

**Snow Family:** import data onto S3 through a physical device, edge computing.  
**OpstHub:** Desktop application to manage Snow family devices.

**Storage Gateway:** Hybrid solution to extend on-premises storage to S3.

### Databases:

#### Relational Databases:

Like Excel spreadsheets

SQL lang. to perform queries

#### Nosql databases:

non-SQL = non-relational dB.

Built for specific data models, flexible schemas for building modern applications.

Benefits: Flexibility, Scalability, High-performance, Highly functional

Eg: key-value, document, graph, in-memory, search dBs.

Eg: JSON, JavaScript Object Notation.

#### Benefits of databases on AWS:

Quick provisioning, High availability, Vertical & Horizontal scaling.

Automated backup & restore, operations and upgrades.

Operating system patching is handled by AWS  
Monitoring, ~~after~~ alerting.

#### AWS RDS:

RDS = Relational Database Service.

SQL as a query language

Create DBs in cloud that are managed by AWS:

- Postgres
- MySQL
- MariaDB
- Oracle
- Microsoft SQL Server
- Aurora.

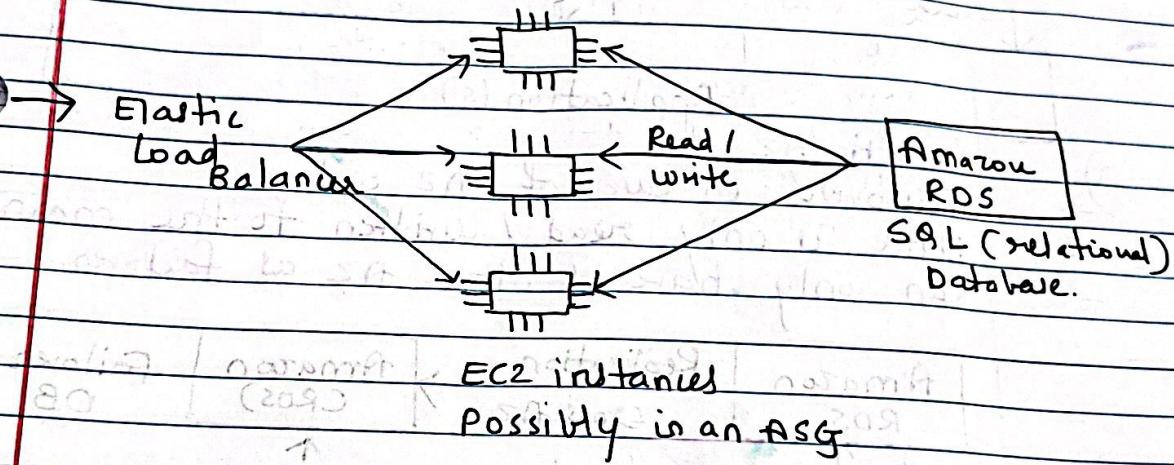
through a  
Snow family  
extend  
chmas  
rmane,  
al

→ Advantage of using RDS versus deploying DB on EC2:

- RDS is a managed service:
  - 1) Automated provisioning, patched service
  - 2) Point in Time restore
  - 3) Monitoring dashboard
  - 4) Read replicas for improved read performance
  - 5) Multi AZ setup for disaster recovery
  - 6) Maintenance windows for upgrades
  - 7) Scaling capability
  - 8) Storing backend by EBS

BUT you can't SSH into your instances.

→ RDS Solution Architecture.



→ Amazon Aurora:

Proprietary technology  
Not open sourced

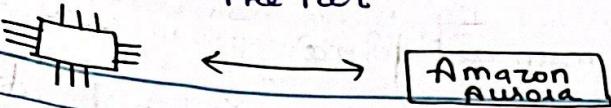
PostgreSQL, MySQL both supported as Aurora DB  
Cloud optimised

5x performance improvement over MySQL on RDS.

