INNOVATE

DATA EDITION

19 August 2021

New data warehousing use cases with Amazon Redshift

Kerry McRae
Solutions Architect
Amazon Web Services

Session overview

- Challenges of data at scale
- Data sharing feature
- Amazon Redshift Data API
- SUPER data type with JSON support
- AQUA for Amazon Redshift
- Demo

Challenges of data analytics at scale

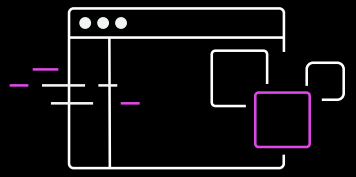
VARIETY

Variety of sources and data types

Multiple analytics needs

Data volume and velocity

PERFORMANCE



Slow performance

Difficult to manage systems

Complex to scale

COST

Increasing and unpredictable cost

Inflexible tools

Security, compliance

Amazon Redshift

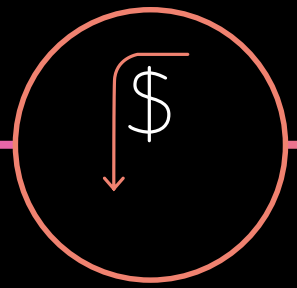
ANALYZE ALL YOUR DATA

Take a **lake house approach** by analyzing all your data across your data warehouse, your Amazon S3 data lake, and operational databases with consistent security and governance policies

PERFORMANCE AT ANY SCALE

Get **better price performance** than other cloud data warehouses with a **self-tuning** system, and boost query performance **with AQUA**

LOWER YOUR COSTS



Start small and pay only for what you use with **predictable** monthly costs, Redshift is at least **50% less expensive** than other cloud data warehouses

Amazon Redshift innovates to meet your needs

Analyze all your data

Lake house with AWS integration

NEW! Amazon Redshift ML

NEW! Data sharing

NEW! Super data type with JSON support

UPDATED! Federated Query

NEW! Lambda UDF

NEW! Partner console integration

Amazon Redshift Spectrum + Lake Formation

Data Lake Export

Performance & scale

Fast and self-tuning

UPDATED! RA3 nodes & managed storage

NEW! AQUA

NEW! Performance tuning: automated (ATO)

UPDATED! Materialized views

NEW! 100K tables

NEW! HyperLogLog

Concurrency scaling

Low cost & best value

Predictable costs

UPDATED! Automatic workload manager

NEW! Cross-AZ cluster recovery

NEW! Data API

On-demand and RIs

Pause and resume

Cost controls

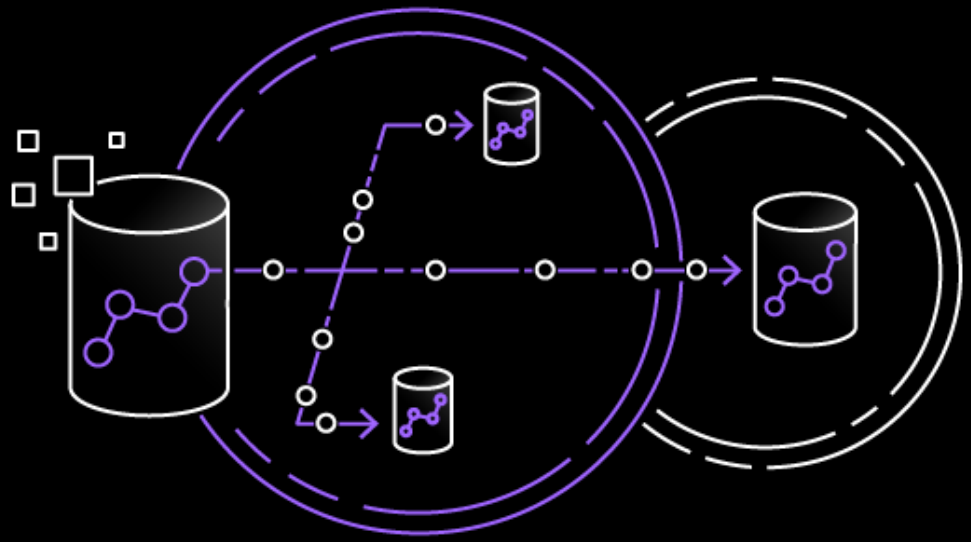
Built-in security features

aws

Amazon Redshift data sharing

Amazon Redshift data sharing

A secure and easy way to share live data across Amazon Redshift clusters within the same or different AWS accounts



