content.aws.training/video/pcgdls/en/x1/1.0.0/jwplayer.html?endpoint=https%3a%2f%2flrs.aws.training%2fTCAPI%2f&auth=Basic%20OjhmZDQ1YW...

PartnerCast Agenda

Session 1:

- Overview of AWS Glue
- Deep dive on AWS Glue components (such as, Glue data catalog, Glue crawlers, Glue ETL, and Glue Workflow), including demonstrations

Session 2: (Today)

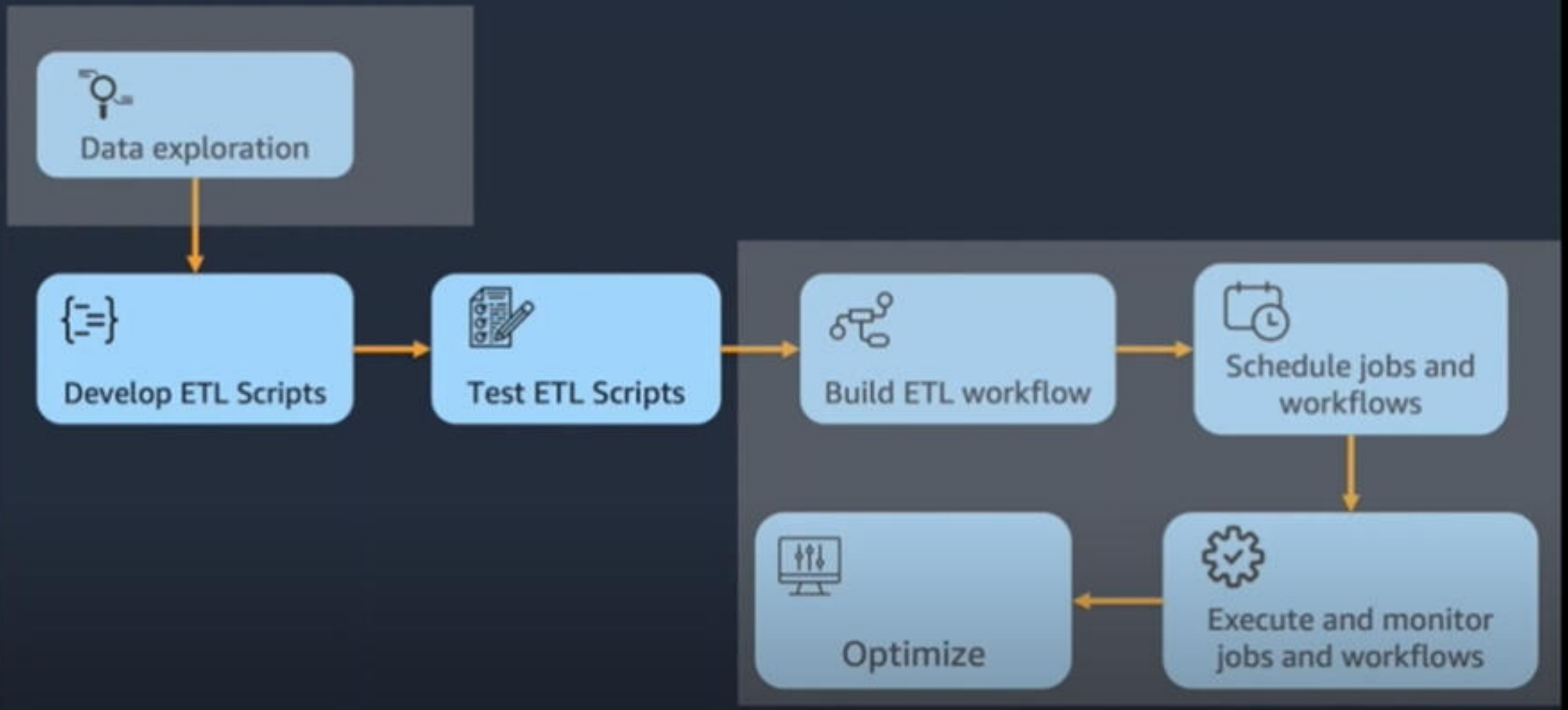
- Dive deep on AWS Glue Studio, and AWS Glue DataBrew
- Overview of AWS Elastic Views
- Build a data lake with ease using AWS Lake Formation, powered by AWS Glue
- AWS Glue common use-cases with reference architectures
- AWS Glue best practices

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

content.aws.training/video/pcgdls/en/x1/1.0.0/jwplayer.html?endpoint=https%3a%2f%2flrs.aws.training%2fTCAPI%2f&auth=Basic%20OjhmZDQ1YW...

Steps in developing and CI/CD ETL processes



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

AWS Glue Studio - Target Persona

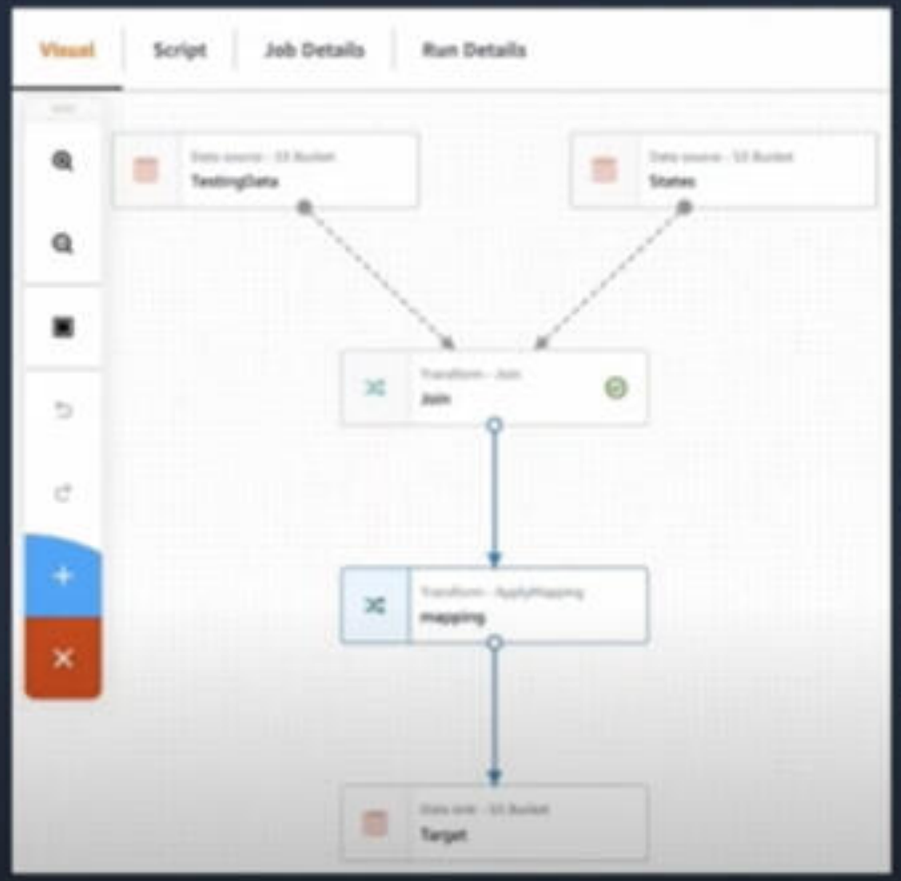
ETL Developers

Users who are experienced with traditional ETL tools with visual interfaces. They define data cleansing and normalization jobs and are familiar with SQL but not Python or Scala. They do not have experience with Apache Spark, so AWS Glue was previously inaccessible to them.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Glue Studio: Visual ETL in AWS Glue

A visual interface that makes it easy to author, run, and monitor AWS Glue ETL jobs



Visual authoring of ETL jobs without writing code

Power of big data processing without managing infrastructure

Manage 1000s of jobs from a single-pane of glass

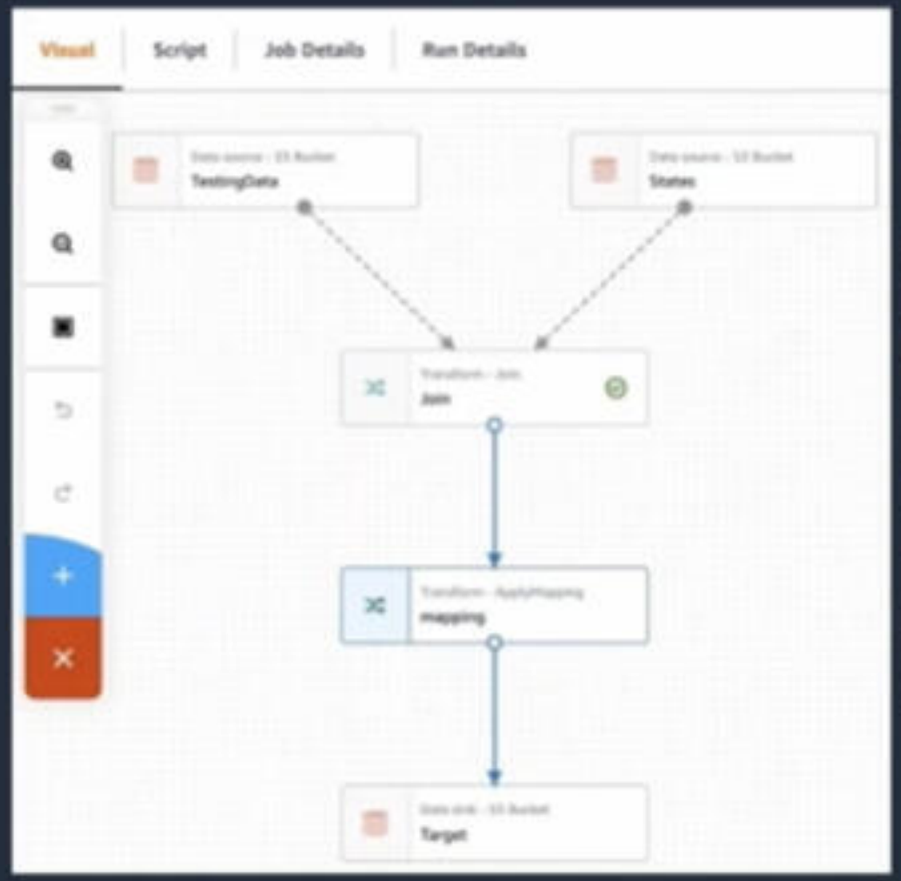
Highly customizable with arbitrary code (python, Scala, and Java)

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Glue Studio: Visual ETL in AWS Glue

A visual interface that makes it easy to author, run, and monitor AWS Glue ETL jobs



Visual authoring of ETL jobs without writing code

Power of big data processing without managing infrastructure

Manage 1000s of jobs from a single-pane of glass

Highly customizable with arbitrary code (python, Scala, and Java)

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

AWS Glue Studio Benefits

Visual ETL in AWS Glue

Rapid development
Author ETL jobs faster and with less debugging

Job Monitoring
See all your job runs through a single pane of glass

Streaming and batch ETL
Use one service to process data from all your sources

Complex Transformations
Use visual transforms or write code snippets to clean and prepare your data for analysis

Structured & semi structured data handling
Glue studio to handle structured and semi structured data like IOT logs

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

AWS Glue Studio Components

Source

S3
Read data from S3 (Parquet, Avro, CSV, JSON, etc). Supports semi structured data processing

Kinesis
Stream data from Kinesis

Kafka
Stream data from Kafka

RDS
Source data from RDS databases

JDBC
Catalog data using JDBC and read via Glue Studio

Transform

ApplyMapping
Change column names, data types, drop columns

Join
Join datasets (inner, outer, left outer, right outer, minus)

Filter
Filter data from datasets

Custom
Add new data elements, write SparkSQL to transform data and aggregate data

Many more..

Target

S3
Write directly to S3

Glue Catalog
Write to Redshift, RDS or any JDBC sources cataloged in Glue Data catalog

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

AWS Glue Studio Components

Source

S3
Read data from S3 (Parquet, Avro, CSV, JSON, etc). Supports semi structured data processing

Kinesis
Stream data from Kinesis

Kafka
Stream data from Kafka

RDS
Source data from RDS databases

JDBC
Catalog data using JDBC and read via Glue Studio

Transform

ApplyMapping
Change column names, data types, drop columns

Join
Join datasets (inner, outer, left outer, right outer, minus)

Filter
Filter data from datasets

Custom
Add new data elements, write SparkSQL to transform data and aggregate data

Many more..

Target

S3
Write directly to S3

Glue Catalog
Write to Redshift, RDS or any JDBC sources cataloged in Glue Data catalog

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

What are customers saying about Glue Studio...

"We recommend AWS Glue Studio to our clients because most of their ETL developers are experienced with visual ETL tools. They want to use big data execution services but also reduce the time spent learning platforms like Apache Spark. AWS Glue Studio makes it faster to write and deploy jobs for ETL because it is easy to onboard newcomers to Glue."

Atish Ray, Managing Director,
ML Engineering & Architecture, Accenture Applied Intelligence

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

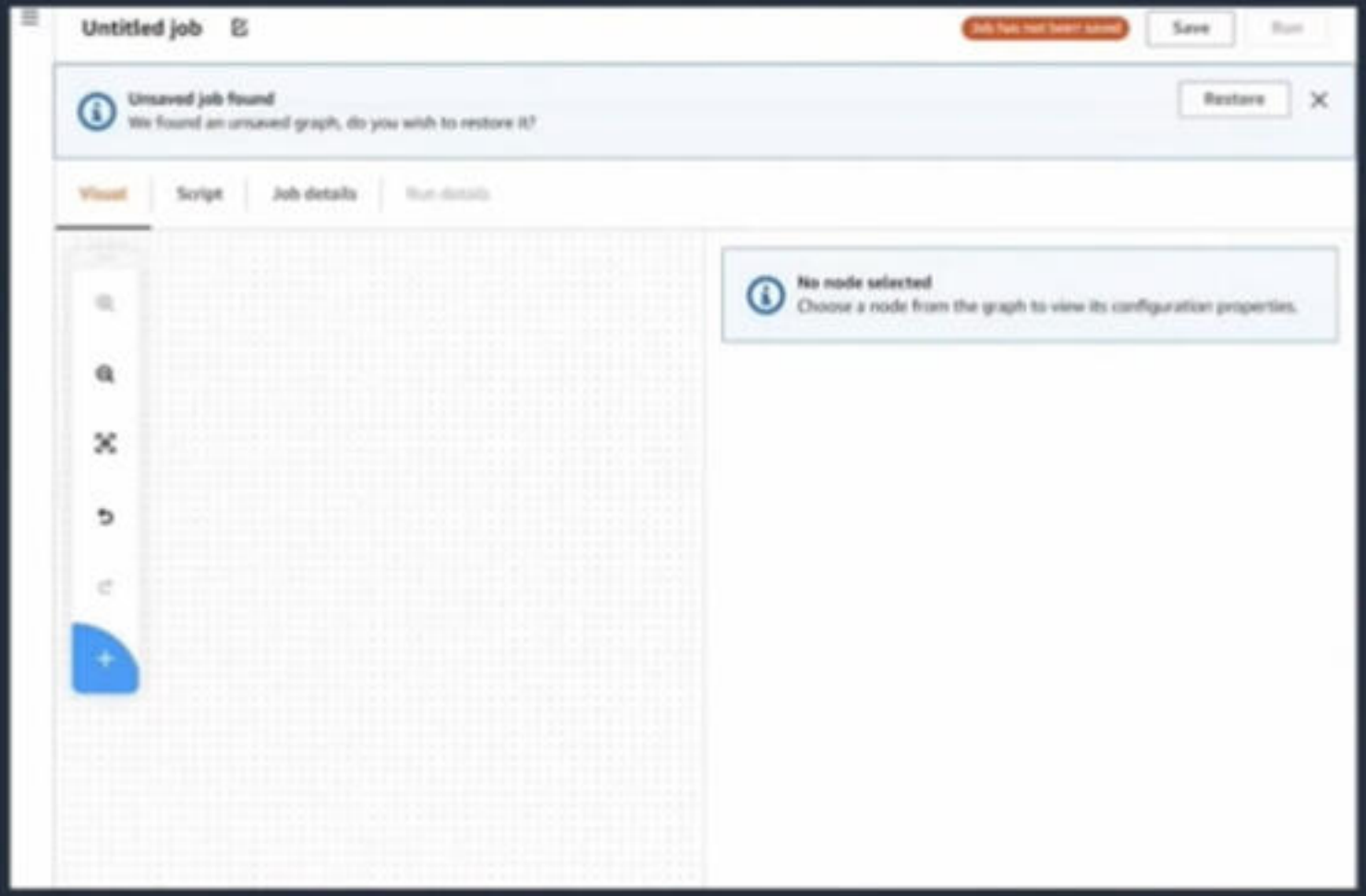
RESOURCES

Blog
https://aws.amazon.com/blogs/big-data/making-etl-easier-with-aws-glue-studio/

AWS Documentation
https://docs.aws.amazon.com/glue/latest/ug/what-is-glue-studio.html

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Glue Studio Demo



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Home Insert Draw Design Transitions Animations Slide Show Review View Share

Paste Cut Copy Format New Slide Layout Reset Section Convert to SmartArt Picture Shapes Text Box Arrange Quick Styles Shape Fill Shape Outline

Intro - 2 min

1

aws partnercast

What's New in Training & Certification

Please Welcome

AWS Glue Deep Learning

2

aws partnercast

AWS Glue Deep Learning
Session - 2

Sudhir Gupta, Sr. Partner Solutions Architect
Harsha Tadiparthi, Sr. Solutions Architect, Analytics

3

PartnerCast Agenda

Session 2 (Today)

Glue Studio Demo

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0553-1744-1198 | N. Virginia | Support

AWS Glue

Data catalog
- Databases
- Tables
- Connections
- Crawlers
- Classifiers
- Schema registries
- Schemas
- Settings

ETL
- AWS Glue Studio
- Workflows
- Jobs
- ML Transforms
- Triggers
- Dev endpoints
- Notebooks

Security
- Security configurations

Tables A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

Add tables | Action | Filter by attributes or search by keyword | Save view | Showing: 1 - 97

Name	Database	Location	Classificatio	Last updated	Deprecated
_parquet__retail_db_categories	retail_db_cdc	retail_db.categories	mysql	2 December 2019 3:18 ...	
_parquet__retail_db_customers	retail_db_cdc	retail_db.customers	mysql	2 December 2019 3:18 ...	
_parquet__retail_db_departments	retail_db_cdc	retail_db.departments	mysql	2 December 2019 3:18 ...	
_parquet__retail_db_order_items	retail_db_cdc	retail_db.order_items	mysql	2 December 2019 3:18 ...	
_parquet__retail_db_orders	retail_db_cdc	retail_db.orders	mysql	2 December 2019 3:18 ...	
_parquet__retail_db_products	retail_db_cdc	retail_db.products	mysql	2 December 2019 3:18 ...	
_raw__retail_db_categories	retail_db_cdc	retail_db.categories	mysql	2 December 2019 11:35...	
_raw__retail_db_customers	retail_db_cdc	retail_db.customers	mysql	2 December 2019 11:35...	
_raw__retail_db_departments	retail_db_cdc	retail_db.departments	mysql	2 December 2019 11:35...	
_raw__retail_db_order_items	retail_db_cdc	retail_db.order_items	mysql	2 December 2019 11:35...	
_raw__retail_db_orders	retail_db_cdc	retail_db.orders	mysql	2 December 2019 11:35...	
_raw__retail_db_products	retail_db_cdc	retail_db.products	mysql	2 December 2019 11:35...	
_temp_parquet__retail_db_categories	retail_db_cdc	s3://datalake-retail-db/c...	Parquet	2 December 2019 3:32 ...	
_temp_parquet__retail_db_customers	retail_db_cdc	s3://datalake-retail-db/c...	Parquet	2 December 2019 3:32 ...	
_temp_parquet__retail_db_departments	retail_db_cdc	s3://datalake-retail-db/c...	Parquet	2 December 2019 3:32 ...	
_temp_parquet__retail_db_order_items	retail_db_cdc	s3://datalake-retail-db/c...	Parquet	2 December 2019 3:33 ...	

AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1#

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] hansha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue

Data catalog
- Databases
- Tables
- Connections
- Crawlers
- Classifiers
- Schema registries
- Schemas
- Settings

ETL
- AWS Glue Studio
- Workflows
- Jobs
- ML Transforms
- Triggers
- Dev endpoints
- Notebooks

Security
- Security configurations

Tables A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

Add tables | Actions | Database : tpcds3tb | Filter or search for tables... | Save view | Showing: 1 - 24

Name	Database	Location	Classificatio	Last updated	Deprecated
call_center	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
catalog_page	tpcds3tb	s3://myworkspace009/t...	parquet	31 October 2020 7:08 P...	
catalog_returns	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
catalog_sales	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
customer	tpcds3tb	s3://myworkspace009/t...	parquet	26 August 2020 8:59 A...	
customer_address	tpcds3tb	s3://myworkspace009/t...	parquet	1 November 2020 11:40...	
customer_demographics	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
date_dim	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
household_demographics	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
income_band	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
inventory	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
item	tpcds3tb	s3://myworkspace009/t...	parquet	7 April 2020 11:15 AM ...	
promotion	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
reason	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
ship_mode	tpcds3tb	s3://myworkspace009/t...	parquet	5 February 2020 9:37 A...	
store	tpcds3tb	s3://myworkspace009/t...	parquet	26 August 2020 9:34 A...	

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/

aws Services Search for services, features, marketplace products, and docs [Option+S] hansha2 @ 0553-1744-1198 N. Virginia Support

Analytics

AWS Glue Studio

Visually create job flows and monitor their performance

AWS Glue Studio is an easy-to-use graphical interface for creating, running, and monitoring AWS Glue extract, transform, and load (ETL) jobs.

Getting started

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
- Monitoring
- **Glue console**
  - Glue catalog
  - Crawlers
  - Security configurations

Documentation

Analytics

AWS Glue Studio

Visually create job flows and monitor their performance

AWS Glue Studio is an easy-to-use graphical interface for creating, running, and monitoring AWS Glue extract, transform, and load (ETL) jobs.

Getting started

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/jobs

aws Services ▾ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▾ N. Virginia ▾ Support ▾

AWS Glue Studio

- Jobs
- Monitoring
- ▾ Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

AWS Glue Studio > Jobs

Jobs Info

Create job Info

Create

Blank graph
Begin a job with an empty canvas.

Source and target added to the graph
Begin a job with source, ApplyMapping transform, and target.

Source: S3 — Glue Data Catalog table with S3 as th...

Target: S3 — S3 bucket by specifying a bucket path...

Your jobs (46) Info

Actions ▾ Run job

Find jobs

1 2 3 4 5

Job name	Type	Last modified
Customer-Deporm	Glue ETL	12/12/2020, 7:42:26 PM

Feedback English (US) ▾

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

No node selected
Choose a node from the graph to view its configuration properties.

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Empty - Placeholder
New node

Node properties

Name
New node

Node type
Choose which type of node to add to the job.
Select node type

Node parents
Choose which nodes will provide inputs for this one.
Select parents

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Empty - Placeholder
customer

Node properties

Name
customer

Node type
Choose which type of node to add to the job.

Select node type

Data source

S3
Glue Data Catalog table with S3 as the data source.

Kinesis

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Data source - S3 buc...
customer

Node properties | Data source properties - S3 2

Output schema

Database

Choose a database

retail_db_cdc
-

sampledb
Sample database

tpcds3tb
-

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Data source - S3 buc...
customer

Node properties | Data source properties - S3 1

Output schema

Database

catalog_page
catalog_returns
Choose a table

Partition predicate - *optional*
Enter a Boolean expression supported by Spark SQL, using only partition columns.

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▾ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▾ N. Virginia ▾ Support ▾

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Data source - S3 buc...
customer

Node properties | Data source properties - S3 1

Output schema

Database

cust

customer_address
customer_demographics
Choose a table

Partition predicate - *optional*
Enter a Boolean expression supported by Spark SQL, using only partition columns.

