Dec 07, 2023

TCSS 562
Cloud Technology Presentation

AMAZON REDSHIFT

Presented By: Angela C Farin, Srishty, Tomoki Kusunoki

BE BOUNDLESS

W

INTRODUCTION

Amazon Redshift is a cloud-based data warehousing service offered by Amazon Web Services (AWS).

Serves as a powerful solution for managing and analyzing large datasets providing rapid query performance through its columnar storage and massively parallel processing (MPP) architecture.

Source: https://en.m.wikipedia.org/wiki/File:Amazon-Redshift-Logo.svg

UNIVERSITY *of* WASHINGTON

HISTORY OF TECHNOLOGY: WHO

The development of Amazon Redshift was led by Amazon Web Services (AWS)

An initial preview beta was released in November 2012 and full release was made available on February 15, 2013

Focus of the development was to create a scalable, high-performance, and cost-effective solution for data warehousing in the cloud
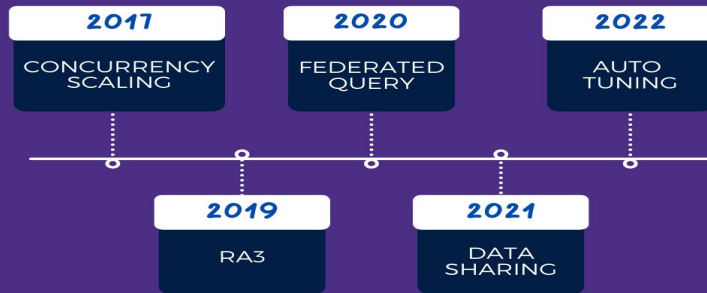
UNIVERSITY *of* WASHINGTON

HISTORY OF TECHNOLOGY: WHY

Traditional on-premises data warehouses faced challenges in handling the increasing volume of data and the demand for faster and more complex analytics.
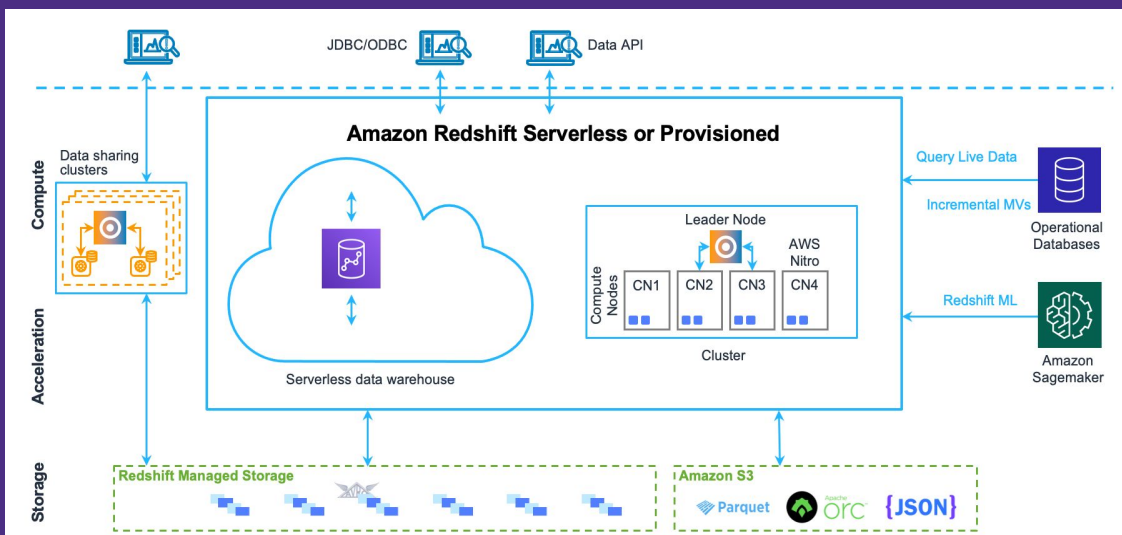
The motivation behind creating Amazon Redshift was to address these challenges by leveraging the scalability and flexibility of cloud computing.

Allows users to focus on deriving insights from their data rather than dealing with the complexities of traditional data warehouse management.

UNIVERSITY *of* WASHINGTON

HISTORY OF TECHNOLOGY: HOW

2017	2020	2022
CONCURRENCY SCALING	FEDERATED QUERY	AUTO TUNING

2019	2021
RA3	DATA SHARING

JDBC/ODBC   Data API

Amazon Redshift Serverless or Provisioned

Compute

Data sharing clusters

Query Live Data

Incremental MVs

Operational Databases

Leader Node

AWS Nitro

Compute Nodes: CN1 CN2 CN3 CN4

Cluster

Redshift ML

Amazon Sagemaker

Acceleration

Serverless data warehouse

Storage

Redshift Managed Storage

Amazon S3

Parquet   Orc   {JSON}

FEATURES

- Column Storage
- Massively Parallel Processing (MPP)
- Automatic Compression and Optimization
- Advanced Analytics Support
- Scalability
- Integration with AWS Ecosystem
- Security
- Concurrency Scaling
- AWS Managed Services Enhancements
- Federated Queries

UNIVERSITY *of* WASHINGTON

USE CASES: NETFLIX, NASDAQ, PINTEREST

- **Netflix** utilizes Amazon Redshift for analyzing viewer preferences, content performance, and streaming patterns. This data-driven approach helps optimize content recommendations and enhance the user experience.