Simple and direct way to share data across Amazon Redshift clusters

Instant, granular, and high-performance access without data copies and data movement

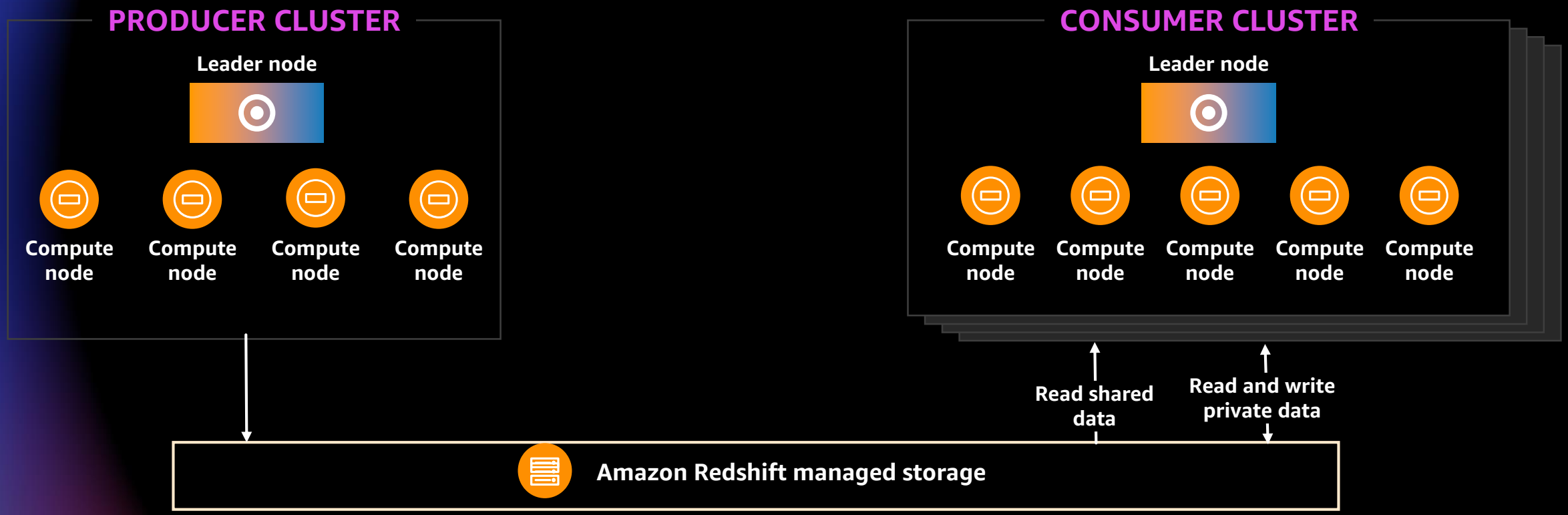
Live and transactionally consistent views of data across all consumers

Secure and governed collaboration within and across organizations and with external parties