Feedback English (US) ▾

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Data source - S3 buc...
customer

Empty - Placeholder
New node

Node properties

Name
custo

Node type
Choose which type of node to add to the job.
Select node type

Node parents
Choose which nodes will provide inputs for this one.
Select parents

customer
S3 - DataSource

Feedback | English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy | Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0353-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Data source - S3 buc... customer

Data source - S3 buc... ca

Node properties | Data source properties - S3 2

Output schema

Database

Choose a database

retail_db_cdc
-

sampledb
Sample database

tpcds3tb
-

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Node properties | Data source properties - S3 2

Output schema

Database

Choose a database

crossaccharsha
-

cvsdemo
-

default
Default Hive database

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Node properties | Data source properties - S3 1

Output schema

Database

dem

customer_demographics

household_demographics

Choose a table

Partition predicate - *optional*

Enter a Boolean expression supported by Spark SQL, using only partition columns.

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Node properties

Name

New node

Node type
Choose which type of node to add to the job.

Select node type

Node parents
Choose which nodes will provide inputs for this one.

Select parents

cd
S3 - DataSource

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Node properties | Transform | Output schema

SelectFields

Field	Data type
c_customer_id	string
c_current_cdemo_sk	int
c_current_hdemo_sk	int
c_current_addr_sk	int
c_first_shipto_date_sk	int

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Node properties

Name

cd-filter

Node type
Choose which type of node to add to the job.

Select node type

Node parents
Choose which nodes will provide inputs for this one.

Select parents

cd
S3 - DataSource

Feedback | English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Node properties

Name

join-c-ca

Node type
Choose which type of node to add to the job.

Select node type

AWS Glue Data Catalog
Glue Data Catalog table.

Transform

ApplyMapping

Feedback English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Node properties 1 | Transform 1

Output schema

Name

- c-filter
  SelectFields - Transform
- ca-filter
  SelectFields - Transform
- cd-filter
  SelectFields - Transform

Unclassified nodes

Select parents

c-filter
SelectFields - Transform

Feedback English (US) ▼

© 2006 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0353-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Node properties

Name
New node

Node type
Choose which type of node to add to the job.
Select node type

Node parents
Choose which nodes will provide inputs for this one.
Select parents

cd-filter
SelectFields - Transform

Feedback | English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy | Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Node properties 1 | Transform 1

Output schema

join-customer-denorm

Node type
Choose which type of node to add to the job.

Join
Join two sources into one output using a column header.

Node parents
Choose which nodes will provide inputs for this one.

Select parents

cd-filter

Feedback | English (US) ▼

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy | Terms of Use

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Node properties

Name

New node

Node type
Choose which type of node to add to the job.

Select node type

Node parents
Choose which nodes will provide inputs for this one.

Select parents

join-customer-denorm
Join - Transform

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/graph

aws Services

Search for services, features, marketplace products, and docs [Option+S]

harsha2 @ 0353-1744-1198 N. Virginia Support

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Unsaved job found

We found an unsaved graph, do you wish to restore it?

Restore

Visual 1 | Script | Job details | Run details

Node properties | Data target properties - S3 1

Output schema

Name

customer-denorm

Node type

Choose which type of node to add to the job.

S3

Output data directly in an S3 bucket.

Node parents

Choose which nodes will provide inputs for this one.

Select parents

join-customer-denorm

Join - Transform

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/script

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Untitled job

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

```
import sys
from awsglue.transforms import *
from awsglue.utils import getResolvedOptions
from pyspark.context import SparkContext
from awsglue.context import GlueContext
from awsglue.job import Job
from awsglue.dynamicframe import DynamicFrame

## @params: [JOB_NAME]
args = getResolvedOptions(sys.argv, ['JOB_NAME'])

sc = SparkContext()
glueContext = GlueContext(sc)
spark = glueContext.spark_session
job = Job(glueContext)
job.init(args['JOB_NAME'], args)
## @type: DataSource
## @args: [database = "tpcds3tb", table_name = "customer_address", transformation_ctx = "DataSource0"]
## @return: DataSource0
## @inputs: []
DataSource0 = glueContext.create_dynamic_frame.from_catalog(database = "tpcds3tb", table_name = "customer_address", transform
## @type: SelectFields
## @args: [paths = ["ca_address_sk", "ca_city", "ca_country", "ca_zip"], transformation_ctx = "Transform4"]
## @return: Transform4
## @inputs: [frame = DataSource0]
Transform4 = SelectFields.apply(frame = DataSource0, paths = ["ca_address_sk", "ca_city", "ca_country", "ca_zip"], transforma
```

https://console.aws.amazon.com/gluestudio/home?region=us-east-1#

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/script

aws Services Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 N. Virginia Support

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

```
## @inputs: []
DataSource2 = glueContext.create_dynamic_frame.from_catalog(database = "tpcds3tb", table_name = "customer_demographics", tran
## @type: SelectFields
## @args: [paths = ["cd_demo_sk", "cd_marital_status", "cd_education_status"], transformation_ctx = "Transform0"]
## @return: Transform0
## @inputs: [frame = DataSource2]
Transform0 = SelectFields.apply(frame = DataSource2, paths = ["cd_demo_sk", "cd_marital_status", "cd_education_status"], tran
## @type: DataSource
## @args: [database = "tpcds3tb", table_name = "customer", transformation_ctx = "DataSource1"]
## @return: DataSource1
## @inputs: []
DataSource1 = glueContext.create_dynamic_frame.from_catalog(database = "tpcds3tb", table_name = "customer", transformation_ct
## @type: SelectFields
## @args: [paths = ["c_customer_sk", "c_current_cdemo_sk", "c_current_addr_sk", "c_birth_day", "c_birth_month", "c_birth_year
## @return: Transform3
## @inputs: [frame = DataSource1]
Transform3 = SelectFields.apply(frame = DataSource1, paths = ["c_customer_sk", "c_current_cdemo_sk", "c_current_addr_sk", "c_
## @type: Join
## @args: [columnConditions = ["="], joinType = left, keys2 = ["ca_address_sk"], keys1 = ["c_current_addr_sk"], transformatio
## @return: Transform1
## @inputs: [frame1 = Transform3, frame2 = Transform4]
Transform3DF = Transform3.toDF()
Transform4DF = Transform4.toDF()
Transform1 = DynamicFrame.fromDF(Transform3DF.join(Transform4DF, (Transform3DF['c_current_addr_sk'] == Transform4DF['ca_addre
## @type: Join
## @args: [columnConditions = ["="], joinType = left, keys2 = ["cd_demo_sk"], keys1 = ["c_current_cdemo_sk"], transformation_
## @return: Transform2
## @inputs: [frame1 = Transform1, frame2 = Transform0]
Transform1DF = Transform1.toDF()
Transform0DF = Transform0.toDF()
Transform2 = DynamicFrame.fromDF(Transform1DF.join(Transform0DF, (Transform1DF['c_current_cdemo_sk'] == Transform0DF['cd_demo
## @type: DataSink
## @args: [connection_type = "s3", format = "glueparquet", connection_options = {"path": "s3://myworkspace009/gluestudio-demo
## @return: DataSink0
## @inputs: [frame = Transform2]
DataSink0 = glueContext.write_dynamic_frame.from_options(frame = Transform2, connection_type = "s3", format = "glueparquet",
job.commit()
```

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/details

aws Services Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 N. Virginia Support

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

20201214-demo

Job has not been saved | Save | Run

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Language

Python 3

Worker type
Set the type of predefined worker that is allowed when a job runs.

G.1X

Number of workers
The number of workers of a defined workerType that are allocated when a job runs. The maximum number of workers you can define are 299 for G.1X, and 149 for G.2X.

10

Job bookmark Info
Specifies how AWS Glue processes job bookmark when the job runs. It can remember previously processed data (Enable), update state information (Pause), or ignore state information (Disable).

Enable

Number of retries

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/blank/details

aws Services Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 N. Virginia Support

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Unsaved job found
We found an unsaved graph, do you wish to restore it?

Restore

Visual | Script | Job details | Run details

Number of workers
The number of workers of a defined workerType that are allocated when a job runs. The maximum number of workers you can define are 299 for G.1X, and 149 for G.2X.

10

Job bookmark Info
Specifies how AWS Glue processes job bookmark when the job runs. It can remember previously processed data (Enable), update state information (Pause), or ignore state information (Disable).

Enable

Number of retries

0

Job timeout (minutes)
Set the execution time. The default is 2,880 minutes (48 hours).

2880

Advanced properties

Script filename

Untitled job.py

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/editor/job/20201214-demo/runs

aws Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0353-1744-1198 ▼ N. Virginia ▼ Support ▼

AWS Glue Studio

- Jobs
  - Editor
  - Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Visual | Script | Job details | Run details

Recent job runs (1) Info

Stop job run

December 14, 2020 11:59 AM

Id
jr_d5e5531b9edb6640f4fd9483f9c472595828978 1ed0aa9885a2f8d0f412993ad

Recent attempt
-

Run status
Running

Glue version
2.0

Logs
Cloudwatch logs

Output Logs
Cloudwatch output logs

Error logs
Cloudwatch error logs

Execution time
4 seconds

Start time
December 14, 2020 11:59 AM

End time
-

https://console.aws.amazon.com/gluestudio/home?region=us-east-1#

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help Mon 12:00 PM

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/job/Untitled%20job/run/jr_d96fe05f7adfa7d9b6cf7f1b6e612f63505942dd9992bf5a41b...

aws Services Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0353-1744-1198 N. Virginia Support

AWS Glue Studio

- Jobs
- Monitoring
- Glue console
  - Glue catalog
  - Crawlers
  - Security configurations
- Documentation

Run Details Info

Job name	Run status	Glue version	Recent attempt
Untitled job	Failed	2.0	3
Start time	**End time**	**Start-up time**	**Execution time**
December 14, 2020 4:53 AM	December 14, 2020 4:54 AM	4 seconds	27 seconds
Trigger name	**Last modified on**	**Security configuration**	**Timeout**
-	December 14, 2020 4:54 AM	-	2880 minutes
Allocated capacity	**Max capacity**	**Number of workers**	**Worker type**
10	10	10	G.1X
Logs	**Output Logs**	**Error logs**	**Error**
Logs	Output Logs	Error Logs	Py4JJavaError: An error occurred while calling o82.getDynamicFrame. java.lang.reflect.InvocationTargetException

Metrics Info

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

AWS Glue DataBrew

Visual Data Preparation and Data Cleaning Service

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Challenges faced by Data Scientists

- Industry estimates indicate that in Analytics and Machine Learning pipelines, up to 80% of time is spent on data preparation activities.

- Challenges with traditional data preparation methods
  - Time consuming – multi-step process to extract, clean, normalize, and load data at scale
  - Manual – hard to operationalize and build repeatable workflows
  - Inefficient – typically done by a few data engineers or programmers in an organization, leading to a bottleneck
  - Expensive – costly user licenses and siloed tools

aws training and certification

Challenges faced by Data Scientists

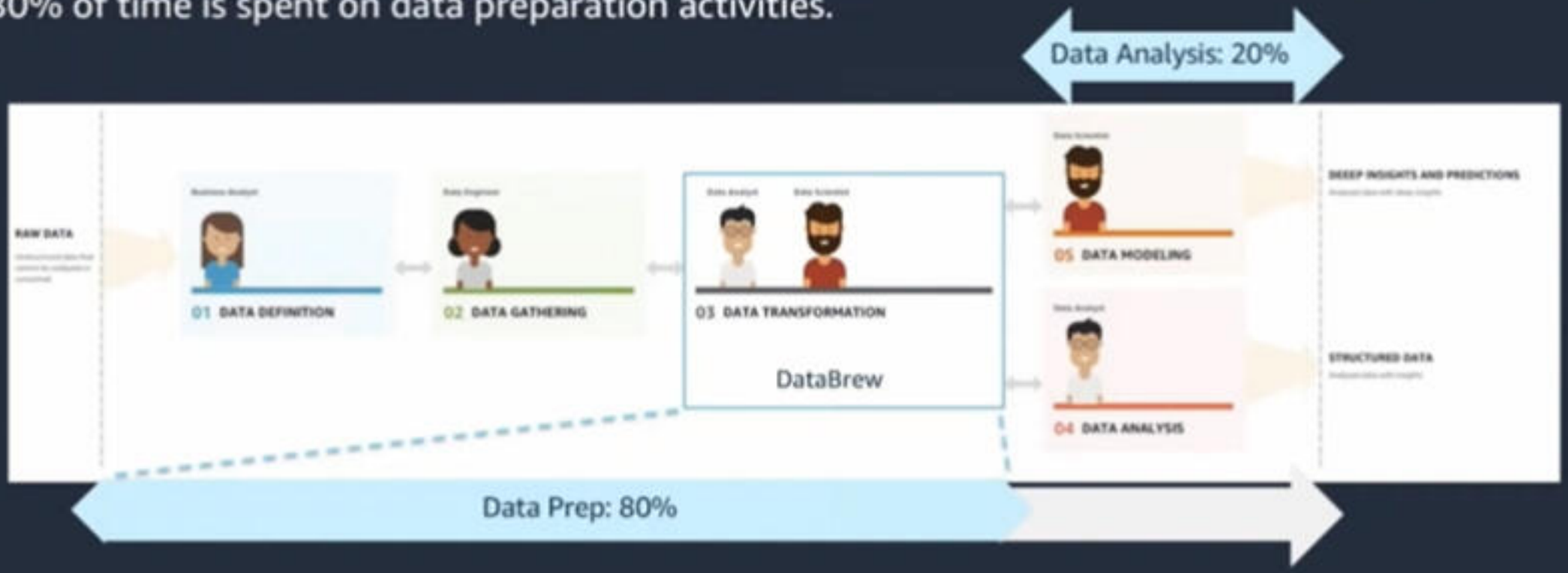
- Industry estimates indicate that in Analytics and Machine Learning pipelines, up to 80% of time is spent on data preparation activities.

- Challenges with traditional data preparation methods
  - Time consuming – multi-step process to extract, clean, normalize, and load data at scale
  - Manual – hard to operationalize and build repeatable workflows
  - Inefficient – typically done by a few data engineers or programmers in an organization, leading to a bottleneck
  - Expensive – costly user licenses and siloed tools

aws training and certification

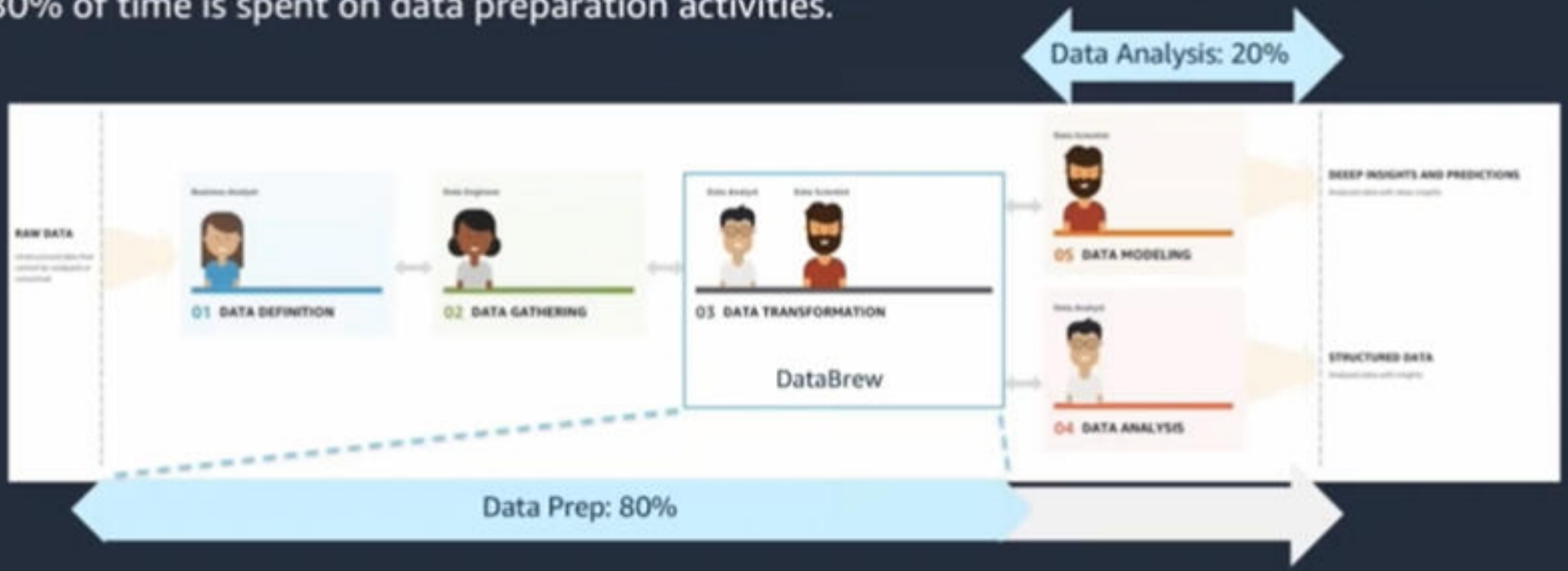
DataBrew: Improve Efficiency and Save Time

Industry estimates indicate that in typical Analytics and Machine Learning use cases, up to 80% of time is spent on data preparation activities.



DataBrew: Improve Efficiency and Save Time

Industry estimates indicate that in typical Analytics and Machine Learning use cases, up to 80% of time is spent on data preparation activities.



Common use cases

- Non-coders, data analysts that need to build analytics and machine learning pipelines. Data Scientists that need to prepare data for ML models.
- Profile and clean raw data as it lands in S3.
- Explore and prepare data in other AWS Services such as Redshift and RDS. Simplify Data Lake and Lake House implementations.
- Prepare and optimize data for visualization in QuickSight and query in Athena.

aws training and certification

AWS Glue DataBrew Functionality

Connect to data sources

Amazon S3, Amazon Redshift, Amazon RDS Aurora or AWS Glue Data Catalog

250 + Built in Transformations

Join, tokenize, split, merge, extract, remove, group, pivot, normalize, one hot or label encode, and more

For users of all technical levels

Visually interact with your data and schema as you build a Recipe without writing any code

Advanced data profiling

Run a profile job in the background to get detailed statistics on your data

Visual Data lineage

View the various stages of data transformation from start to end

Integrate with data pipelines

SDK/API access to integrate with existing data pipelines

Serverless usage at scale

Operate at massive scale in a serverless capacity. Pay only for what you use

Advanced use cases

Use for traditional Data Preparation, Machine Learning, Analytics/Reporting and more

aws training and certification

DataBrew: How does it work?

DATA SOURCES

Data lake

Local files

Amazon S3 buckets

AWS Glue Data Catalog

Amazon Redshift

JDBC/ RDS Aurora

Permissions

AMAZON GLUE DATABREW

Select Data | Build Recipes | Join data | Schedule

remove nulls | fill missing values with aggregates} convert to uppercase, lowercase, sentence case | schema type mismatches | move columns to a position | merge columns | split on delimiters | add suffix, prefix | change date patterns | delete rows | split columns using positions | create column based on extracted values | create a flag column | replace with values | join datasets – inner join, outer join, left join, right join | union datasets | filter data | remove stop words | stemming | clean by removing special characters | remove special characters | expand contradictions | one-hot-encode columns | pattern matching | create columns based on functions | pivot | & more

Clean data stores

Re-use Recipes

Data profiles

Data quality

Data lineage

Reporting

Notebook

Model

ETL/ML pipelines

Pricing

Jobs are $0.48 per node hour and Interactive Sessions are $1 per 30 min session. 5 nodes are allocated to a job as a default. (1 node is 4 vCPUs, 16GB memory). Visit pricing page for more details.

Pricing scenario

- If a DataBrew job runs for 10 minutes with 6 Glue DataBrew nodes.
  Cost = 6 nodes * 1/6 hour at $0.48 per node hour = $0.48

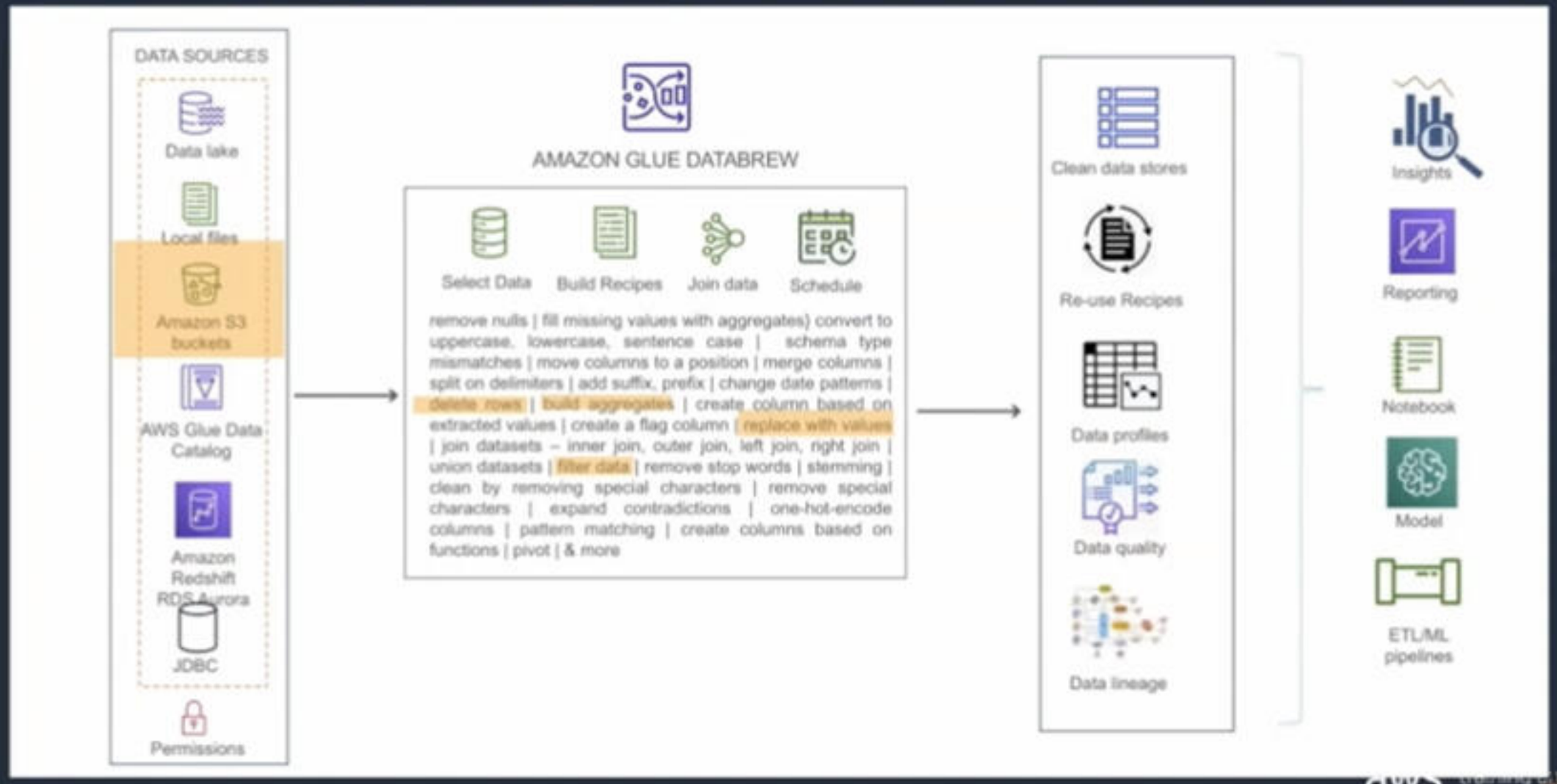
Sample Jobs with test data:

- One hot encode on 30M rows, 10Gb csv dataset took ~30 mins and cost ~$1.5.
- 13T row pivot operation on 6000 parquet files, 100Mb each. Job took ~3 days, $170.
- Math functions such as state if true or false depending on whether a value is even or not for 12B rows from 100 10Mb parquet files took ~2 hours and cost ~$5.
- Fun Fact: 13 Trillion row pivot, single column dataset ran successfully.