3x performance of PostgreSQL on RDS

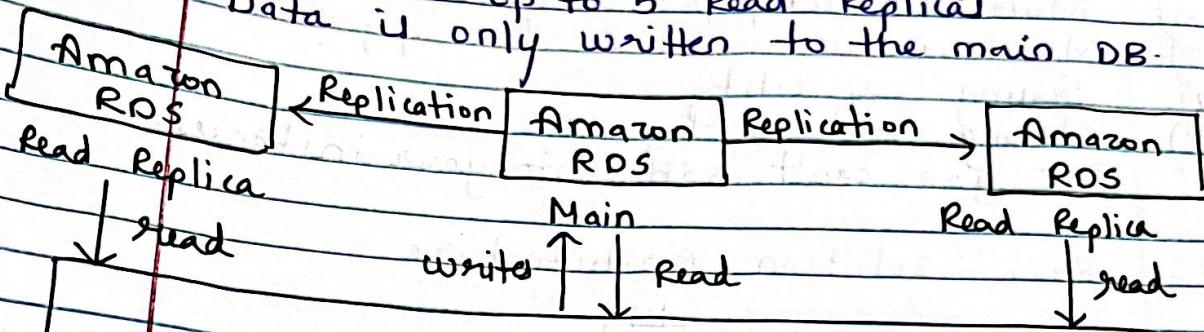
Storage: Grows in increments of 10GB upto 64TB  
Costs more than RDS (20% more)

But More efficient  
in the free tier.



→ 1) RDS Deployments: Read Replicas, multi-AZ  
Read Replicas:

Scale the read workload of your DB  
Can create up to 5 Read Replica  
Data is only written to the main DB.

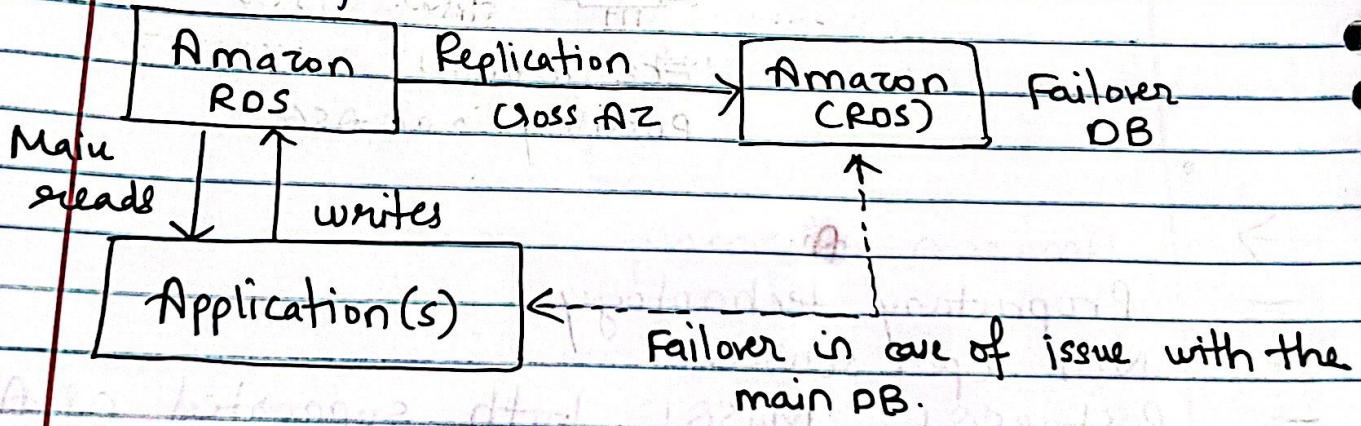


2)

Multi-AZ:

Failover in case of AZ outage

Data is only read / written to the main DB.  
Can only have 1 other AZ as failover.



