Amazon Aurora

VS

Amazon Redshift

مدينة زويل
ZEWAIL CITY

BY:
Rahma Hassan
Riham Abdo

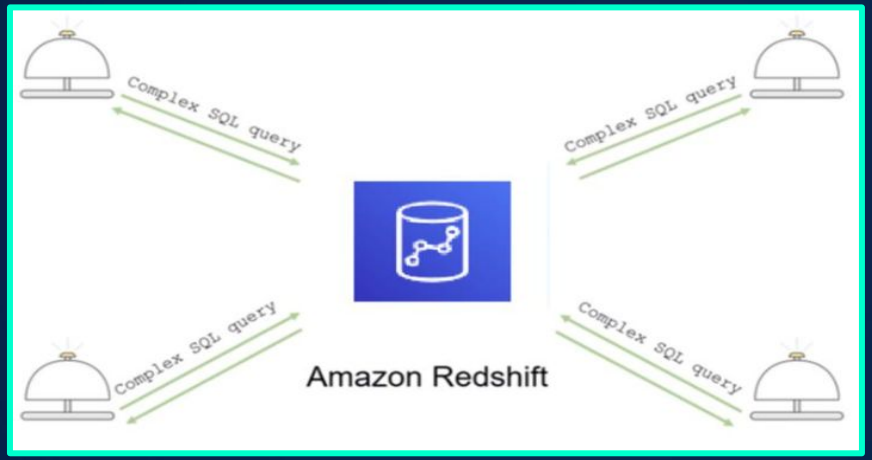
TABLE OF CONTENTS

01

Intro to Redshift

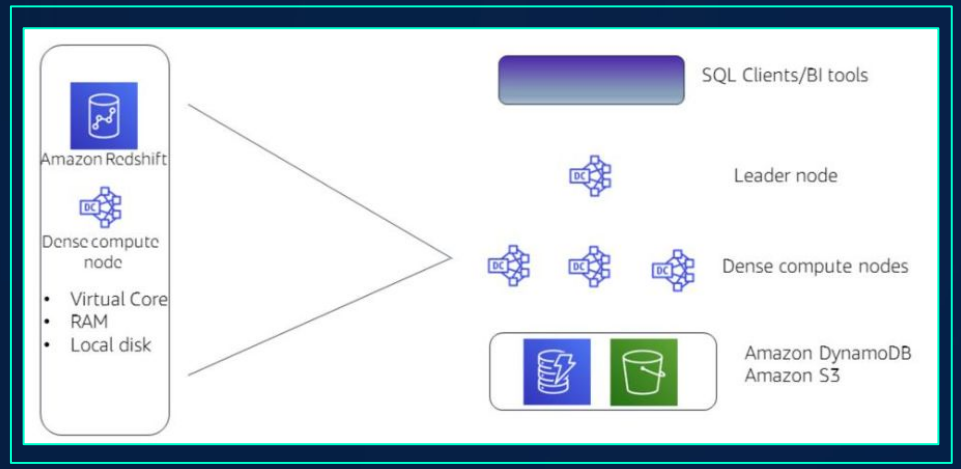
Introduction to Redshift

Amazon Redshift is a fast, fully managed data warehouse that makes it simple and cost-effective to analyze all your data by using standard SQL and your existing business intelligence (BI) tools.



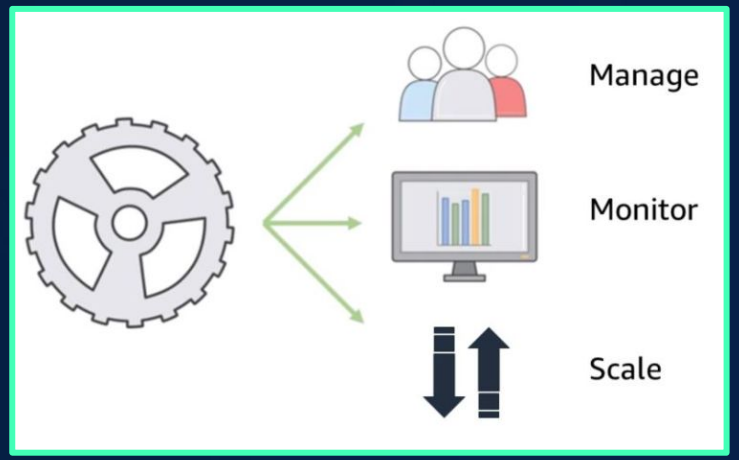
Parallel Processing Architecture

The leader node manages communications with client programs and all communication with compute nodes. It parses and develops plans to carry out database operations—specifically, the series of steps that are needed to obtain results for complex queries. The leader node compiles code for individual elements of the plan and assigns the code to individual compute nodes. The compute nodes run the compiled code and send intermediate results back to the leader node for final aggregation.