*Note: these are examples with test data. Please verify actual benchmarks by testing in your environment.

aws training and certification

Glue Databrew Demo



PowerPoint File Edit View Insert Format Arrange Tools Slide Show Window Help

Partnercast - AWS Glue - Session 2_final

Home Insert Draw Design Transitions Animations Slide Show Review View

Paste Cut Copy Format New Slide Layout Reset Section Picture Shapes Text Box Arrange Quick Styles Shape Fill Shape Outline

Databrew - 30 min

Sample Jobs with test data:

- One fact arcade on 30M rows, 10Gb csv dataset took ~30 mins and cost ~$1.5.
- 137 row pivot operation on 6000 parquet files, 100Mb each, job took ~5 days, $170.
- Math functions such as state if true or false depending on whether a value is even or not for 128 rows from 100 10Mb parquet files took ~2 hours and cost ~$5.
- Fun Fact: 13 Trillion row pivot, single column dataset ran successfully.

*Note: these are examples with test data. Please verify actual benchmarks by testing in your environment.

27

Demo
Glue Databrew

28

Glue Databrew Demo

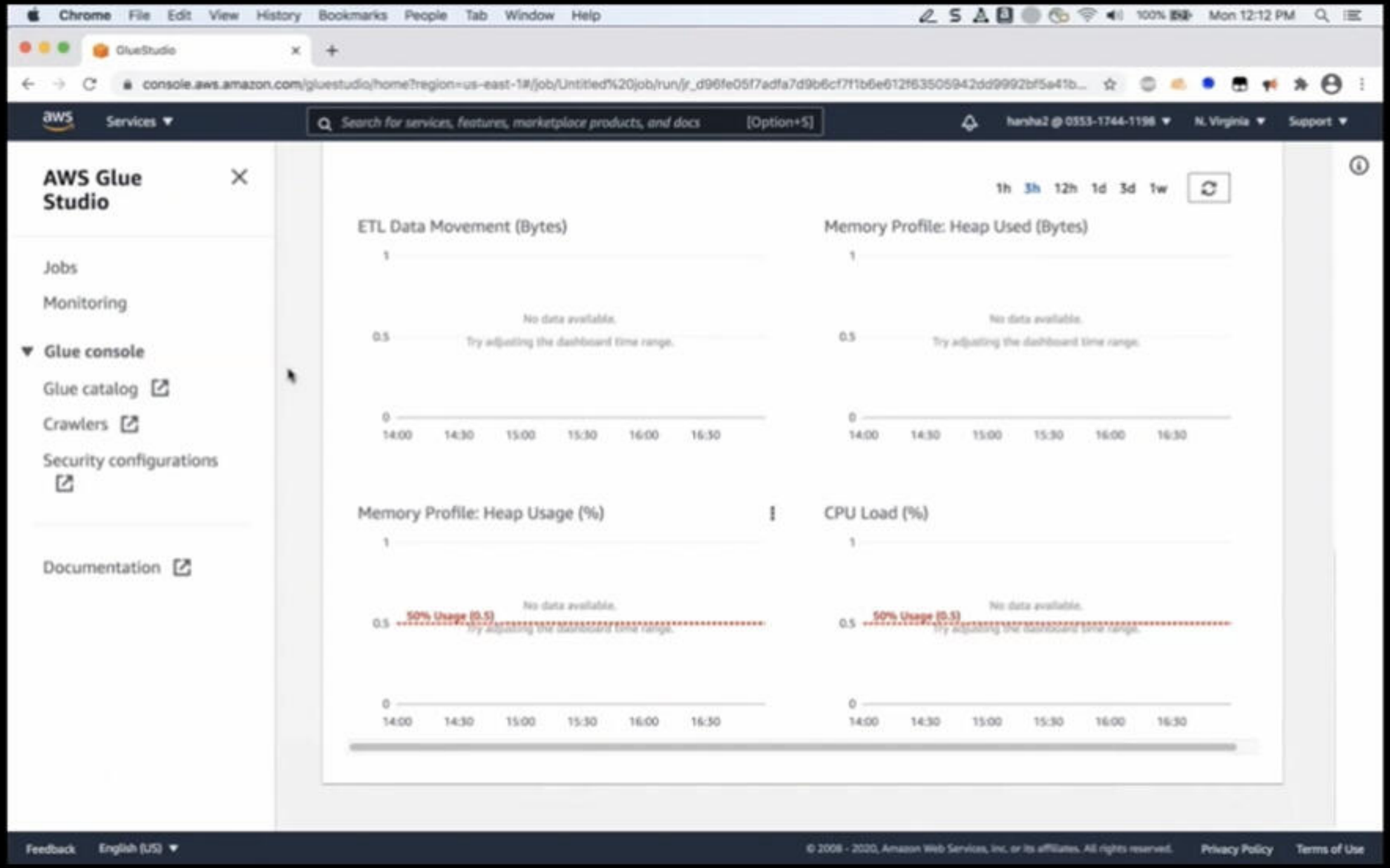
Google Chrome

Glue Databrew Demo

Elastic View and lake f

Slide 28 of 81

80%


Chrome File Edit View History Bookmarks People Tab Window Help
Mon 12:12 PM
GlueStudio
console.aws.amazon.com/gluestudio/home?region=us-east-1#/job/Untitled%20job/run/jr_d96fe05f7adfa7d9b6cf7f1b6e612f63505942dd9992bf5a41b...
Services
Search for services, features, marketplace products, and docs
[Option+S]
harsha2 @ 0553-1744-1198
N. Virginia
Support
AWS Glue Studio
Jobs
Monitoring
Glue console
Glue catalog
Crawlers
Security configurations
Documentation
1h 3h 12h 1d 3d 1w
ETL Data Movement (Bytes)
Memory Profile: Heap Used (Bytes)
Memory Profile: Heap Usage (%)
CPU Load (%)
No data available.
Try adjusting the dashboard time range.
50% Usage (0.5)
1
0.5
0
14:00
14:30
15:00
15:30
16:00
16:30
Feedback
English (US)
© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/job/Untitled%20job/run/jr_d96fe05f7adfa7d9b6cf7f1b6e612f63505942dd9992bf5a41b...

aws Services

Search for services, features, marketplace products, and docs [Option+S]

harsha2 @ 0553-1744-1198 N. Virginia Support

Favorites

- AWS Glue
- AWS Glue DataBrew
- AWS Lake Formation
- Amazon Redshift
- Athena
- CloudFormation
- DynamoDB
- EC2
- EMR
- IAM
- Resource Groups & Tag Ed...
- S3

Recently visited

- AWS Glue
- Console Home

All services

Compute
- EC2
- Lightsail
- Lambda
- Batch
- Elastic Beanstalk
- Serverless Application Repo...
- AWS Outposts
- EC2 Image Builder

Storage
- S3
- EFS
- FSx
- S3 Glacier
- Storage Gateway
- AWS Backup

Database
- RDS
- DynamoDB
- ElastiCache
- Neptune
- Amazon QLDB
- Amazon DocumentDB
- Amazon Keyspaces

Customer Enablement
- AWS IQ
- Support
- Managed Services
- Activate for Startups

Blockchain
- Amazon Managed Blockchain

Satellite
- Ground Station

Quantum Technologies
- Amazon Braket

Management & Governance
- AWS Organizations
- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Systems Manager
- AWS AppConfig

Machine Learning
- Amazon SageMaker
- Amazon Augmented AI
- Amazon CodeGuru
- Amazon Comprehend
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Kendra
- Amazon Lex
- Amazon Personalize
- Amazon Polly
- Amazon Rekognition
- Amazon Textract
- Amazon Transcribe
- Amazon Translate
- AWS DeepComposer
- AWS DeepLens
- AWS DeepRacer
- AWS Panorama
- Amazon Monitron
- Amazon HealthLake
- Amazon Lookout for Vision
- Amazon Lookout for Equip...
- Amazon Lookout for Metrics

Analytics

Front-end Web & Mobile
- AWS Amplify
- Mobile Hub
- AWS AppSync
- Device Farm

AR & VR
- Amazon Sumerian

Application Integration
- Step Functions
- Amazon AppFlow
- Amazon EventBridge
- Amazon MQ
- Simple Notification Service
- Simple Queue Service
- SWF
- Managed Apache Airflow

AWS Cost Management
- AWS Cost Explorer
- AWS Budgets
- AWS Marketplace Subscripti...

Customer Engagement
- Amazon Connect

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

GlueStudio

console.aws.amazon.com/gluestudio/home?region=us-east-1#/job/Untitled%20job/run/jr_d96fe05f7adfa7d9b6cf7f1b6e612f63505942dd9992bf5a41b...

aws Services

Search for services, features, marketplace products, and docs [Option+S]

harsha2 @ 0553-1744-1198 N. Virginia Support

Favorites

- AWS Glue
- AWS Glue DataBrew
- AWS Lake Formation
- Amazon Redshift
- Athena
- CloudFormation
- DynamoDB
- EC2
- EMR
- IAM
- Resource Groups & Tag Ed...
- S3

Recently visited

- AWS Glue
- Console Home

All services

Compute
- EC2
- Lightsail
- Lambda
- Batch
- Elastic Beanstalk
- Serverless Application Repo...
- AWS Outposts
- EC2 Image Builder

Storage
- S3
- EFS
- FSx
- S3 Glacier
- Storage Gateway
- AWS Backup

Database
- RDS
- DynamoDB
- ElastiCache
- Neptune
- Amazon QLDB
- Amazon DocumentDB
- Amazon Keyspaces

Customer Enablement
- AWS IQ
- Support
- Managed Services
- Activate for Startups

Blockchain
- Amazon Managed Blockchain

Satellite
- Ground Station

Quantum Technologies
- Amazon Braket

Management & Governance
- AWS Organizations
- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Systems Manager
- AWS AppConfig

Machine Learning
- Amazon SageMaker
- Amazon Augmented AI
- Amazon CodeGuru
- Amazon Comprehend
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Kendra
- Amazon Lex
- Amazon Personalize
- Amazon Polly
- Amazon Rekognition
- Amazon Textract
- Amazon Transcribe
- Amazon Translate
- AWS DeepComposer
- AWS DeepLens
- AWS DeepRacer
- AWS Panorama
- Amazon Monitron
- Amazon HealthLake
- Amazon Lookout for Vision
- Amazon Lookout for Equip...
- Amazon Lookout for Metrics

Analytics

Front-end Web & Mobile
- AWS Amplify
- Mobile Hub
- AWS AppSync
- Device Farm

AR & VR
- Amazon Sumerian

Application Integration
- Step Functions
- Amazon AppFlow
- Amazon EventBridge
- Amazon MQ
- Simple Notification Service
- Simple Queue Service
- SWF
- Managed Apache Airflow

AWS Cost Management
- AWS Cost Explorer
- AWS Budgets
- AWS Marketplace Subscripti...

Customer Engagement
- Amazon Connect

https://console.aws.amazon.com/databrew/home?region=us-east-1

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

https://console.aws.amazon.com

console.aws.amazon.com/databrew/home?region=us-east-1

aws Services

Search for services, features, marketplace products, and docs [Option+S]

harsha2 @ 0553-1744-1198 N. Virginia Support

Favorites

- AWS Glue
- AWS Glue DataBrew
- AWS Lake Formation
- Amazon Redshift
- Athena
- CloudFormation
- DynamoDB
- EC2
- EMR
- IAM
- Resource Groups & Tag Ed...
- S3

Recently visited

- AWS Glue
- Console Home

All services

Compute
- EC2
- Lightsail
- Lambda
- Batch
- Elastic Beanstalk
- Serverless Application Repo...
- AWS Outposts
- EC2 Image Builder

Storage
- S3
- EFS
- FSx
- S3 Glacier
- Storage Gateway
- AWS Backup

Database
- RDS
- DynamoDB
- ElastiCache
- Neptune
- Amazon QLDB
- Amazon DocumentDB
- Amazon Keyspaces

Customer Enablement
- AWS IQ
- Support
- Managed Services
- Activate for Startups

Blockchain
- Amazon Managed Blockchain

Satellite
- Ground Station

Quantum Technologies
- Amazon Braket

Management & Governance
- AWS Organizations
- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Systems Manager
- AWS AppConfig

Machine Learning
- Amazon SageMaker
- Amazon Augmented AI
- Amazon CodeGuru
- Amazon Comprehend
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Kendra
- Amazon Lex
- Amazon Personalize
- Amazon Polly
- Amazon Rekognition
- Amazon Textract
- Amazon Transcribe
- Amazon Translate
- AWS DeepComposer
- AWS DeepLens
- AWS DeepRacer
- AWS Panorama
- Amazon Monitron
- Amazon HealthLake
- Amazon Lookout for Vision
- Amazon Lookout for Equip...
- Amazon Lookout for Metrics

Analytics

Front-end Web & Mobile
- AWS Amplify
- Mobile Hub
- AWS AppSync
- Device Farm

AR & VR
- Amazon Sumerian

Application Integration
- Step Functions
- Amazon AppFlow
- Amazon EventBridge
- Amazon MQ
- Simple Notification Service
- Simple Queue Service
- SWF
- Managed Apache Airflow

AWS Cost Management
- AWS Cost Explorer
- AWS Budgets
- AWS Marketplace Subscripti...

Customer Engagement
- Amazon Connect

Waiting for console.aws.amazon.com...

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#create-project

aws Services Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0353-1744-1198 N. Virginia Support

Create new recipe

20201214-demo-recipe

The recipe name must contain 1-255 characters. Valid characters are alphanumeric (A-Z, a-z, 0-9), hyphen (-), period (.), and space.

Import steps from recipe

Import recipe steps from an existing recipe into your project. The existing recipe that you chose will not be edited.

Select a dataset

Select the dataset that you want to work on

My datasets Your imported datasets

Sample files Explore example files for your dataset

New dataset Import new dataset

Find datasets

1

Dataset name	Data type	Source	Create date
chess-games	xlsx	S3	22 days ago November 22, 2020, 6:58:28 pm

Sampling - optional

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#create-project

aws Services Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0353-1744-1198 N. Virginia Support

DATASETS
PROJECTS
RECIPES
JOBS

Tags - *optional*

Metadata that you can define and assign to AWS resources. Each tag is a simple label consisting of a customer-defined key (name) and an optional value. Using tags can make it easier for you to manage, search for, and filter resources by purpose, owner, environment, or other criteria.

Permissions Info

DataBrew needs permission to connect to data on your behalf. Use an IAM role with the **required policy** attached.

Role name

Choose the role that has access to connect to your data. Refresh to see the latest updates.

my-glue-databrew-role

As soon as you create a DataBrew project, the project opens and costs begin to accrue to your AWS account. Pricing details

Cancel Create project

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services ▾ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0353-1744-1198 ▾ N. Virginia ▾ Support ▾

Created project "20201214-demo".

20201214-demo

Dataset: chess-games Sample: First n sample (0 rows)

Create job

Viewing 0 rows

GRID SCHEMA PROFILE

Your session will be ready soon!

Provisioning compute

0%

Recipe (0)

20201214-demo-recipe

Working version

Build your recipe

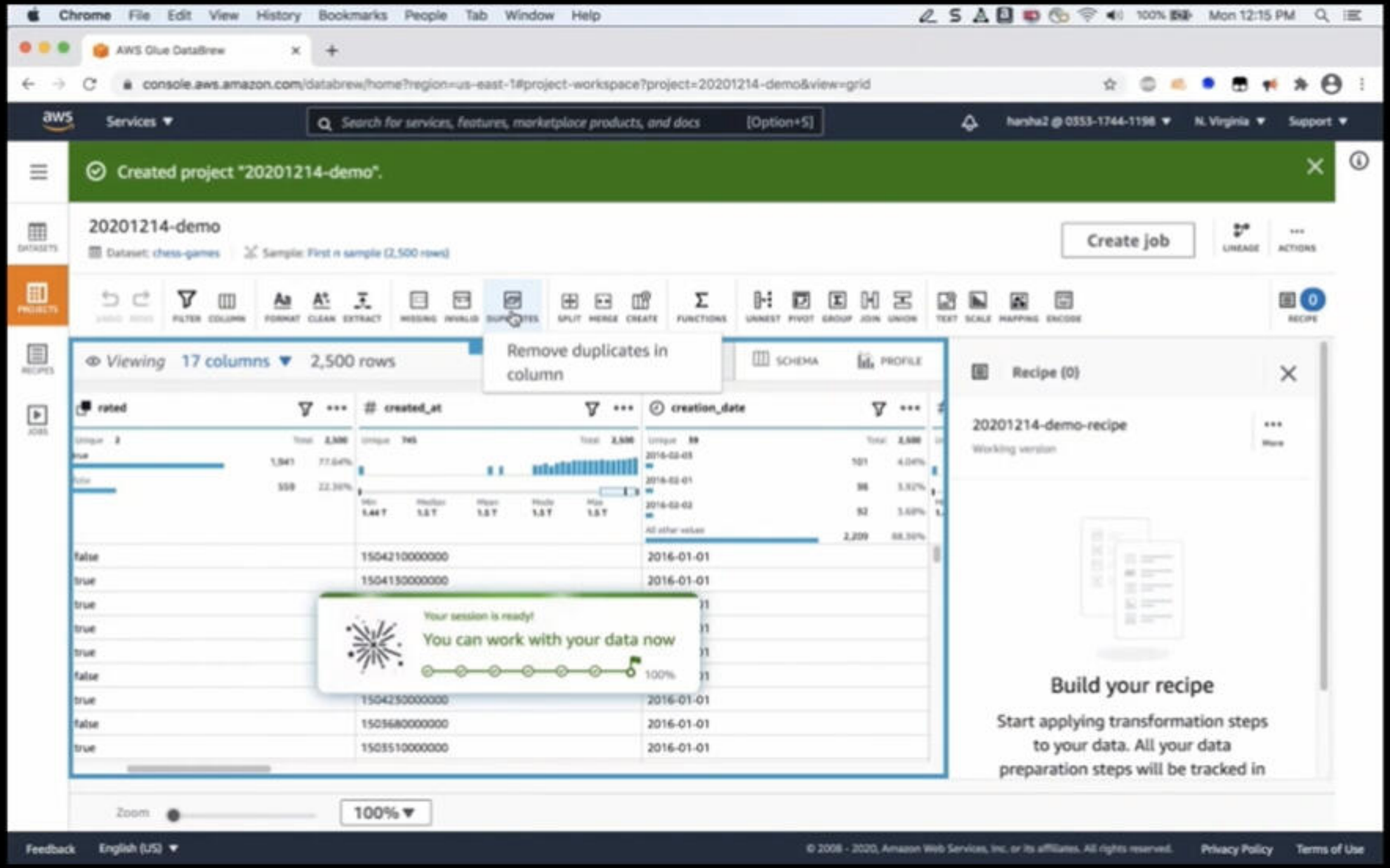
Start applying transformation steps to your data. All your data preparation steps will be tracked in

Zoom 100%

Feedback English (US) ▾

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use


Chrome File Edit View History Bookmarks People Tab Window Help
AWS Glue DataBrew
console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid
Services
Search for services, features, marketplace products, and docs [Option+S]
N. Virginia
Support
Created project "20201214-demo".
20201214-demo
Dataset: chess-games
Sample: First n sample (2,500 rows)
Create job
Viewing 17 columns 2,500 rows
Remove duplicates in column
SCHEMA
PROFILE
rated
created_at
creation_date
Your session is ready!
You can work with your data now
Recipe (0)
20201214-demo-recipe
Build your recipe
Start applying transformation steps to your data. All your data preparation steps will be tracked in
Zoom
100%
Feedback
English (US)
Privacy Policy
Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

Created project "20201214-demo".

20201214-demo

Dataset: chess-games | Sample: First n sample (601 rows)

Create job | LINEAGE | ACTIONS

UNDO | REDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | UNNEST | PIVOT | GROUP | JOIN | UNION | TEXT | SCALE | MAPPING | ENCODE | RECIPE

Viewing 17 columns 601 rows | GRID | SCHEMA | PROFILE

ABC black_id	# black_rating	ABC moves
shivangithegenius	1079	e4 e5 Nf3 Nc6 d4 exd4 Bc4 Bc5 c3 dxc3 Bxf7+ Kxf...
shivangithegenius	978	c4 e5 Nc3 Bc5 g3 Nc6 Bg2 Nf6 Nf3 b6 Nxe5 Bb7 N...
shivangithegenius	978	e4 c5 Nf3 d6 d4 cxd4 Nxd4 Nf6 Nc3 g6 Be3 Bg7 f3...
shivangithegenius	978	e4 c5 Nf3 d6 d4 cxd4 Nxd4 Nf6 Nc3 g6 Be3 Bg7 f3...
shivangithegenius	978	e4 d5 exd5 Qxd5 Qf3 Qe5+ Ne2 Nf6 d4 Qd5 Qxd5 ...
oldpaths	1581	d4 Nf6 c4 e6 Nf3 c5 d5 exd5 cxd5 c4 Nc3 Bb4 Bg5 ...
oldpaths	1599	e4 e6 d4 d5 exd5 exd5 Bd3 Bd6 h3 Ne7 Nf3 O-O O...
avelizzB	1739	e4 e6 d4 d5 Nc3 c5 exd5 exd5 Nf3 Nc6 Bb5 Nf6 O-...
avelizzB	1729	e4 e6 d4 d5 e5 c5 c3 Qb6 Nf3 Nc6 a3 Bd7 Be2 Be7...

Column actions

Recipe (1)

20201214-demo-recipe

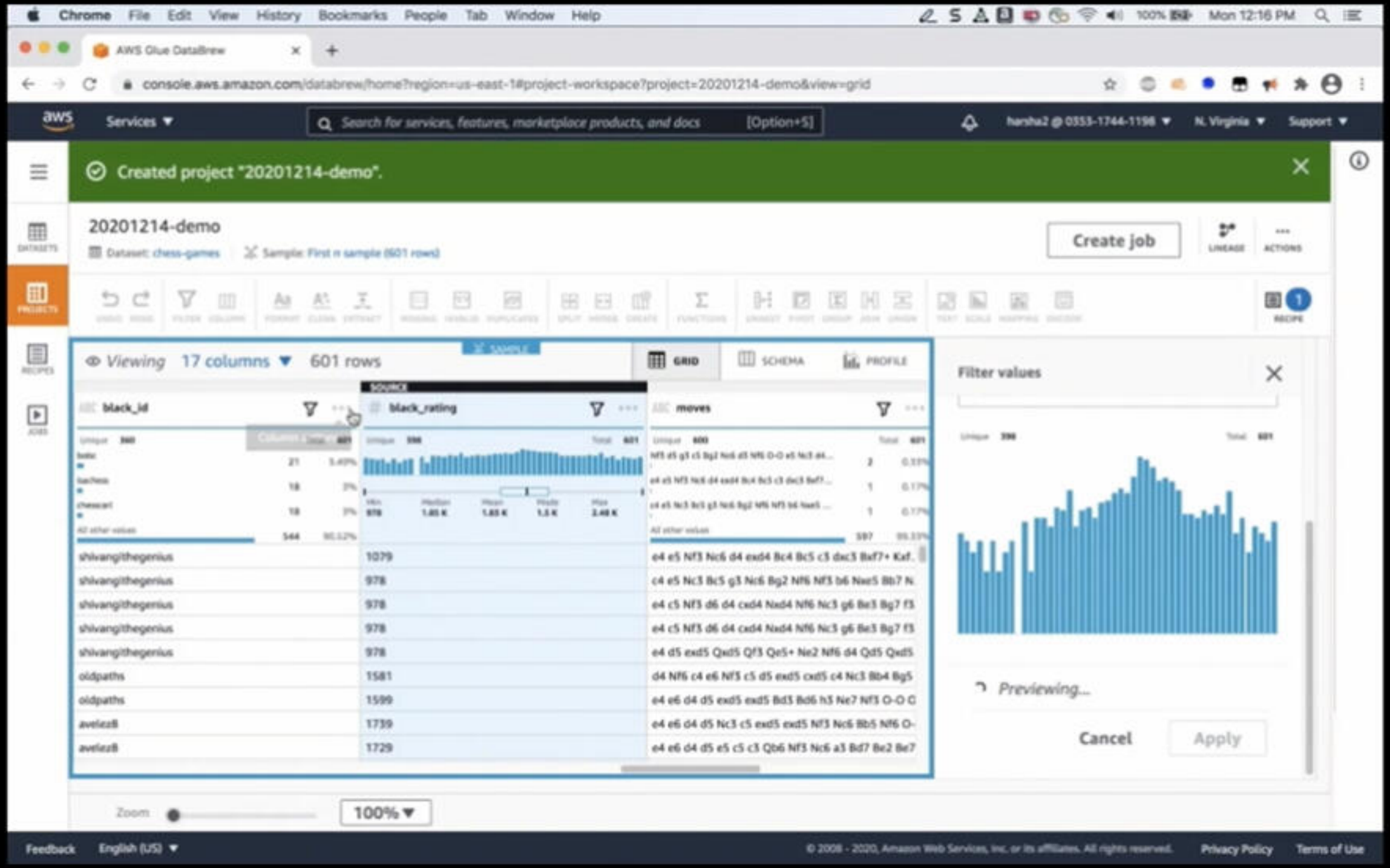
Working version | Publish | More

Applied steps (1) | Clear all | Add step

1. Filter values by white_rating

Zoom 100%

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use


Chrome File Edit View History Bookmarks People Tab Window Help
AWS Glue DataBrew
console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid
Services
Search for services, features, marketplace products, and docs [Option+S]
N. Virginia
Support
Created project "20201214-demo".
20201214-demo
Dataset: chess-games
Sample: First n sample (601 rows)
Create job
LINEAGE
ACTIONS
DATASETS
PROJECTS
RECIPES
JOBS
RECIPE
Viewing 17 columns 601 rows
GRID
SCHEMA
PROFILE
black_id
black_rating
moves
Filter values
Previewing...
Cancel
Apply
Zoom
100%
Feedback
English (US)
Privacy Policy
Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=profile

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

Created project "20201214-demo".

