

# AWS Certification – Database Services – Cheat Sheet

## RDS

- provides Relational Database service
- supports MySQL, MariaDB, PostgreSQL, Oracle, Microsoft SQL Server, and the new, MySQL-compatible Amazon Aurora DB engine
- as it is a managed service, shell (root ssh) access is not provided
- manages backups, software patching, automatic failure detection, and recovery
- supports user initiated manual backups and snapshots
- daily automated backups with database transaction logs enables Point in Time recovery up to the last five minutes of database usage
- snapshots are user-initiated storage volume snapshot of DB instance, backing up the entire DB instance and not just individual databases that can be restored as a independent RDS instance
- support encryption at rest using KMS as well as encryption in transit using SSL endpoints
- for encrypted database
  - logs, snapshots, backups, read replicas are all encrypted as well
  - cross region replicas and snapshots does not work

across region

- Multi-AZ deployment
  - provides high availability and automatic failover support and is NOT a scaling solution
  - maintains a synchronous standby replica in a different AZ
  - transaction success is returned only if the commit is successful both on the primary and the standby DB
  - Oracle, PostgreSQL, MySQL, and MariaDB DB instances use Amazon technology, while SQL Server DB instances use SQL Server Mirroring
  - snapshots and backups are taken from standby & eliminate I/O freezes
  - during automatic failover, its seamless and RDS switches to the standby instance and updates the DNS record to point to standby
  - failover can be forced with the Reboot with failover option
- Read Replicas
  - uses the PostgreSQL, MySQL, and MariaDB DB engines' built-in replication functionality to create a separate Read Only instance
  - updates are asynchronously copied to the Read Replica, and data might be stale
  - can help scale applications and reduce read only load
  - requires automatic backups enabled
  - replicates all databases in the source DB instance
  - for disaster recovery, can be promoted to a full fledged

database

- can be created in a different region for MySQL, Postgres and MariaDB, for disaster recovery, migration and low latency across regions
- RDS does not support all the features of underlying databases, and if required the database instance can be launched on an EC2 instance
- RMAN (Recovery Manager) can be used for Oracles backup and recovery when running on an EC2 instance

## DynamoDB

- fully managed NoSQL database service
- synchronously replicates data across three facilities in an AWS Region, giving high availability and data durability
- runs exclusively on SSDs to provide high I/O performance
- provides provisioned table reads and writes
- automatically partitions, reallocates and re-partitions the data and provisions additional server capacity as data or throughput changes
- provides Eventually consistent (by default) or Strongly Consistent option to be specified during a read operation
- creates and maintains indexes for the primary key attributes for efficient access of data in the table
- supports secondary indexes
  - allows querying attributes other than the primary key attributes without impacting performance.
  - are automatically maintained as sparse objects
- Local vs Global secondary index
  - shares partition key + different sort key **vs** different

partition + sort key

- search limited to partition **vs** across all partition
- unique attributes **vs** non unique attributes
- linked to the base table **vs** independent separate index
- only created during the base table creation **vs** can be created later
- cannot be deleted after creation **vs** can be deleted
- consumes provisioned throughput capacity of the base table **vs** independent throughput
- returns all attributes for item **vs** only projected attributes
- Eventually or Strongly **vs** Only Eventually consistent reads
  - size limited to 10Gb per partition **vs** unlimited
- supports cross region replication using DynamoDB streams which leverages Kinesis and provides time-ordered sequence of item-level changes and can help for lower RPO, lower RTO disaster recovery
- Data Pipeline jobs with EMR can be used for disaster recovery with higher RPO, lower RTO requirements
- supports triggers to allow execution of custom actions or notifications based on item-level updates

## ElastiCache

- managed web service that provides in-memory caching to deploy and run Memcached or Redis protocol-compliant cache clusters
- ElastiCache with Redis,
  - like RDS, supports Multi-AZ, Read Replicas and

## Snapshots

- Read Replicas are created across AZ within same region using Redis's asynchronous replication technology
- Multi-AZ differs from RDS as there is no standby, but if the primary goes down a Read Replica is promoted as primary
- Read Replicas cannot span across regions, as RDS supports
- cannot be scaled out and if scaled up cannot be scaled down
- allows snapshots for backup and restore
- AOF can be enabled for recovery scenarios, to recover the data in case the node fails or service crashes. But it does not help in case the underlying hardware fails
- Enabling Redis Multi-AZ as a Better Approach to Fault Tolerance
- ElastiCache with Memcached
  - can be scaled up by increasing size and scaled out by adding nodes
  - nodes can span across multiple AZs within the same region
  - cached data is spread across the nodes, and a node failure will always result in some data loss from the cluster
  - supports auto discovery
  - every node should be homogenous and of same instance type

- ElastiCache Redis vs Memcached
  - complex data objects **vs** simple key value storage
  - persistent **vs** non persistent, pure caching
  - automatic failover with Multi-AZ **vs** Multi-AZ not supported
  - scaling using Read Replicas **vs** using multiple nodes
  - backup & restore supported **vs** not supported
- can be used state management to keep the web application stateless

## Redshift

- fully managed, fast and powerful, petabyte scale data warehouse service
- uses replication and continuous backups to enhance availability and improve data durability and can automatically recover from node and component failures
- provides Massive Parallel Processing (MPP) by distributing & parallelizing queries across multiple physical resources
- columnar data storage improving query performance and allowing advance compression techniques
- only supports Single-AZ deployments and the nodes are available within the same AZ, if the AZ supports Redshift clusters
- spot instances are NOT an option