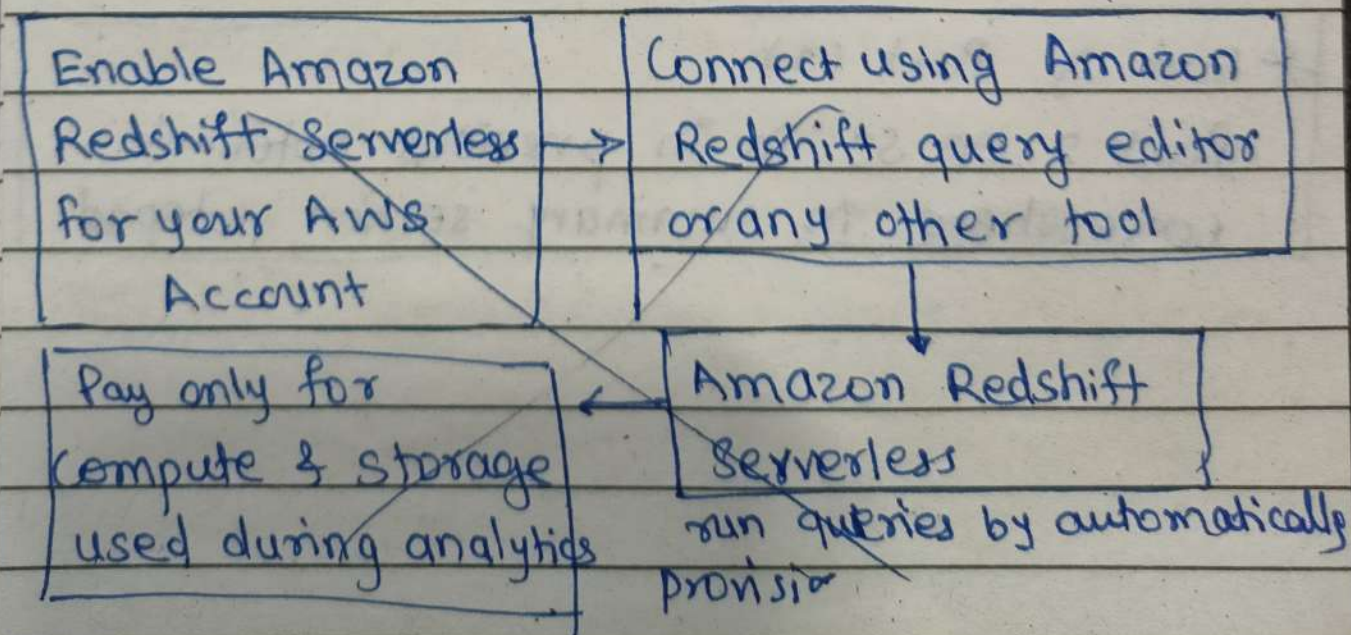# Redshift Overview

- Redshift is based on PostgreSQL, but it's not used for OLTP.

- It's OLAP - Online Analytical Processing (analytics and data warehousing)

- Load data once every hour, not every second.

- 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), highly available

- Pay as you go based on the instances provisioned.

- has SQL interface for performing the queries.

- BI tools such as AWS Quicksight or Tableau integrate with it.

Redshift serverless

- Automatically provisions & scales data warehouse underlying capacity.

- Run analytics workloads without managing data warehouse infrastructure.

- Pay only for what you use (save costs)

- Use Cases: Reporting, dashboarding appln's, real-time analytics.....

Enable Amazon Redshift Serverless for your AWS Account	Connect using Amazon Redshift query editor or any other tool
Pay only for compute & storage used during analytics	Amazon Redshift Serverless run queries by automatically provisio

user		Connect using Amazon Redshift query editor or any other tool

Enable Amazon Redshift serverless for your Aws Acc.

Payment		Amazon Redshift Serverless

Pay only for compute & storage used during analysis.

run queries by automatically provisioned & scale capacity based on workload

EMR Overview

- EMR stands for "Elastic MapReduce".

- EMR helps creating Hadoop clusters (Big Data) to analyze & process vast amount of data.

- The clusters can be made of hundreds of EC2 instances.

- Also supports Apache sparks, HBase, Presto, Flink.....

- EMR takes care of all the provisioning & configuration

- Auto-scaling and integrated with spot instances

- Use Cases: data processing, machine learning, web indexing, big data....

Athena Overview

- Serverless query service to perform analytics against S3 objects

- Uses standard SQL language to query the files.

- Supports CSV, JSON, ORC, Avro, & Parquet (built on Presto)

- Pricing 5$ per TB of data scanned.

- Use compressed or columnar data for cost - savings (less scans)

- Use - Cases : Business Intelligence /analytics/ reporting, analyze & query VPC flowlogs, ELB logs, Cloudtrail trails, etc....

- Exam Tip :- analyze data in S3 using serverless SQL, use Athena.

Amazon Quicksight

- Serverless machine learning-powered business intelligent service to create interactive dashboards.

- Fast, automatically scalable, embeddable, with person pricing.

- Use Cases: 1] Business analytics
              2] Business visualizations
              3] Perform ad-hoc analysis
              4] Get business insights using data.

- Integrated with RDS, Aurora, Athena, Redshift, S3.....

DocumentDB (NOSQL DB)

- Aurora is an "AWS - implementation" of PostgreSQL/MySQL........

- MongoDB is used to store, query, & index JSON data.