20201214-demo

Dataset: chess-games | Sample: First n sample (399 rows)

Create job | LINEAGE | ACTIONS

chess-games (Input)
S3 chess-games.xlsx 4.6 MB

View dataset

RECIPE 2

Data profile overview | Column statistics

Summary

TOTAL ROWS
20,000

TOTAL COLUMNS
17

DATA TYPES

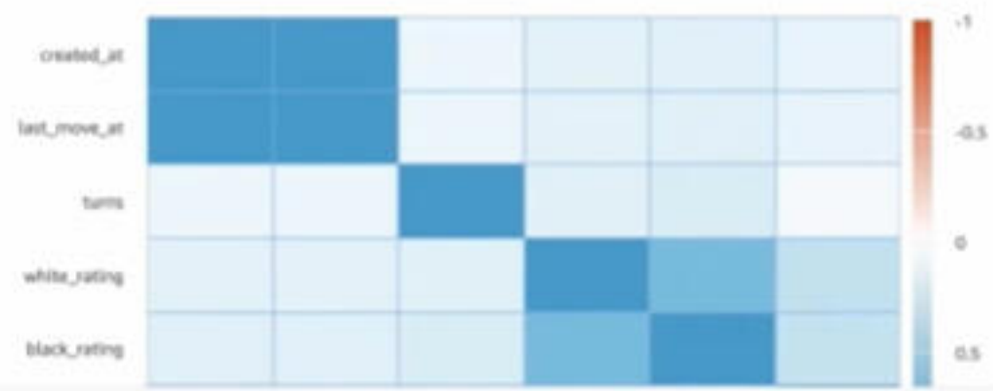
BIG INTEGER	BOOLEAN	STRING	DATE
6 columns	1 columns	9 columns	1 columns

MISSING CELLS

VALID CELLS | MISSING CELLS

Correlations

Correlation coefficient (r) defines how closely two variables are related. It ranges from -1.0 to +1.0 , where 0 means there is no relationship between the variables.



Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | hansha2 @ 0353-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (399 rows)

Create job | LINEAGE | ACTIONS

DATASETS | PROJECTS | RECIPES | JOBS

UNDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | UNNEST | PIVOT | GROUP | JOIN | UNION | TEXT | SCALE | MAPPING | ENCODE | RECIPE 2

Viewing 17 columns

ABC id

Unique 399

1SnlkNRf 1

xw41tjna 1

1zRjgcqm 1

All other values 396

1SnlkNRf

xw41tjna

1zRjgcqm

vfnLULoh

ut6LJGPI

cprATt9o

deKf5OPy

vQ1psjxd

k1qVSHHz

QQ5iIM2V

GpgFo5Bm

Leading and trailing quotation marks

Add

Add prefix (leading characters)

Add suffix (trailing characters)

Leading and trailing single quotation marks

Leading and trailing double quotation marks

Replace

Between delimiters

Between positions

Replace value or pattern

GRID | SCHEMA | PROFILE

created_at

Total 399 | Unique 174 | Total 399

350 87.72%

49 12.28%

Min 1.48 T | Median 1.5 T | Mean 1.5 T | Mode 1.5 T, 1.5 T | Max 1.5 T

1504140000000

1503950000000

1503800000000

1503690000000

1503630000000

1503620000000

1503620000000

1503600000000

1503550000000

1503460000000

1503450000000

Recipe (2)

20201214-demo-recipe

Working version

Publish | More

Applied steps (2) | Clear all | Add step

1. Filter values by white_rating

2. Filter values by black_rating

Zoom 100%

Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | hansha2 @ 0353-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (399 rows)

Create job | LINEAGE | ACTIONS

DATASETS | PROJECTS | RECIPES | JOBS

UNDO | REDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | UNNEST | PIVOT | GROUP | JOIN | UNION | TEXT | SCALE | MAPPING | ENCODE | RECIPE 3

SAMPLE

Viewing 17 columns 399 rows | GRID | SCHEMA | PROFILE

ABC id	rated	# created_at	creation_date	# last_move_at
0dm3RaQv	true	1503120000000	2016-01-04	1503120000000
0zirgLwT	true	1501480000000	2016-01-15	1501480000000
15LAUrJ2	false	1504230000000	2016-01-09	1504230000000
1uRmJzvQ	true	1502810000000	2016-01-25	1502810000000
2shZQ2X6	true	1504170000000	2016-01-18	1504170000000
HDH18gtj	true	1504280000000	2016-01-09	1504280000000
IXaGMTT8	true	1483900000000	2016-02-04	1483910000000
KOXOGIOh	true	1500400000000	2016-01-13	1500400000000
MEaajEWP	true	1499130000000	2016-01-15	1499130000000
MZRiITT5	true	1495840000000	2016-01-19	1495840000000
Miwgmdaf	true	1496560000000	2016-01-19	1496570000000

Zoom 100%

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services ▼ Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 ▼ N. Virginia ▼ Support ▼

20201214-demo

Dataset: chess-games Sample: First n sample (399 rows)

Create job

LINEAGE ACTIONS

DATASETS PROJECTS RECIPES JOBS

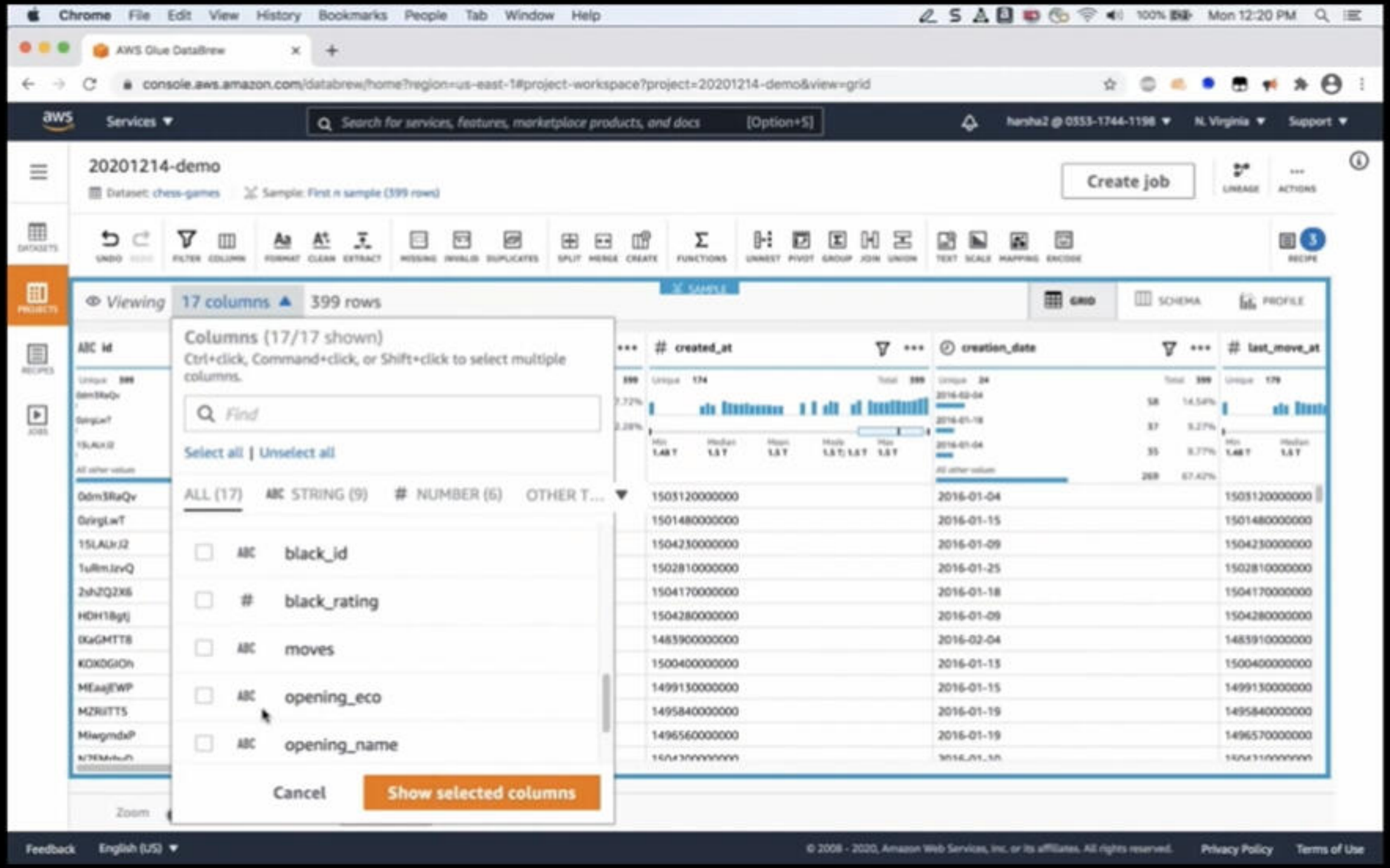
UNDO REDO FILTER COLUMN FORMAT CLEAN EXTRACT MISSING INVALID DUPLICATES SPLIT MERGE CREATE FUNCTIONS UNNEST PIVOT GROUP JOIN UNION TEXT SCALE MAPPING ENCODE RECIPE 3

Viewing 17 columns 399 rows SAMPLE GRID SCHEMA PROFILE

ABC id	rated	# created_at	creation_date	# last_move_at
0dm3RaQv	true	1503120000000	2016-01-04	1503120000000
0zirgLwT	true	1501480000000	2016-01-15	1501480000000
15LAUrJ2	false	1504230000000	2016-01-09	1504230000000
1uRmJzvQ	true	1502810000000	2016-01-25	1502810000000
2shZQ2X6	true	1504170000000	2016-01-18	1504170000000
HDH18gtj	true	1504280000000	2016-01-09	1504280000000
IXaGMTT8	true	1483900000000	2016-02-04	1483910000000
KOXOGIOh	true	1500400000000	2016-01-15	1500400000000
MEaajEWP	true	1499130000000	2016-01-15	1499130000000
MZRIITT5	true	1495840000000	2016-01-19	1495840000000
Miwgmdaf	true	1496560000000	2016-01-19	1496570000000

Zoom 100%

Feedback English (US) ▼ © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use


Chrome File Edit View History Bookmarks People Tab Window Help
AWS Glue DataBrew
console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid
Services
Search for services, features, marketplace products, and docs [Option+S]
N. Virginia
Support
20201214-demo
Dataset: chess-games
Sample: First n sample (399 rows)
Create job
Viewing 17 columns 399 rows
GRID
SCHEMA
PROFILE
Columns (17/17 shown)
Ctrl+click, Command+click, or Shift+click to select multiple columns.
Find
Select all | Unselect all
ALL (17)
STRING (9)
NUMBER (6)
OTHER T...
black_id
black_rating
moves
opening_eco
opening_name
Cancel
Show selected columns
id
created_at
creation_date
last_move_at
Feedback
English (US)
Privacy Policy
Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0553-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (7 rows)

Create job | LINEAGE | ACTIONS

UNDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | UNNEST | PIVOT | GROUP | JOIN | UNION | TEXT | SCALE | MAPPING | ENCODE | RECIPE 5

DATASETS | PROJECTS | RECIPES | JOBS

Viewing -10/3 columns | 7 rows | SAMPLE | GRID | SCHEMA | PROFILE

ABC winner	ABC victory_status	# winner_count
black	mate	33
white	mate	35
black	resign	144
white	resign	126
black	outoftime	12
draw	outoftime	4
white	outoftime	15

Recipe (5)

Working version | Publish | More

Filter values step applied successfully.

Applied steps (5) | Clear all | Add step

1. Filter values by white_rating
2. Filter values by black_rating
3. Remove duplicates from id
4. Group by winner, victory_status and create winner_count from COUNT(winner)
5. Filter values by victory_status

Zoom 100%

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (7 rows)

Create job | LINEAGE | ACTIONS

UNDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | PIVOT | GROUP | JOIN | UNION | TEXT | MAPPING | ENCODE | RECIPE

DATASETS | PROJECTS | RECIPES | JOBS

Viewing -10/3 columns | 7 rows | SAMPLE | GRID | SCHEMA | PROFILE

ABC winner	ABC victory_status	# winner_count
black	mate	33
white	mate	35
black	resign	144
white	resign	126
black	outoftime	12
draw	outoftime	4
white	outoftime	15

Column details

ABC String victory_status

The statistics below are only on the sample

Column statistics

Recommendations

Data quality

VALID VALUES 7 100%

MISSING VALUES 0 0%

Value distribution

Unique 3 Total 7

Zoom 100%

Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (7 rows)

Create job | LINEAGE | ACTIONS

UNDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | PIVOT | GROUP | JOIN | UNION | TEXT | MAPPING | ENCODE | RECIPE 5

DATASETS | PROJECTS | RECIPES | JOBS

Viewing -10/3 columns | 7 rows | SAMPLE | GRID | SCHEMA | PROFILE

ABC winner	ABC victory_status	# winner_count
black	mate	33
white	mate	35
black	resign	144
white	resign	126
black	outoftime	12
draw	outoftime	4
white	outoftime	15

Replace values

Replace values Info

Replace values of column by delimiters or positions

Source column

Name of the source column

victory_status

Specify values to replace

- Between delimiters
- Between positions
- Value or pattern

Value to be replaced

- Enter custom value
- Enter regex value

Zoom 100%

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0553-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (7 rows)

Create job | LINEAGE | ACTIONS

UNDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | PIVOT | GROUP | JOIN | UNION | TEXT | MAPPING | ENCODE | RECIPE 5

DATASETS | PROJECTS | RECIPES | JOBS

Viewing -10/3 columns | 7 rows | SAMPLE | GRID | SCHEMA | PROFILE

winner	victory_status	winner_count
black	mate	33
white	mate	35
black	resign	144
white	resign	126
black	outoftime	12
draw	outoftime	4
white	outoftime	15

Replace values

Value or pattern

Value to be replaced

Enter custom value | Enter regex value

mate

Enter any character, including a space.

Replace with value

ch

Preview changes

Cancel | Apply

Zoom 100%

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (7 rows)

Create job | LINEAGE | ACTIONS

UNDO REDO FILTER COLUMN FORMAT CLEAN EXTRACT MISSING INVALID DUPLICATES SPLIT MERGE CREATE FUNCTIONS UNNEST PIVOT GROUP JOIN UNION TEXT SCALE MAPPING ENCODE RECIPE 5

DATASETS | PROJECTS | RECIPES | JOBS

Viewing -10/3 columns | 7 rows | SAMPLE | GRID | SCHEMA | PROFILE

SOURCE

ABC winner	ABC victory_status	# winner_count
black	mate	33
white	mate	35
black	resign	144
white	resign	126
black	outoftime	12
draw	outoftime	4
white	outoftime	15

Replace values

Value or pattern

Value to be replaced

Enter custom value | Enter regex value

mate

Enter any character, including a space.

Replace with value

checkmate

Preview changes

Cancel | Apply

Zoom 100%

Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0553-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (7 rows)

Create job | LINEAGE | ACTIONS

UNDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | PIVOT | GROUP | JOIN | UNION | TEXT | MAPPING | ENCODE | RECIPE 6

DATASETS | PROJECTS | RECIPES | JOBS

Viewing -10/3 columns

ABC winner

Unique 5

black 3

white 3

draw 1

black
white
black
white
black
draw
white

Leading and trailing quotation marks

Add

Add prefix (leading characters)

Add suffix (trailing characters)

Leading and trailing single quotation marks

Leading and trailing double quotation marks

Replace

Between delimiters

Between positions

Replace value or pattern

GRID | SCHEMA | PROFILE

winner_count

Total 7 | Unique 7 | Total 7

5 42.86%

2 28.57%

2 28.57%

Min 4 | Median 33 | Mean 52.71 | Mode None | Max 144

33
35
144
126
12
4
15

Column details

ABC String victory_status

The statistics below are only on the sample

Column statistics

Recommendations

Data quality

VALID VALUES 7 100%

MISSING VALUES 0 0%

Value distribution

Unique 3 | Total 7

Zoom 100%

Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-workspace?project=20201214-demo&view=grid

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

20201214-demo

Dataset: chess-games | Sample: First n sample (7 rows)

Create job | LINEAGE | ACTIONS

UNDO | FILTER | COLUMN | FORMAT | CLEAN | EXTRACT | MISSING | INVALID | DUPLICATES | SPLIT | MERGE | CREATE | FUNCTIONS | UNNEST | PIVOT | GROUP | JOIN | UNION | TEXT | SCALE | MAPPING | ENCODE | RECIPE 7

DATASETS | PROJECTS | RECIPES | JOBS

Viewing -10/3 columns | 7 rows | SAMPLE | GRID | SCHEMA | PROFILE

ABC winner	ABC victory_status	# winner_count
black	checkmate	33
white	checkmate	35
black	resigned from game	144
white	resigned from game	126
black	outoftime	12
draw	outoftime	4
white	outoftime	15

Recipe (7)

Applied steps (7) | Clear all | Add step

1. Filter values by white_rating
2. Filter values by black_rating
3. Remove duplicates from id
4. Group by winner, victory_status and create winner_count from COUNT(winner)
5. Filter values by victory_status
6. Replace text mate with checkmate in victory_status
7. Replace text resign with resigned from game in victory_status

Zoom 100%

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help 100% Mon 12:24 PM

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-data-lineage?project=20201214-demo

aws Services Search for services, features, marketplace products, and docs [Option+S] harsha2 @ 0553-1744-1198 N. Virginia Support

DataBrew > Projects > 20201214-demo > Project data lineage

DATASETS

PROJECTS

RECIPES

JOBS

Data lineage

Project: 20201214-demo

OPEN PROJECT

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-data-lineage?project=20201214-demo

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

DataBrew > Projects > 20201214-demo > Project data lineage

Data lineage

Project: 20201214-demo

OPEN PROJECT

Lineage | CloudTrail logs

Zoom 100%

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-data-lineage?project=20201214-demo

20201214-demo-recipe

Recipe name
20201214-demo-recipe

Recipe version
0.1

Published version
-

Version description
-

Recipe steps (7)

1. Filter values by white_rating
2. Filter values by black_rating
3. Remove duplicates from id
4. Group by winner , victory_status and create winner_count from COUNT(winner)
5. Filter values by victory_status
6. Replace text mate with checkmate in victory_status
7. Replace text resign with resigned from game in victory_status

View recipe

Close

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#project-data-lineage?project=20201214-demo

20201214-demo-recipe

20201214-demo-recipe	0.1
Published version	Version description
-	-

Recipe steps (7)

1. Filter values by white_rating
2. Filter values by black_rating
3. Remove duplicates from id
4. Group by winner , victory_status and create winner_count from COUNT(winner)
5. Filter values by victory_status
6. Replace text mate with checkmate in victory_status
7. Replace text resign with resigned from game in victory_status

View recipe

Close

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#recipes

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

DataBrew > Recipes

▸ What is a recipe

Recipes (1) Info

Download as YAML | Download as JSON | Create job with this recipe | Actions

Find recipes | Published | 1

Recipe name	Version description	Associated projects	Published date	Published by	Tags
demo-11-22-2020-recipe Version 2.0	version 2	demo-11-22-2020	22 days ago November 22, 2020, 7:51:40 pm	harsha2	-

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#recipes

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0353-1744-1198 | N. Virginia | Support

DataBrew > Recipes

What is a recipe

Recipes (1) Info

Download as YAML | Download as JSON | Create job with this recipe | Actions

Find recipes | Published

Delete
Duplicate
Manage tags

Recipe name	Version description	Associated projects	Published date	Published by
demo-11-22-2020-recipe Version 2.0	version 2	demo-11-22-2020	22 days ago November 22, 2020, 7:51:40 pm	harsha2

Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#recipes

Services

Search for services, features, marketplace products, and docs [Option+S]

harsha2 @ 0353-1744-1198 N. Virginia Support

DataBrew > Recipes

DATASETS

PROJECTS

RECIPES

JOBS

What is a recipe

Recipes (1) Info

Download as YAML | Download as JSON | Create job with this recipe | Actions

Find recipes

Published

Delete

Duplicate

Manage tags

	Recipe name	Version description	Associated projects	Published date	Published by	
☑	demo-11-22-2020-recipe Version 2.0	version 2	demo-11-22-2020	22 days ago November 22, 2020, 7:51:40 pm	harsha2	-

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#recipes

Services | Search for services, features, marketplace products, and docs [Option+S] | harsha2 @ 0553-1744-1198 | N. Virginia | Support

DataBrew > Recipes

What is a recipe

Recipes (1) Info

Download as YAML | Download as JSON | Create job with this recipe | Actions

Find recipes | Published

Delete
Duplicate
Manage tags

Recipe name	Version description	Associated projects	Published date	Published by	
demo-11-22-2020-recipe Version 2.0	version 2	demo-11-22-2020	22 days ago November 22, 2020, 7:51:40 pm	harsha2	-

Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Glue DataBrew

console.aws.amazon.com/databrew/home?region=us-east-1#jobs?tab=recipe

Services

Search for services, features, marketplace products, and docs [Option+S]

harsha2 @ 0553-1744-1198 N. Virginia Support

DataBrew > Jobs

Recipe jobs | Profile jobs | Schedules

Recipe jobs (0) Info

Run job to output data

Test your recipe against the entire dataset. You can specify the output formats and any compression to apply.

Create job

Microsoft PowerPoint

Feedback Terms of Use

AWS Glue Elastic Views (Preview)

Combine and replicate data across multiple data stores using SQL

Easy to use: Use familiar SQL to create materialized views of data data in multiple different data stores.

Fully managed: Copies data automatically from source data stores to target data stores.

Near real-time: Continuously monitors source data stores for changes and updates target data stores within seconds.

Serverless: Scales capacity up or down to support your workloads. Pay for what you use.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

AWS Glue Elastic Views (Preview)

Combine and replicate data across multiple data stores using SQL

Easy to use: Use familiar SQL to create materialized views of data data in multiple different data stores.

