

AWS re:Invent

NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT308

Seamless data sharing using Amazon Redshift

BP Yau

Sr Product Manager – Amazon Redshift
Amazon Web Services

Asser Moustafa

Sr Analytics Specialist Solutions Architect
Amazon Web Services



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

Agenda

Introduction

Demo

Workshop

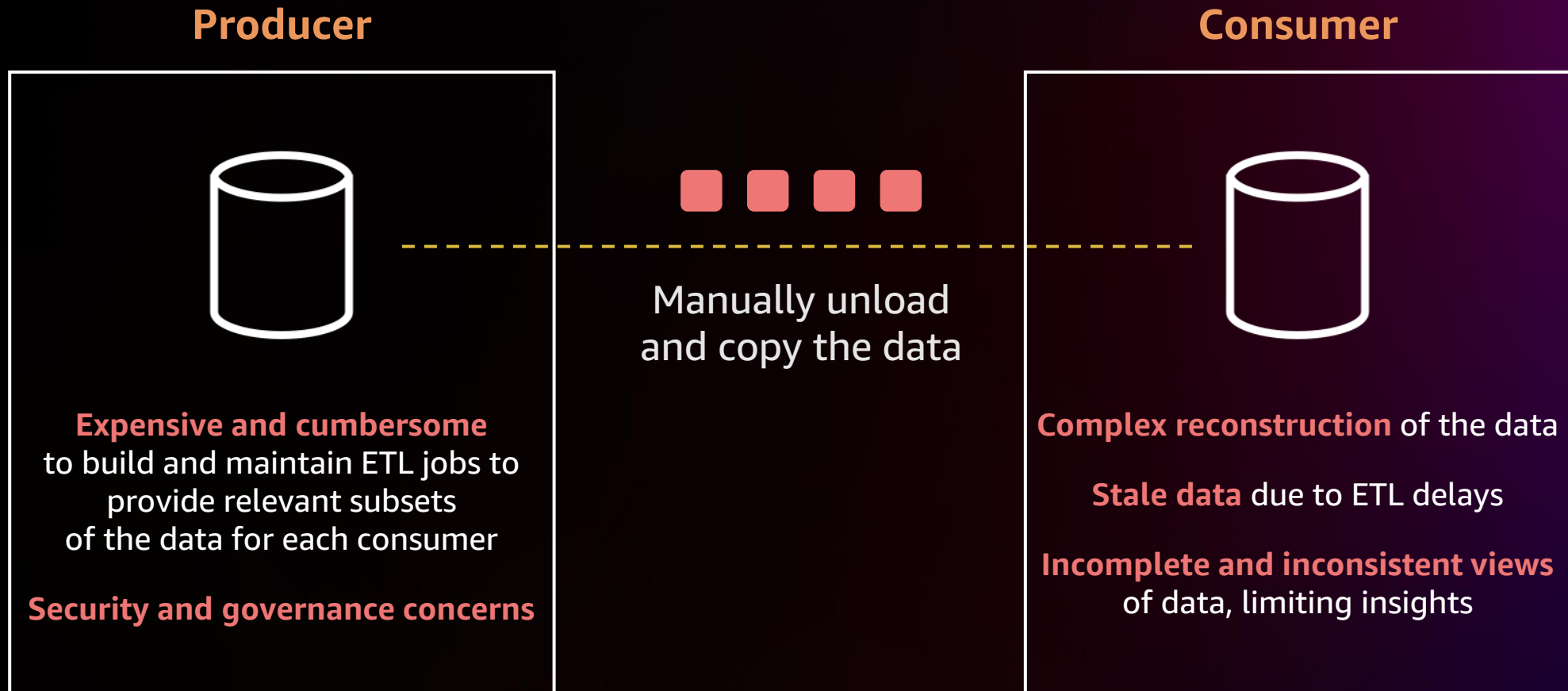
