Amazon Redshift Data Sharing

Build scalable multi-tenant architectures

Milind Oke
Analytics Specialist Solutions Architect
Amazon Web Services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Introduction to Amazon Redshift

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Why Amazon Redshift for your data needs?

FAST, EASY, AND SECURE ANALYTICS AT SCALE

Easy analytics for everyone

Focus on getting from data to insights in seconds without worrying about infrastructure

Analyze all your data

Get insights running real-time and predictive analytics on complex, scaled data across your operational databases, data lake, and data warehouse

Best price performance at any scale

Gain up to 3x better price performance than other cloud data warehouses, and dynamically scale to improve query speed for complex and critical workloads

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Amazon Redshift innovates to meet your needs

Easy analytics for everyone

- NEW! Serverless
- NEW! Query Editor v2
- Updated! Automated DW Management
- NEW! Automatic Materialized Views
- Data API
- Updated! Redshift Advisor
- Updated! Cloud Formation templates
- NEW! Grafana Plugin

Analyze all your data

- Updated! Data sharing
- NEW! ADX Integration
- NEW! Streaming Ingestion
- Updated! Amazon Redshift ML
- Updated! Federated Query
- Updated! Geospatial enhancements
- SUPER data type with JSON

Performance at any scale

- Updated! RA3 nodes & managed storage
- AQUA
- Updated! Workload Management enhancements
- NEW! Concurrency scaling for Writes
- Updated! SQL Enhancements & migration support
- Security, Governance, & Compliance

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

What is the Amazon Redshift service?

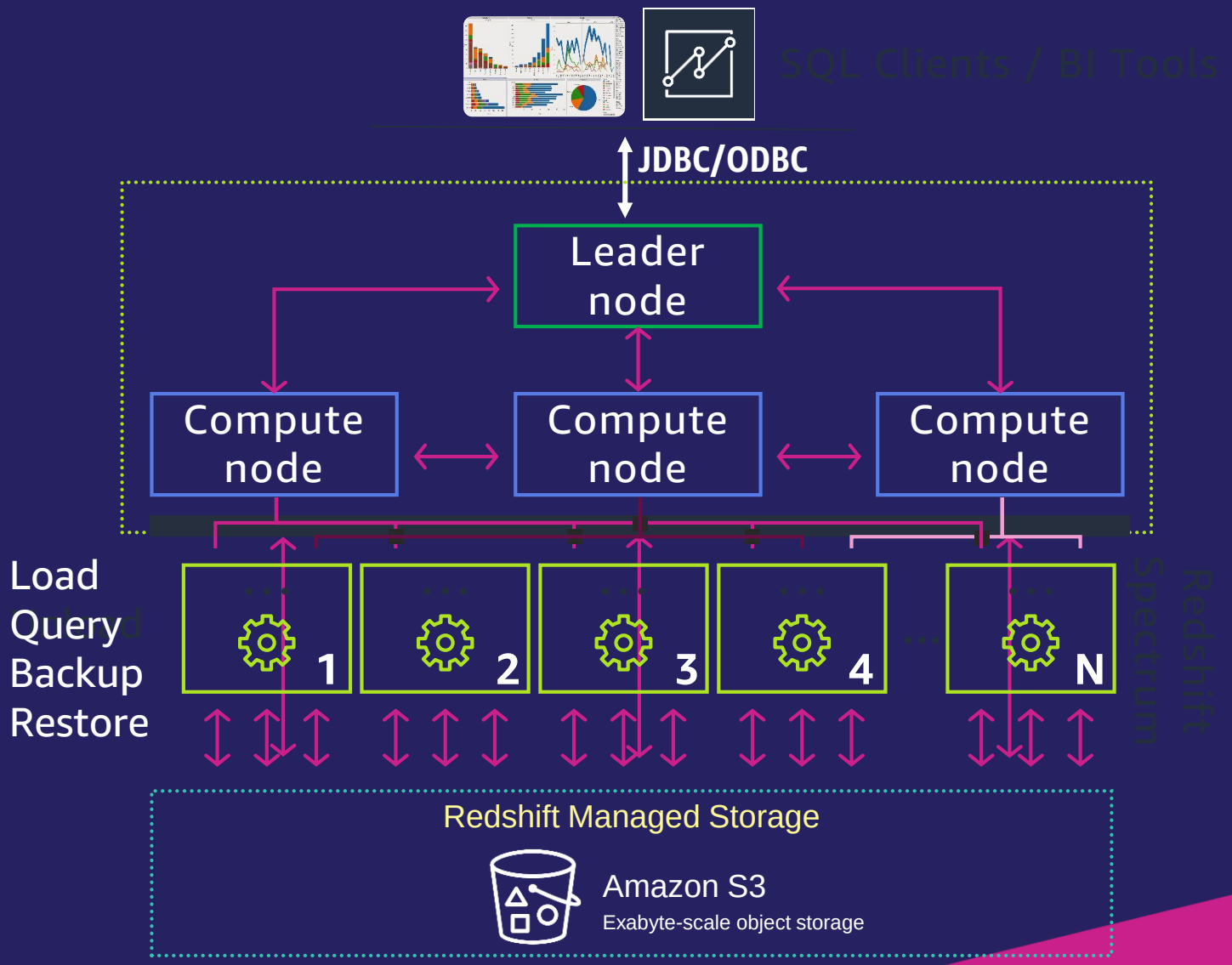
Not a standalone data warehouse, but a data warehouse that breaks down the silos to keep data "free"