aws

Data sharing builds on Amazon Redshift managed storage

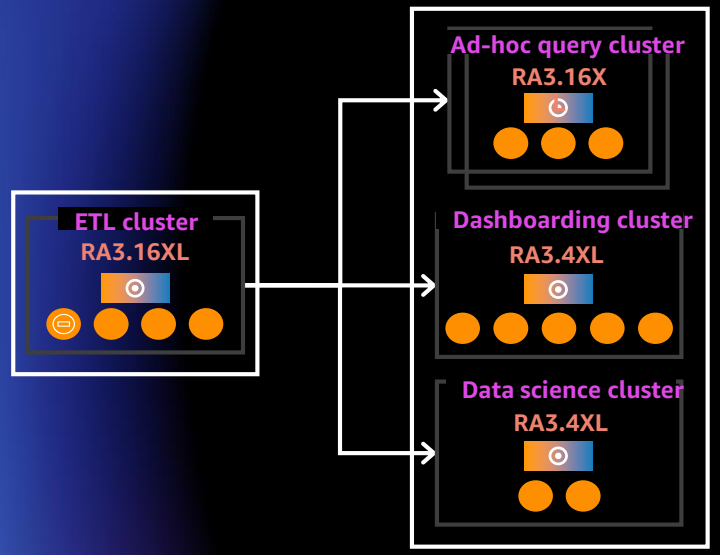
HIGH PERFORMANCE DATA ACCESS WHILE PRESERVING WORKLOAD ISOLATION



Producer pays for Amazon Redshift managed storage and consumers pay for consumer cluster

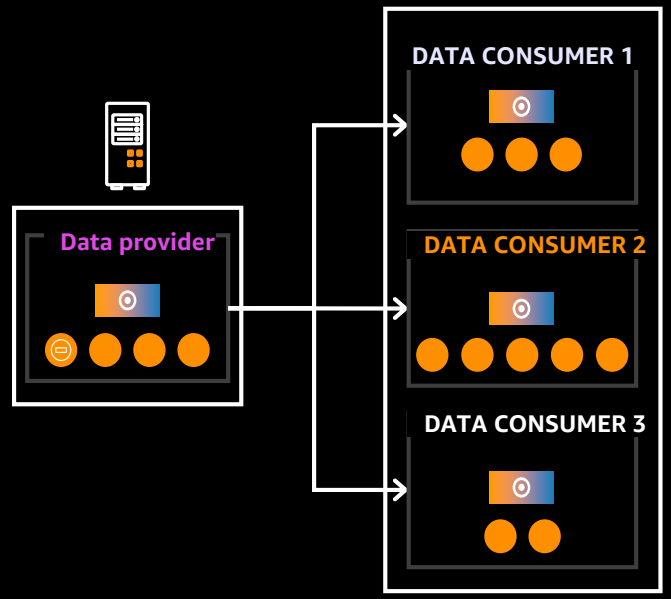
Workloads accessing shared data are isolated from each other and the producer

Data sharing use cases

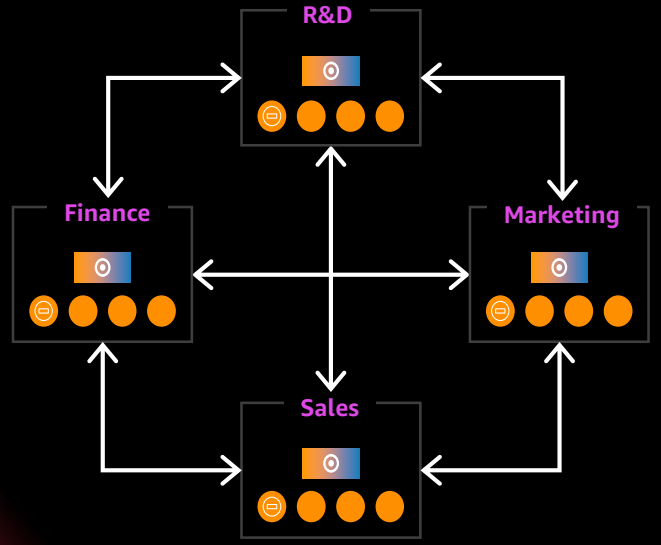


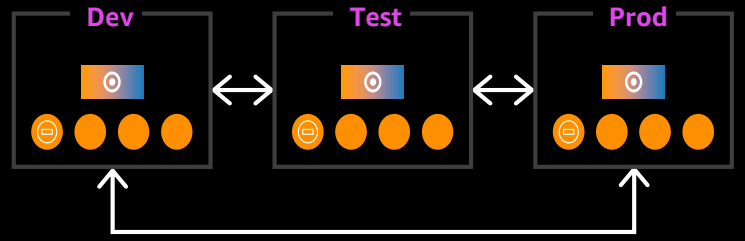
Supporting different kinds of business-critical workloads

