
AWS
re:Invent

ANT203

What's new with Amazon Redshift

Eugene Kawamoto
Director, Product Management, Amazon Redshift
AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Amazon Redshift overview

Key investment areas

Key innovations and new features

Summary



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.


More data is created every *hour* today
than in an entire year just 20 years ago
(Source: IDC Seagate Rethink Data Survey, January 2020)

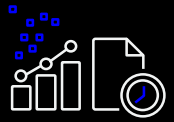
Challenges of data analytics at scale



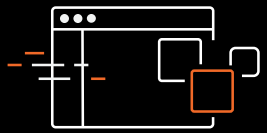
Variety of sources and data types



Multiple analytics needs



Data volume and velocity



Slow performance



Difficult to manage systems



Complex to scale



Increasing and unpredictable cost



Inflexible tools

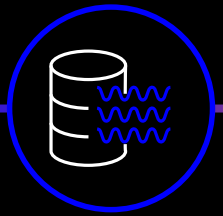


Security, compliance

Amazon Redshift

THE MOST WIDELY USED CLOUD DATA WAREHOUSE, WITH TENS OF THOUSANDS OF CUSTOMERS

ANALYZE ALL YOUR DATA



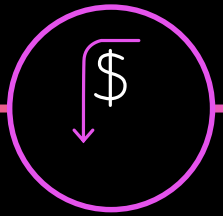
Take a **lake house approach** by analyzing all your data across your data warehouse, your Amazon S3 data lake, and operational databases with consistent security and governance policies

PERFORMANCE AT ANY SCALE



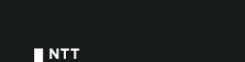
Get up to **3x better price performance** than other cloud data warehouses with a **self-tuning** system, and boost queries up to **10x with AQUA**

LOWER YOUR COSTS



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

Tens of thousands of customers process exabytes of data with Amazon Redshift daily



NTT DOCOMO

Moved >10 PB of data from on-premises to cloud



FOX Corp.

Taking a lake house approach with RA3 nodes and Amazon S3



Yelp

Enabling a data-driven organization with concurrency scaling



Jack in the box

Improved ops by moving off of on-premises DW



Warner Bros. Games

Performance, scale, cost-effective





































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
- EARLY Q1! — AQUA
- NEW! — Performance tuning: automated
- UPDATED! — Materialized views
- NEW! — 100K tables
- NEW! — HyperLogLog
- Concurrency scaling