Fully managed: Copies data automatically from source data stores to target data stores.

Near real-time: Continuously monitors source data stores for changes and updates target data stores within seconds.

Serverless: Scales capacity up or down to support your workloads. Pay for what you use.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

AWS Glue Elastic Views (Preview)

Combine and replicate data across multiple data stores using SQL

Easy to use: Use familiar SQL to create materialized views of data data in multiple different data stores.

Fully managed: Copies data automatically from source data stores to target data stores.

Near real-time: Continuously monitors source data stores for changes and updates target data stores within seconds.

Serverless: Scales capacity up or down to support your workloads. Pay for what you use.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

AWS Glue Elastic Views (Preview)

Combine and replicate data across multiple data stores using SQL

Easy to use: Use familiar SQL to create materialized views of data data in multiple different data stores.

Fully managed: Copies data automatically from source data stores to target data stores.

Near real-time: Continuously monitors source data stores for changes and updates target data stores within seconds.

Serverless: Scales capacity up or down to support your workloads. Pay for what you use.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Supported source and target data stores

During preview:

Source : Amazon DynamoDB
Target: Amazon S3, Amazon Redshift and Amazon Elasticsearch

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

When to use AWS Glue Elastic View vs AWS DMS

Use DMS

- Migrating database from on-premises to AWS
- Database replication

Use Glue Elastic Views

- Data is in AWS
- Combine and replicate from your data sources into another data store.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

When to use AWS Glue Elastic View vs AWS DMS

Use DMS

- Migrating database from on-premises to AWS
- Database replication

Use Glue Elastic Views

- Data is in AWS
- Combine and replicate from your data sources into another data store.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Use cases at preview

- Create a searchable index of DynamoDB data in Elasticsearch

- Integrate operational data in DynamoDB in a data lake

- Run analytics on DynamoDB data in Redshift

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Use cases at preview

- Create a searchable index of DynamoDB data in Elasticsearch

- Integrate operational data in DynamoDB in a data lake

- Run analytics on DynamoDB data in Redshift

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Use cases at preview

- Create a searchable index of DynamoDB data in Elasticsearch

- Integrate operational data in DynamoDB in a data lake

- Run analytics on DynamoDB data in Redshift

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Getting started with AWS Glue Elastic Views

Signup for PREVIEW

https://pages.awscloud.com/AWS-Glue-Elastic-Views-Preview.html

Easily build a materialized view in a few steps

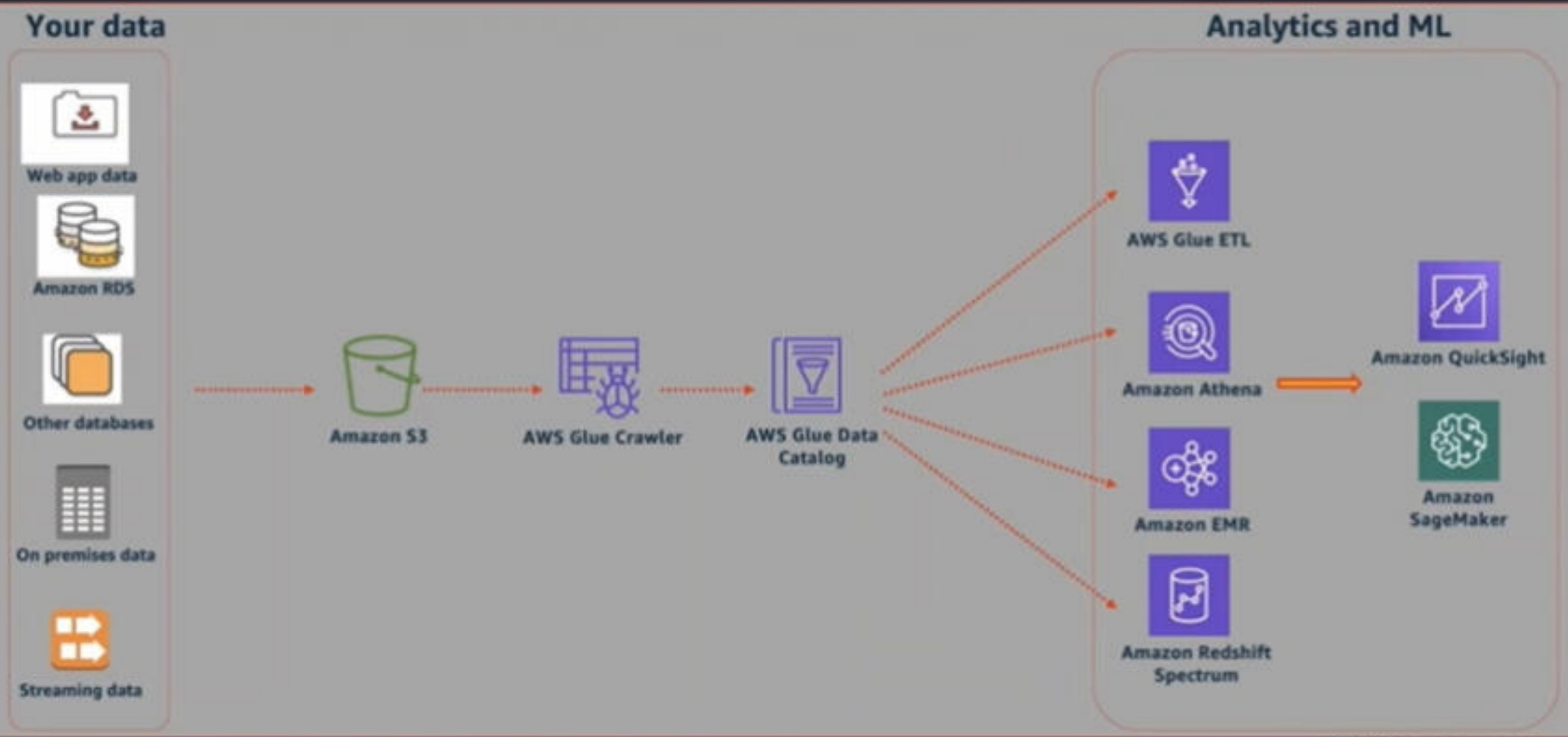
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Build AWS Data Lake with ease using AWS Lake Formation, powered by Glue

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

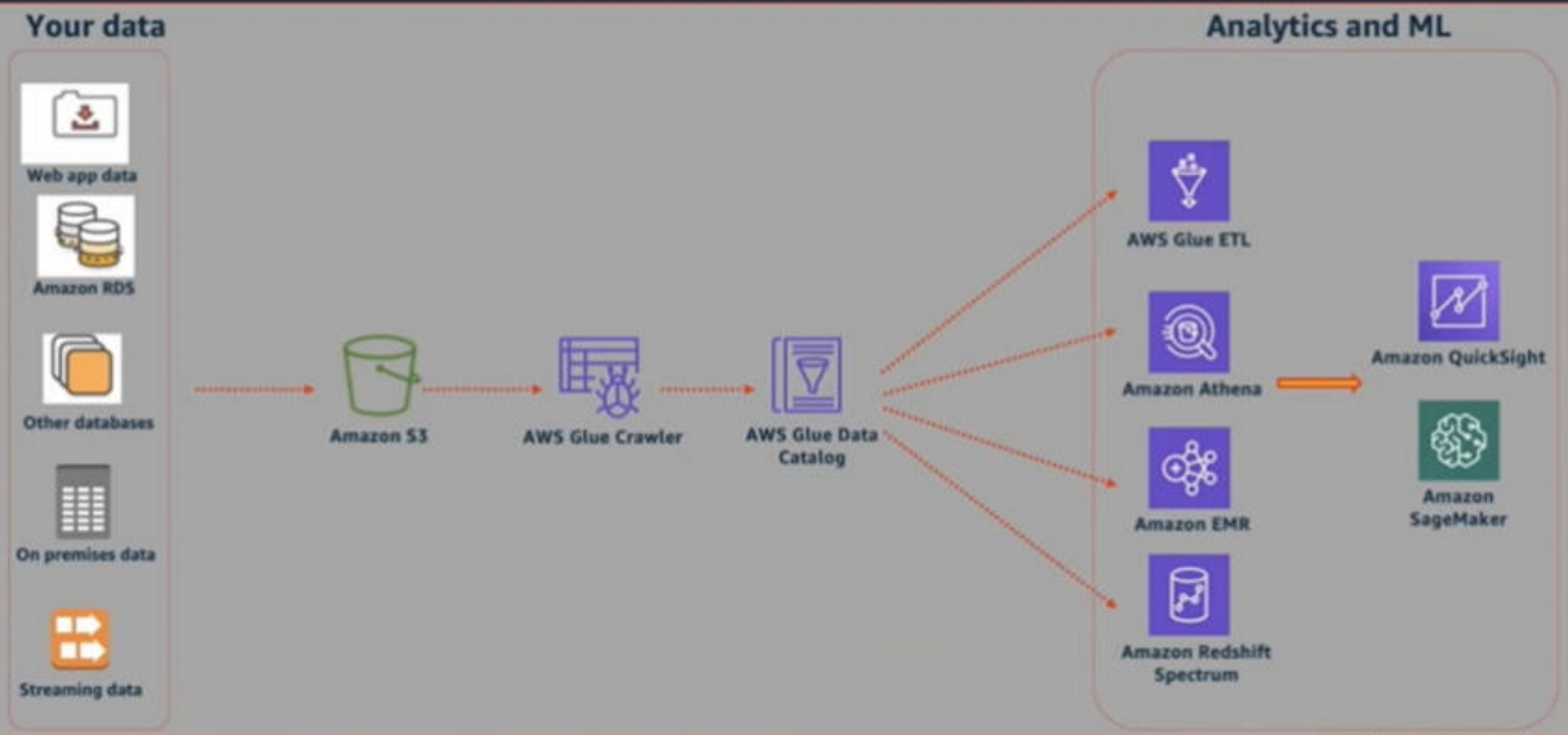
Data Lakes on AWS



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

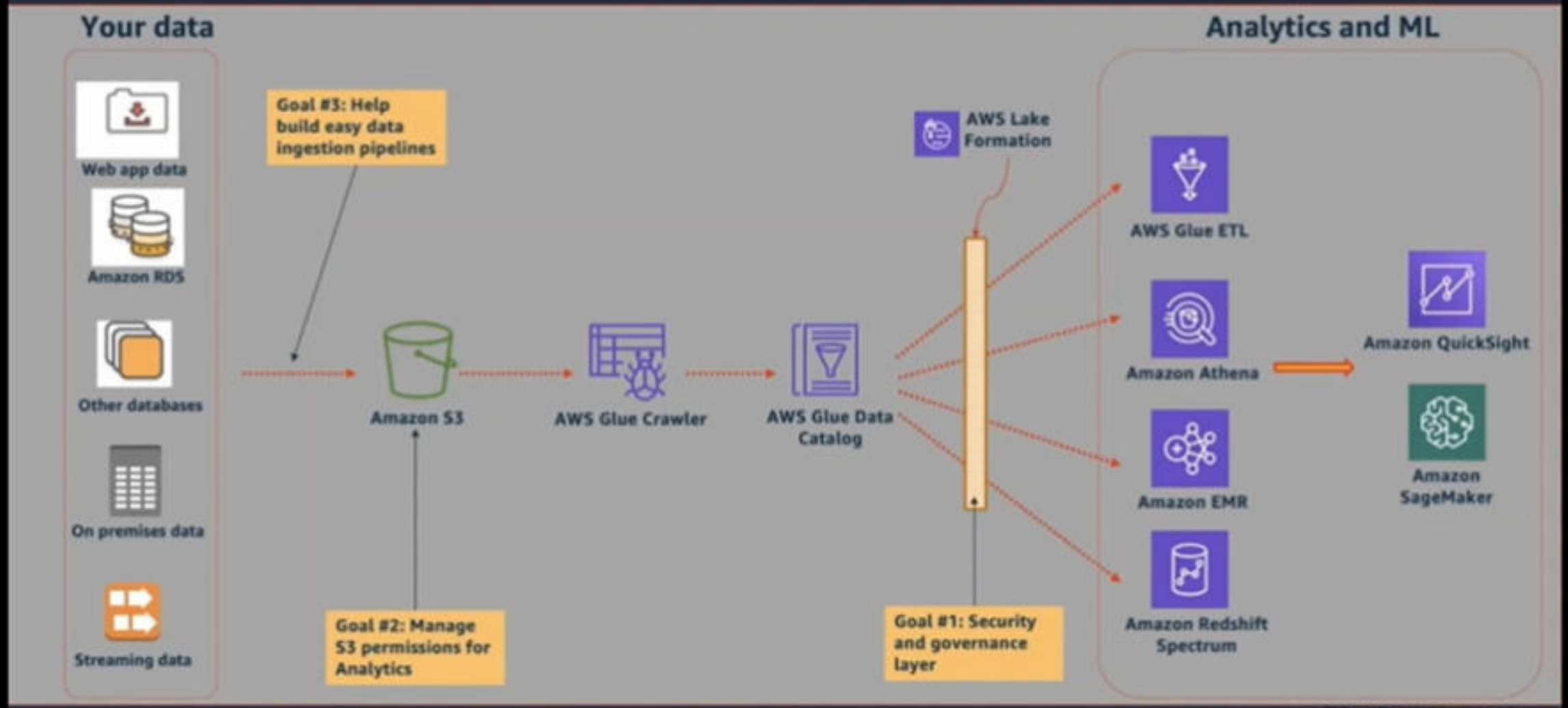
Data Lakes on AWS



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

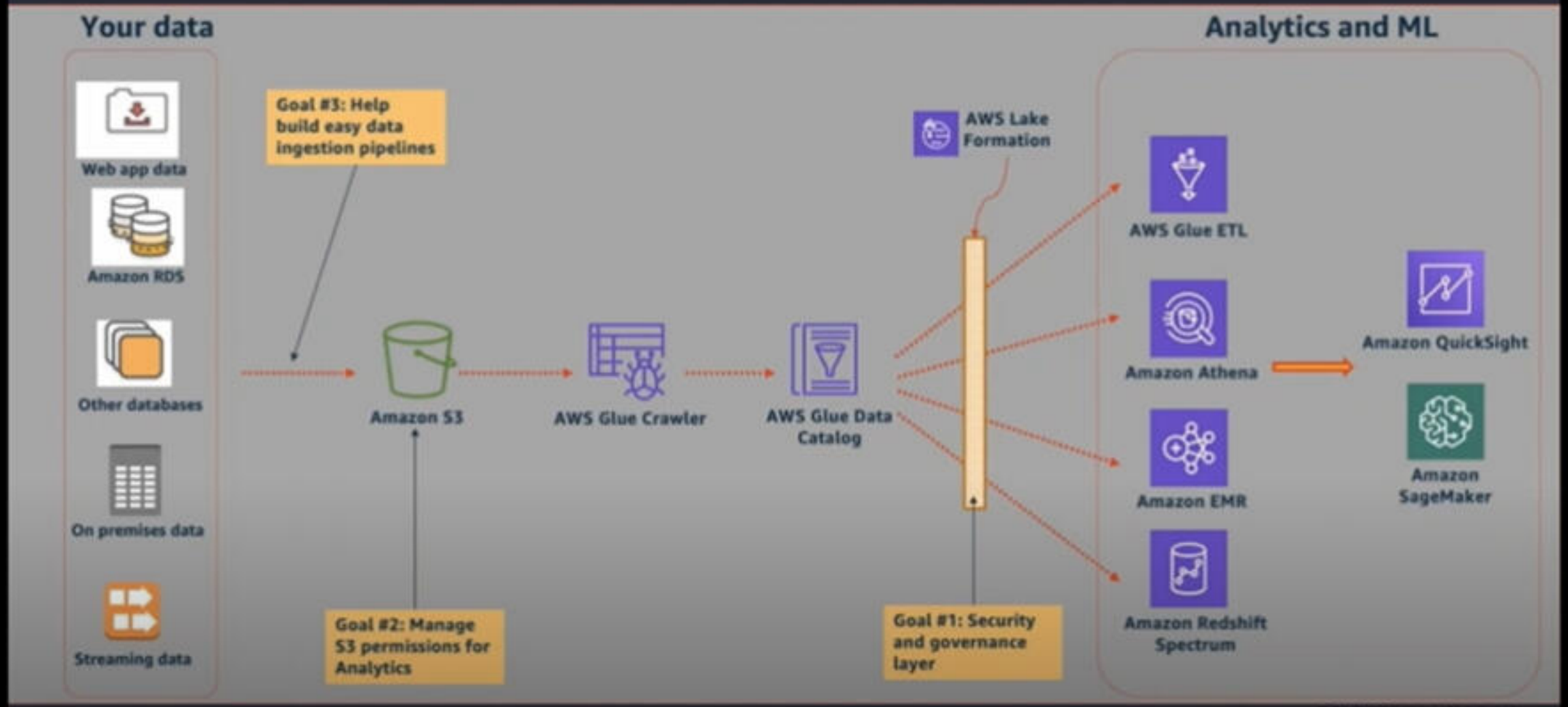
Data Lakes on AWS



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

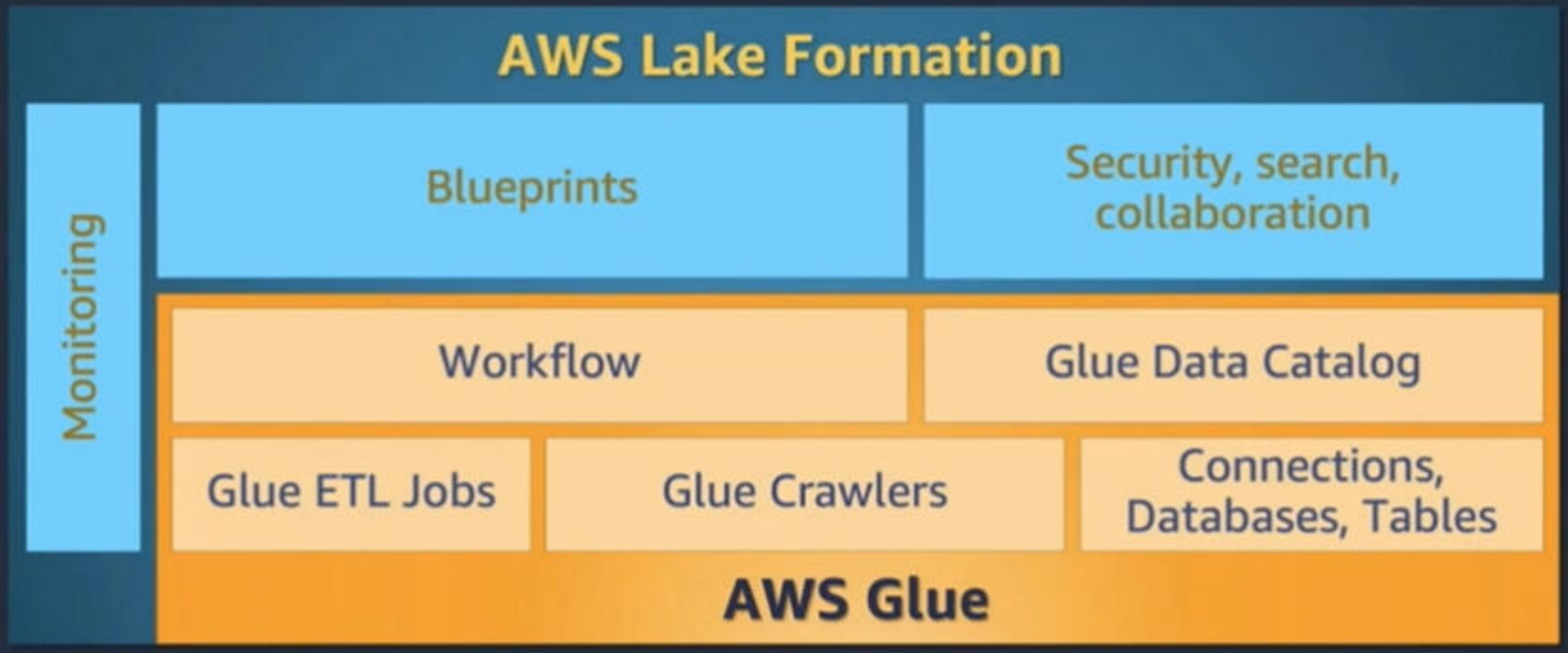
Data Lakes on AWS



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

AWS Glue Powers AWS Lake Formation



aws training and certification

Desktop 1

GoToWebinar

Connected to GoToWebinar

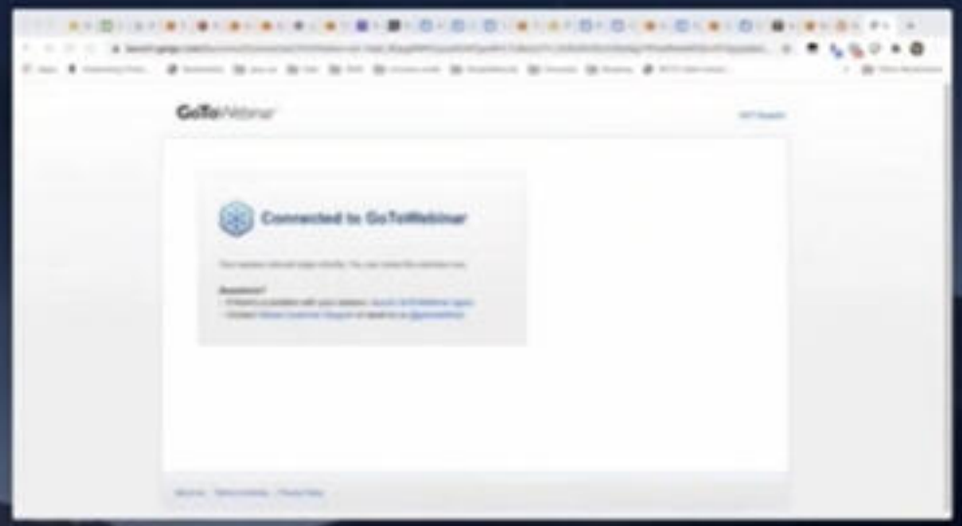
Demo

Lake Formation

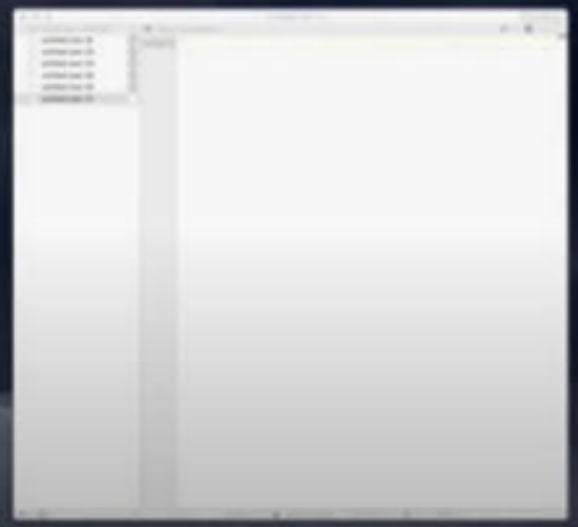
AWS Lake Formation

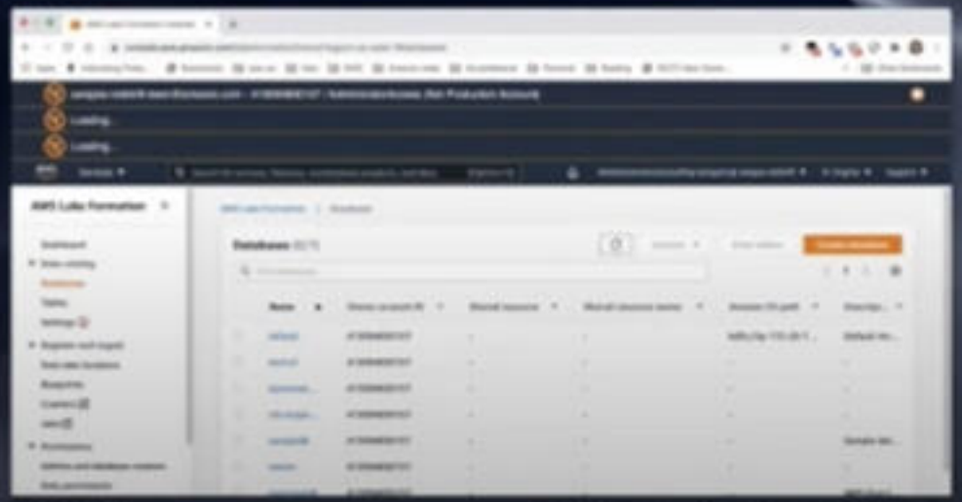
Adobe Acrobat Reader DC

Desktop 1


GoToWebinar
Connected to GoToWebinar

Demo
Lake Formation




AWS Lake Formation

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#databases

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Databases

Databases (0/7)

Actions | View tables | Create database

Find databases

Name	Owner account ID	Shared resource	Shared resource owner	Amazon S3 path	Descript...
default	413094830157	-	-	hdfs://ip-172-20-7...	Default Hiv...
demo2	413094830157	-	-	-	-
dynamod...	413094830157	-	-	-	-
rds-snape...	413094830157	-	-	-	-
sampledb	413094830157	-	-	-	Sample dat...
slalom	413094830157	-	-	-	-
spectrumdb	413094830157	-	-	-	AWS Glue C...