Delivering data as a service



Enabling cross-group collaboration





Sharing data between environments

Amazon Redshift Data API

Amazon Redshift Data API

Simplifies data access from web services based applications

The Data API simplifies access to Amazon Redshift with no database connections and credentials to manage

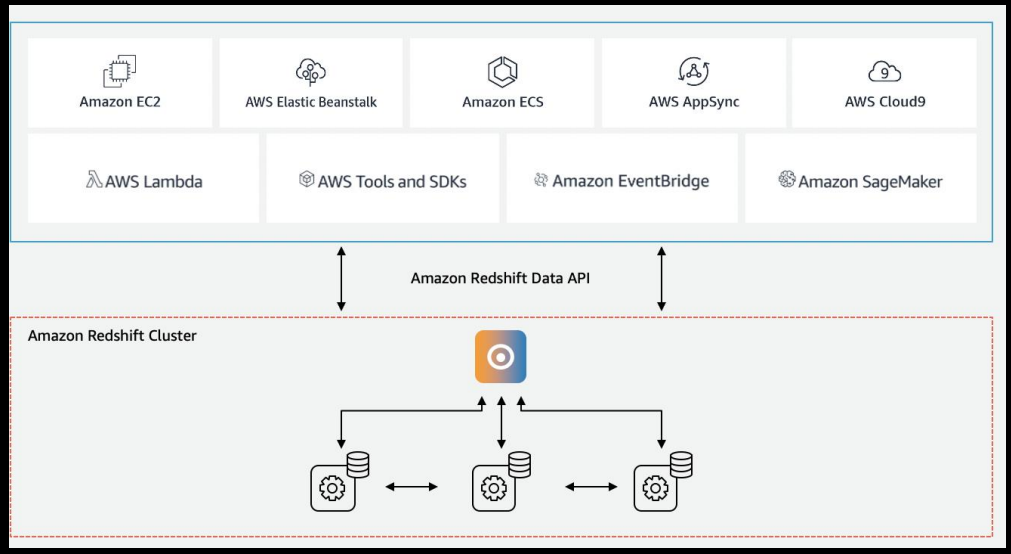
Use AWS SDK from languages such as Python, Go, Java, Node.js, PHP, Ruby, and C++ by invoking an API endpoint

Query, load and unload data from CLI/SDK

Easy to build web-services based applications and integrate with services like AWS Lambda, AWS AppSync, and AWS Cloud9

Results stored for 24 hours and is retrieved asynchronously

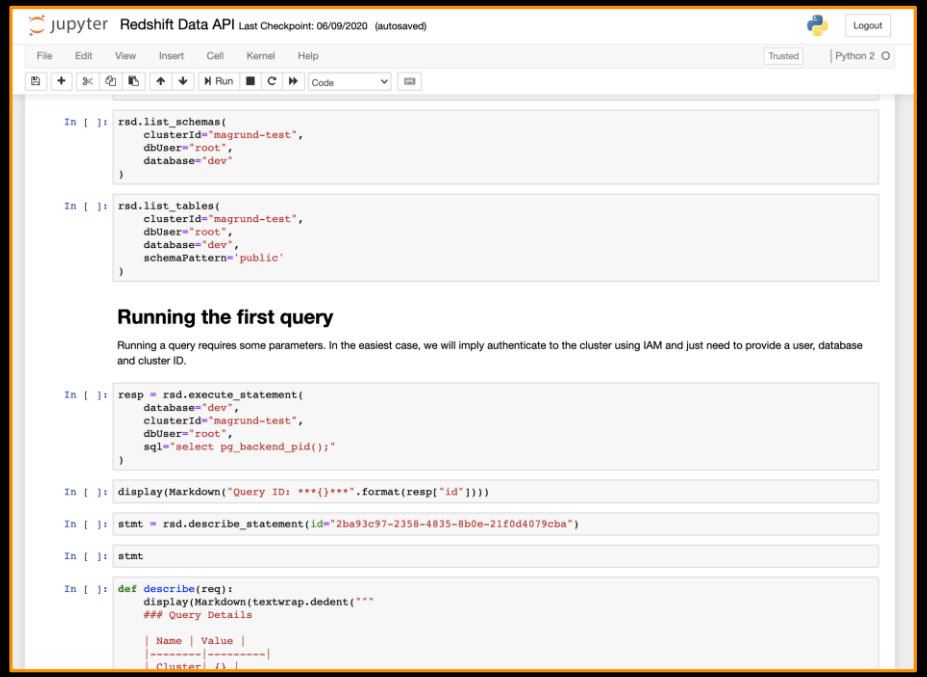
```
aws redshift-data execute-statement
--database [DATABASE]
--query [QUERY]
--secret-arn [CREDENTIALS_ARN]
```