Fully Managed	Automated maintenance & workload management; cost-effective cloud data warehouse	Scale	Query GBs to EBs; auto scaling; independent compute/storage scaling
Superior Speed	Extensive machine learning based optimizations and features	Highly-rated & Most Popular	Tens of thousands of deployments; highly-rated by agencies such as Gartner
Highly-resilient	Service SLA: 99.9%	RDS, ML, and Data Lake Integration	Query data in-place in your data lake and RDS; ACID and ANSI SQL
Secure	End-to-end encryption; SSO; compliance with SOC 1/2/3, HIPAA, FedRamp & more	Commitment to Open Formats	Query open formats in place Import/Export Parquet & CSV; Query ORC, Avro, JSON, …

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Redshift cluster architecture

- Leader node
  - SQL endpoint
  - Stores metadata
  - Coordinates parallel SQL processing & ML optimizations
  - Leader node is no-charge for clusters with 2+ nodes
- Compute nodes
  - Local, columnar storage
  - Executes queries in parallel
  - Load, unload, backup, restore from S3
- Amazon Redshift Spectrum nodes
  - Execute queries directly against data lake
- Massively parallel, shared nothing architecture



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Amazon Redshift Data Sharing

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Data Sharing

A secure and easy way to share data across your Amazon Redshift clusters

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

Live and consistently updating views of data across all consumers

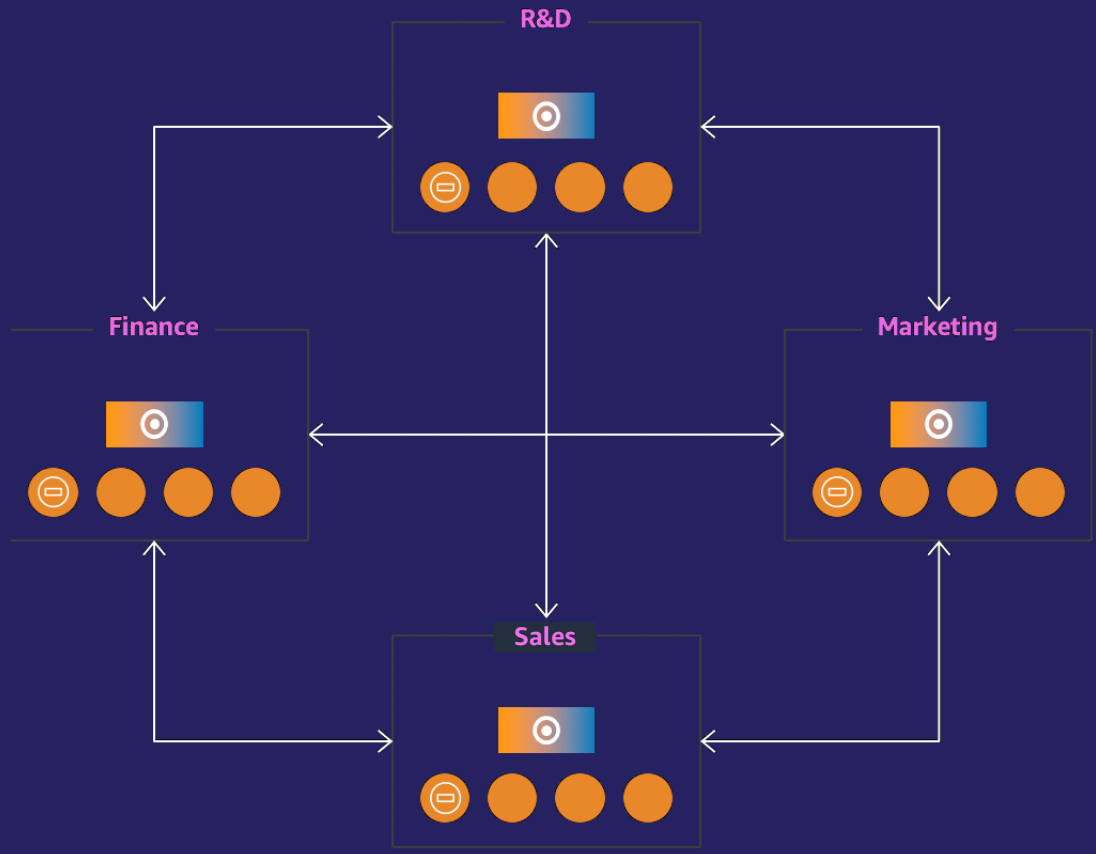
Secure and governed collaboration within and across organizations