Best practices and considerations



Introduction

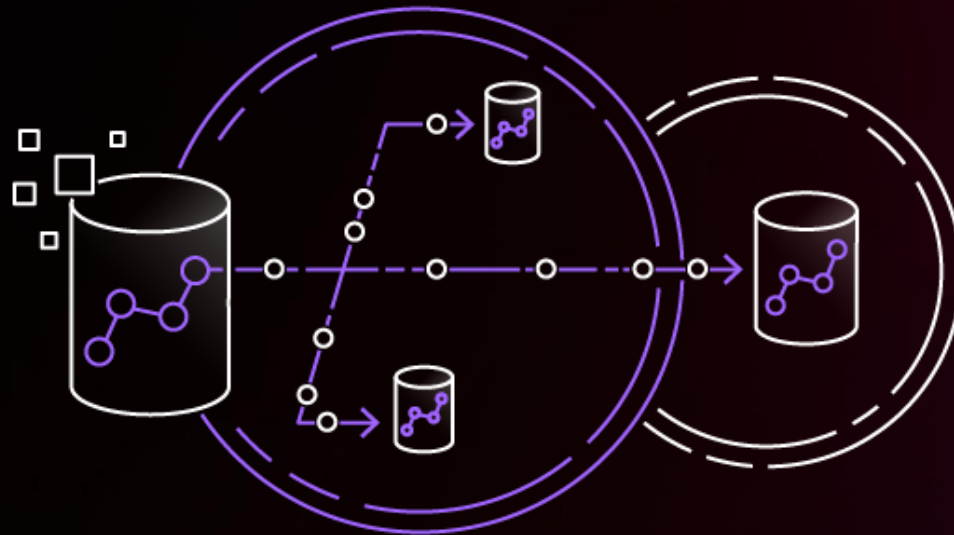


Sharing data in organizations is complex and offers inconsistent views to users



Amazon Redshift data sharing

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



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

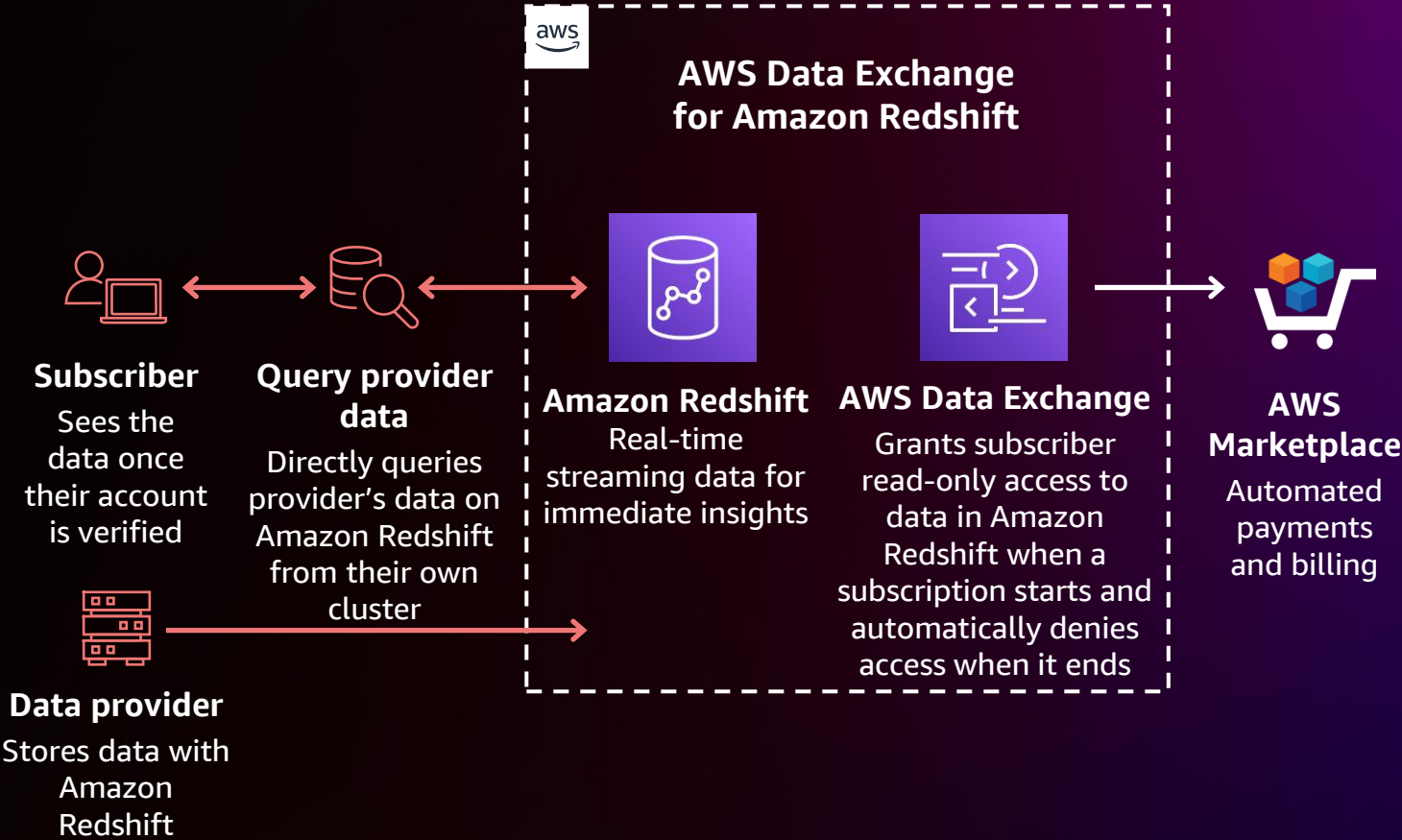
Find, subscribe to, and query third-party data