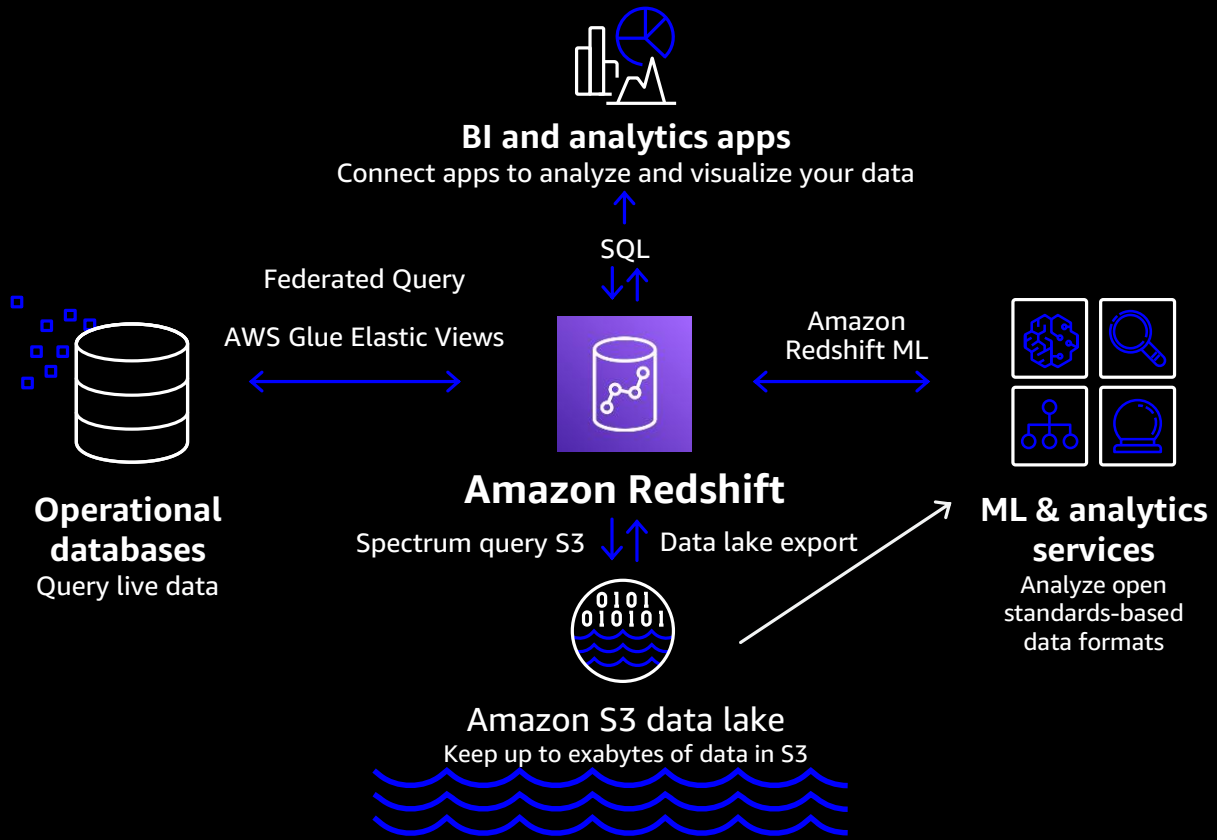
Low cost & best value

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

Analyze all your data

WITH A LAKE HOUSE APPROACH TO ANALYTICS

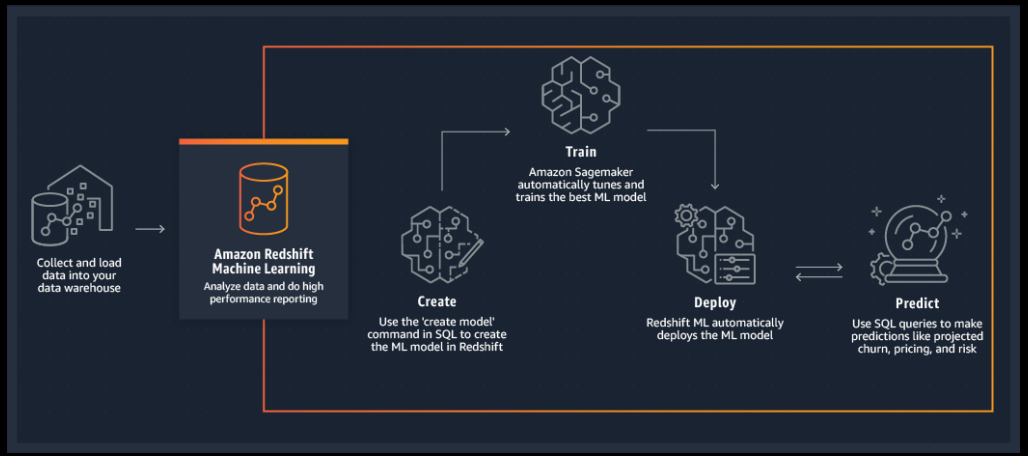


Amazon Redshift ML

PREVIEW

EASILY CREATE AND TRAIN ML MODELS USING SQL QUERIES WITH AMAZON SAGEMAKER

- Use case: Product recommendations, fraud prevention, reduce customer churn
- Create, train, and apply ML models using SQL
- Automatic selection of ML algorithms or select your algorithm with XGBoost
- Automatic pre-processing, creation, training, deployment of your model
- Deploy inference models locally in Amazon Redshift; run an inference as invoking a user-defined function as part of SQL statements