Workloads accessing shared data are isolated from each other

NEW Cross-account* data sharing

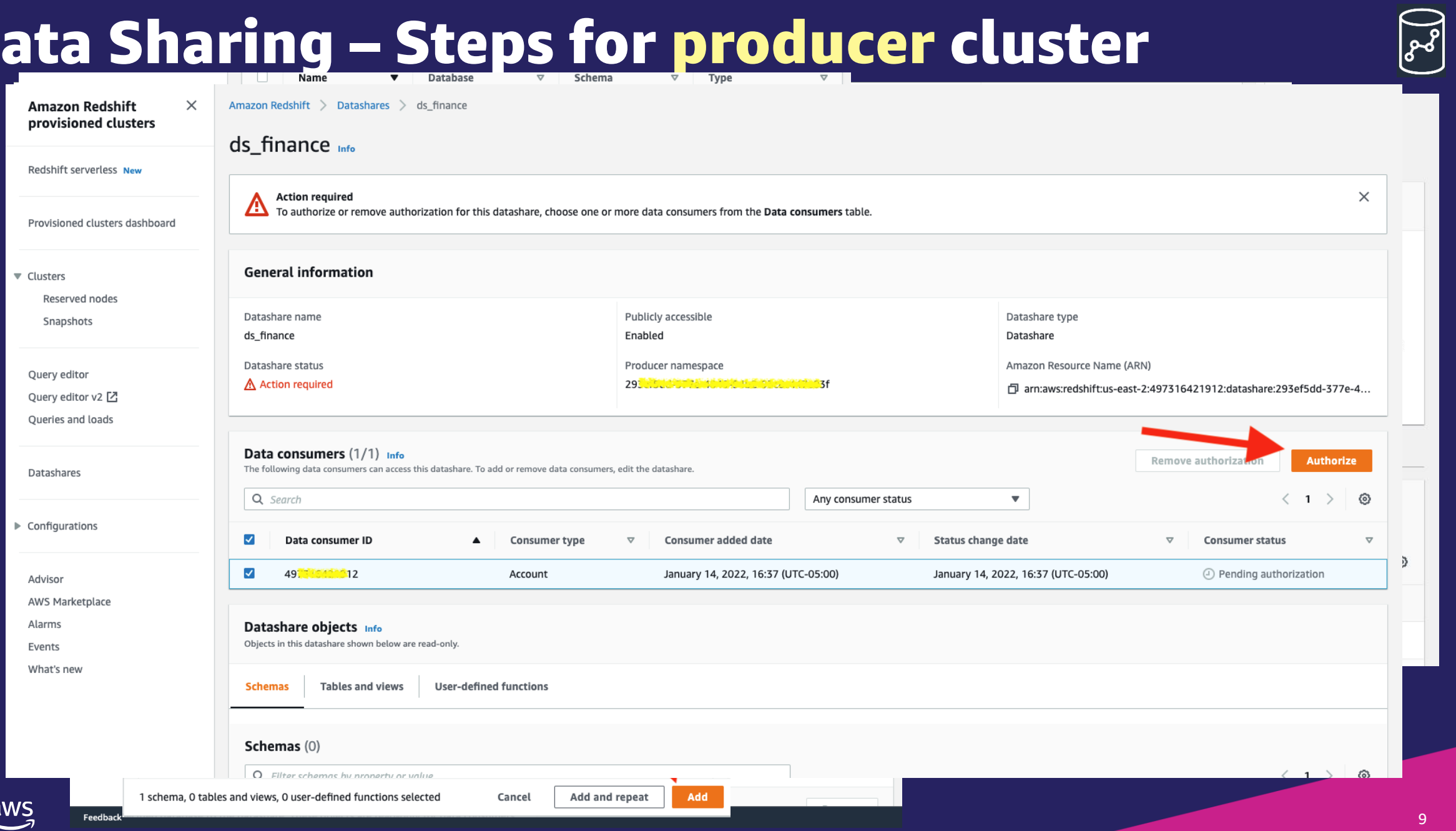
NEW PREVIEW Cross-Region* data sharing

NEW Performance Enhancements: Result Caching & Concurrency scaling for consumer clusters