- **Nasdaq** employs Amazon Redshift to analyze financial market data in real-time. This includes tracking stock prices, trade volumes, and market trends, enabling timely decision-making for traders and investors.

- **Pinterest** utilizes Amazon Redshift to analyze user engagement, content interactions, and trends on its platform. This data-driven approach helps enhance user recommendations and content discovery.

UNIVERSITY *of* WASHINGTON

USE CASES: SAMSUNG & YELP

- **Samsung** utilizes Amazon Redshift for analyzing data from connected devices, IoT sensors, and customer interactions. This data-driven approach aids in improving product functionality, anticipating maintenance needs, and enhancing user experiences.

- **Yelp** leverages Amazon Redshift for analyzing user reviews, business performance, and customer feedback. This data-driven insight assists businesses in optimizing their online presence and service offerings.

ADVANTAGES

- Scalability
- High Performance
- Cost-Effectiveness
- Integration with AWS Ecosystem
- AWS Managed Services
- Security Features
- Concurrent Query Scaling
- Federated Query

DISADVANTAGES

- Cost at Scale
- Complexity of Optimization
- Data Loading Complexity
- Limited Flexibility in Node Types
- Warm-Up Time for Concurrency Scaling
- Data Transfer Costs
- Dependency on Network Connectivity
- Limited Support for Unstructured Data
- Incremental Backups

USABILITY

- **Seamless Integration**: Amazon Redshift seamlessly integrates with various data sources, including data lakes, facilitating a unified and comprehensive approach to data analytics.
- **Learning Curve**: While basic operations are accessible, mastering advanced features like query optimization and cluster configuration may have a steeper learning curve.
- **Programming APIs**: Amazon Redshift supports standard SQL queries, making it familiar for those with SQL expertise and easy integration with programming languages like Python, Java.

COST DISCUSSION

- **Pay-as-You-Go Pricing**: Costs include charges for compute nodes, storage, data transfer, and any additional features or services utilized.
- **Compute Node Pricing**: Pricing could range from $0.25 to $6.80 per node per hour, depending on the node type and region.
- **Storage Costs**: Storage costs typically range from $0.125 to $1.25 per gigabyte per month, depending on the region.
- **Data Transfer Costs**: no charge for data transferred between Amazon Redshift and Amazon S3 within the same AWS Region for backup, restore, load, and unload operations. For all other data transfers into and out of Amazon Redshift, billing at standard AWS data transfer rates.

Read More on Pricing:
https://aws.amazon.com/redshift/pricing/#:~:text=With%20Amazon%20Redshift%2C%20you%20can,over%2Dprovisioning%20compute%20or%20storage

---

COST EXAMPLE

SCENARIO:
- Business Profile: Mid-sized e-commerce platform.
- Data Volume: Handling sales data, customer interactions, and product analytics.
- Workload: Moderate to high query concurrency during peak sales periods.
- Storage Requirement: Storing transactional data for the past 12 months.

CONFIGURATION:
- Compute Nodes: dc2.8xlarge instances (for high performance during peak periods).
- Storage: 10 TB provisioned storage.
- Data Transfer: Moderate data transfer to the internet.

COST EXAMPLE

Estimated Monthly Cost Breakdown:

- **Compute Nodes (dc2.8xlarge)**:
  - 2 nodes x $4.00 per node per hour = $8.00 per hour
  - Estimated monthly compute cost: $8.00 x 24 hours x 30 days = $5,760.00
- **Storage (10 TB)**: $0.125 per GB per month x 10,000 GB = $1,250.00
- **Data Transfer**: Assuming 500 GB data transfer to the internet.
  $0.02 per GB x 500 GB = $10.00

Total Estimated Monthly Cost: $5,760.00 (Compute) + $1,250.00 (Storage) + $10.00 (Data Transfer) = $7,020.00

UNIVERSITY *of* WASHINGTON

15

CONCLUSIONS

- Used by organizations seeking a scalable, high-performance, and cost-effective cloud-based data warehousing solution.
- Widely adopted across industries, including retail, finance, healthcare, and more.
- Suited for businesses of varying scales, from small enterprises to large corporations.
- Anticipated to continue its prominent role in cloud-based data analytics.

UNIVERSITY *of* WASHINGTON

16

REFERENCES

- https://aws.amazon.com/redshift/pricing/
- https://docs.aws.amazon.com/redshift/latest/dg/cost.html
- https://aws.amazon.com/pm/redshift/
- https://aws.amazon.com/blogs/big-data/amazon-redshift-announcements-at-aws-reinvent-2023-to-enable-analytics-on-all-your-data/

DEMONSTRATION

QUESTIONS?

THANK YOU!

BE BOUNDLESS