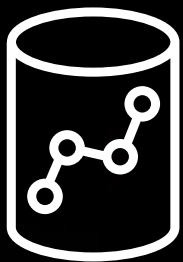
Data providers can package their Amazon Redshift data as data products, with pricing terms and conditions

Subscribers can search and subscribe to products

Get automated “live access” to the data in your Amazon Redshift cluster or data lake

Providers update their data directly at the source, without having to ETL or generate a new file





Amazon Redshift Serverless

Get insights from data in seconds
without having to manage data
warehouse infrastructure



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

Get insights from data more easily

Deliver consistently high performance

Optimize costs

YOU
focus on
insights

Automatic scaling

Compute provisioning

Automated patching

Automatic failover

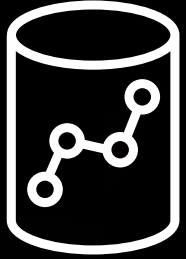
Advanced monitoring

Backup and recovery

Routine maintenance

Security and industry compliance

AWS
takes care
of the rest



Amazon Redshift Serverless

Common use cases



Load and get started
with querying



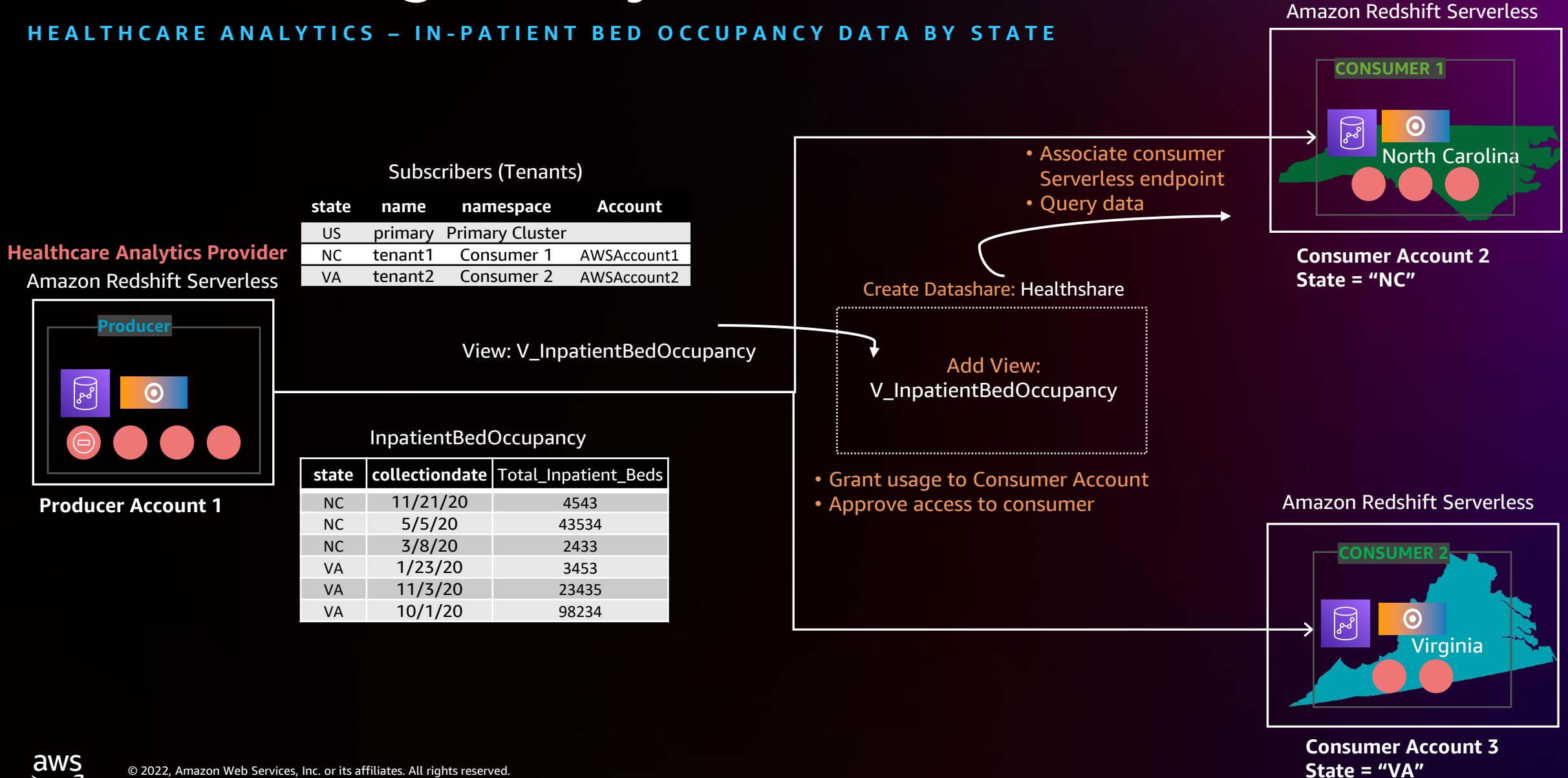
Variable and spiky
workloads



Periodic workloads

Data sharing: Analytics as a service

HEALTHCARE ANALYTICS – IN-PATIENT BED OCCUPANCY DATA BY STATE



Workshop



Seamless data sharing using Amazon Redshift

Workshop goals

- Perform workload isolation for ETL and BI workloads
- Permissions management for data shares
- Monitoring and security of data shares
- [Optional] Enable cross group/account collaboration using Amazon Redshift data sharing

▼ Amazon Redshift Data Sharing Labs

► Before You Begin

(skip this section for re:Invent)