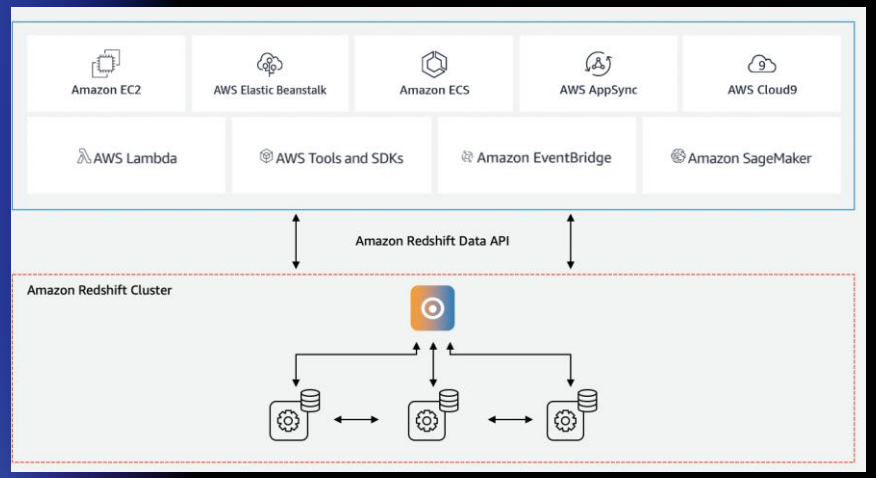
Amazon Data API - Authentication

- Authorize a user to access the Data API by adding a predefined AWS IAM policy
- *AmazonRedshiftDataFullAccess* managed policy provides full access to Amazon Redshift Data API operations
- No persistent connection; instead use secure HTTP endpoint and SDKs
- Asyncronous calls with two authentication methods:
  - AWS Secrets Manager: use secret-arn secret value stored in AWS Secrets Manager.
  - Temporary credentials: provide your cluster-identifier, database, and db-user values.

Improved notebook experience with Amazon Data API



Common use cases



- Integrating web services-based applications to access data
- Long-running queries
- Run query once and retrieve results multiple times
- Building ETL pipelines with AWS Step Functions, AWS Lambda, and stored procedures
- Simplified access for Amazon SageMaker notebooks for easy access by data scientists and analysts
- Enables event-driven pipelines and scheduled query execution

SUPER data type with JSON support

Native semi-structured data support

New data type: **SUPER**

Easy, efficient, and powerful JSON processing

Fast row-oriented data ingestion

Fast column-oriented analytics with materialized views over SUPER/JSON

Access to schema-less nested data with easy-to-use SQL extensions powered by the PartiQL query language

id INTEGER	name SUPER	phones SUPER
1	{"given":"Jane", "family":"Doe"}	[{"type":"work", "num":"9255550100"}, {"type":"cell", "num": 6505550101}]
2	{"given":„Richard", "family":„Roe"},	[{"type":"work", "num": 5105550102}]

```
SELECT name.given AS firstname, ph.num
FROM customers c, c.phones ph
WHERE ph.type = 'cell';

firstname | num
----------+---------------
"Jane"    | 6505550101
```

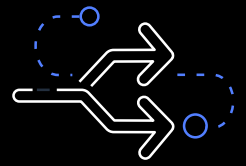
AQUA (Advanced Query Accelerator) for Amazon Redshift