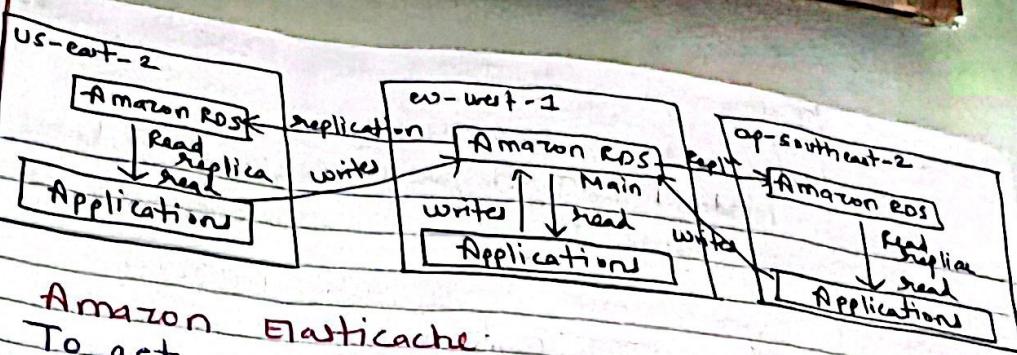
→ RDS Deployments: Multi-Region

Multi-region (Read Replicas)

Disaster recovery

Local performance

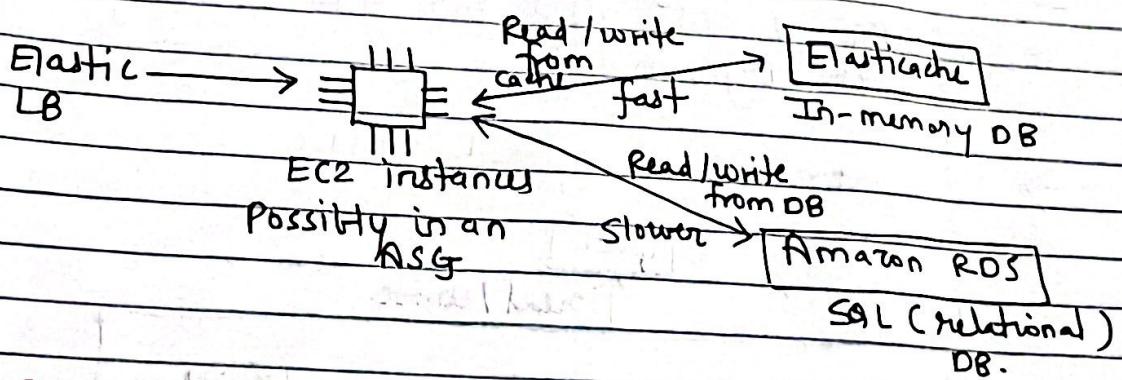
Replication cost



### Amazon ElastiCache:

To get managed Redis or Memcached In-memory database with high performance, low latency. Helps reduce load off dBs for read intensive workloads. AWS takes care of OS maintenance/patching, optimizations, setup, configuration, monitoring, failure recovery and backups.

### ElastiCache solution architecture



### DynamoDB:

Fully managed, highly available. Replication across 3 AZ.

NoSQL DB

Scales to massive workloads, distributed "sharded" DB

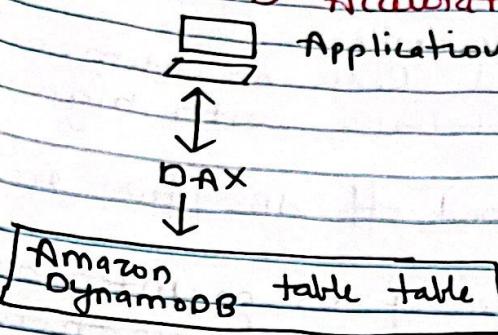
Millions of requests per seconds, trillions of rows, 100s of TB of storage