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

© 2020, Amazon W

certification

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#databases

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Databases

Databases (0/7)

Actions | View tables | Create database

Find databases

Name	Owner account ID	Shared resource	Shared resource owner	Amazon S3 path	Descript...
default	413094830157	-	-	hdfs://ip-172-20-7...	Default Hiv...
demo2	413094830157	-	-	-	-
dynamod...	413094830157	-	-	-	-
rds-snape...	413094830157	-	-	-	-
sampledb	413094830157	-	-	-	Sample dat...
slalom	413094830157	-	-	-	-
spectrumdb	413094830157	-	-	-	AWS Glue C...

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Dashboard

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage	**Create a database**	**Grant permissions**
Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.	Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.	Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.
Register location	Create database	Grant permissions

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdministratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Dashboard

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1

Register your Amazon S3 storage

Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.

Register location

Stage 2

Create a database

Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.

Create database

Stage 3

Grant permissions

Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.

Grant permissions

Recent access activity (0/50)

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdministratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Dashboard

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1

Register your Amazon S3 storage

Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.

Register location

Stage 2

Create a database

Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.

Create database

Stage 3

Grant permissions

Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.

Grant permissions

Recent access activity (0/50)

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Register location | Create database | Grant permissions

Recent access activity (0/50)

Recent access activity for your data lake in AWS Cloudtrail. Events can take several minutes to appear in Cloudtrail and are limited to the last 90 days.

View event

Find events

Event name	Principal	Alert time
GetDataLakeSettings	sudhig-tsengard	Tue, Dec 15, 2020, 3:54 AM UTC
GetDataLakeSettings	sudhig-tsengard	Tue, Dec 15, 2020, 3:39 AM UTC
ListPermissions	sudhig-tsengard	Tue, Dec 15, 2020, 3:39 AM UTC
GetDataLakeSettings	sudhig-tsengard	Tue, Dec 15, 2020, 3:39 AM UTC
ListResources	sudhig-tsengard	Tue, Dec 15, 2020, 3:34 AM UTC

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1

Register your Amazon S3 storage

Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.

Register location

Stage 2

Create a database

Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.

Create database

Stage 3

Grant permissions

Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.

Grant permissions

Recent access activity (0/50)

View event

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

Register location

Amazon S3 location

Register an Amazon S3 path as the storage location for your data lake.

Amazon S3 path

Choose an Amazon S3 path for your data lake.

e.g.: s3://bucket/prefix/ | Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

Register location

Amazon S3 location

Register an Amazon S3 path as the storage location for your data lake.

Amazon S3 path

Choose an Amazon S3 path for your data lake.

e.g.: s3://bucket/prefix/ | Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

© 2020, Am... ...ertification

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

Apps Interesting Finds... Bookmarks aws-sa flats DWS Amazon-wide StudyMaterial Personal Reading IRCTC Next Gener... Other Bookmarks

Register location

Amazon S3 location

Amazon S3 path

Choose an Amazon S3 location in region us-east-1

S3

Cancel Select

IAM role

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

Choose an Amazon S3 location in region us-east-1

- aws-glue-scripts-413094830157-us-east-1
- aws-glue-temporary-413094830157-us-east-1
- aws-logs-413094830157-us-east-1
- awspsa-aod-recordings-2018
- awspsa-redshift-cloudtrail-bucket
- awspsa-redshift-guarddutylogs
- awspsa-redshift-lab
- awspsa-sampledata
- cf-templates-10521p1nfd0yo-us-east-1
- cloudtrail-awslogs-413094830157-0kyvvn5u-isengard-do-not-delete
- config-bucket-413094830157
- config-bucket-trustedservice-do-not-delete-413094830157
- demo-sudhig-workshop
- dms-j6mc3i6cf4dfdqkau2jqt4tuoa
- dynamodb-exports-413094830157-us-east-1
- rs-aod-recordings-2018

Cancel | Select

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

Apps Interesting Finds... Bookmarks aws-sa flats DWS Amazon-wide StudyMaterial Personal Reading IRCTC Next Gener... Other Bookmarks

Register location

Amazon S3 location

Choose an Amazon S3 location in region us-east-1

- awspsa-aod-recordings-2018
- awspsa-redshift-cloudtrail-bucket
- awspsa-redshift-guarddutylogs
- awspsa-redshift-lab
- awspsa-sampledata
- cf-templates-10521p1nfd0yo-us-east-1
- cloudtrail-awslogs-413094830157-0kyvvn5u-isengard-do-not-delete
- config-bucket-413094830157
- config-bucket-trustedservice-do-not-delete-413094830157
- demo-sudhig-workshop
- dms-j6mc3i6cf4dfdqkau2jqt4tuoa
- dynamodb-exports-413094830157-us-east-1
- rs-aod-recordings-2018
- rslab-sct-migration
- shared-content-awspsa
- tpcdhawspsa

Cancel Select

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Choose an Amazon S3 path for your data lake.

s3://demo-sudhig-workshop

Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

⚠ Do not select the service linked role if you plan to use EMR.

Cancel Register location

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

Services

Search for services, features, marketplace products, and docs [Option+S]

AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Choose an Amazon S3 path for your data lake.

s3://demo-sudhig-workshop

Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

Find IAM roles

AWSServiceRoleForLakeFormationDataAccess

AWSServiceRoleForLakeFormationDataAccess

Cancel

Register location

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

Services

Search for services, features, marketplace products, and docs [Option+S]

AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Choose an Amazon S3 path for your data lake.

s3://demo-sudhig-workshop

Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

⚠ Do not select the service linked role if you plan to use EMR.

Cancel

Register location

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-bucket

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Choose an Amazon S3 path for your data lake.

s3://demo-sudhig-workshop

Browse

Review location permissions - strongly recommended

Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

Review location permissions

IAM role

To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

⚠ Do not select the service linked role if you plan to use EMR.

Cancel Register location

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-list

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Your Amazon S3 path: s3://demo-sudhig-workshop has been registered successfully.

AWS Lake Formation > Data lake locations

Data lake locations (0/2)

Actions | Register location

Find data lake storage

Amazon S3 path	IAM role	Last modified
s3://demo-sudhig-workshop	AWSServiceRoleForLakeFormation...	Tue, Dec 15, 2020, 4:11 AM UTC
s3://tpcdhawspsa	AWSServiceRoleForLakeFormation...	Wed, Nov 18, 2020, 4:58 PM UTC

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-list

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Your Amazon S3 path: s3://demo-sudhig-workshop has been registered successfully.

AWS Lake Formation > Data lake locations

Data lake locations (0/2)

Actions | Register location

Find data lake storage

Amazon S3 path	IAM role	Last modified
s3://demo-sudhig-workshop	AWSServiceRoleForLakeFormation...	Tue, Dec 15, 2020, 4:11 AM UTC
s3://tpcdhawspsa	AWSServiceRoleForLakeFormation...	Wed, Nov 18, 2020, 4:58 PM UTC

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-list

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Data lake locations

Data lake locations (0/2)

Actions | Register location

Find data lake storage

Amazon S3 path	IAM role	Last modified
s3://demo-sudhig-workshop	AWSServiceRoleForLakeFormation....	Tue, Dec 15, 2020, 4:11 AM UTC
s3://tpcdhawspsa	AWSServiceRoleForLakeFormation....	Wed, Nov 18, 2020, 4:58 PM UTC

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-list

awspsa-redshift-team@amazon.com - 413094830157 / AdministratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Data lake locations

Data lake locations (0/2)

Actions | Register location

Find data lake storage

Amazon S3 path	IAM role	Last modified
s3://demo-sudhig-workshop	AWSServiceRoleForLakeFormation....	Tue, Dec 15, 2020, 4:11 AM UTC
s3://tpcdhawspsa	AWSServiceRoleForLakeFormation....	Wed, Nov 18, 2020, 4:58 PM UTC

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-list

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Data lake locations

Data lake locations (0/2)

Actions | Register location

Find data lake storage

1

Amazon S3 path	IAM role	Last modified
s3://demo-sudhig-workshop	AWSServiceRoleForLakeFormation....	Tue, Dec 15, 2020, 4:11 AM UTC
s3://tpcdhawspsa	AWSServiceRoleForLakeFormation....	Wed, Nov 18, 2020, 4:58 PM UTC

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#register-list

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Data lake locations

Data lake locations (0/2)

Actions | Register location

Find data lake storage

Amazon S3 path	IAM role	Last modified
s3://demo-sudhig-workshop	AWSServiceRoleForLakeFormation....	Tue, Dec 15, 2020, 4:11 AM UTC
s3://tpcdhawspsa	AWSServiceRoleForLakeFormation....	Wed, Nov 18, 2020, 4:58 PM UTC

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Dashboard

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage	**Create a database**	**Grant permissions**
Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.	Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.	Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.
Register location	Create database	Grant permissions

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdministratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Dashboard

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage	**Create a database**	**Grant permissions**
Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.	Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.	Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.
Register location	Create database	Grant permissions

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Database details

Create a database in the AWS Glue Data Catalog.

Database
Create a database in my account.

Resource link
Create a resource link to a shared database.

Name

demo

Location - *optional*

Choose an Amazon S3 path for this database, which eliminates the need to grant data location permissions on catalog table paths that are this location's children

e.g.: s3://bucket/prefix/

Browse

Description - *optional*

Enter a description

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

Apps Interesting Finds... Bookmarks aws-sa flats DWS Amazon-wide StudyMaterial Personal Reading IRCTC Next Gener... Other Bookmarks

Choose an Amazon S3 location in region us-east-1

S3

Cancel Select

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

Apps Interesting Finds... Bookmarks aws-sa flats DWS Amazon-wide StudyMaterial Personal Reading IRCTC Next Gener... Other Bookmarks

Choose an Amazon S3 location in region us-east-1

- aws-logs-413094830157-us-east-1
- awspsa-aod-recordings-2018
- awspsa-redshift-cloudtrail-bucket
- awspsa-redshift-guarddutylogs
- awspsa-redshift-lab
- awspsa-sampledata
- cf-templates-10521p1nfd0yo-us-east-1
- cloudtrail-awslogs-413094830157-0kyvvn5u-isengard-do-not-delete
- config-bucket-413094830157
- config-bucket-trustedservice-do-not-delete-413094830157
- demo-sudhig-workshop
- dms-j6mc3il6cf4dfdqkau2jqt4tuoa
- dynamodb-exports-413094830157-us-east-1
- rs-aod-recordings-2018
- rslab-sct-migration
- shared-content-awspsa
- tpcdhawspsa

Cancel Select

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

Apps Interesting Finds... Bookmarks aws-sa flats DWS Amazon-wide StudyMaterial Personal Reading IRCTC Next Gener... Other Bookmarks

Choose an Amazon S3 location in region us-east-1

- aws-logs-413094830157-us-east-1
- awspsa-aod-recordings-2018
- awspsa-redshift-cloudtrail-bucket
- awspsa-redshift-guarddutylogs
- awspsa-redshift-lab
- awspsa-sampledata
- cf-templates-10521p1nfd0yo-us-east-1
- cloudtrail-awslogs-413094830157-0kyvvn5u-isengard-do-not-delete
- config-bucket-413094830157
- config-bucket-trustedservice-do-not-delete-413094830157
- demo-sudhig-workshop
- dms-j6mc3i6cf4dfdqkau2jqt4tuoa
- dynamodb-exports-413094830157-us-east-1
- rs-aod-recordings-2018
- rslab-sct-migration
- shared-content-awspsa
- tpcdhawspsa

Cancel Select

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

demo-catdb

Location - *optional*

Choose an Amazon S3 path for this database, which eliminates the need to grant data location permissions on catalog table paths that are this location's children

s3://demo-sudhig-workshop

Browse

Description - *optional*

demo

Descriptions can be up to 2048 characters long.

Default permissions for newly created tables

This setting maintains existing AWS Glue Data Catalog behavior. You can still set individual permissions, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See **Changing Default Settings for Your Data Lake.**

☑ Use only IAM access control for new tables in this database

Cancel Create database

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

Services

Search for services, features, marketplace products, and docs [Option+S]

AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Database
Create a database in my account.

Resource link
Create a resource link to a shared database.

Name

demo-catdb

Location - *optional*
Choose an Amazon S3 path for this database, which eliminates the need to grant data location permissions on catalog table paths that are this location's children

s3://demo-sudhig-workshop

Browse

Description - *optional*

demo

Descriptions can be up to 2048 characters long.

Default permissions for newly created tables
This setting maintains existing AWS Glue Data Catalog behavior. You can still set individual permissions, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See Changing Default Settings for Your Data Lake.

Use only IAM access control for new tables in this database

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

Apps Interesting Finds... Bookmarks aws-sa flats DWS Amazon-wide StudyMaterial Personal Reading IRCTC Next Gener... Other Bookmarks

Choose an Amazon S3 location in region us-east-1

<S3

Cancel Select

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

Apps Interesting Finds... Bookmarks aws-sa flats DWS Amazon-wide StudyMaterial Personal Reading IRCTC Next Gener... Other Bookmarks

Choose an Amazon S3 location in region us-east-1

< S3

- aws-athena-query-results-413094830157-us-east-1 >
- aws-athena-query-results-us-east-1-413094830157 >
- aws-glue-scripts-413094830157-us-east-1 >
- aws-glue-temporary-413094830157-us-east-1 >
- aws-logs-413094830157-us-east-1 >
- awspsa-aod-recordings-2018 >
- awspsa-redshift-cloudtrail-bucket >
- awspsa-redshift-guarddutylogs >
- awspsa-redshift-lab >
- awspsa-sampledata >
- cf-templates-10521p1nfd0yo-us-east-1 >
- cloudtrail-awslogs-413094830157-0kyvvn5u-isengard-do-not-delete >
- config-bucket-413094830157 >
- config-bucket-trustedservice-do-not-delete-413094830157 >
- demo-sudhig-workshop >

Cancel Select

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-database

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Database
Create a database in my account.

Resource link
Create a resource link to a shared database.

Name

demo-catdb

Location - optional
Choose an Amazon S3 path for this database, which eliminates the need to grant data location permissions on catalog table paths that are this location's children

s3://awspsa-sampledata

Browse

Description - optional

demo

Descriptions can be up to 2048 characters long.

Default permissions for newly created tables
This setting maintains existing AWS Glue Data Catalog behavior. You can still set individual permissions, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See Changing Default Settings for Your Data Lake.

Use only IAM access control for new tables in this database

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#databases

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Your database: demo-catdb has been created successfully.

AWS Lake Formation > Databases

Databases (0/8)

Actions | View tables | Create database

Find databases

1

Name	Owner account ID	Shared resource	Shared resource owner	Amazon S3 path	Descript...
default	413094830157	-	-	hdfs://ip-172-20-7...	Default Hiv...
demo-catdb	413094830157	-	-	s3://awspsa-sampl...	demo
demo2	413094830157	-	-	-	-
dynamod...	413094830157	-	-	-	-
rds-snape...	413094830157	-	-	-	-

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#database-details/demo-catdb?catalogId=413094830157

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Your database: demo-catdb has been created successfully.

AWS Lake Formation > Databases > demo-catdb

demo-catdb

Actions | View tables | Edit | Delete

Database details

Name
demo-catdb

Description
demo

Amazon S3 path
s3://awspsa-sampledata

Default permissions for newly created tables
Use only IAM access control for new tables in this database

AWS accounts and organizations with access to all tables

Find AWS accounts and AWS organizations with access

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#database-details/demo-catdb?catalogId=413094830157

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Your database: demo-catdb has been created successfully.

AWS Lake Formation > Databases > demo-catdb

demo-catdb

Actions | View tables | Edit | Delete

Database details

Name
demo-catdb

Description
demo

Amazon S3 path
s3://awspsa-sampledata

Default permissions for newly created tables
Use only IAM access control for new tables in this database

AWS accounts and organizations with access to all tables

Find AWS accounts and AWS organizations with access

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#database-details/demo-catdb?catalogId=413094830157

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Your database: demo-catdb has been created successfully.

AWS Lake Formation > Databases > demo-catdb

demo-catdb

Actions | View tables | Edit | Delete

Database details

Name
demo-catdb

Description
demo

Amazon S3 path
s3://awspsa-sampledata

Default permissions for newly created tables
Use only IAM access control for new tables in this database

AWS accounts and organizations with access to all tables

Find AWS accounts and AWS organizations with access

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#database-details/demo-catdb?catalogId=413094830157

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Your database: demo-catdb has been created successfully.

AWS Lake Formation > Databases > demo-catdb

demo-catdb

Actions | View tables | Edit | Delete

Database details

Name
demo-catdb

Description
demo

Amazon S3 path
s3://awspsa-sampledata

Default permissions for newly created tables
Use only IAM access control for new tables in this database

AWS accounts and organizations with access to all tables

Find AWS accounts and AWS organizations with access

1

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

https://console.aws.amazon.com/lakeformation/home?region=us-east-1#dashbo...

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Grant permissions

Choose the access permissions to grant.

My account
User or role from this AWS account.

External account
AWS account or AWS organization outside of my account.

IAM users and roles
Add one or more IAM users or roles.

Choose IAM principals to add

Users

- ANT371User
  User
- awspsa
  User
- datalake_user
  User
- labuser1
  User

☐ Create table ☐ Alter ☐ Drop ☐ Describe

☐ Super

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Grant permissions

Choose the access permissions to grant.

My account
User or role from this AWS account.

External account
AWS account or AWS organization outside of my account.

IAM users and roles
Add one or more IAM users or roles.

Choose IAM principals to add

awspsa
User

SAML and Amazon QuickSight users and groups
Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database
Add one or more databases.

Choose databases

Database permissions
Choose the specific access permissions to grant.

- [ ] Create table
- [ ] Alter
- [ ] Drop
- [ ] Describe

- [ ] Super

This permission is the union of the individual permissions above and supersedes them. See here

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

awspsa ✕
User

SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb ✕
415094830157

Table - *optional*

Add one or more tables.

Choose tables

Database permissions

Choose the specific access permissions to grant.

☐ Create table ☐ Alter ☐ Drop ☐ Describe

☑ Super

This permission is the union of the individual permissions above and supersedes them. See here

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

awspsa ✕
User

SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb ✕
415094830157

Table - optional

Add one or more tables.

Choose tables

* All tables ✕

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☑ Super

This permission is the union of the individual permissions above and supersedes them. See here

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb
415094830157

Table - optional

Add one or more tables.

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [ ] Select
- [ ] Describe

- [x] **Super**

This permission is the union of the individual permissions above and supersedes them. See here

Grantable permissions

Choose the permissions that may be granted to others.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [ ] Select
- [ ] Describe

- [ ] **Super**

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

Database

Add one or more databases.

Choose databases

demo-catdb
415094830157

Table - optional

Add one or more tables.

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [x] Select
- [ ] Describe

- [x] **Super**

This permission is the union of the individual permissions above and supersedes them. See here

Grantable permissions

Choose the permissions that may be granted to others.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [ ] Select
- [ ] Describe

- [ ] **Super**

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa
User

SAML and Amazon QuickSight users and groups
Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database
Add one or more databases.

Choose databases

demo-catdb
415094830157

Table - optional
Add one or more tables.

Choose tables

* All tables

Table permissions
Choose the specific access permissions to grant.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [x] Select
- [x] Describe

- [x] Super

This permission is the union of the individual permissions above and supersedes them. See here

Grantable permissions
Choose the permissions that may be granted to others.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [ ] Select
- [ ] Describe

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Table - *optional*

Add one or more tables.

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [ ] Select
- [ ] Describe

- [x] Super

This permission is the union of the individual permissions above and supersedes them. See here

Grantable permissions

Choose the permissions that may be granted to others.

- [ ] Alter
- [ ] Insert
- [ ] Drop
- [ ] Delete
- [ ] Select
- [ ] Describe

- [ ] Super

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Resource does not exist or requester is not authorized to access requested permissions.

Cancel Grant

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

Table - optional

Add one or more tables.

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☐ Super

Required field

This permission is the union of the individual permissions above and supersedes them. See here

Required field

Grantable permissions

Choose the permissions that may be granted to others.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☐ Super

Required field

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Required field

Resource does not exist or requester is not authorized to access requested permissions.

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

User

SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb
415094830157

Table - optional

Add one or more tables.

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☐ Super

Required field

This permission is the union of the individual permissions above and supersedes them. See here

Required field

Grantable permissions

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☑ **Super**

This permission is the union of the individual permissions above and supersedes them. See here

Grantable permissions

Choose the permissions that may be granted to others.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☐ **Super**

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Resource does not exist or requester is not authorized to access requested permissions.

Cancel Grant

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

labuser16 ✕
User

SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb ✕
415094830157

Table - optional

Add one or more tables.

Choose tables

* All tables ✕

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

labuser16 ✕
User

SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb ✕
415094830157

Table - optional

Add one or more tables.

Choose tables

* All tables ✕

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Grant permissions

Choose the access permissions to grant.

- My account
  User or role from this AWS account.
- External account
  AWS account or AWS organization outside of my account.

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

sudhig
User

SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb
413094830157

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Grant permissions

Choose the access permissions to grant.

- My account
  User or role from this AWS account.
- External account
  AWS account or AWS organization outside of my account.

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

sudhig
User

SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database

Add one or more databases.

Choose databases

demo-catdb
413094830157

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

Alter Insert Drop Delete Select Describe

Super

Required field

This permission is the union of the individual permissions above and supersedes them. See here

Required field

Grantable permissions

Choose the permissions that may be granted to others.

Alter Insert Drop Delete Select Describe

Super

Required field

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Required field

Resource does not exist or requester is not authorized to access requested permissions.

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☑ **Super**

This permission is the union of the individual permissions above and supersedes them. See here

Grantable permissions

Choose the permissions that may be granted to others.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☐ **Super**

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Resource does not exist or requester is not authorized to access requested permissions.

Cancel Grant

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

Choose tables

* All tables

Table permissions

Choose the specific access permissions to grant.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☑ **Super**

This permission is the union of the individual permissions above and supersedes them. See here

Grantable permissions

Choose the permissions that may be granted to others.

☐ Alter ☐ Insert ☐ Drop ☐ Delete ☐ Select ☐ Describe

☐ **Super**

This permission allows the principal to grant any of the above permissions and supersedes those grantable permissions.

Resource does not exist or requester is not authorized to access requested permissions.

Cancel Grant

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdministratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Register your Amazon S3 storage

Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.

Register location

Create a database

Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.

Create database

Grant permissions

Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.

Grant permissions

Recent access activity (0/50)

Recent access activity for your data lake in AWS Cloudtrail. Events can take several minutes to appear in Cloudtrail and are limited to the last 90 days.