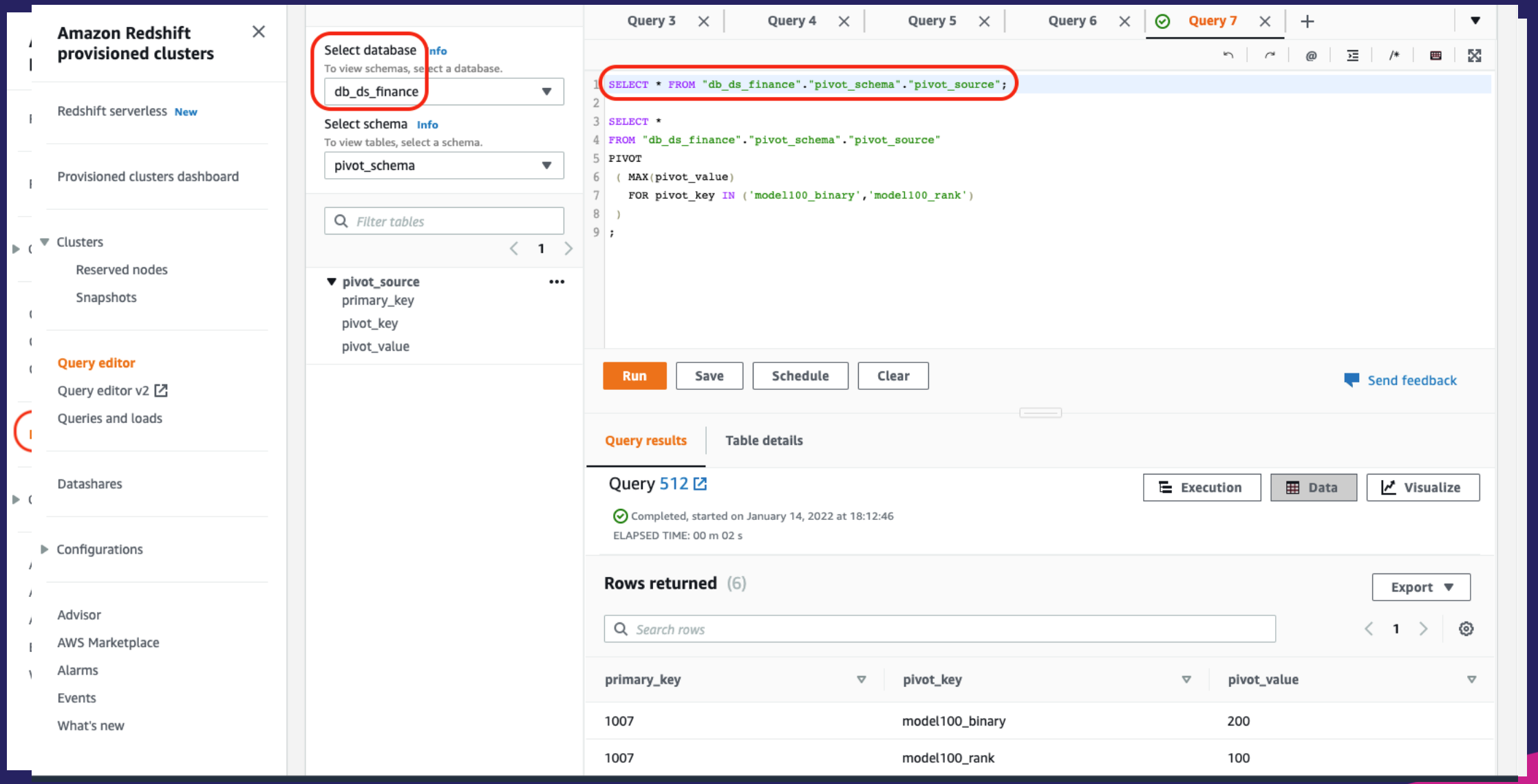


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Data Sharing – Steps for producer cluster



Data Sharing – Steps for consumer cluster

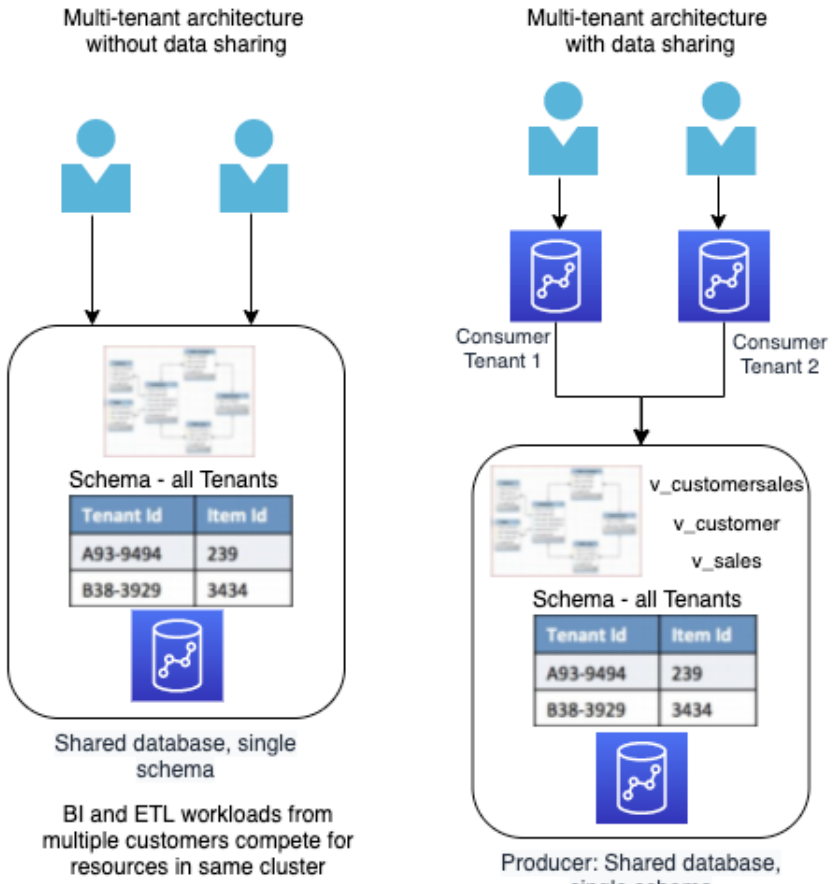


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Multi-tenant patterns in Amazon Redshift using Data Sharing

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Amazon Redshift Data Sharing via Pool Model



Pool Model

Data is stored in a single database schema for all tenants, and a new column is used to scope and control access to individual tenant data. Access to the multi-tenant data is controlled using views.