RA3 nodes with managed storage

SCALE COMPUTE AND STORAGE INDEPENDENTLY

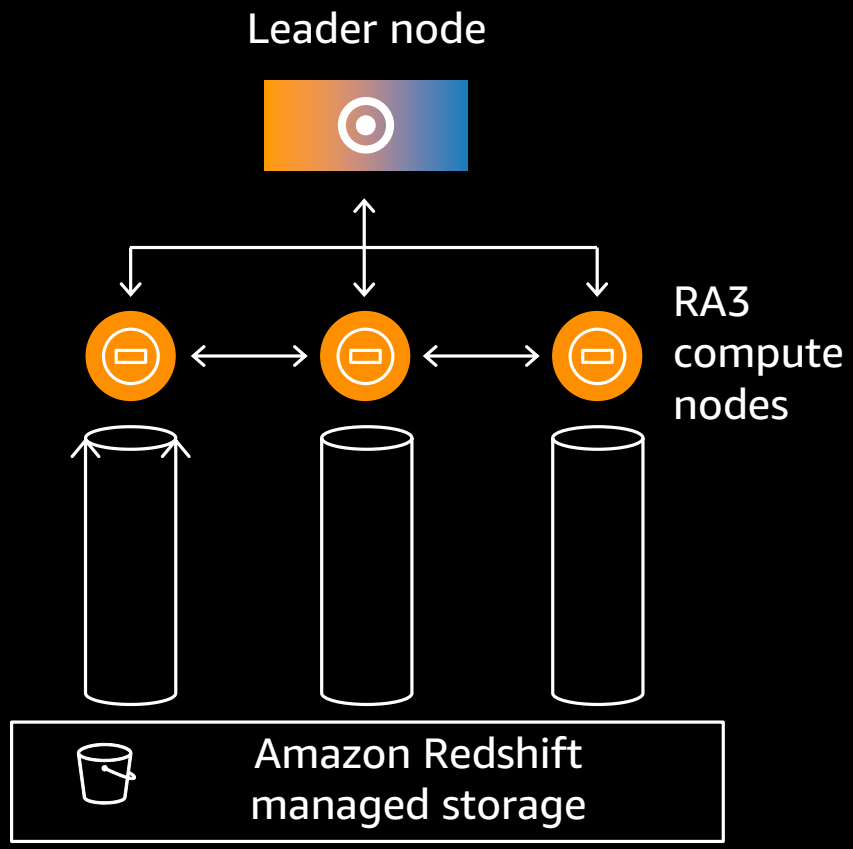
Managed storage

Large high-speed cache



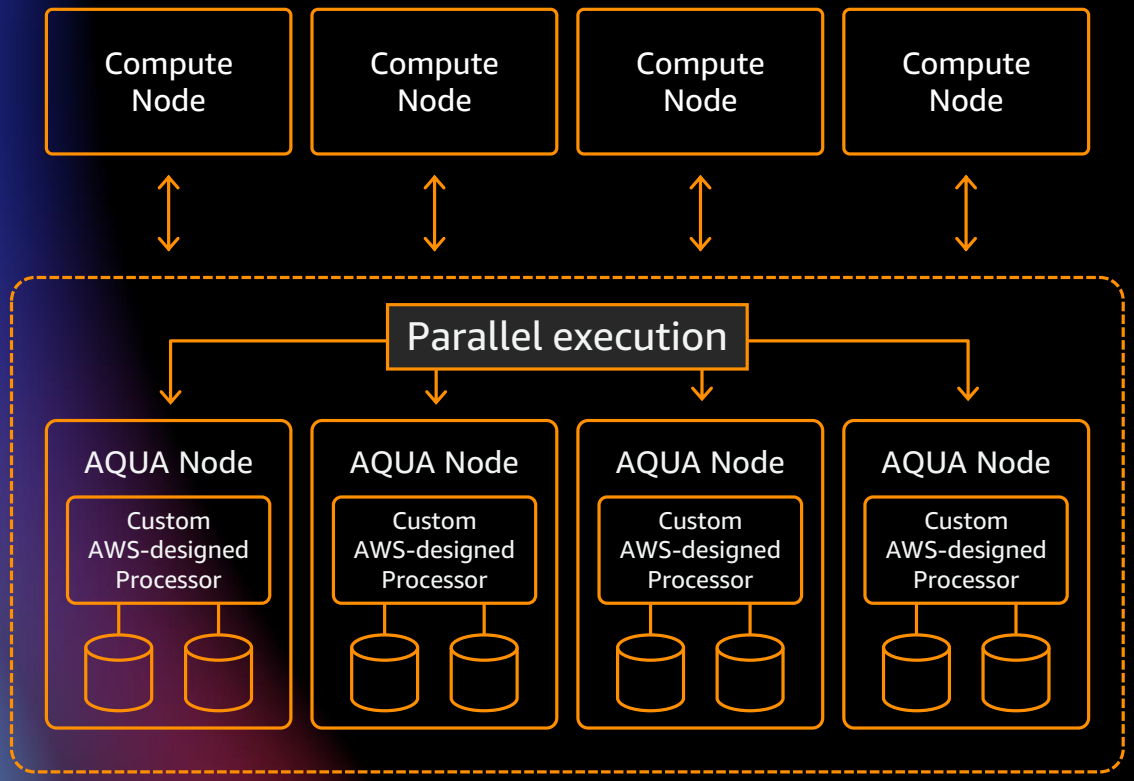
High-bandwidth networking

- Size of data warehouse only based on steady state compute needs
- Scale and pay independently for compute and storage
