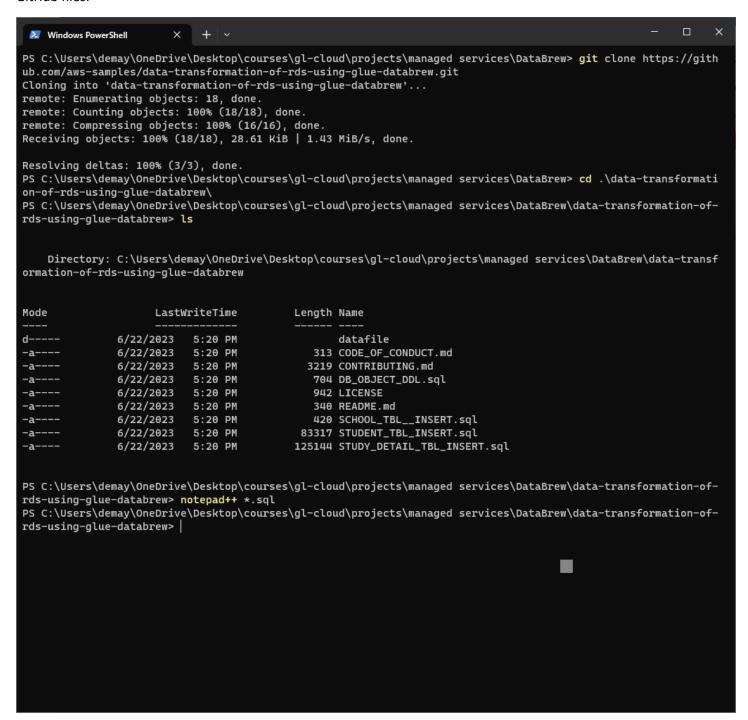From Olympus, Managed Services on AWS course. Under Advanced DIY

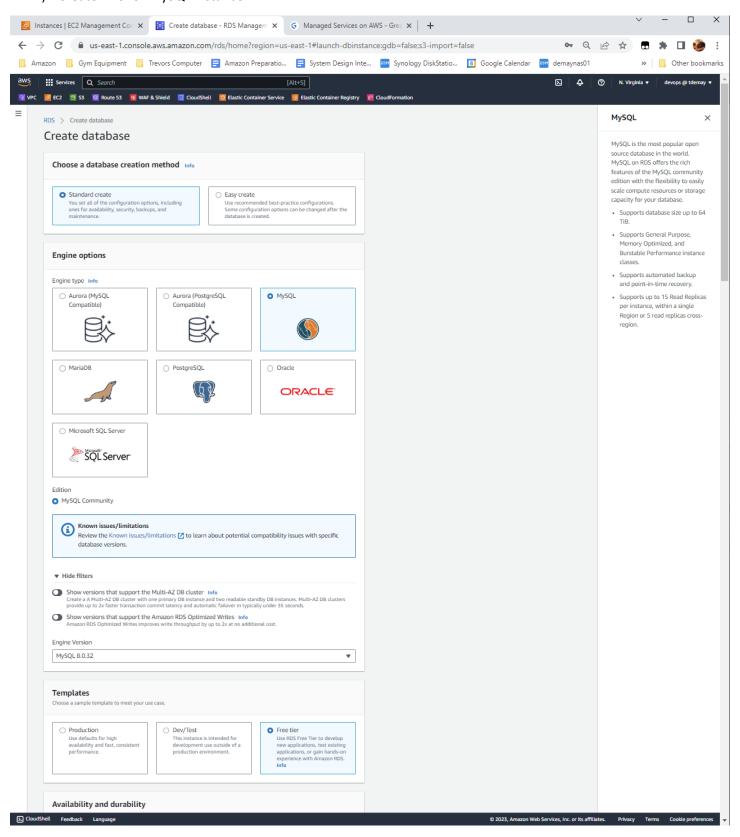https://olympus.mygreatlearning.com/courses/89455?module_id=598785