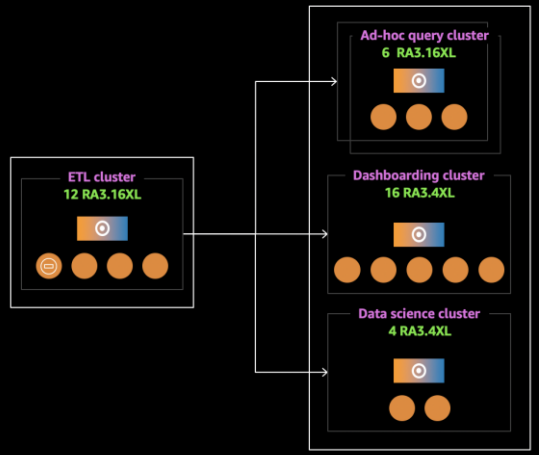


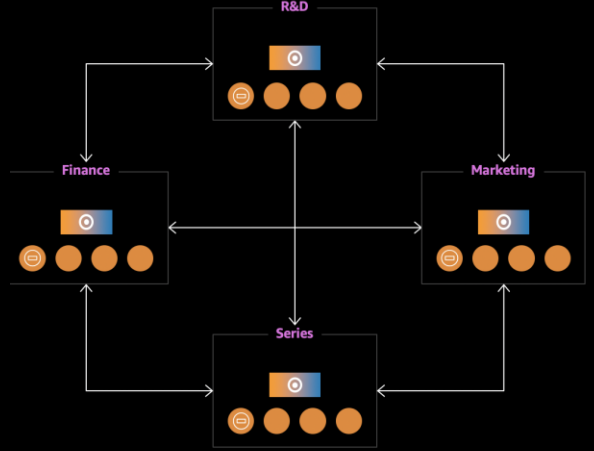

```
CREATE MODEL demo_ml.customer_churn
FROM (SELECT c.age, c.zip, c.monthly_spend,
c.monthly_cases, c.active FROM
customer_info_table c)
TARGET c.active;
```

Data sharing

PREVIEW

A SECURE AND EASY WAY TO SHARE DATA ACROSS AMAZON REDSHIFT CLUSTERS





- Instant, granular, high-performance data access without data copies / movement
- Live and consistently updating views of data across all consumers
- Secure and governed collaboration within and across organizations and with external parties

- Workloads accessing shared data are isolated from each other
- Use cases: Cross-group collaboration and sharing, workload isolation and chargeability, data as a service
- Sharing to other AWS analytic services – coming soon

yelp

"Data sharing feature seamlessly allows multiple Amazon Redshift clusters to query data located in our RA3 clusters and their managed storage. This eliminates our concerns with delays in making data available for our teams, reduces the amount of data duplication and associated backfill headache. We now can concentrate even more of our time making use of our data in Amazon Redshift and enable better collaboration instead of data orchestration."

Steven Moy, Yelp

Native semi-structured data support

PREVIEW

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';