Perform Workload Isolation using Amazon Redshift Data Sharing

Permissions Management for DataShares

Monitoring and Security Cleanup

► Enable Cross Group Collaboration using Amazon Redshift Data Sharing

Getting started with this workshop

Pre-provisioned

As a participant, you will have access to an AWS account with any optional pre-provisioned infrastructure and IAM policies needed to complete this workshop

Temporary

The AWS account will be available only for the duration of this workshop; you will lose access to the account thereafter

Region-focused

The optional pre-provisioned infrastructure will be deployed to a specific region; check your workshop content to determine whether other regions will be used

Terms and conditions

Be sure to review the terms and conditions of the event; do not upload any personal or confidential information in the account

Step 1: Sign in using your preferred method



<https://catalog.workshops.aws/join>

aws workshop studio

Workshop Studio > Sign in

Sign in

Choose a preferred sign-in method

Email one-time password (OTP)

Enter your personal or corporate email to receive a one-time password

Login with Amazon

Login with your Amazon.com retail account

Amazon employee

Login with your Amazon Corporate account. Only for Amazon Employees.

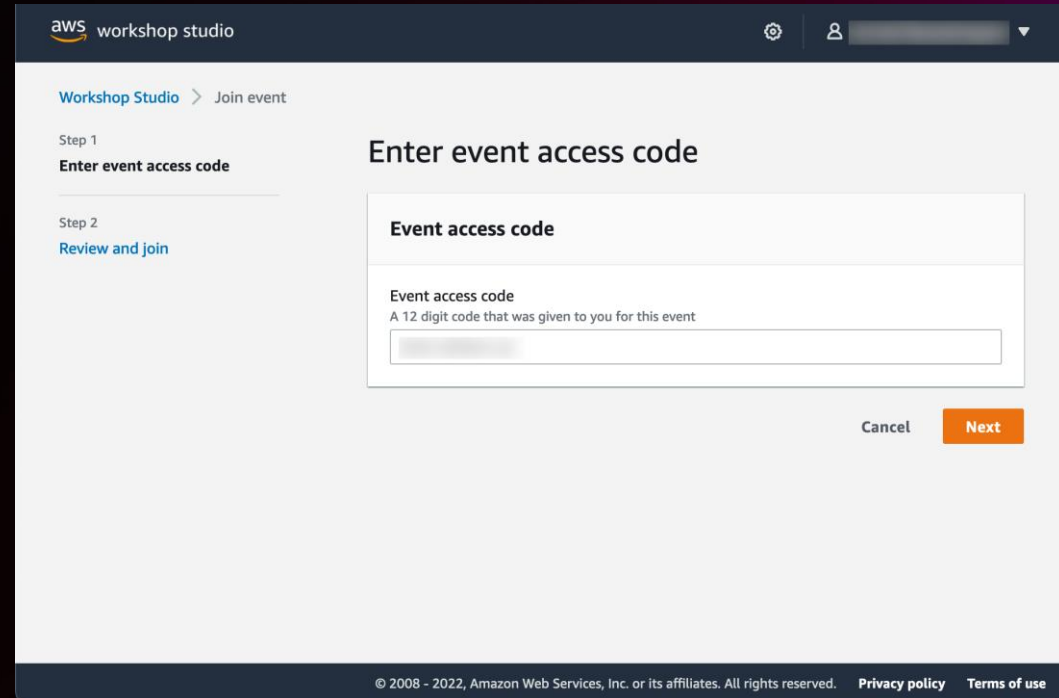
© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. [Privacy policy](#) [Terms of use](#)

Step 2: Enter event access code

ENTER THE 12-CHARACTER EVENT ACCESS CODE (IF YOU WERE GIVEN A ONE-CLICK JOIN LINK, YOU CAN SKIP THIS STEP)