Automation and Scaling

Redshift can scale automatically in a matter of minutes in the case of the newer generation nodes.
Automatic scaling is achieved by adding more nodes. A cluster can only be created using the same kind of nodes.



02

Intro to Aurora

Aurora

-It is an Amazon Web Services fully managed relational database engine that is compatible with MySQL and PostgreSQL.

-It's about five times faster than traditional MySQL databases and three times faster than traditional PostgreSQL databases.

-If your company uses one of these database engines, you can migrate your data to Aurora without changing a single line of code.

-Aurora has features:
1- distributed fault-tolerant

2-self-healing storage system .

- All standard database administration chores, such as hardware provisioning, data backup, and update installation, are fully automated.

Aurora Scalability Feature

-Aurora can grow up to 64 TB of storage.

-Aurora also has a serverless functionality that allows for a completely on-demand experience, with the database scaling down automatically in low-load situations and vice versa.

-Hence,payment is better for customers:
In this mode, customers only need to pay for the time the database is active.

Aurora Replication Feature

-The data is going to be replicated over multiple regions automatically.

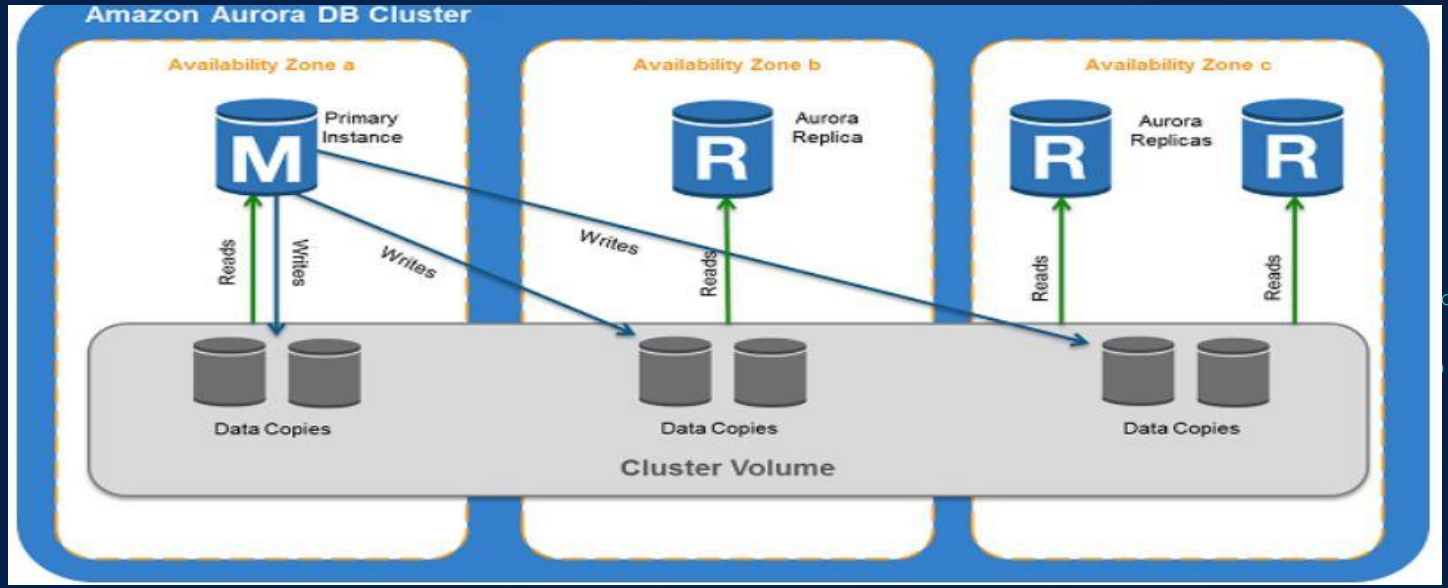
-No data is being lost.

-If you have any problem with master instance, Amazon will direct to another replicas without affecting your data.