Fast & consistent

low latency retrieval

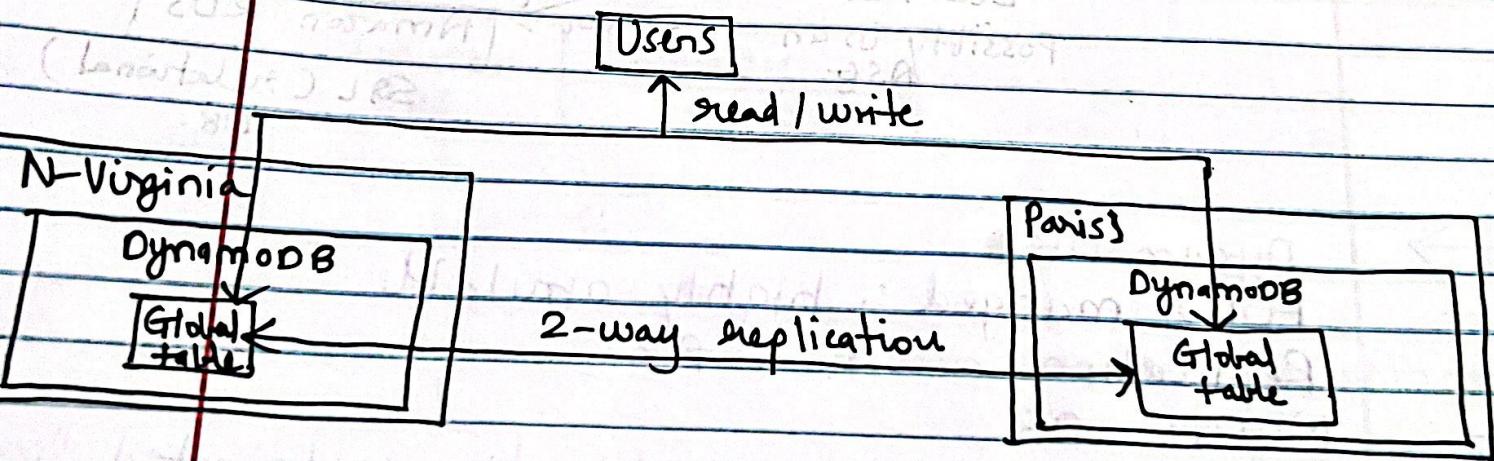
Integrated with IAM for security, authorization, administration  
Low cost, auto-scaling  
Standard & Infrequent Access (IA) table class  
key/Value db.

### DynamoDB Accelerator (DAX)



In-memory cache  
10x performance improvement  
Secure, highly scalable & highly available  
DAX is only used for and is integrated with DynamoDB.

### DynamoDB - Global Tables.



### Redshift

PostgreSQL

Not used for OLTP

Its OLAP - Online Analytical Processing (Analytics and data warehousing).  
Load data once every hour; not sec

- Scale better performance than others  
- Scale to PBs of data  
- Columnar storage of data  
- Massively Parallel & very Execution (MPP)

- Highly available

- Pay as you go  
- SQL interface

- BI tools - AWS Quicksight, Tableau.

### Amazon EMR

Elastic MapReduce

Helps creating Hadoop Clusters (Big Data) to analyze and process vast amount of data

Hadoop Clusters - Hundreds of EC2 instances

Supports Apache spark, HBase, Presto, Flink ..

Takes care of all provisioning & configuration

Auto-scaling

Integrated with spot instances

Use cases: Data Processing, ML, web-indexing, big data ..

### Amazon Athena

Serverless query service to analyze data

Uses SQL

Supports CSV, JSON, ORC, Avro, Parquet

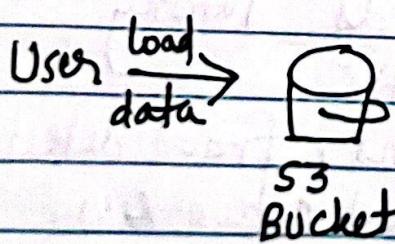
\$5 per TB of data scanned

Use compressed or columnar data for cost-savings

Use cases: BI / Analytics / Reporting / analyze & query VPC flow logs, ELB logs, Cloud Trail trails

Analyze data in S3 using serverless SQL, we

Athena



Amazon  
Athena

Reporting  
&  
dashboards | Amazon  
Quicksight

Amazon Quicksight:  
Serverless ML powered BI service to create  
interactive dashboards

Fast, automatically scalable, embeddable, with  
per-session pricing.  
Use cases:  
1) Business Analytics  
2) Building visualizations  
3) Performing ad-hoc analysis  
4) Get business insights using data  
Integrated with RDS, Aurora, Athena, Redshift,  
S3.

→

DocumentDB

AWS implementation of MongoDB (noSQL)

MongoDB - used to store, query, and index  
JSON data

Similar 'deployment concepts' as Aurora

Replication across 3 AZ

Automatically grows in increments of 10GB,  
upto 64TB

Automatically scales to workloads with  
millions of requests per seconds.