- Automatic, no changes to any workflows, no need to manage storage



AQUA (Advanced Query Accelerator)

NEW HARDWARE-ACCELERATED CACHE THAT DELIVERS BETTER QUERY PERFORMANCE THAN OTHER CLOUD DATA WAREHOUSES



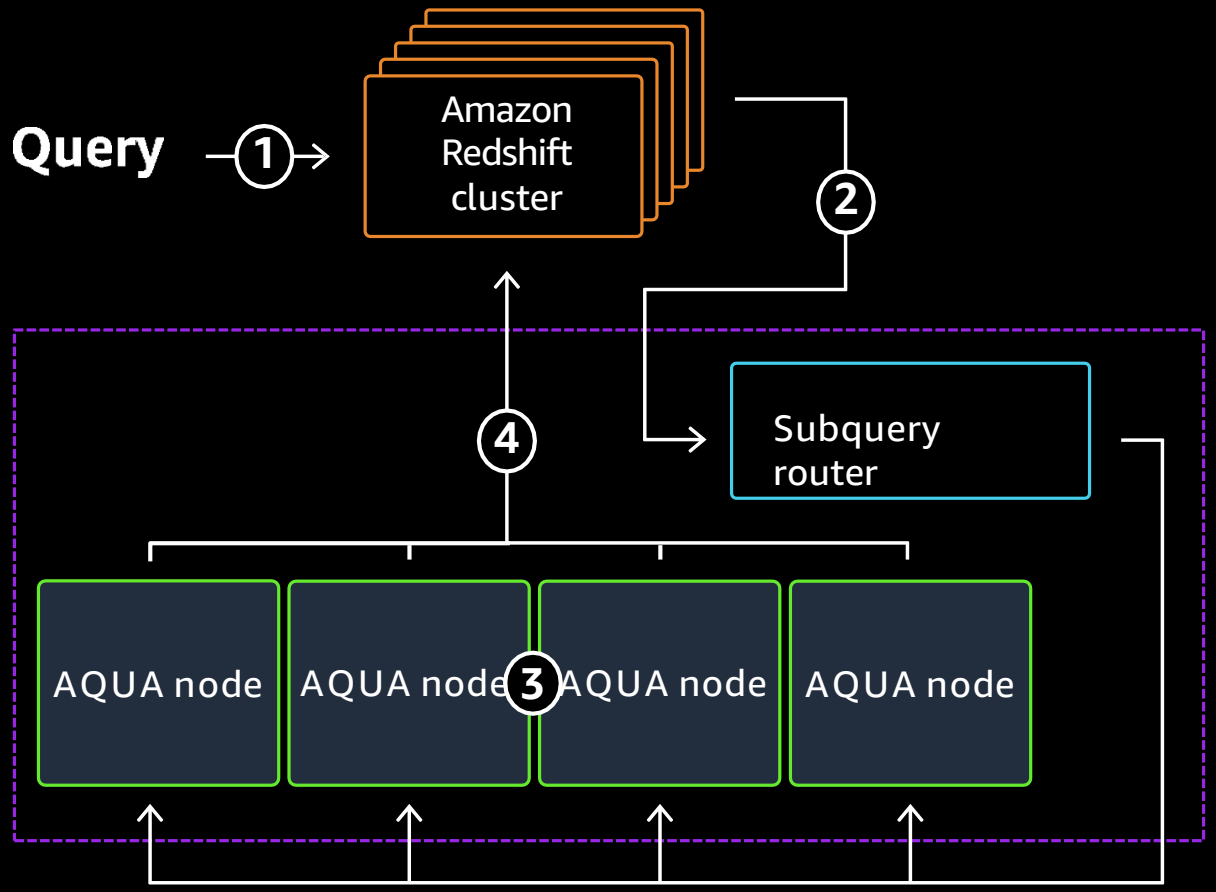
Minimize data movement over the network by pushing down operations to AQUA Nodes