1. Create new table TENANT and capture the consumer clusters namespace
2. Add TENANT ID to all tables in DWH
3. Create views for each table with logic to filter rows by each tenant's namespace
4. Create data share and add views to shares
5. Consumers will see row level filtered results

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Amazon Redshift provisioned clusters

Redshift serverless New

Provisioned clusters dashboard

Clusters
Reserved nodes
Snapshots

Query editor
Query editor v2
Queries and loads

Datashares

Configurations

Advisor
AWS Marketplace
Alarms
Events
What's new

Get consistent, fast query performance for thousands of concurrent queries and users. View tutorial

Amazon Redshift > Provisioned clusters dashboard

Provisioned clusters dashboard Info

Try Amazon Redshift Serverless (Preview) | Purchase reserved nodes | Create cluster

Resources overview

Resource data for US East (N. Virginia) Region.

Total nodes	On-demand nodes	Reserved nodes	Reserved nodes available	Automated snapshots	Manual snapshots
4	4	0	0 (0 of 0 used)	2	0

Cluster overview (2)

Available

Cluster	Status
redshift-dmw-producer	Available

View all clusters

Cluster metrics

Any clusters | Last hour | View in CloudWatch

Number of queries | Database connections | Disk space used | CPU utilization

10

Datashares

Authorize other AWS accounts to access datashares created in this AWS account. Associate or decline datashares from other AWS accounts.

Require authorization	Require association
0	0

Alarms (0)

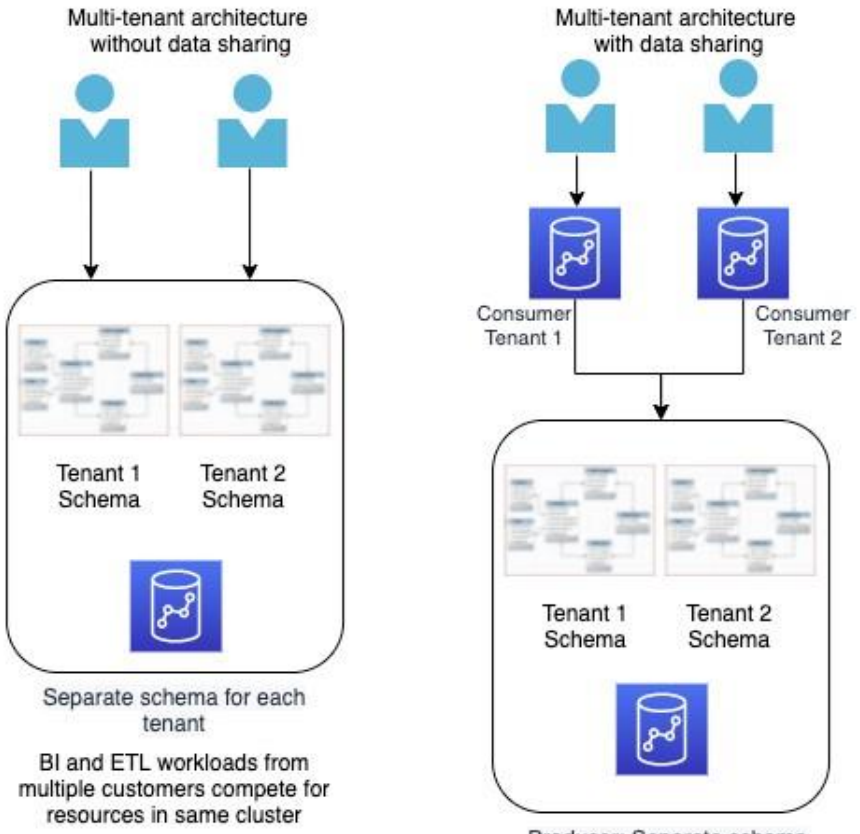
View in CloudWatch

Alarm name
No ongoing alarms

Events (37)