→

Amazon Neptune

Graph database

Eq: Social network dataset

Available across 3 AZ, upto 15 read replicas

Complex & hand queries

Can store upto billions of relations and query  
the graph with milliseconds latency

Replication across multiple AZ

Great for knowledge graphs, fraud detection,  
recommendation engines, social networking

- 
- Amazon QLDB
    - Quantum Ledger DB
    - Recording financial transactions
    - Fully managed, serverless, high available, replication across 3 AZ
    - Used to review history of all the changes made to your application data over time.
    - Immutabe system
    - 2-3x better performance than common ledger blockchain frameworks, manipulate data using SQL
    - Difference with Amazon Managed Blockchain - no decentralization component, in accordance with financial regulation rules.

→

Amazon Managed Blockchain

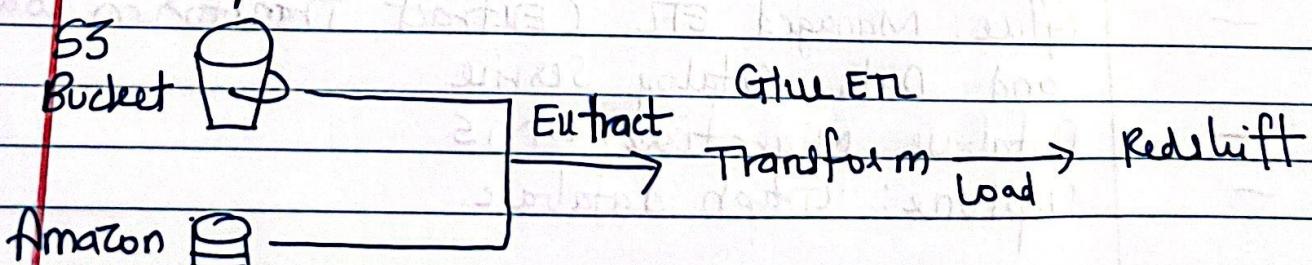
Blockchain makes it possible to build applications where multiple parties can execute transactions without the need for a trusted, central authority.

Amazon Managed Blockchain is a managed service to:

- 1) Join public blockchain networks
- 2) Or create your own scalable private network compatible with the frameworks Hyperledger Fabric & Ethereum.

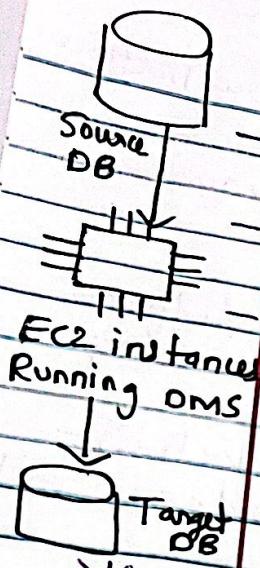
→ AWS Glue

- Managed extract, Transform, Load (ETL) service
- Prepare and transform data for analytics.
- Fully serverless service.



- Glue Data Catalog - catalog of datasets
  - ↳ Can be used by Athena, Redshift, EMR

## Data Migration service



Quickly and securely migrate DBs to AWS, resilient, self-healing. The source database remains available during the migration.

Supports:

- 1) Homogeneous migrations: ex Oracle to Oracle
- 2) Heterogeneous migrations: ex Microsoft SQL Server to Aurora.

Databases & Analytics Summary in AWS \*

Relational Databases: - OLTP - RDS, Aurora(SQ)

Differences between Multi-AZ, Read Replicas, Multi-Region:

In-memory DB: ElastiCache

Key/Value DB: DynamoDB (serverless) & DAX (cache for DynamoDB)

Warehouse - OLAP: Redshift (SQL)

Hadoop cluster: EMR

Attena: Query data on Amazon S3 (serverless & SQL)

Quicksight: Dashboard on your data (serverless)

Document DB: Aurora for MongoDB (JSON - NoSQL)

Amazon GLDB: Financial Transactions ledger

Immutable journal, cryptographically verifiable

Amazon managed Blockchain: Managed Hyper-ledger Fabric & Ethereum blockchain.

Glue: Managed ETL (Extract Transform Load) and Data Catalog Service

Database Migration: DMS

Neptune: Graph database.