Aurora Architecture

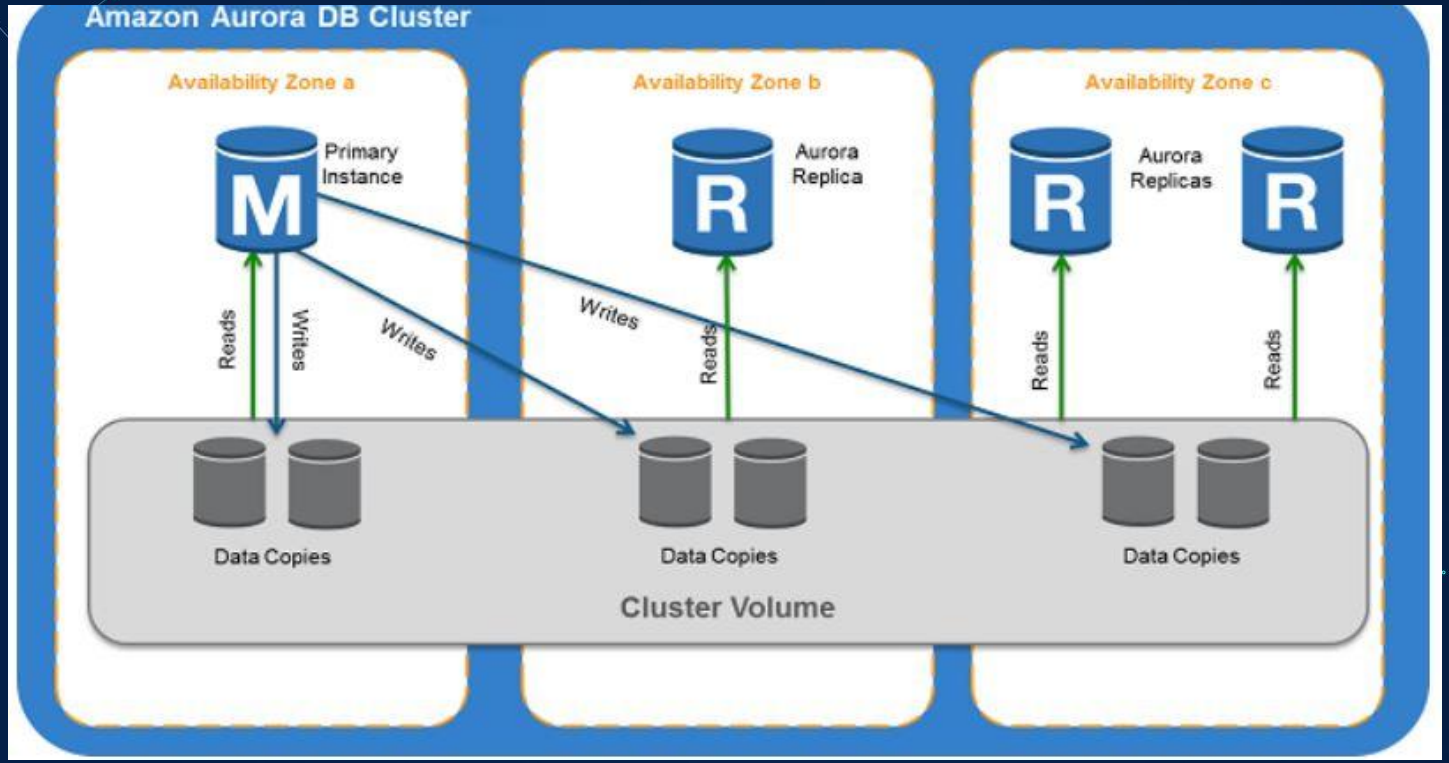
-Aurora architecture is based on a cluster volume that maintains all of the data for all of the database instances in that cluster.
-A cluster volume is effectively virtual database storage that spans many availability zones.

-The underlying storage volume is built on top of numerous cluster nodes spread across multiple availability zones.