View event

Find events

Event name	Principal	Alert time

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#tables

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Tables

Tables

Actions | Create table using a crawler | Create table

Find table by properties

Name	Database	Owner account ID	Shared resource	Shared resource owner	Location

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

https://console.aws.amazon.com/lakeformation/home?region=us-east-1#tables

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Blueprint type

Configure a blueprint to create a workflow.

- Database snapshot
  Bulk load data to your data lake from MySQL, PostgreSQL, Oracle, and Microsoft SQL Server databases.
- Incremental database
  Load new data to your data lake from MySQL, PostgreSQL, Oracle, and SQL Server databases.
- AWS CloudTrail
  Bulk load data from AWS CloudTrail sources.
- Classic Load Balancer logs
  Load data from Classic Load Balancer logs.
- Application Load Balancer logs
  Load data from Application Load Balancer logs.

Import source

Configure the workflow source.

Database connection

Choose the connection to the data source. Create a connection in AWS Glue

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Blueprint type

Configure a blueprint to create a workflow.

- Database snapshot
  Bulk load data to your data lake from MySQL, PostgreSQL, Oracle, and Microsoft SQL Server databases.
- Incremental database
  Load new data to your data lake from MySQL, PostgreSQL, Oracle, and SQL Server databases.
- AWS CloudTrail
  Bulk load data from AWS CloudTrail sources.
- Classic Load Balancer logs
  Load data from Classic Load Balancer logs.
- Application Load Balancer logs
  Load data from Application Load Balancer logs.

Import source

Configure the workflow source.

Database connection

Choose the connection to the data source. Create a connection in AWS Glue

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

Blueprint type

Configure a blueprint to create a workflow.

- **Database snapshot**
  Bulk load data to your data lake from MySQL, PostgreSQL, Oracle, and Microsoft SQL Server databases.
- **Incremental database**
  Load new data to your data lake from MySQL, PostgreSQL, Oracle, and SQL Server databases.
- **AWS CloudTrail**
  Bulk load data from AWS CloudTrail sources.
- **Classic Load Balancer logs**
  Load data from Classic Load Balancer logs.
- **Application Load Balancer logs**
  Load data from Application Load Balancer logs.

Import source

Configure the workflow source.

Database connection

Choose the connection to the data source. Create a connection in AWS Glue

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Database connection

Choose the connection to the data source. Create a connection in AWS Glue

Database connection

Find database connections

lab

snapexp_demo

test-public

tpc-rs

ma/table (case sensitive).

Exclude patterns - *optional*

Specify a glob pattern to limit the workflow. Tables that match this pattern will be excluded. Learn more

Exclude pattern

Enter an exclude pattern

Remove exclusion

Add exclusion

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdministratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Database connection

Choose the connection to the data source. Create a connection in AWS Glue

tpc-rs

Source data path

Enter the path from which to ingest data. For JDBC databases with schema support, enter database/schema/table (case sensitive). Substitute the percent (%) wildcard for schema or table.

database-name/table-name

Exclude patterns - *optional*

Specify a glob pattern to limit the workflow. Tables that match this pattern will be excluded. Learn more

Exclude pattern

Enter an exclude pattern

Remove exclusion

Add exclusion

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Exclude patterns - *optional*

Specify a glob pattern to limit the workflow. Tables that match this pattern will be excluded. Learn more

Exclude pattern

Enter an exclude pattern

Remove exclusion

Add exclusion

Import target

Configure the target of the workflow.

Target database

Choose a database in the AWS Glue Data Catalog. Create database

Choose a database

Target storage location

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Enter an exclude pattern

Remove exclusion

Add exclusion

Import target

Configure the target of the workflow.

Target database

Choose a database in the AWS Glue Data Catalog. Create database

demo-catdb

Target storage location

Choose a data lake location or other Amazon S3 path.

s3://awspsa-sampledata

Browse

Data format

Choose the output data format.

Parquet

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Data format

Choose the output data format.

Parquet

Import frequency

Schedule the workflow.

Frequency

Choose how often to run the workflow.

Run on demand

Import options

Configure the workflow.

Workflow name

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

Import options

Configure the workflow.

Workflow name

wf1

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (_), and must be less than 256 characters long.

IAM role

Choose an IAM role

Table prefix

The table prefix that is used for catalog tables that are created.

Enter a table prefix

Table prefixes may contain lower case letters (a-z), numbers (0-9), hyphens (-), or underscores (_).

Maximum capacity - *optional*

Sets the number of data processing units (DPUs) that can be allocated when this job runs. A DPU is a relative measure of processing power

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

© 2020, Am... ...ertification

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

Import options

Configure the workflow.

Workflow name

wf1

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (_), and must be less than 256 characters long.

IAM role

Choose an IAM role

Table prefix

The table prefix that is used for catalog tables that are created.

Enter a table prefix

Table prefixes may contain lower case letters (a-z), numbers (0-9), hyphens (-), or underscores (_).

Maximum capacity - *optional*

Sets the number of data processing units (DPUs) that can be allocated when this job runs. A DPU is a relative measure of processing power

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

Import options

Configure the workflow.

Workflow name

wf1

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (_), and must be less than 256 characters long.

IAM role

LakeFormationWorkflowRole

Table prefix

The table prefix that is used for catalog tables that are created.

Enter a table prefix

Table prefixes may contain lower case letters (a-z), numbers (0-9), hyphens (-), or underscores (_).

Maximum capacity - *optional*

Sets the number of data processing units (DPUs) that can be allocated when this job runs. A DPU is a relative measure of processing power that consists of 4 vCPUs of compute capacity and 16 GB of memory.

Enter a maximum capacity

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#create-workflow

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

IAM role

LakeFormationWorkflowRole

Table prefix

The table prefix that is used for catalog tables that are created.

demo_

Table prefixes may contain lower case letters (a-z), numbers (0-9), hyphens (-), or underscores (_).

Maximum capacity - *optional*

Sets the number of data processing units (DPUs) that can be allocated when this job runs. A DPU is a relative measure of processing power that consists of 4 vCPUs of compute capacity and 16 GB of memory.

Enter a maximum capacity

Concurrency - *optional*

Sets the maximum number of concurrent runs that are allowed for this job. An error is returned when this threshold is reached. The default is 5.

5

Cancel Create

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#workflows

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

your data lake, either as bulk load snapshot, or incrementally load new data over time.

Balancer, and Application Load Balancer logs

Use blueprint

Creating workflow(s): wf1.
Your workflow(s) will be created in a few minutes.

Workflows (0/1)

Workflows are instances of ingestion blueprints in Lake Formation.

Actions Use blueprint

Find workflows

1

Name	Created on	Last updated	Last run status
demo-wf	Mon, Dec 14, 2020, 7:02 AM UTC	Mon, Dec 14, 2020, 7:02 AM UTC	COMPLETED

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#workflows

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

your data lake, either as bulk load snapshot, or incrementally load new data over time.

Balancer, and Application Load Balancer logs

Use blueprint

Creating workflow(s): wf1.
Your workflow(s) will be created in a few minutes.

Workflows (0/1)

Workflows are instances of ingestion blueprints in Lake Formation.

Actions Use blueprint

Find workflows

1

Name	Created on	Last updated	Last run status
demo-wf	Mon, Dec 14, 2020, 7:02 AM UTC	Mon, Dec 14, 2020, 7:02 AM UTC	COMPLETED

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=workflows

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

Loading...

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

AWS Glue

Data catalog

- Databases
  - Tables
  - Connections
- Crawlers
  - Classifiers
- Schema registries
  - Schemas
- Settings

ETL

- AWS Glue Studio
  New
- Workflows

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=workflows

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

Loading...

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

AWS Glue

Data catalog

- Databases
  - Tables
  - Connections
- Crawlers
  - Classifiers
- Schema registries
  - Schemas
- Settings

ETL

- AWS Glue Studio
  New
- Workflows

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=workflows

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

Loading...

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

AWS Glue

Data catalog

- Databases
  - Tables
  - Connections
- Crawlers
  - Classifiers
- Schema registries
  - Schemas
- Settings

ETL

- AWS Glue Studio
  New
- Workflows

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=workflows

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

AWS Glue

Data catalog

Databases

Tables

Connections

Crawlers

Classifiers

Schema registries

Schemas

Settings

ETL

AWS Glue Studio

New

Workflows

Workflows > demo-wf > Run: wr_959001b0fb17f796cf376da31d233cdc44ae8395affb0255f9a410d5879b22f8 > Graph

Stop run

Workflow name: demo-wf

Previous run id: -

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

© 2020, Ama... ertification

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=workflows

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Glue

Data catalog

Databases

Tables

Connections

Crawlers

Classifiers

Schema registries

Schemas

Settings

ETL

AWS Glue Studio

New

Workflows

Workflows > demo-wf > Run: wr_959001b0fb17f796cf376da31d233cdc44ae8395affb0255f9a410d5879b22f8 > Graph

Stop run

Workflow name: demo-wf

Previous run: -

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/glue/home?region=us-east-1#etl:tab=workflows

Apps | Interesting Finds... | Bookmarks | aws-sa | flats | DWS | Amazon-wide | StudyMaterial | Personal | Reading | IRCTC Next Gener... | Other Bookmarks

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws | Services | Search for services, features, marketplace products, and docs [Option+S] | AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

AWS Glue

Data catalog

Databases

Tables

Connections

Crawlers

Classifiers

Schema registries

Schemas

Settings

ETL

AWS Glue Studio

Workflows

Workflows > demo-wf > Run: wr_95f001b0fb17f796cf376da31d233cdc44ae8395affb0255f9a410d5879b22f8 > Graph

Stop run

Workflow name: demo-wf

Previous run: -

Feedback | English (US) | © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. | Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#workflow-details/demo-wf

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions
  - Data locations
  - External data filtering

Workflow details

Name demo-wf	Last updated Mon, Dec 14, 2020, 7:02 AM UTC
Last run status COMPLETED	Created on Mon, Dec 14, 2020, 7:02 AM UTC

Workflow runs (1)

Find Workflow runs

Name	Started on	Run ID
demo-wf	Mon, Dec 14, 2020, 7:03 AM UTC	wr_959001b0fb17f796cf376da31d233cdc44ae8395affb0255f9a410d5879b22f8

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#catalog-settings

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions
  - Data locations
  - External data filtering

Data lake administrators (0/1)

Administrators can view all metadata in the AWS Glue Data Catalog. They can also grant and revoke permissions on data resources to principals, including themselves.

Grant

Find administrators

Name	Type
sudhig	IAM user

Database creators

Choose IAM principals permitted to create databases in your AWS Glue Data Catalog.

Revoke Grant

Find database creators

1

Principal	Principal type	Permissions	Grantable
No permissions			

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#access-control

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions
  - Data locations
  - External data filtering

AWS Lake Formation > Permissions

Data permissions (2)

Choose a database or table for which to review, grant or revoke user permissions.

Revoke | Grant

Find by properties

1

Principal	Principal type	Resource type	Resource	Owner account ID	Permissions	Grantable	RAM Resource Share
Adminisrator rAccess	IAM role	Database	demo-catdb	413094830157	Super, Alter, Create table, Describe, Drop	Super, Alter, Create table, Describe, Drop	-
IAMAllowed Principals	Group	Database	demo-catdb	413094830157	Super	-	-

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

console.aws.amazon.com/lakeformation/home?region=us-east-1#access-control

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

Services

Search for services, features, marketplace products, and docs [Option+S]

AdministratorAccess/sudhig-tsengard @ awspsa-redshift | N. Virginia | Support

- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions
  - Data locations
  - External data filtering

AWS Lake Formation > Permissions

Data permissions (2)

Choose a database or table for which to review, grant or revoke user permissions.

Revoke | Grant

Find by properties

Principal	Principal type	Resource type	Resource	Owner account ID	Permissions	Grantable	RAM Resource Share
Adminisrator rAccess	IAM role	Database	demo-catdb	413094830157	Super, Alter, Create table, Describe, Drop	Super, Alter, Create table, Describe, Drop	-
IAMAllowed Principals	Group	Database	demo-catdb	413094830157	Super	-	-

Feedback | English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy | Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#databases

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Databases

Databases (0/8)

Actions | View tables | Create database

Find databases

Name	Owner account ID	Shared resource	Shared resource owner	Amazon S3 path	Descript...
default	413094830157	-	-	hdfs://ip-172-20-7...	Default Hiv...
demo-catdb	413094830157	-	-	s3://awspsa-sampl...	demo
demo2	413094830157	-	-	-	-
dynamod...	413094830157	-	-	-	-
rds-snape...	413094830157	-	-	-	-
sampledb	413094830157	-	-	-	Sample dat...
slalom	413094830157	-	-	-	-

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#databases

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Databases (0/8)

Actions | View tables | Create database

Find databases

Name	Owner account ID	Shared resource	Shared resource owner	Amazon S3 path	Descript...
default	413094830157	-	-	hdfs://ip-172-20-7...	Default Hiv...
demo-catdb	413094830157	-	-	s3://awspsa-sampl...	demo
demo2	413094830157	-	-	-	-
dynamod...	413094830157	-	-	-	-
rds-sample...	413094830157	-	-	-	-
sampledb	413094830157	-	-	-	Sample dat...
slalom	413094830157	-	-	-	-
spectrumdb	413094830157	-	-	-	AWS Glue C...

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#databases

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

Databases (0/8)

Actions | View tables | Create database

Find databases

Name	Owner account ID	Shared resource	Shared resource owner	Amazon S3 path	Descript...
default	413094830157	-	-	hdfs://ip-172-20-7...	Default Hiv...
demo-catdb	413094830157	-	-	s3://awspsa-sampl...	demo
demo2	413094830157	-	-	-	-
dynamod...	413094830157	-	-	-	-
rds-snape...	413094830157	-	-	-	-
sampledb	413094830157	-	-	-	Sample dat...
slalom	413094830157	-	-	-	-
spectrumdb	413094830157	-	-	-	AWS Glue C...

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

AWS Lake Formation

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions

AWS Lake Formation > Dashboard

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage	**Create a database**	**Grant permissions**
Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.	Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.	Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.
Register location	Create database	Grant permissions

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Chrome File Edit View History Bookmarks People Tab Window Help

AWS Lake Formation Console | AWS Glue Console

console.aws.amazon.com/lakeformation/home?region=us-east-1#dashboard

awspsa-redshift-team@amazon.com - 413094830157 / AdminisratorAccess (Not Production Account)

Loading...

Loading...

aws Services Search for services, features, marketplace products, and docs [Option+S] AdministratorAccess/sudhig-tsengard @ awspsa-redshift N. Virginia Support

- Dashboard
- Data catalog
  - Databases
  - Tables
  - Settings
- Register and ingest
  - Data lake locations
  - Blueprints
  - Crawlers
  - Jobs
- Permissions
  - Admins and database creators
  - Data permissions
  - Data locations
  - External data filtering

AWS Lake Formation > Dashboard

Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage	**Create a database**	**Grant permissions**
Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake.	Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows.	Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users.
Register location	Create database	Grant permissions

Feedback English (US)

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Privacy Policy Terms of Use

Use-cases and Reference Architecture
AWS Glue

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

How customers use **AWS Glue**

- Build Data Lakes and Lake Houses for scalable data analysis
- Process petabytes of data both in batch and real-time using Apache Spark
- Migrate from expensive traditional ETL solutions to gain flexibility and reduce costs
- Prepare data for Machine Learning
- Catalog data assets to make them available to AWS Analytics services

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

How customers use **AWS Glue**

- Build Data Lakes and Lake Houses for scalable data analysis
- Process petabytes of data both in batch and real-time using Apache Spark
- Migrate from expensive traditional ETL solutions to gain flexibility and reduce costs
- Prepare data for Machine Learning
- Catalog data assets to make them available to AWS Analytics services

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

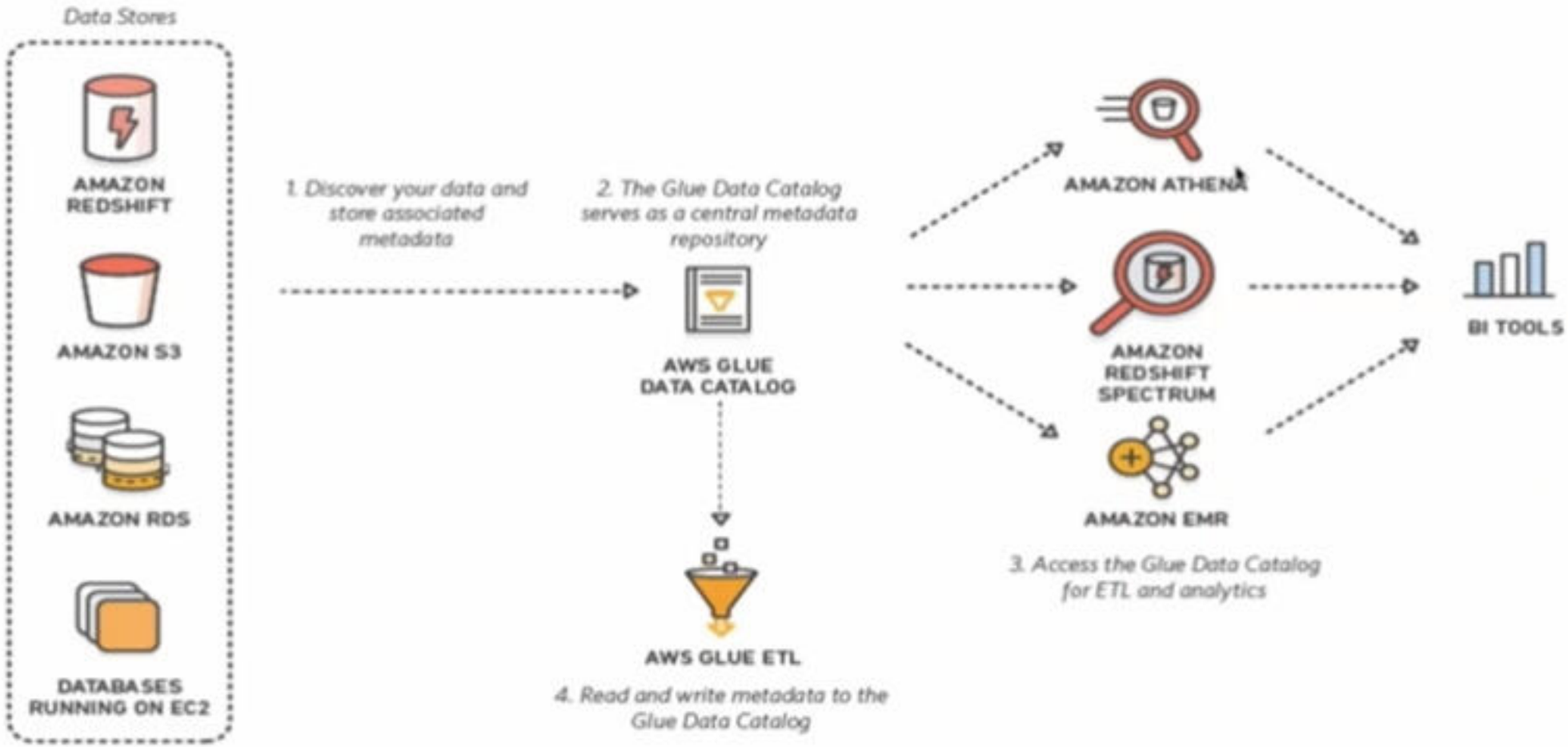
How customers use **AWS Glue**

- Build Data Lakes and Lake Houses for scalable data analysis
- Process petabytes of data both in batch and real-time using Apache Spark
- Migrate from expensive traditional ETL solutions to gain flexibility and reduce costs
- Prepare data for Machine Learning
- Catalog data assets to make them available to AWS Analytics services

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

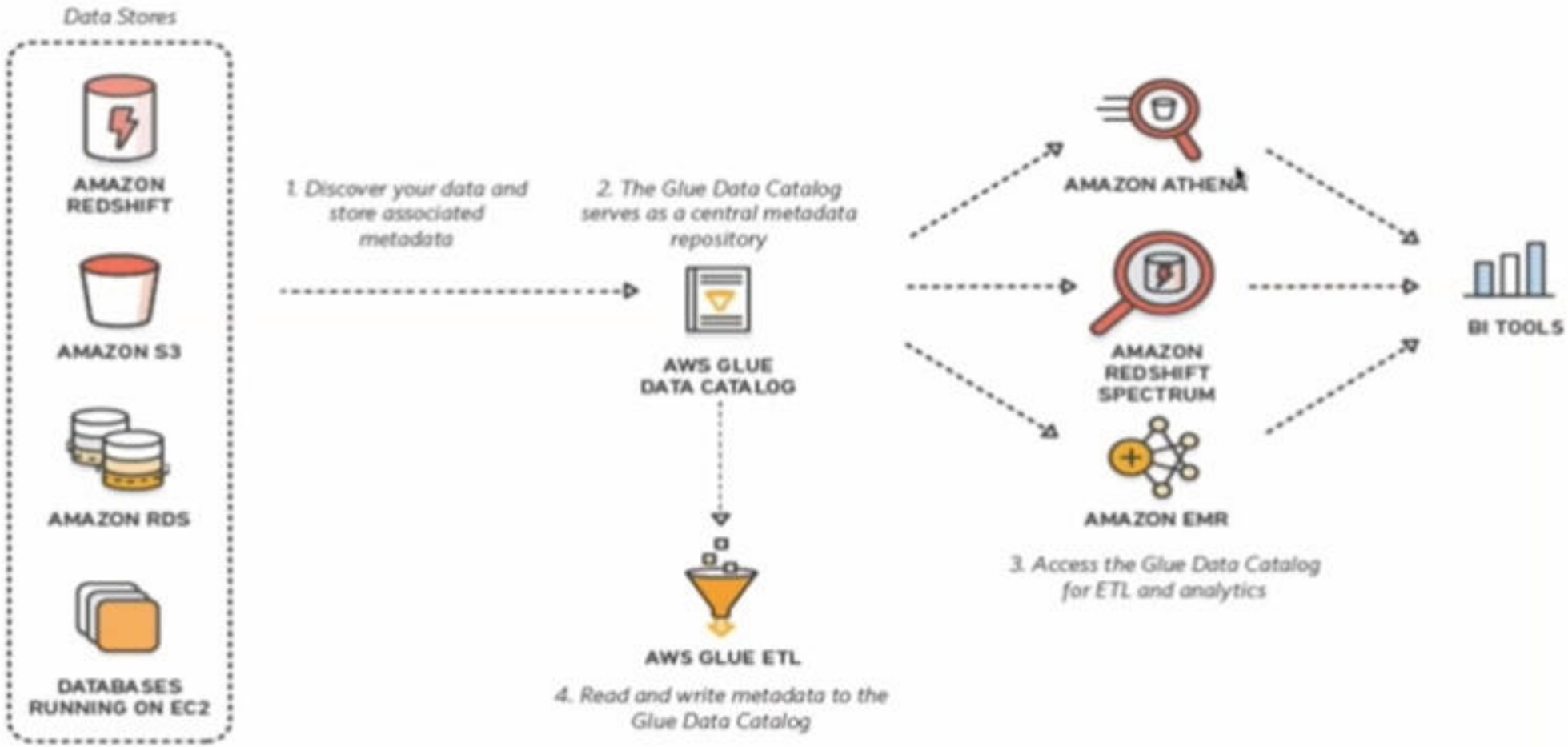
aws training and certification

Understand your data assets



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Understand your data assets



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Instantly query your data lake on Amazon S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Instantly query your data lake on Amazon S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Instantly query your data lake on Amazon S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