AQUA Nodes with custom AWS-designed analytics processors to make operations (compression, encryption, filtering, and aggregations) faster than traditional CPUs

Available only with RA3, no code changes required, no additional cost

Query Lifecycle in AQUA

1. Amazon Redshift receives a query; builds and optimizes a query tree (execution plan)
2. Ships subqueries with SCAN / AGG to AQUA
3. The AQUA subquery router receives the query and distributes it across multiple AQUA nodes to process in parallel
4. AQUA nodes process the query, including hydrating data from Amazon S3, and send the results back to Amazon Redshift



Demo

Additional resources

AWS BLOG POSTS

Using the Amazon Redshift Data API to interact with Amazon Redshift clusters

Sharing Amazon Redshift data securely across Amazon Redshift clusters for workload isolation

Work with semi-structured data using Amazon Redshift SUPER

Visit the AWS Data Resource Hub

Dive deeper with these resources, get inspired and learn how you can use data to make better decisions and innovate faster.

- Building a winning data strategy
- The new leadership mindset for data & analytics
- Harness data to reinvent your organization
- Put your data to work with a modern analytics approach
- Breaking free from on-premises database constraints
- Cloud storage adoption: From cost optimization to agility & innovation
- A strategic playbook for data, analytics, and machine learning
- … and more!

https://tinyurl.com/aws-data-resource

Visit resource hub

AWS Training and Certification

Empower your teams with comprehensive training

By building skills with AWS Training and Certification, businesses and individuals can see the bigger picture understanding the reasoning behind every data point. As training progresses and teams become data-fluent, previously hidden insights come into view.

Leverage free digital training

Learn how to harness the world's most valuable resource: data. Access digital and virtual instructor-led courses on data analytics and databases built by the experts at AWS and start your learning journey to become data-driven.

Take a digital course »

Get certified

Earn industry-recognized credibility and set tangible goals for success with industry-recognized certifications, like *AWS Certified Data Analytics – Specialty.*

Learn more »

Ramp-up your skills

Deep dive into new topics and focus on knowledge gaps at your own pace with the *AWS Ramp-Up Guide: Database* and *AWS Ramp-Up Guide: Data Analytics*. With a wide range of whitepapers, blog posts, videos, webinars and peer resources available for data professionals to leverage for independent learning.

Download ramp-up guides »

Thank you for attending AWS Innovate – Data Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.


aws-apj-marketing@amazon.com


twitter.com/AWSCloud


facebook.com/AmazonWebServices


youtube.com/user/AmazonWebServices


slideshare.net/AmazonWebServices


twitch.tv/aws

Thank you!