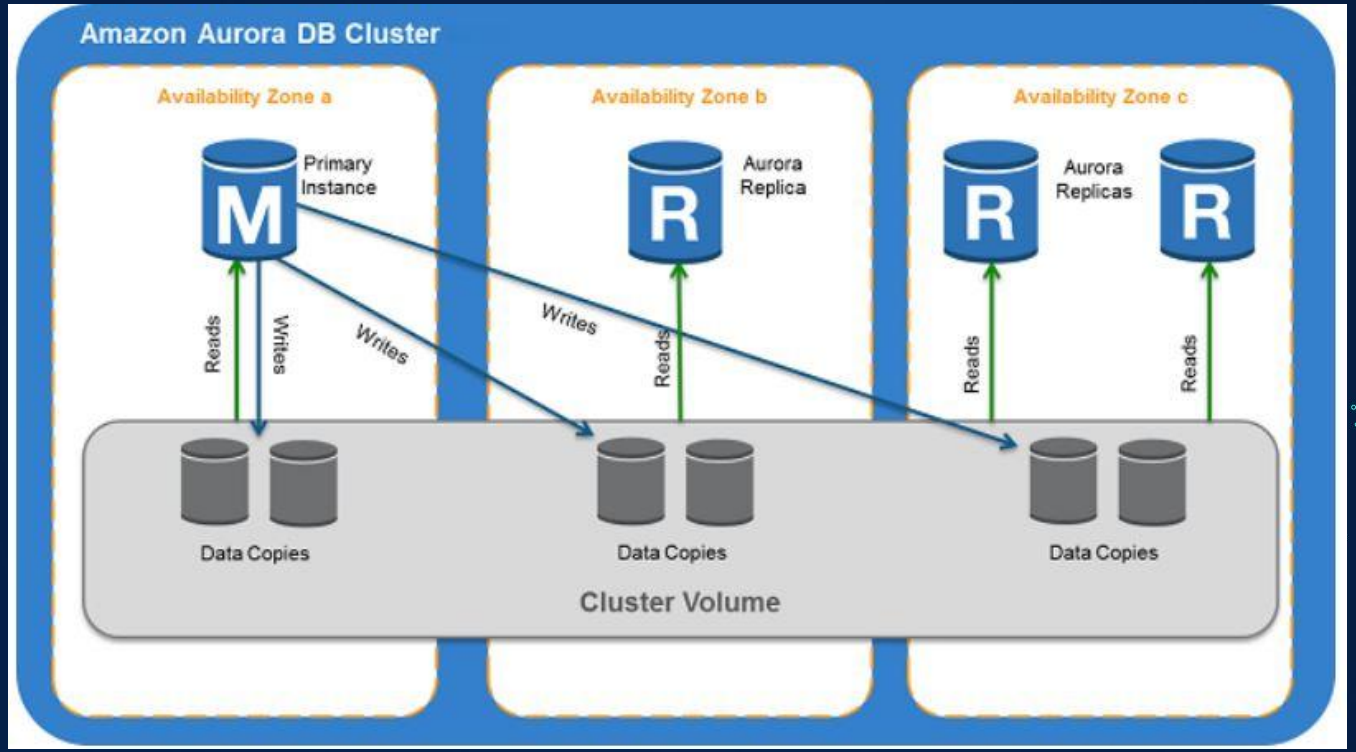


Primary DB instance – Supports read and write operations on the cluster volume, as well as all data modifications.

There is only one primary DB instance in each Aurora DB cluster.



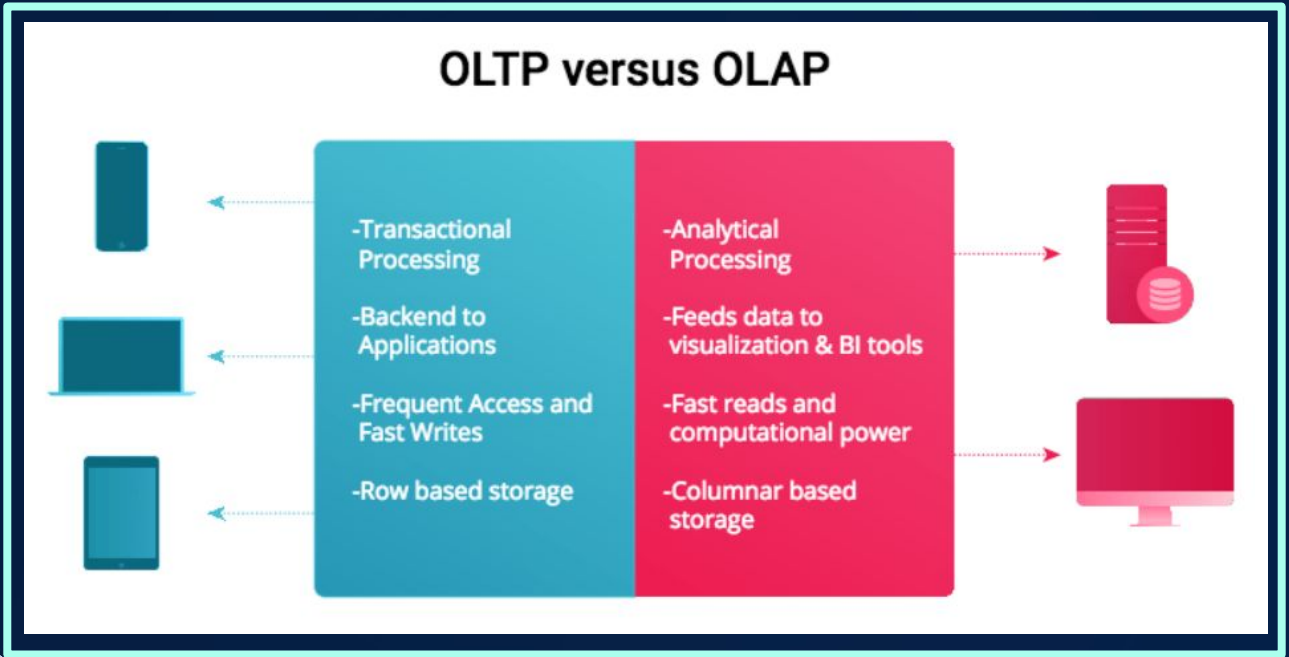
Aurora Replica – This replica connects to the same storage volume but only allows read operations.In addition to the core DB instance, each Aurora DB cluster can have up to 15 Aurora Replicas.Keep Aurora Replicas in distinct Availability Zones to ensure high availability.