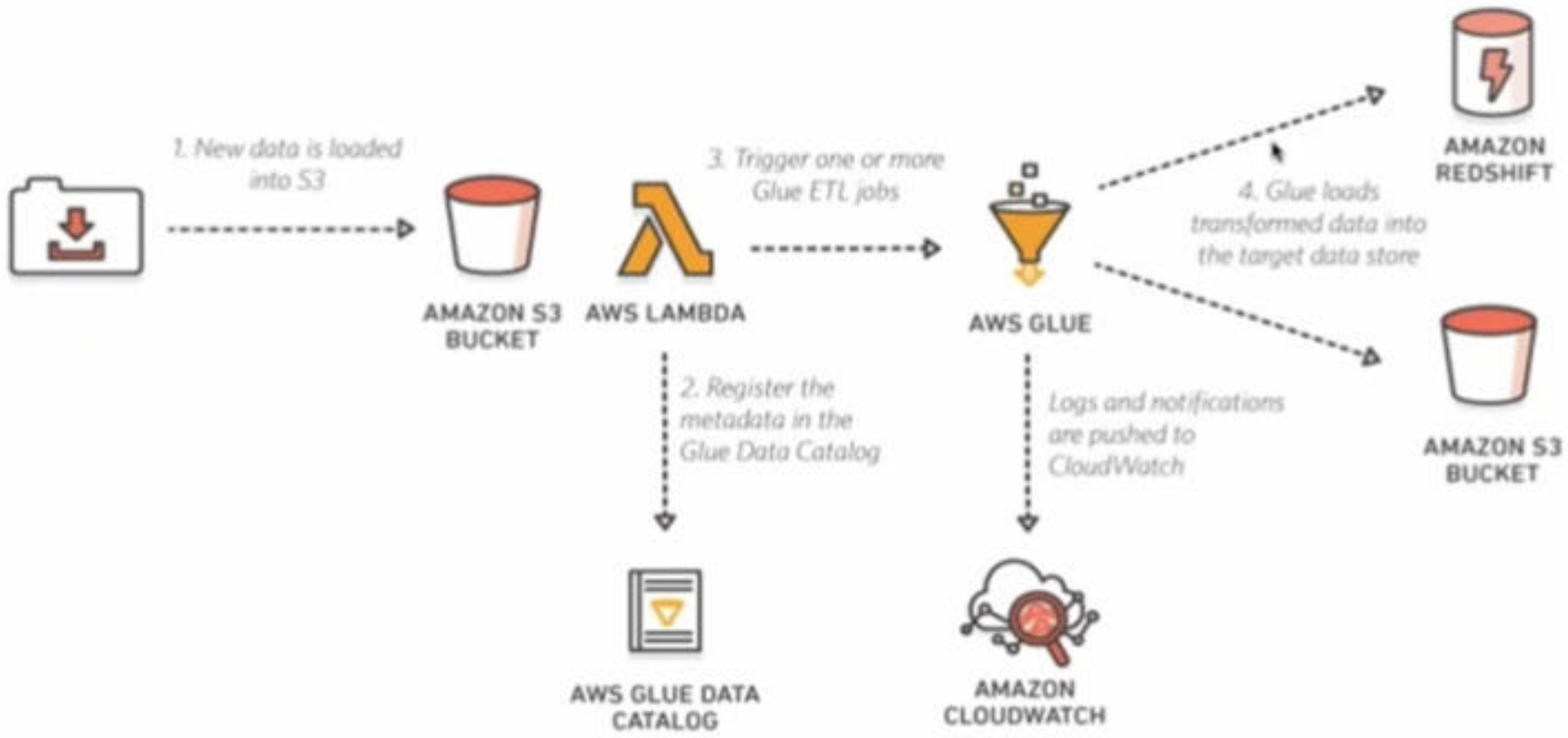
ETL data into your data warehouse

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

ETL data into your data warehouse

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Build event-driven ETL pipelines



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

ALICE uses AWS Glue to solve complex data migration challenges

ALICE

Challenge

ALICE acquired a large competitor, GoConcierge with a global customer base of over 1k hotels. Their challenge was to upgrade their customers without getting in the way of their operation. They needed a technology that was highly versatile and flexible in transforming one data structure into another.

Solution

The ALICE architecture leverages AWS Glue to load the data from the source database, transform it into the target data model, and to build new foreign keys for re-establishing the relationship within data set and between data set.

Benefits

They have migrated over 500 properties, and it continues to be stable. An average hotel takes one hour to run end-to-end. Glue made this all a very simple process and has been at the foundation of successfully upgrading our customers to the ALICE platform.

> "It is amazing how seamlessly Glue jobs can be integrated into a technology landscape."

Michael Dreikorn, Tech Lead

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

ALICE uses AWS Glue to solve complex data migration challenges

ALICE

Challenge

ALICE acquired a large competitor, GoConcierge with a global customer base of over 1k hotels. Their challenge was to upgrade their customers without getting in the way of their operation. They needed a technology that was highly versatile and flexible in transforming one data structure into another.

Solution

The ALICE architecture leverages AWS Glue to load the data from the source database, transform it into the target data model, and to build new foreign keys for re-establishing the relationship within data set and between data set.

Benefits

They have migrated over 500 properties, and it continues to be stable. An average hotel takes one hour to run end-to-end. Glue made this all a very simple process and has been at the foundation of successfully upgrading our customers to the ALICE platform.

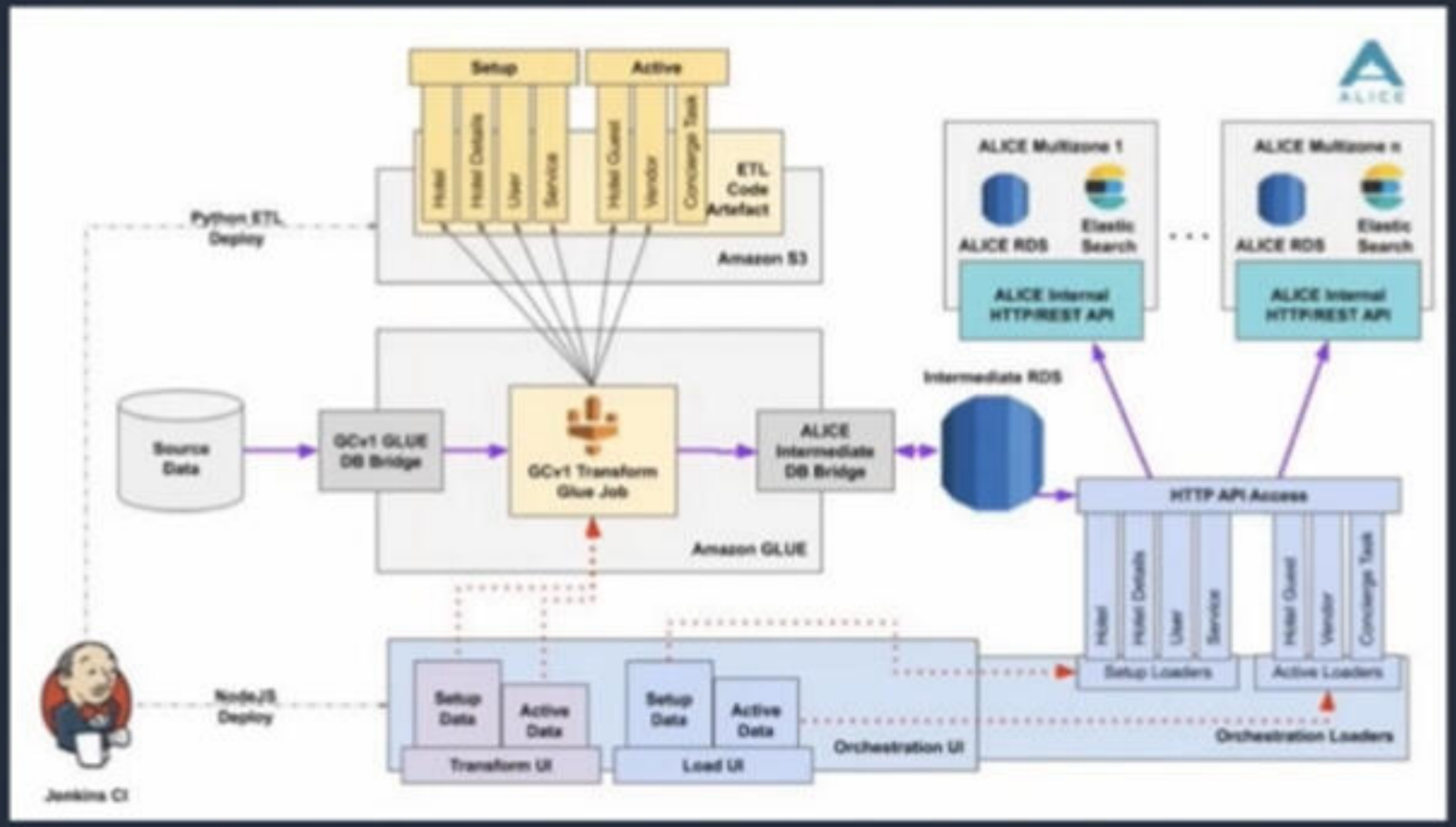
> "It is amazing how seamlessly Glue jobs can be integrated into a technology landscape."

Michael Dreikorn, Tech Lead

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

ALICE uses AWS Glue to solve complex data migration challenges



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

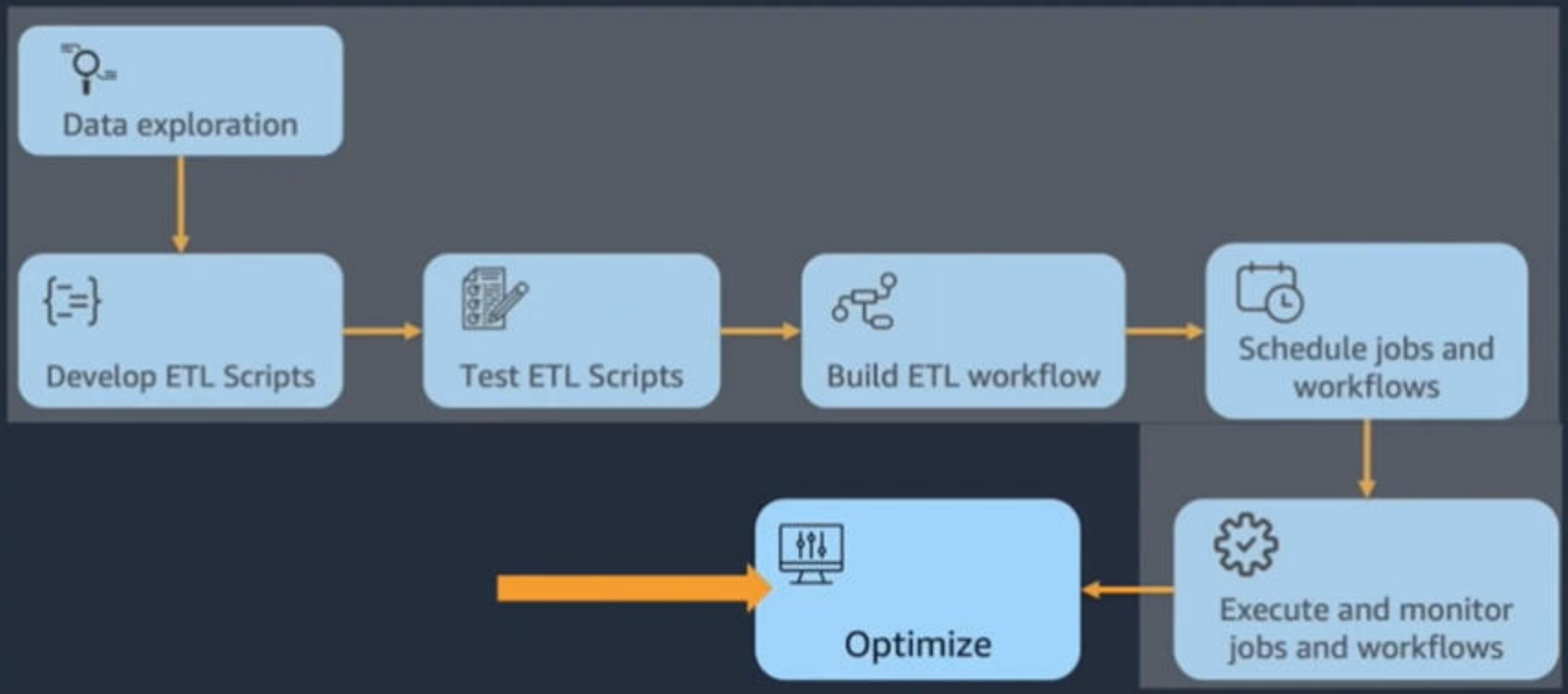
aws training and certification

AWS Glue Best Practices

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Steps in developing and CI/CD ETL processes



© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Optimize

a. Optimize Glue Crawler runs https://tinyurl.com/y4hm42qf

b. Optimize Glue jobs – no. of nodes (DPUs) and instances https://tinyurl.com/y9k4542k

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

How many crawlers do you need?

Crawler to AWS Glue table ratio

- Owner of the AWS Glue table owns the crawler
- How often do you have to crawl?
- Crawler mode
  - Update the entire table
  - Add new columns
  - Don't update the schema (add/remove partitions only)

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

How many crawlers do you need?

Crawler to AWS Glue table ratio

- Owner of the AWS Glue table owns the crawler
- How often do you have to crawl?
- Crawler mode
  - Update the entire table
  - Add new columns
  - Don't update the schema (add/remove partitions only)
- How many 'tables' do you have to crawl?
- How much data do you have to crawl?
- How many files do you have to crawl?
  - Too many small files?

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

How many crawlers do you need?

Crawler to AWS Glue table ratio

- Owner of the AWS Glue table owns the crawler
- How often do you have to crawl?
- Crawler mode
  - Update the entire table
  - Add new columns
  - Don't update the schema (add/remove partitions only)
- How many 'tables' do you have to crawl?
- How much data do you have to crawl?
- How many files do you have to crawl?
  - Too many small files?

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Save money on Crawlers

Pay-as-you-go pricing, longer the crawlers run, more expensive it gets.

1. If schema is not changing and no partitions being added, don't run the crawler. Just because new data is added/removed doesn't mean you have to run crawlers

2. Use fewer files to build tables and bootstrap the data lake – data lake bootstrapper.

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Save money on Crawlers – Use No crawler

3. Use AWS Glue APIs or Hive DDLs to manage AWS Glue data catalog tables.

4. Manage partitions also through AWS Glue APIs

5. Use 'Show Create Table' statement to migrate DDLs instead of running crawlers again.

6. Use a 'Create Table' DDL to create the table first and then attach a Crawler to the table for schema/partition management

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

PartitionIndex for partitioned tables

Improve query performance for highly partitioned data

For highly partitioned data query performance can degrade

PartitionIndexes allow retrieval of relevant partitions quickly

Improves query performance and reduces data transfer

Enables interactive exploration of very large data sets

Powered by Glue Data Catalog

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Development Environment and Testing best practices

- Setup AWS Glue libraries(v1.0) available through public Amazon S3 buckets
- Consider packaging the libraries as Docker container for portability
- Setup an IDE like PyCharm / Jupyter
- Explore using Glue development endpoint, as alternative
- Build unit test suites leveraging the local libraries
    - *Mock data or use sample files*
    - *Pytest or ScalaTest*
    - *Modularize the code for streamlined testing*
- Integrate with your source code repository locally
- Leverage the full open source Spark APIs and AWS Glue samples

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

General Coding best practices

- PySpark is different from Python
  - *Interfaces with the JVM and runs distributed code*
  - *Perform minimal data manipulation on the driver code*
- Know your data
  - *Allows customizing your code early. For example to decide the optimal join strategy*
- Do not overuse Spark *caching*
- Prefer using *in-built functions* instead of developing custom UDF
- Consider using Pandas UDF instead of Python UDF, if possible
- For complicated custom UDF, consider developing using Scala and exposing it as SQL function to use from PySpark
- Get familiar with the Spark UI and query plans
- Isolate the data rows with errors through the *error_as_dynamic_frame()* method for Dynamic Frames. Useful for future debugging

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Deployment Process (CI / CD) best practices

- Maintain the Glue crawler / job definition on your source code repo
  - *json file or CloudFormation templates*
- Depending on your git lifecycle practices, build/create the updated codebase
  - *Example: create python library, jar, configuration files, modified job definition etc.)*
  - *Execute the test cases on local sample data gets executed during this step*
- Deploy the artifacts to a staging environment on AWS
  - *Creating/updating the Glue crawler/jobs*
  - *Moving the generated libraries and scripts to S3*
  - *Running manual/automated integration tests, data validation*
  - *Approve production deployment*
- Deploy the artifacts to production. Refer this blog for an example end-to-end flow

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Deployment Environment best practices

- Confirm the network communication paths are available for your resources
  - *Subnet config, Firewall , DNS configs*
  - *Available IP addresses for higher DPU jobs*
- Ensure Glue has the right access
  - *Check ACLs, Egress Security Groups*
- VPC Flow logs can be used to troubleshoot connectivity issues
- Explore using S3 access via Endpoint
- Consider using AWS Glue with VPC Endpoints
- Start with the standard worker type
- Use the G.1X or G.2X worker types for memory intensive jobs

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Operational best practices - Auditing and Monitoring

- Enable job metrics
  - *During debugging and optimization*
  - *Metrics may be turned off to save on cost for stable production jobs*
- Enable Continuous Logging
  - *Driver and Executor logs on CloudWatch*
- Spark History event logs
  - *Enables you to use Spark UI*
  - *Best is to leave them on for production jobs*
- References
  - *Job Monitoring*
  - *Optimize Memory Management*

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Performance optimization: best practices for Data Processing - Input

- While reading many small files
  - *leverage the groupFile and groupSize features of Dynamic Frames*
  - *use the Glue S3 List Implementation for optimized performance*
  - *Decreases memory pressure*
- Avoid reading gzip files as input sources or writing out for future consumption
- Enable reading in parallel from JDBC tables using a combination of the different *hash* options to partition the data
- Use *pushdown predicate* while reading partitioned data for better performance
- Consider using Job Bookmarks for incremental processing
  - *use tranformationContext while reading input source*
  - *use Job.init() and Job.commit()*
  - *last modified timestamp for S3 file should be later than last successful job run*
  - *primary key / incremental column present for JDBC data source*

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Performance optimization: best practices for Data Processing - Output

- Prefer storing data using columnar formats like Parquet / ORC
  - *write-once read-multiple (WORM) data storage*
  - *target 128 MB+ file size*
- Consider partitioning the data based on your query patterns
  - *Avoid over-partitioning by using very high cardinality column (hundred thousands)*
  - *Consider tweaking the DataFrame partitions to create large files (~ 128 MB) while writing out*
  - *Explore bucketing the data for better downstream query performance*
- Check the amount of data shuffled during job execution by looking at the metrics, Spark UI, logs
- Use the optimized Apache Parquet writer (*glueparquet* format option)
  - *Computes and merges the schema dynamically at runtime*
  - *Supports schema evolution through deletion and addition of columns*
  - *Allows providing Parquet block_size and page_size*

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Job Execution best practices

- Familiarize with AWS Glue special job parameters
- Update jobs to have appropriate resources, timeouts and retries based on historical job run trends
- Consider using Glue workflows for orchestration
  - *Combine jobs and crawlers into complex data flows*
  - *Use Glue Python Shell job Pre/Post processing*
- Enable EMRFS S3-optimized committer for writing Parquet data
  - Set the parameter *--enable-s3-parquet-optimized-committer* to *true*
  - *Improves performance*
  - *Helps avoid issues with Amazon S3 eventual consistency*
- References
  - *Best Practices to scale Spark Jobs*
  - *Load Data Incrementally*

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Security and Governance

- Segment the jobs by IAM role / business purpose
- Able to turn on encryption for the data , Glue data catalog, Job bookmarks
    - *Integrated with AWS Key Management Services (KMS)*
- Use VPC endpoints
- Use Glue Security Configuration for sharing security settings
- Consider using AWS Lake Formation for centralized and granular access control
    - *Integrated with Amazon Athena, Amazon Redshift Spectrum, Amazon EMR (Beta)*

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Troubleshooting - Scenario

- Slow jobs and OOM Errors
  - Check the driver and executor logs for additional details
  - Check Spark UI to determine Spark Job stage and operations with OOM error
  - Confirm the data is uniformly partitioned in Spark
  - Check metrics from CloudWatch and use appropriate worker instances
  - Refer to OOM Debugging , Straggler Debugging and Memory Management

- S3 Throttling errors
  - Check the driver and executor logs S3 specific error
  - Repartition to write bigger files
  - Use EMRFS parquet committer
  - Use less number of instances to decrease parallel writes to S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Troubleshooting - Scenario

- Slow jobs and OOM Errors
  - Check the driver and executor logs for additional details
  - Check Spark UI to determine Spark Job stage and operations with OOM error
  - Confirm the data is uniformly partitioned in Spark
  - Check metrics from CloudWatch and use appropriate worker instances
  - Refer to OOM Debugging , Straggler Debugging and Memory Management

- S3 Throttling errors
  - Check the driver and executor logs S3 specific error
  - Repartition to write bigger files
  - Use EMRFS parquet committer
  - Use less number of instances to decrease parallel writes to S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Troubleshooting - Scenario

- Slow jobs and OOM Errors
  - Check the driver and executor logs for additional details
  - Check Spark UI to determine Spark Job stage and operations with OOM error
  - Confirm the data is uniformly partitioned in Spark
  - Check metrics from CloudWatch and use appropriate worker instances
  - Refer to OOM Debugging , Straggler Debugging and Memory Management

- S3 Throttling errors
  - Check the driver and executor logs S3 specific error
  - Repartition to write bigger files
  - Use EMRFS parquet committer
  - Use less number of instances to decrease parallel writes to S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Troubleshooting - Scenario

- Slow jobs and OOM Errors
  - Check the driver and executor logs for additional details
  - Check Spark UI to determine Spark Job stage and operations with OOM error
  - Confirm the data is uniformly partitioned in Spark
  - Check metrics from CloudWatch and use appropriate worker instances
  - Refer to OOM Debugging , Straggler Debugging and Memory Management

- S3 Throttling errors
  - Check the driver and executor logs S3 specific error
  - Repartition to write bigger files
  - Use EMRFS parquet committer
  - Use less number of instances to decrease parallel writes to S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Troubleshooting - Scenario

- Slow jobs and OOM Errors
  - Check the driver and executor logs for additional details
  - Check Spark UI to determine Spark Job stage and operations with OOM error
  - Confirm the data is uniformly partitioned in Spark
  - Check metrics from CloudWatch and use appropriate worker instances
  - Refer to OOM Debugging , Straggler Debugging and Memory Management

- S3 Throttling errors
  - Check the driver and executor logs S3 specific error
  - Repartition to write bigger files
  - Use EMRFS parquet committer
  - Use less number of instances to decrease parallel writes to S3

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

Troubleshooting - Scenario

- Crawler running for too long
  - Consider whether you need to run a crawler
  - Use exclude patterns
  - Run multiple crawlers (maybe on each partition)
  - Check if you have large number of small files
  - If you have large number of data sources to crawl, consider using the Glue Catalog API to define those tables in the data catalog

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

Troubleshooting - Scenario

- Crawler running for too long
  - Consider whether you need to run a crawler
  - Use exclude patterns
  - Run multiple crawlers (maybe on each partition)
  - Check if you have large number of small files
  - If you have large number of data sources to crawl, consider using the Glue Catalog API to define those tables in the data catalog

© 2020, Am

Upcoming Sessions

Mastering Amazon Redshift 5-sessions series:

- Feb 8 – Feb 12, 2021 at 8:00am PT (AMER/EMEA)
  https://register.gotowebinar.com/register/3981980727766618384

- Feb 8 – Feb 11 & Feb 15, 2021 at 7:00pm PT
  (APJC) https://register.gotowebinar.com/register/3811476161116374797

Bookmark AWS PartnerCast page: https://aws.amazon.com/partners/training/partnercast/

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark.

aws training and certification

aws training and certification

Certificate of Completion

Hem Bahadur Gurung

Has successfully completed

AWS PartnerCast: AWS Glue Deep Learning Series

Director, Training and Certification	60 minutes	10 September, 2021
	Duration	Completion Date