© 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Amazon Redshift Data Sharing via Bridge Model



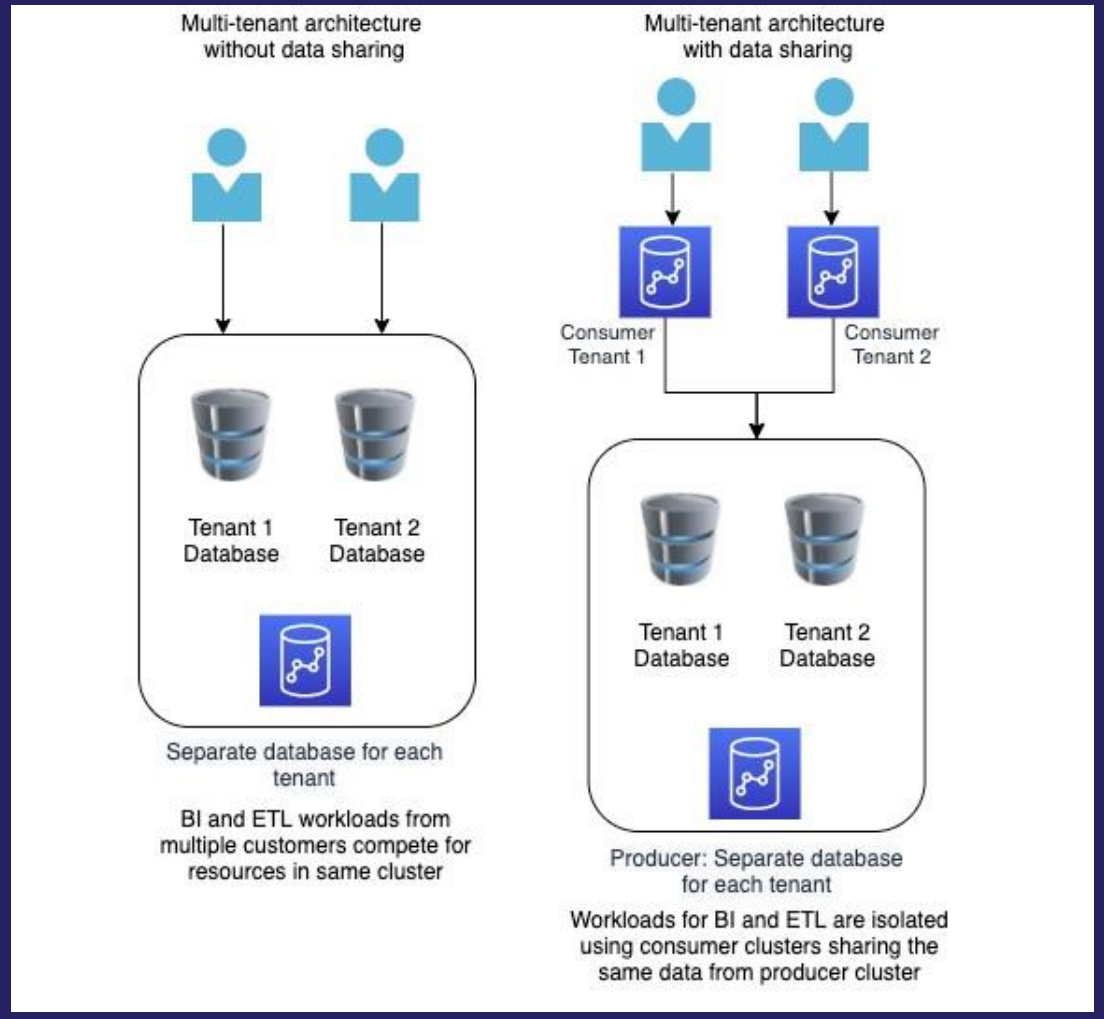
Bridge Model

Storage and access to data for each tenant is controlled at individual schema level in the same database.

