Amazon Redshift

Cloud Data Warehouse

Guilherme Nogueira

What is Redshift?

1	Cloud-based data warehouse (AWS)
2	Columnar storage for performance
3	Massively Parallel Processing (MPP)
4	SQL-based, built for analytics

How Redshift Works

Massively Parallel Processing (MPP)	Columnar Data Storage
Leader node parses queries and creates exec plans	Stores data by columns, not rows
Distributes tasks to compute nodes	Reads only relevant columns (faster I/O)
Compute nodes run in parallel (slices)	Advanced compression saves space
Results aggregated and sent back	COPY command for fast S3 ingestion

Internal Architecture

Leader Node

Coordinates queries, creates execution plans, communicates with clients

Compute Nodes

Execute compiled code in parallel, process data in slices

RA3 Nodes

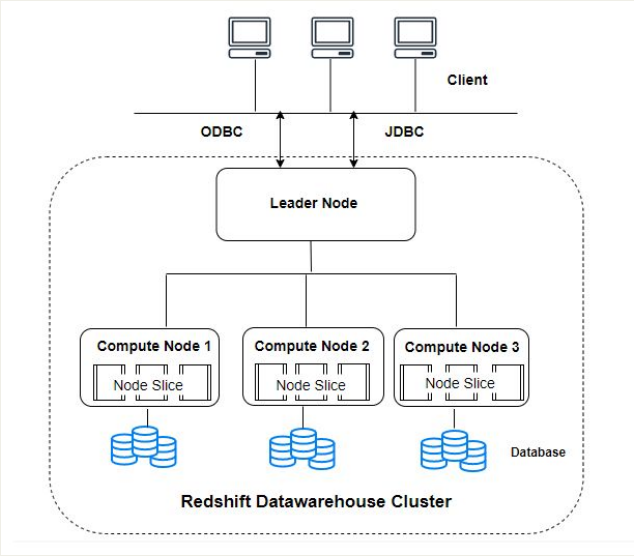
Decouple compute and storage for independent scaling

Redshift Managed Storage (RMS)

Uses high-performance SSD cache + S3

Node Slices

Partitions of compute nodes that process in parallel



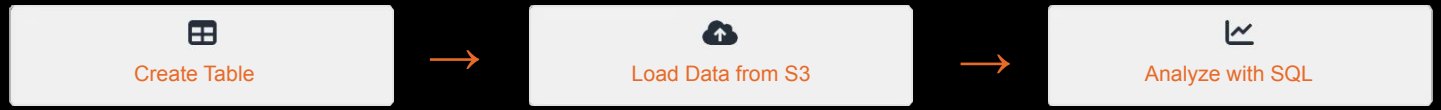
Trade-offs

PROS	CONS
High performance with MPP	Needs tuning (sort/dist keys)
Decoupled compute/storage (RA3)	Not fully serverless
SQL-native, AWS-integrated	Limited JSON/nested data support

Comparison with Other Tools

Redshift (AWS)	BigQuery (GCP)	Snowflake
PostgreSQL-compatible	Standard SQL	Multi-cloud
Usage-based pricing	Query-based billing	Rich SQL
Partial serverless (RA3)	Fully serverless	Serverless by design

Demo Overview



Thank you!

Guilherme Nogueira

```sql
-- ------------------------------------------------
-- Create a sample table to store user data
-- ------------------------------------------------
CREATE TABLE IF NOT EXISTS users (
    id INT,
    name VARCHAR(50),
    email VARCHAR(100),
    created_at TIMESTAMP
);
```

```sql
-- ------------------------------------------------
-- Load CSV data from an S3 bucket using COPY
-- - This command reads data directly from S3
-- - You must replace the S3 path and IAM role ARN
-- ------------------------------------------------
COPY users
FROM 's3://demo-redshift-csv/sample-users_global.csv'
IAM_ROLE 'arn:aws:iam::xxxxxx:role/service-role/AmazonRedshift-CommandsAccessRole-20250730T204952'
FORMAT AS CSV
IGNOREHEADER 1;
```

```sql
-- ------------------------------------------------
-- Run a query to verify that the data was loaded
-- - This will return all rows ordered by timestamp
-- ------------------------------------------------
SELECT *
FROM users
ORDER BY created_at;
```

```sql
-- ------------------------------------------------
-- Example of a simple analytical query
-- - Count how many users were created per day
-- ------------------------------------------------
SELECT
  DATE(created_at) AS signup_date,
  COUNT(*) AS total_users
FROM users
GROUP BY signup_date
ORDER BY signup_date;
```