<https://catalog.workshops.aws/join>

A screenshot of the AWS Workshop Studio web interface. The page title is 'Enter event access code'. On the left, there is a sidebar with 'Step 1: Enter event access code' and 'Step 2: Review and join'. The main content area has a heading 'Enter event access code' and a sub-heading 'Event access code'. Below this, it says 'A 12 digit code that was given to you for this event' and provides a text input field. At the bottom right of the form are 'Cancel' and 'Next' buttons. The footer contains copyright information: '© 2008 - 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.' and links to 'Privacy policy' and 'Terms of use'.

d29e-01211d-2c

Step 3: Review terms and join event

Event Access Code: d29e-01211d-2c



<https://catalog.workshops.aws/join>

aws workshop studio

Workshop Studio > Join event

Step 1
[Enter event access code](#)

Step 2
Review and join

Review and join

Event details

Name	Start time	Duration	Level
Seamless Data Sharing using Amazon Redshift event	10/16/2022 10:30 AM	2 hours	-

Description

Test event for content Seamless Data Sharing using Amazon Redshift

Terms and Conditions

Read and accept before joining the event

1. By using AWS Workshop Studio for the relevant event, you agree to the AWS Event Terms and Conditions and the AWS Acceptable Use Policy. You acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relevant event. If you find residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. AWS reserves the right to terminate the account and delete the contents at any time.
2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by AWS, and (b) copy, import, export or otherwise create derivate works of materials provided by AWS, including but not limited to, data sets.
3. AWS is under no obligation to enable the transmission of your materials through Event Engine and may, in its discretion, edit, block, refuse to post, or remove your materials at any time.
4. Your use of AWS Workshop Studio will comply with these terms and all applicable laws, and your access to AWS Workshop Studio will immediately and automatically terminate if you do not comply with any of these terms or conditions.

☒ I agree with the Terms and Conditions

Cancel Previous **Join event**

Step 4: Get started with the workshop

Event Access Code: d29e-01211d-2c



<https://catalog.workshops.aws/join>

Event in progress
Ends in 1 hour 36 minutes 42 seconds.

[Event dashboard](#) > Seamless Data Sharing Using Amazon Redshift

Seamless Data Sharing using Amazon Redshift event

Event information

Start time	Duration
10/16/2022 10:30 AM	2

Description
Test event for content Seamless Data Sharing using Amazon Redshift

Workshop

Title	Complexity level	AWS services	Topics
Seamless Data Sharing using Amazon Redshift	400	Amazon Redshift	Analytics

Description
Organizations today with multiple groups across business using Data warehousing solutions are looking at easy way to share data with each other. With data sharing, customers can now share the data across multiple redshift clusters without copying data through ETL jobs. Data sharing, provides instant, granular, and high performance access to the clusters data. In this workshop, we will discuss and implement different business use cases that can be solved with Amazon Redshift Data sharing.

[Get started >](#)