03

What is the Difference?

OLTP (Aurora) VS OLAP (Redshift)



What to consider when making a decision?

1 Scaling

Redshift

- Scales by adding nodes
- Automatic
- Concurrency Scaling

Aurora

- Scales vertically through instance upgrade
- Scales horizontally by through eard replicas

What to consider when making a decision?

2

Storage Capacity

Redshift

- Can scale to petabytes of data.
- Can support upto 60 user-defined database per cluster.

Aurora

- Has a hard limit of 64 TB.
- Limited number of database instances of 40.

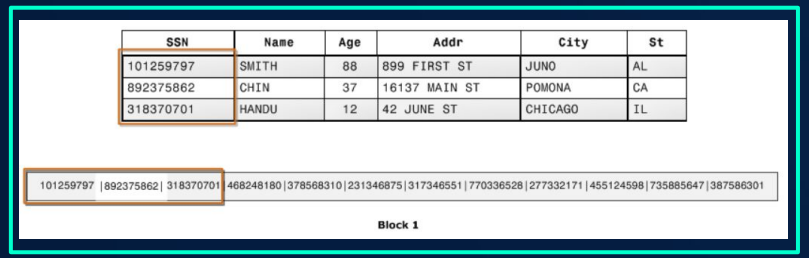
What to consider when making a decision?

3

Data Structure

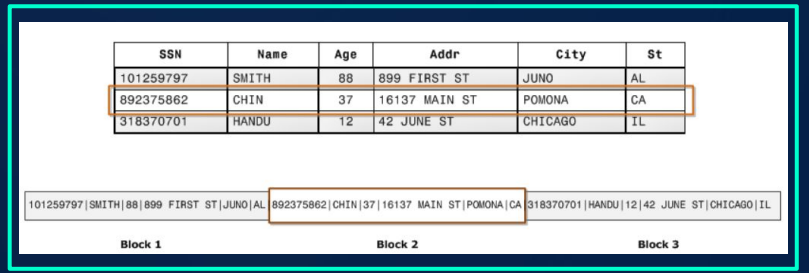
Redshift

- Uses columnar storage.
- Misses out on many data types.



Aurora

- Row-oriented storage.
- Supports complete data types in both MySQL and Postgres.



What to consider when making a decision?

4

Performance

Redshift

- Faster with large amount of data.
- Optimized for OLAP workloads.
- Supports only eventual consistency.

Aurora

- Offers better performance than traditional MySQL instance.
- Optimized for OLTP workloads.

What to consider when making a decision?

5

Security, Maintenance, and Pricing

- About security, there is nothing much to differentiate between the two services since both being part of the AWS portfolio.
- Both are completely managed services and require very little maintenance.
- Redshift is generally higher in price. However, pricing can not be compared independent of the customer use case

References

[1] AWS Academy.

[2] S. on D. Warehouse, V. S. on D. Integration, and B. I. on D. Warehouse, “Amazon Redshift vs Aurora: An in-depth comparison: Hevo Blog,” Hevo Blog - Transformative ideas and real insights on all things Data, 04-Apr-2022. [Online]. Available: https://hevodata.com/blog/amazon-redshift-vs-aurora/#f1. [Accessed: 24-May-2022].

[3] “What's the difference between Amazon Redshift and Aurora? - treasure data blog,” Treasure Data Blog - Enterprise Customer Data Platform, 18-Aug-2019. [Online]. Available: https://blog.treasuredata.com/blog/2016/02/10/whats-the-difference-between-aws-redshift-aurora/. [Accessed: 24-May-2022].

[4]https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/Aurora.Overview.html

THANKS

Any questions?