1. Create a schema for first tenant
2. Create separate schema for second tenant
3. Create all the tables in each schema individually
4. Create a datashare and add first schema
5. Create separate datashare and add the second schema

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Amazon Redshift Data Sharing via Silo Model



Silo Model

Storage and access control to data for each tenant is maintained in separate databases.

1. Create a database for first tenant
2. Create separate database for second tenant
3. Create the schema with its tables in first database
4. Create separate schema with its tables in second database
5. Create a datashare and add first database
6. Create separate datashare and add the second database

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Considerations for choosing between Pool, Bridge, Silo

Things to consider	Pool	Bridge	Silo
Level of separation for tenant's data	Views	Schema	Database
Complexity of ETL pipelines	Low	Low	Medium
Quotas and Limits	100,000 tables* per cluster	9,900 schemas per database	60 databases per cluster
Chargeback to consumer accounts	Yes	Yes	Yes
Scalability	High	High	High

* for clusters with RA3.4XL, and RA3.16XL node types

Refer blogs for details
https://aws.amazon.com/blogs/big-data/implementing-multi-tenant-patterns-in-amazon-redshift-using-data-sharing/
https://aws.amazon.com/blogs/aws/cross-account-data-sharing-for-amazon-redshift/
https://aws.amazon.com/blogs/big-data/sharing-amazon-redshift-data-securely-across-amazon-redshift-clusters-for-workload-isolation/

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

AWS Data Exchange for Amazon Redshift

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

AWS Data Exchange for Amazon Redshift

FIND, SUBSCRIBE TO, AND QUERY THIRD-PARTY DATA IN AN AMAZON REDSHIFT WAREHOUSE IN MINUTES

AWS Data Exchange

+

Amazon Redshift

Data Provider

Automatic entitlements, billing, and payment management

Multiple data delivery methods on a single product

Data Consumer

Query data in minutes, with no ETL required

Use the latest and greatest data as soon as its updated

FACTSET

"We're really excited about the integration of Amazon Redshift and AWS Data Exchange. This will further remove friction associated with data onboarding. It's going to make things like billing more transparent and easier to manage for our clients. And it's going to really improve the user experience for those interested in browsing content and finding the data that they need, and sourcing it from trusted partners like FactSet."

Peter Dorsey,
VP of Cloud Strategy,
FactSet

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

AWS Management Console

us-west-2.console.aws.amazon.com/console/home?region=us-west-2#

milindo+FieldDemonstrationProgram@amazon.com - 097186925072 / Admin (Not Production Account)

Services | Search for services, features, blogs, docs, and more [Option+S] | Oregon | Admin/milindo-Isengard @ 0971-8692-5072

AWS Data Exchange | EC2 | AWS Glue | IAM | Amazon Redshift | S3 | Athena | CloudFormation | VPC | Database Migration Service | RDS | CodeCommit | Secrets Manager

Console Home Info

Actions

Recently visited Info

- AWS Data Exchange
- Amazon Redshift
- IAM
- Key Management Service
- Systems Manager
- Trusted Advisor
- AWS Well-Architected Tool
- VPC
- Amazon SageMaker
- IAM
- Personal Health Dashboard
- CloudFormation
- AWS SQL Workbench
- Billing

Welcome to AWS

Getting started with AWS
Learn the fundamentals and find valuable information to get the most out of AWS.

Training and certification
Learn from AWS experts and advance your skills and knowledge.

What's new with AWS?
Discover new AWS services, features, and Regions.

AWS Health Info

Open issues
0 Past 7 days

Scheduled changes
0 Upcoming and past 7 days

Other notifications
0 Past 7 days

Cost and usage Info

Current month costs
$28.38

Forecasted month end costs
$51.06 Down 99% from last month

Last month costs
$4,683.62

Top costs for current month

Amazon QuickSight	$15.55
Amazon SageMaker	$12.54
Amazon Simple Storage Service	$0.23
Amazon GuardDuty	$0.04
Amazon Redshift	$0.03

© 2022, Amazon Web Services, Inc. or its affiliates.

Thank you!

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.