Step 5: Access the workshop

Event Access Code: d29e-01211d-2c



<https://catalog.workshops.aws/join>

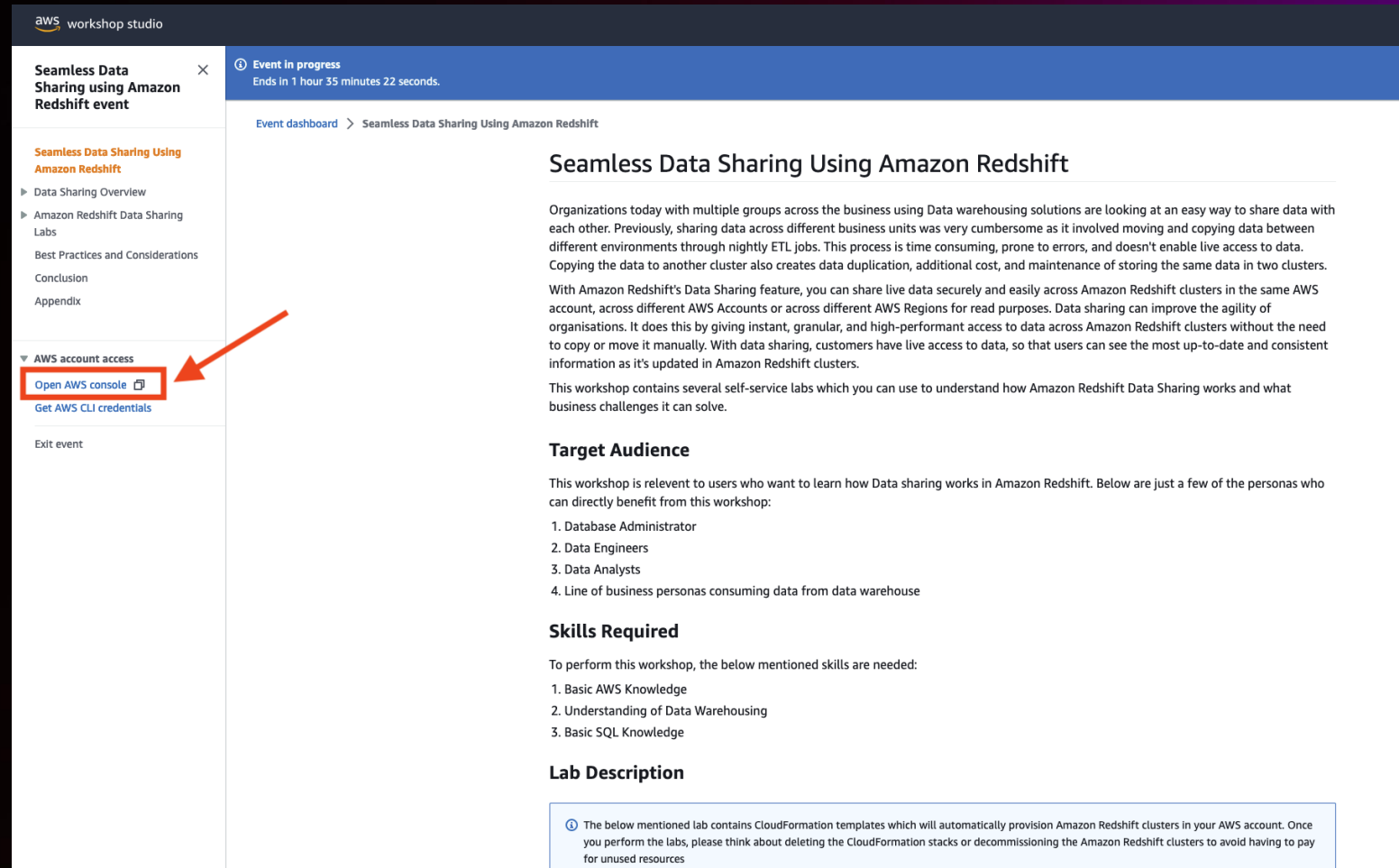
A screenshot of the AWS Workshop Studio interface. The left sidebar shows a navigation menu with the following items: Data Sharing Overview, Amazon Redshift Data Sharing Labs (expanded), Before You Begin (with sub-items: Local System Setup, AWS Account and IAM User, Workshop at an AWS Event, Congratulations on setting up!), Perform Workload Isolation using Amazon Redshift Data Sharing (highlighted with a red box and a red arrow pointing to it), Permissions Management for DataShares, Monitoring and Security, Cleanup, Enable Cross Group Collaboration using Amazon Redshift Data Sharing, Best Practices and Considerations, and Conclusion. The main content area shows the title 'Perform Workload Isolation using Amazon Redshift Data Sharing' and a 'Content' section with a list of topics: Overview, Prerequisite, Create a DataShare, Consuming and Querying DataShare, Result Set Caching on Consumer Redshift Cluster, Create External Schema from DataShare, Workload Isolation, and Before You Leave. Below this is an 'Overview' section with text explaining the lab's purpose and a placeholder for a sample diagram.

Step 6: Access AWS account

Event Access Code: d29e-01211d-2c



<https://catalog.workshops.aws/join>

The screenshot shows the AWS Workshop Studio interface. On the left is a sidebar with a table of contents: 'Seamless Data Sharing Using Amazon Redshift event', 'Seamless Data Sharing Using Amazon Redshift', 'Data Sharing Overview', 'Amazon Redshift Data Sharing Labs', 'Best Practices and Considerations', 'Conclusion', 'Appendix', 'AWS account access' (expanded), 'Open AWS console' (highlighted with a red box and an arrow), 'Get AWS CLI credentials', and 'Exit event'. The main content area has a blue header bar with 'Event in progress' and a timer 'Ends in 1 hour 35 minutes 22 seconds.' Below this is a breadcrumb 'Event dashboard > Seamless Data Sharing Using Amazon Redshift'. The main title is 'Seamless Data Sharing Using Amazon Redshift'. The text describes how organizations use data warehousing and how Amazon Redshift's Data Sharing feature allows for secure, live data access across different AWS accounts and regions. It lists the target audience (Database Administrators, Data Engineers, Data Analysts, and Line of business personas) and the skills required (Basic AWS Knowledge, Understanding of Data Warehousing, Basic SQL Knowledge). A lab description at the bottom states that the lab contains CloudFormation templates for provisioning Amazon Redshift clusters.

aws workshop studio

Seamless Data Sharing Using Amazon Redshift event

Seamless Data Sharing Using Amazon Redshift

Data Sharing Overview

Amazon Redshift Data Sharing Labs

Best Practices and Considerations

Conclusion

Appendix

AWS account access

Open AWS console

Get AWS CLI credentials

Exit event

Event in progress
Ends in 1 hour 35 minutes 22 seconds.