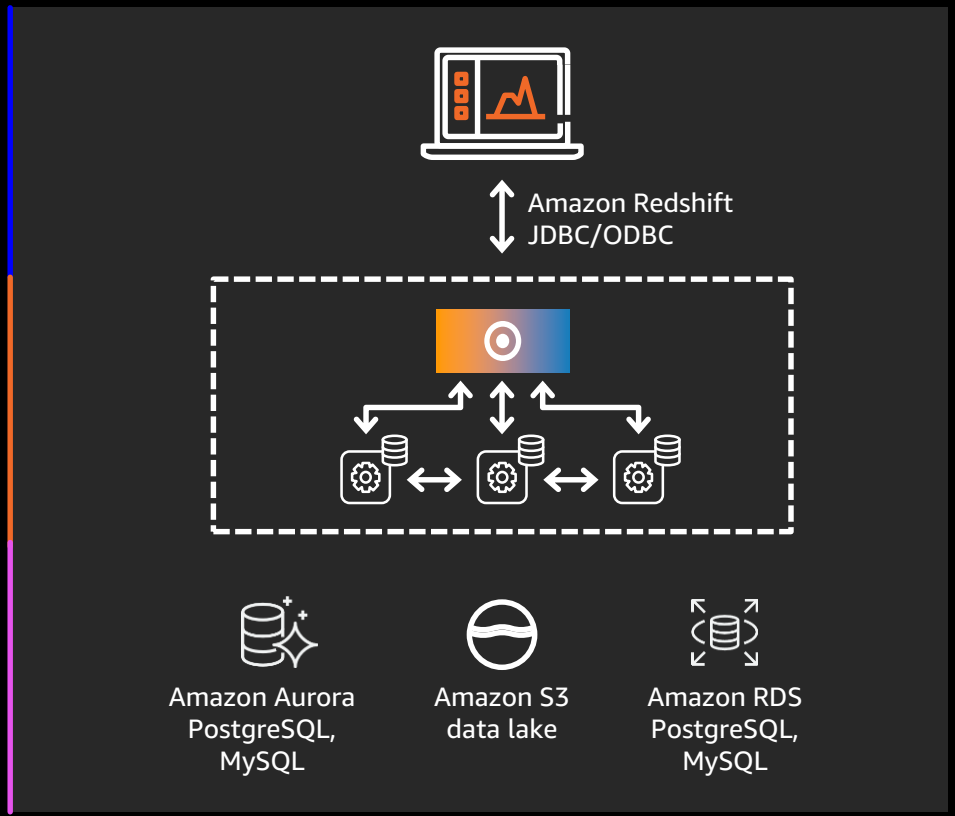
firstname | num
----------+---------------
"Jane"    | 6505550101
```

Amazon Redshift Federated Query

UPDATED

UNIFIED ANALYTICS ACROSS DATABASES, DATA WAREHOUSE, AND DATA LAKE

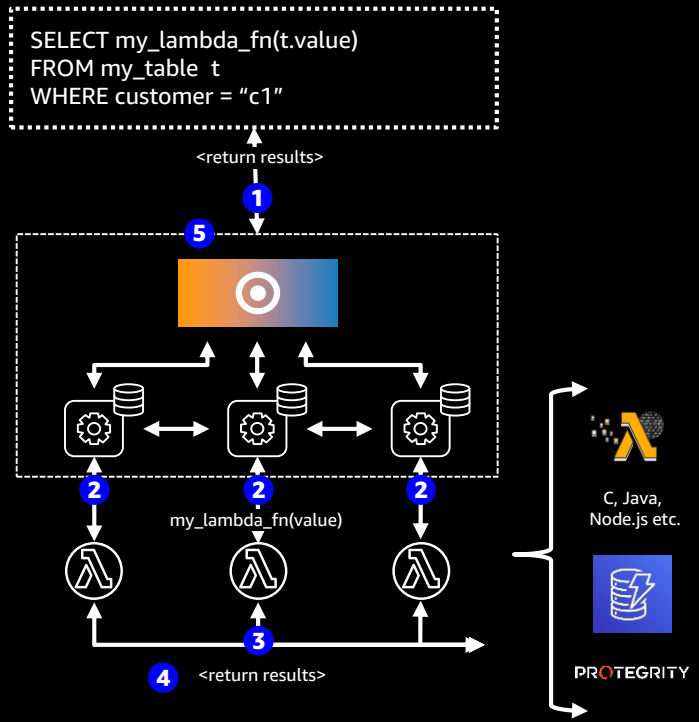
- Use case: Integrate operational data with DW and data lake for real-time analytics
- Analytics on operational data without data movement and ETL delays
- Query and join data from one or more Amazon RDS and Aurora PostgreSQL databases
- Flexible and easy way to ingest data avoiding complex ETL pipelines
- Intelligent distribution of computation to remote sources to optimize performance
- Preview — Amazon RDS and Aurora MySQL support



Lambda UDFs

NEW

INTEGRATE WITH EXTERNAL SERVICES USING AWS LAMBDA



Invoke AWS Lambda programs as UDFs in Amazon Redshift SQL queries

Simple integration with external services

- Tokenization with third-party vendors like Protegrity
- More languages runtimes (C++, Java, etc.)
- Access Amazon DynamoDB, Amazon SageMaker, etc.

Concurrent and batch processing

Cost controls and error controls

Native partner integration

PREVIEW

INTEGRATE WITH SELECT PARTNERS FROM THE AMAZON REDSHIFT CONSOLE











Ingest business data and gain insights in minutes

Deep links into partner platforms streamline account setup

Centralized dashboard for all partner integration statuses

Reach out to redshift-partners@amazon.com to integrate your product into the Amazon Redshift console today!

Get the best price performance and scale

Up to **3x**

Amazon Redshift has better price performance than other cloud data warehouses