- Similar "deployment concepts" as Aurora.

- Fully managed, highly available with replication across 3 AZ.

- DocumentDB storage automatically grows in increaments of 10 GB.

- Automatically scales to workloads with millions of requests per seconds.

Amazon Neptune

- Fully managed graph database.

- A popular graph datasets would be social network.
    - Users have friends
    - Posts have comments
    - Comments have likes from users.
    - Users share & like posts....

- Highly available across 3 AZ, with upto 15 read replicas.

- Build & run applications working with highly connected datasets - optimized for these complex & hard queries.

- Can store upto billions of relations & query the graph with milliseconds latency.

- Highly available with replications across multiple AZs.

- Great for knowledge graphs(wikipedia), fraud detection, recommendation, social network

Amazon Timestream

- Fully managed, fast, scalable, serverless time series database.

- Automatically scales up/down to adjust capacity.

- Store and analyse trillions of events per day

- 1000s times faster & 1/10th the cost of relational databases.

- Built-in time series analytics functions (helps you identify patterns in your data in real-time.

Amazon QLDB

- QLDB stands for " Quantum Ledger Database"

- A ledger is a book 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.

- Immutable system: no entry can be removed or modified, cryptographically verifiable.

- 2-3x better performance than common ledger blockchain: no decentralization component, in accordance with financial relu regulation rules.
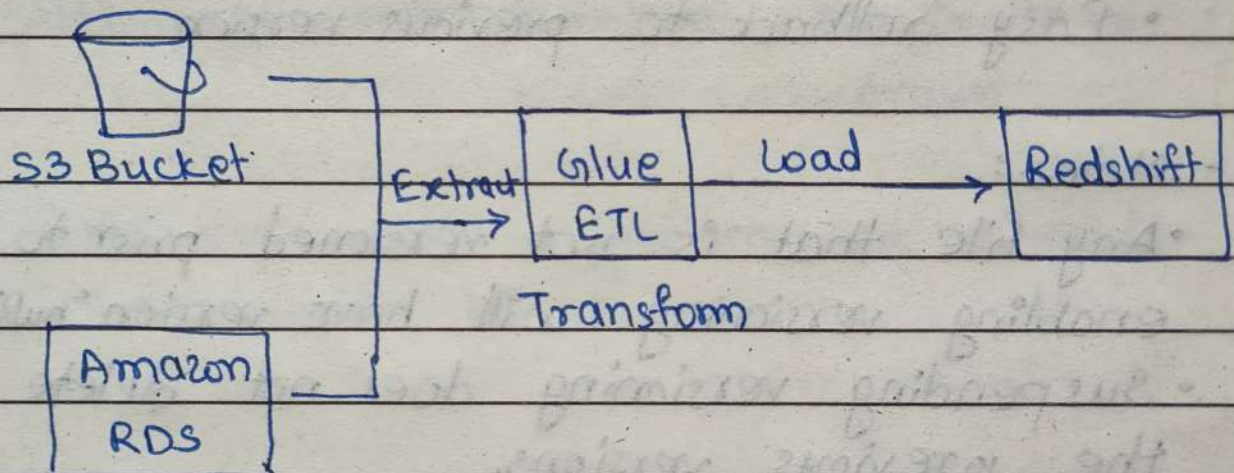
Managed Blockchain (Decentralized)

- Blockchain makes it possible to build appl^n's where multiple parties can execute transactions without the need of a trusted central Authority.

- Amazon managed Blockchain is a managed service to:
  - Join public blockchain networks.
  - Or create your own scalable private network.

- Compatible with the frameworks Hyperledger Fabric & Ethereum.

AWS Glue

- Managed ==extract, transform & load (ETL)== service.

- Useful to prepare & transform data for analytics.

- Fully ==serverless== serice service.

```
   ___
  |   |
  |___|        ┌─────────┐   Load   ┌──────────┐
 S3 Bucket     │  Glue   │ ───────→ │ Redshift │
        Extract│  ETL    │          └──────────┘
         ──→   └─────────┘
 ┌────────┐     Transform
 │ Amazon │
 │  RDS   │
 └────────┘
```

- Glue Data Catalog: catalog of datasets
  - This can be used by Athena, Redshift, EMR

DMS - Data Migration Service

- Quickly and securely migrate databases 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.

Database & Analytics Summary in AWS

① Relational Databases - OLTP : RDS & Aurora
                                    (SQL)

② Differences between Multi-AZ, Read-replicat
   Multi - Region

③ In - memory : Elasticache

④ Key/Vaue Databases : DynamoDB (serverless)
   & DAX (cache for dynamoDB) (NoSQL)
   (can handle millions op request per sec)

⑤ Warehouse - OLAP : Redshift (SQL)
                            (Analytics)

⑥

⑥ Hadoop cluster : EMR (Big data)

⑦ Athena : query data on Amazon S3
                (serverless & SQL)

⑧ QuickSight : dashboards on your data
                (serverless)

⑨ DocumentDB: "Aurora for MongoDB"
(JSON - NoSQL)

⑩ Amazon QLDB: financial Transactions ledger (immutable journal, cryptographically verificable)

⑪ Amazon Managed Blockchain: managed Hyperledger fabric & Etherum blockchain (decentralized Blockchain)

⑫ Glue: Managed ETL (extract Transform Load) and Data catalog Service.

⑬ Database Migration : DMS

⑭ Neptune : graph database

⑮ Timestream : time - series database