Event dashboard > Seamless Data Sharing Using Amazon Redshift

Seamless Data Sharing Using Amazon Redshift

Organizations today with multiple groups across the business using Data warehousing solutions are looking at an easy way to share data with each other. Previously, sharing data across different business units was very cumbersome as it involved moving and copying data between different environments through nightly ETL jobs. This process is time consuming, prone to errors, and doesn't enable live access to data. Copying the data to another cluster also creates data duplication, additional cost, and maintenance of storing the same data in two clusters. With Amazon Redshift's Data Sharing feature, you can share live data securely and easily across Amazon Redshift clusters in the same AWS account, across different AWS Accounts or across different AWS Regions for read purposes. Data sharing can improve the agility of organisations. It does this by giving instant, granular, and high-performant access to data across Amazon Redshift clusters without the need to copy or move it manually. With data sharing, customers have live access to data, so that users can see the most up-to-date and consistent information as it's updated in Amazon Redshift clusters.

This workshop contains several self-service labs which you can use to understand how Amazon Redshift Data Sharing works and what business challenges it can solve.

Target Audience

This workshop is relevant to users who want to learn how Data sharing works in Amazon Redshift. Below are just a few of the personas who can directly benefit from this workshop:

1. Database Administrator
2. Data Engineers
3. Data Analysts
4. Line of business personas consuming data from data warehouse

Skills Required

To perform this workshop, the below mentioned skills are needed:

1. Basic AWS Knowledge
2. Understanding of Data Warehousing
3. Basic SQL Knowledge

Lab Description

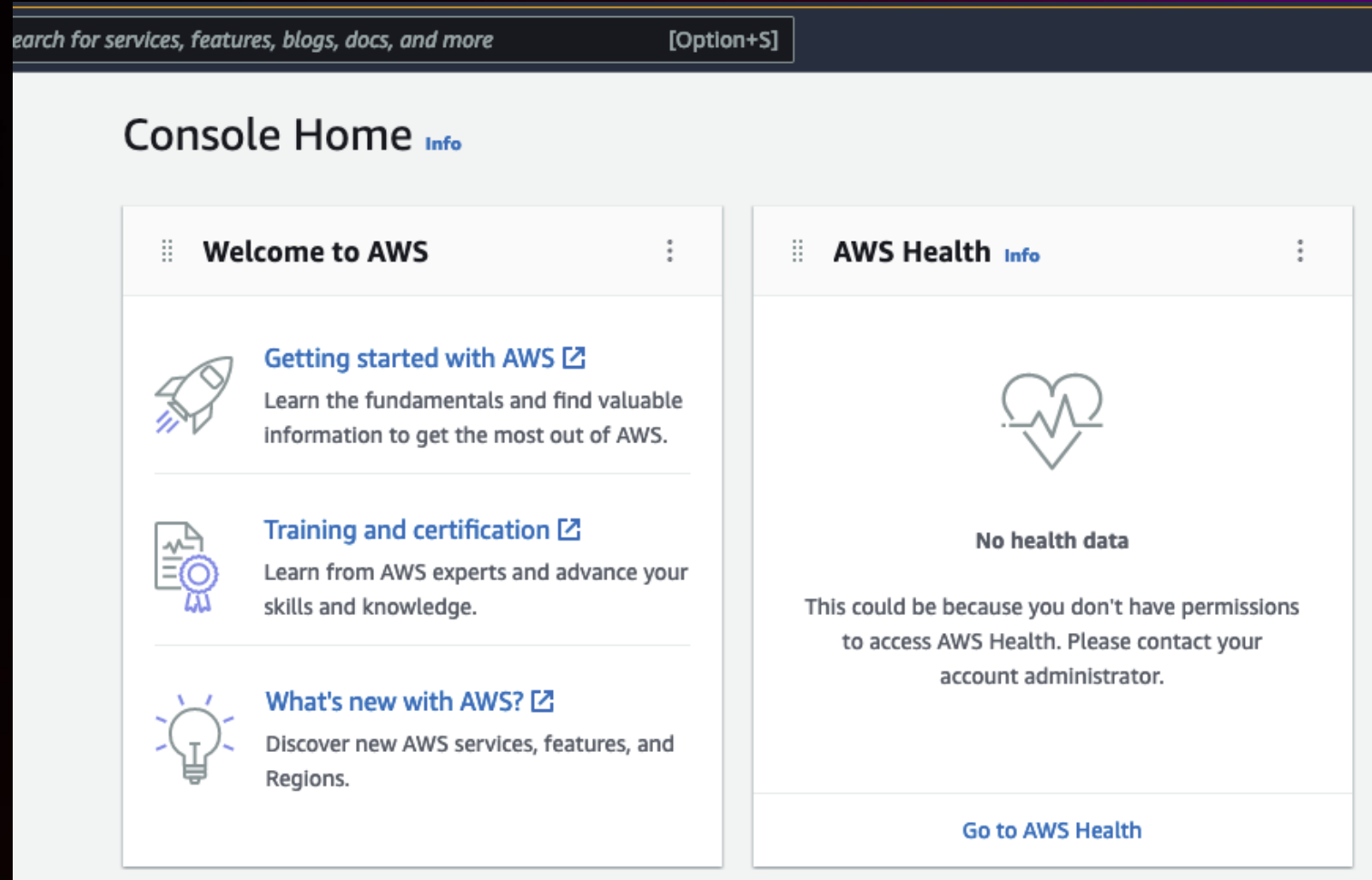
The below mentioned lab contains CloudFormation templates which will automatically provision Amazon Redshift clusters in your AWS account. Once you perform the labs, please think about deleting the CloudFormation stacks or decommissioning the Amazon Redshift clusters to avoid having to pay for unused resources