- Customers like Duolingo, Yelp, and Codecademy are doubling performance and saving costs with RA3
- Amazon Redshift has up to 3x better price performance than other cloud data warehouses*
- Scales data volume from 1TB to 1PB+ with predictable cost and performance
- Boost queries up to 10x with AQUA for Amazon Redshift

RA3 nodes with managed storage

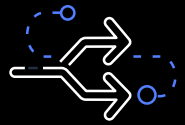
SCALE COMPUTE AND STORAGE INDEPENDENTLY



Managed storage

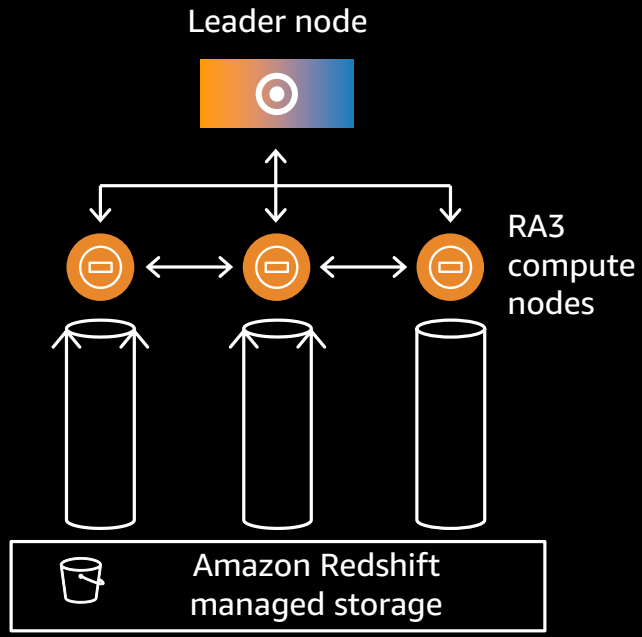


Large high-speed cache



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



UPDATED

RA3 compute nodes

MULTIPLE NODE SIZES AVAILABLE TO PROVISION THE RIGHT AMOUNT OF COMPUTE

RA3.16xl

48 vCPU, 384 GiB Memory, 8 GB IO

On-demand price— $13.04/node/hour

RA3.4xl

12 vCPU, 96 GiB Memory, 2 GB IO

On-demand price— $3.26/node/hour

NEW

RA3.xlplus

4 vCPU, 32 GiB Memory, 650 MB IO

On-demand price— $1.09/node/hour

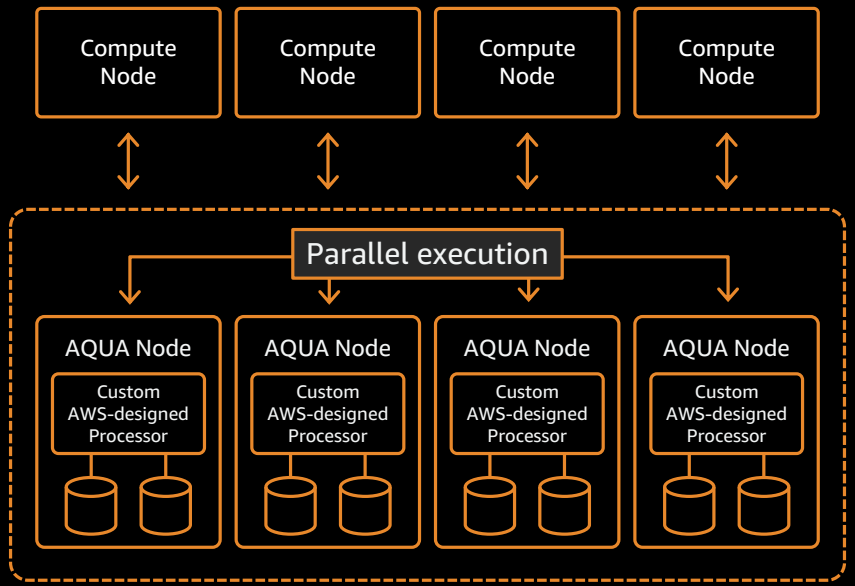
Scale to tens of PB of data; pay as you go at $0.024/GB-month for storage

Save more with Reserved Instances: 33% discount over OD with 1-year and 60% with 3-year

AQUA (Advanced Query Accelerator)

EARLY Q1

NEW HARDWARE-ACCELERATED CACHE THAT DELIVERS UP TO 10X 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

Concurrency scaling

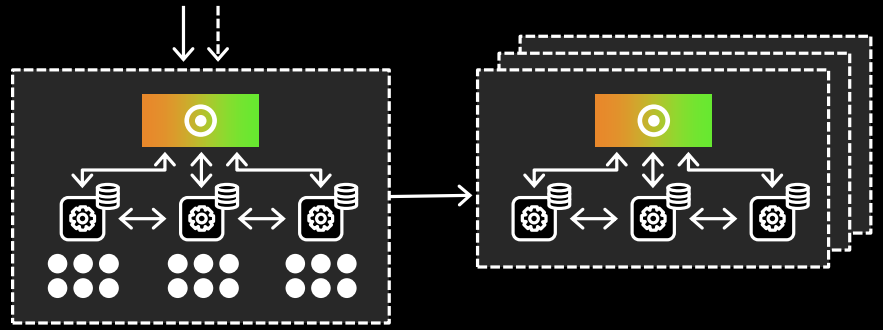
COMPUTE ELASTICITY AND SCALABILITY TO HANDLE UNPREDICTABLE USER DEMAND

Scale out to multiple Amazon Redshift clusters from a single endpoint in seconds

Support virtually unlimited concurrent users and queries while maintaining SLAs

