w/o managing data warehouse infrastructure
- Pay only for what you use

AWS EMR: (Elastic MapReduce)

- Creates Hadoop clusters (Big data) to analyze and process vast amount of data
- Clusters can be made of hundreds of EC2 instances
- Also supports Apache Spark, Hbase, Presto, Flink etc.
- Data security integrated with IAM instances
- Use cases: ML, Data processing etc.

AMAZON ATHENA:

- Serverless query service to perform analytics against S3 objects
- Uses standard SQL language
- Pricing $5 per TB of data scanned

AMAZON QUICKSIGHT:

- Serverless BI powered by service to create interactive dashboards
- Fast & automatically available
- Pay service pricing
- Integrated with RDS, Aurora, Athena, Redshift, S3

DOCUMENT DB:

- Like Aurora → AWS implementation for Postgres DB / MongoDB
- DocumentDB is the same for Mongo (MongoDB Database)
- MongoDB → store, query & index JSON
- Similar development concept as Aurora
- Fully managed & highly available.

AMAZON NEPTUNE:

- Fully managed Graph database
- A popular graph database → Neo4j database

AMAZON TIMESTREAM:

- Fully managed, fast & scalable, serverless time series database
- Automatically tiered up/down
- Built in time series analytics functions
- Now time series & JSON data of RDS

AMAZON QLDB:

- QLDB - Quantum Ledger Database
- A ledger is a book recording financial transactions
- Used to review the history of all the changes made
- Immutable system: no entry can be removed/modified - cryptographically verifiable.
- Difference with Amazon Managed Blockchain, QLDB is centralized ledger

AMAZON MANAGED BLOCKCHAIN:

- A Blockchain makes it possible to build apps where multiple parties make transactions only need for a trusted central authority.
- AWS is a managed service for:
  - Join public blockchain networks
  - Create your own network.
- Compatible with frameworks Hyperledger Fabric & Ethereum

AWS DMS:

- Managed ETL Service
- Fully serverless service
- Used for migrations & transform analytics

SMS:

- SMS - Database Migration Service.
- Quickly & securely migrate DBs to AWS
- The Source database remains available during migration
- Supports:
  - Homogeneous migration: get Oracle to Oracle
  - Heterogeneous migration: get MySQL to Aurora

SUMMARY:

- Relational Databases → RDS + EC2 & Aurora (MySQL)
- In-memory Database → Elasticache
- Key/value Database → Dynamodb & DAX
- Warehouse → OLAP → Redshift (SQL)
- Hadoop Clusters → EMR
- Analytics → Query data on Amazon S3 → Quicksight → Dashboards
- Document DB → Mongo DB (SQL-like)
- Amazon RDS → Postgres
- Amazon Managed Blockchain → Hyperledger Fabric & Ethereum
- Graph → Neo4j
- DMS → Database Migration System
- Neptune → Graph Database
- Timestream → Time series database.