Step 7: Access AWS console

Event Access Code: d29e-01211d-2c



<https://catalog.workshops.aws/join>



Best practices and considerations



Best practices/considerations

PERFORMANCE

Materialized view ownership

Producer

- Centralized management of MVs

Consumer

- Customized view on consumer based on its use case

Cross-region data sharing

- Data transfer cost on consumer – if cache, no data transfer
- Performance will vary compared to in-region due to network throughput

Table maintenance

If running vacuum manually, use **vacuum re-cluster** wherever possible (does not run merge and full delete), especially in use cases where:

- Large objects on producer where frequent maintenance is needed as a result of frequent DMLs
- Benefits the data sharing workloads by reducing the block metadata sync times

Best practices/considerations

SECURITY

INCLUDE**NEW** datashare property

- When set to true, producer will automatically add future objects to an existing datashare, which helps reduce datashare objects management, but disabling it provides greater control over datashare objects

Fine-grained access control on consumer

- LBV or MV on consumer on shared objects
- Central access control with AWS Lake Formation– **NEW**

Audit data share usage and changes

- **SVL_DATASHARE_CHANGE_LOG** – Records the activity and usage of datashares on the consumer cluster (https://docs.aws.amazon.com/redshift/latest/dg/r_SVL_DATASHARE_CHANGE_LOG.html)
- **SVL_DATASHARE_USAGE_CONSUMER** – Records the activity and usage of datashares on the consumer cluster (https://docs.aws.amazon.com/redshift/latest/dg/r_SVL_DATASHARE_USAGE_CONSUMER.html)
- **SVL_DATASHARE_USAGE_PRODUCER** – Records the activity and usage of datashares on the producer cluster (https://docs.aws.amazon.com/redshift/latest/dg/r_SVL_DATASHARE_USAGE_PRODUCER.html)

Best practices/considerations

PRODUCER-CONSUMER DEPENDENCIES

Workload isolation, encryption, read consistency

- Queries on the consumer cluster will have no impact in terms of performance or activity on the producer cluster
- Data sharing works seamlessly with homogenous encryption configurations
- All the queries involving shared objects on the consumer cluster follow read-committed transaction consistency while checking for visible data for that transaction

Metadata access

Use restrictive filtering while querying the system views for metadata (such as `SVV_ALL_COLUMNS`, `SVV_ALL_SCHEMAS`, `SVV_ALL_TABLES`), for example, instead of the query

```
select * from svv_all_tables;
```

which will try to collect metadata for all the shared and local objects, making it very heavy in terms of metadata scans, especially for shared objects, use the following query to achieve a similar result

```
SELECT table_name, column_name, data_type FROM svv_all_tables WHERE table_name = <
tablename > AND schema_name = < schemaname > AND database_name = < databasename > ORDER
BY ordinal_position;
```

You can also use the `SVV_DATASHARE*` system views to exclusively see shared object-related information

Best practices/considerations

COMPLIMENTARY FEATURES

Real-time analytics and data sharing with streaming data

- Streaming ingestion with KDS on producer
- Data exposed through MV and can be shared to consumer for real-time analytics

Concurrency scaling

- Concurrency scaling is supported for data sharing queries

Redshift Serverless integration

- The new serverless platform provides out-of-box data sharing support for both provisioned/serverless producers/consumers

AWS Data Exchange (ADX)

- Monetize dataset
- Data as service
- Public dataset without ETL

A few other notable complimentary features

- Spectrum
- Federated query



Thank you!

BP Yau
boonyau@amazon.com

Asser Moustafa
aserm@amazon.com

Thank you from the entire workshop team for attending the workshop



BP
Yau



Asser
Moustafa



Rajesh
Francis



Brandon
Schur



Saman
Irfan



Natasha
McCann



Sohaib
Katariwala



Josh
Tow



Mamta
Vaidya



Viral
Shah



Rafael
Rodrigues



Dilip
Rajan



Please complete the session
survey in the **mobile app**