Per-second billing for additional clusters used

Cost controls and free one-hour usage per day



Amazon Redshift automates performance tuning

NEW

ML-BASED OPTIMIZATIONS TO GET STARTED EASILY AND GET THE FASTEST PERFORMANCE QUICKLY

Automates physical data design and optimization

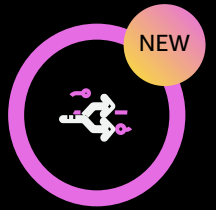
Optimizes for peak performance as data and workloads scale

Leverages machine learning to adapt to shifting workloads

Automated performance tuning



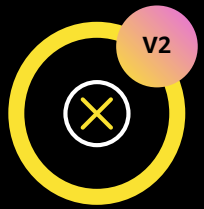
Automatic vacuum delete



Automatic distribution keys



Automatic sort keys



Auto workload manager



Automatic table sort



MV auto-refresh and rewrite

Materialized views auto refresh and query rewrite

UPDATED

SPEED UP QUERY PERFORMANCE BY ORDERS OF MAGNITUDE WITH PRECOMPUTED RESULTS

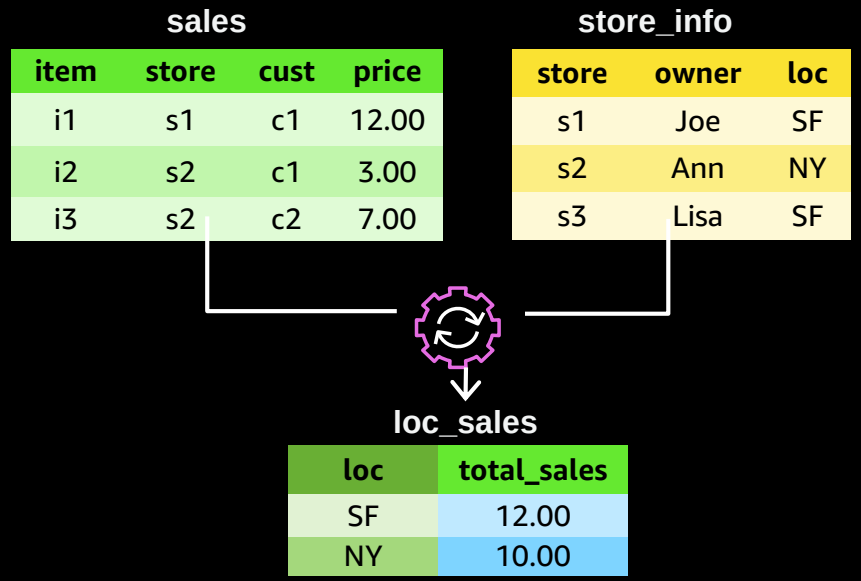
Simplify and accelerate iterative and predictable workloads, such as ETL, BI/dashboarding queries

MVs can be based on one or more Amazon Redshift tables or external tables (Spectrum, Federated)

Efficient incremental maintenance

Scheduled, automatic, or manually timed refresh

Amazon Redshift auto query rewrite optimizes queries by replacing native tables with materialized views



"The Amazon Redshift materialized view auto query rewrite feature reduced dashboard load times from 8 minutes to just 500 ms. The best part is that this is completely transparent for Tableau and the business user."

Arman Nasrollahi, Home24

Low cost and best value

WITH PREDICTABLE PRICING AND A FAST PACE OF INNOVATION

At least **50%**

less expensive than other cloud data warehouses

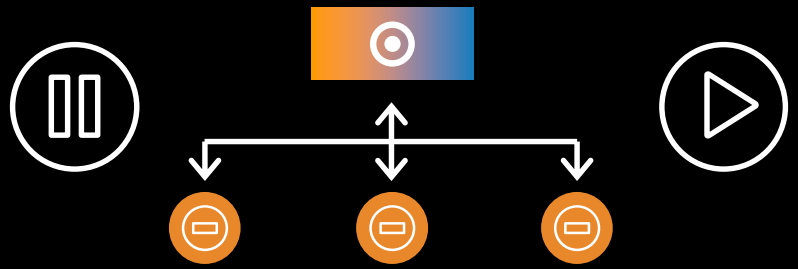
Start small at $0.25 per hour and pay as you go, or save up to 60% with Reserved Instances

At least 50% less expensive than other cloud data warehouses, with predictable costs

Use pause and resume, and set cost controls

Cluster pause and resume

EASILY START AND STOP A CLUSTER TO SAVE COST FOR INTERMITTENT WORKLOADS



Pause a development or QA cluster during non-business hours

Run an ETL cluster only during load and unload

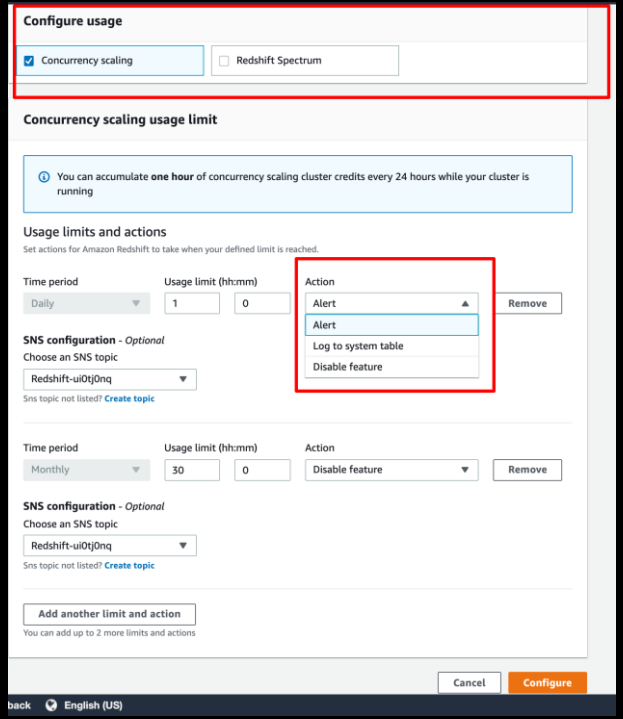
Cost-effectively manage a cluster with infrequently accessed data

Pause and resume cluster operations with 1 click in the console or via API

Pause and resume using a user-defined schedule

Cost controls

CONTROL YOUR SPEND BY CREATING USAGE LIMITS FOR CONCURRENCY SCALING AND AMAZON REDSHIFT SPECTRUM



Create daily, weekly, and monthly limits

Up to 4 limits per feature

Log to system table, Alert, Disable feature

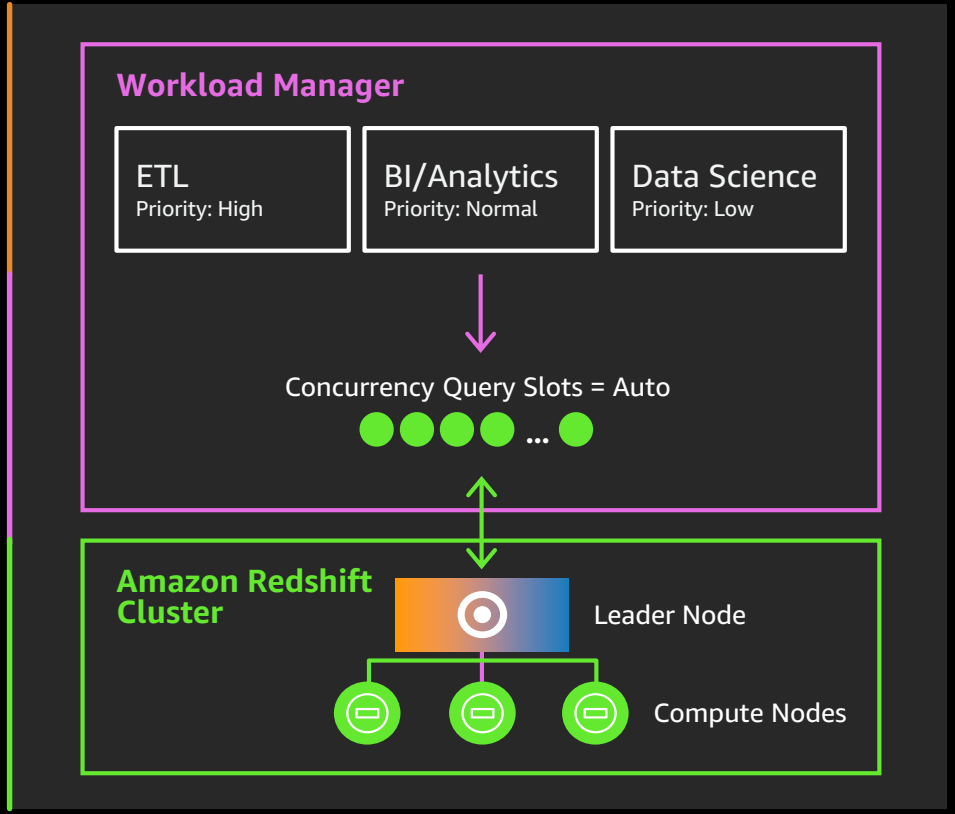
Automatically generate Amazon CloudWatch alarm

Automatic WLM and query priorities

UPDATED

OPTIMIZE RESOURCES FOR MAXIMUM QUERY THROUGHPUT AND CONSISTENT PERFORMANCE

- Ensure that high-priority workloads get preferential treatment
- Prevent expensive queries from consuming system resources
- Uses machine learning to optimize query throughput
- Ensures highly efficient resource utilization for maximum ROI
- Fair sharing of resources



Built-in security and compliance

SECURITY AND COMPLIANCE FEATURES WITH NO EXTRA COSTS WITH AMAZON REDSHIFT

Authentication

- IAM integration
- ID federation
- Azure AD, AD, Okta, Ping Federate
- Multifactor authentication

Access control

- Column-level privileges for Amazon Redshift and data lake

Audit

- AWS CloudTrail integration
- Macie integration

Encryption

- AWS KMS integration
- Encrypted data in motion, data at rest
- Tokenization with Lambda UDFs and third-party tools

Helps achieve compliance

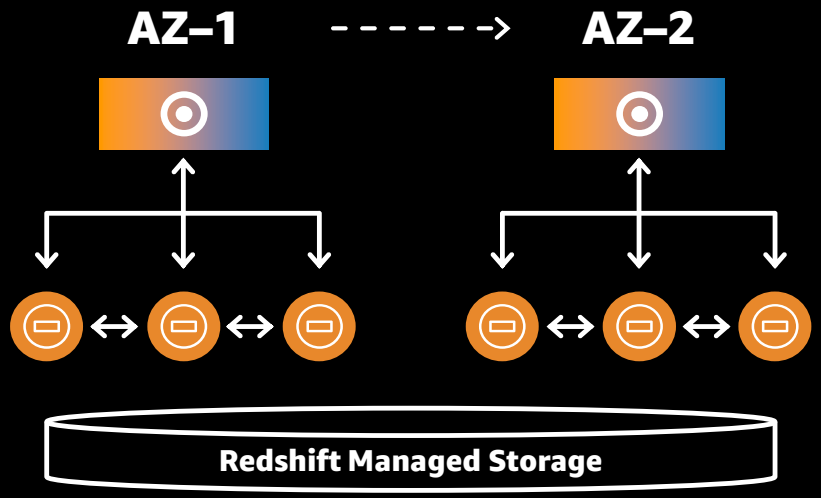
- SOC
- PCI
- FedRAMP
- HIPAA and others

Cross-AZ cluster recovery

NEW

RELOCATE YOUR CLUSTER IN ANOTHER AZ IN RESPONSE TO FAILURE

- Recovery with zero data loss (RP = Zero)
- No need to restore from a snapshot
- On-demand failover
- Cluster is created in another AZ on-demand, so cost of a standby replica cluster is avoided
- Supported on the RA3 instance family



Amazon Redshift is the most widely used cloud data warehouse



Tens of thousands of customers use Amazon Redshift and process exabytes of data per day

It’s still Day 1, and the pace of innovation is accelerating

We’ll continue to innovate and add new features around key investment areas

Next steps

Already using Amazon Redshift?

Ask your account team for a free Amazon Redshift optimization session

Learn more about what's new
https://aws.amazon.com/redshift/whats-new/

Getting started with Amazon Redshift?

Ask your account team for a 10-minute demo

Learn more about starting a proof of concept
https://docs.aws.amazon.com/redshift/proof-of-concept-playbook.html

Or request help from AWS
https://pages.awscloud.com/redshift-proof-of-concept-request

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.