Data preparation using an Amazon RDS for MySQL database with AWS Glue DataBrew

https://aws.amazon.com/blogs/big-data/data-preparation-using-an-amazon-rds-for-mysql-database-with-aws-glue-databrew/

GitHub files:

Prelab steps

1) Create RDS for MySQL instance

Instances | EC2 Management Cor ×   Create database - RDS Managen ×   G   Managed Services on AWS - Gre   ×   +

◁ ▷ C   🔒 us-east-1.console.aws.amazon.com/rds/home?region=us-east-1#launch-dbinstance:gdb=false;s3-import=false

📁 Amazon   📁 Gym Equipment   📁 Trevors Computer   📄 Amazon Preparatio...   📄 System Design Inte...   DSM Synology DiskStatio...   8 Google Calendar   DSM demaynas01   »   📁 Other bookmarks

aws   ▦ Services   🔍 Search   [Alt+S]   ⌨   🔔   ⑦   N. Virginia ▾   devops @ tdemay ▾

🟦 VPC   🟧 EC2   🟧 S3   🟪 Route 53   🟥 WAF & Shield   🟦 CloudShell   🟧 Elastic Container Service   🟧 Elastic Container Registry   🟧 CloudFormation

---

☰

Availability and durability

Deployment options Info
The deployment options below are limited to those supported by the engine you selected above.

⦿ Multi-AZ DB Cluster - *new*
    Creates a DB cluster with a primary DB instance and two readable standby DB instances, with each DB instance in a different
    Availability Zone (AZ). Provides high availability, data redundancy and increases capacity to serve read workloads.

⦿ Multi-AZ DB instance (not supported for Multi-AZ DB cluster snapshot)
    Creates a primary DB instance and a standby DB instance in a different AZ. Provides high availability and data redundancy, but
    the standby DB instance doesn't support connections for read workloads.

⦿ Single DB instance (not supported for Multi-AZ DB cluster snapshot)
    Creates a single DB instance with no standby DB instances.

Settings

DB instance identifier Info
Type a name for your DB instance. The name must be unique across all DB instances owned by your AWS account in the current AWS
Region.

```
databrew-rds
```

The DB instance identifier is case-insensitive, but is stored as all lowercase (as in "mydbinstance"). Constraints: 1 to 60 alphanumeric
characters or hyphens. First character must be a letter. Can't contain two consecutive hyphens. Can't end with a hyphen.

▾ **Credentials Settings**

Master username Info
Type a login ID for the master user of your DB instance.

```
root
```

1 to 16 alphanumeric characters. First character must be a letter.

☐ **Manage master credentials in AWS Secrets Manager**
    Manage master user credentials in Secrets Manager. RDS can generate a password for you and
    manage it throughout its lifecycle.

> ⓘ If you manage the master user credentials in Secrets Manager, some RDS features aren't supported.
> Learn more 🔗

☐ **Auto generate a password**
    Amazon RDS can generate a password for you, or you can specify your own password.

Master password Info

```
••••••••
```

Constraints: At least 8 printable ASCII characters. Can't contain any of the following: / (slash), '(single quote), "(double quote) and @
(at sign).

Confirm master password Info

```
••••••••
```

Instance configuration

The DB instance configuration options below are limited to those supported by the engine that you selected above.

> ⓘ **Amazon RDS Optimized Writes** - *new* Info
>   ◯ Show instance classes that support Amazon RDS Optimized Writes

DB instance class Info

⦿ Standard classes (includes m classes)

⦿ Memory optimized classes (includes r and x classes)

⦿ Burstable classes (includes t classes)

```
db.t2.micro                                            ▾
1 vCPUs   1 GiB RAM   Not EBS Optimized
```

◯ Include previous generation classes

MySQL   ✕

MySQL is the most popular open
source database in the world.
MySQL on RDS offers the rich
features of the MySQL community
edition with the flexibility to easily
scale compute resources or storage
capacity for your database.

• Supports database size up to 64
  TiB.

• Supports General Purpose,
  Memory Optimized, and
  Burstable Performance instance
  classes.

• Supports automated backup
  and point-in-time recovery.

• Supports up to 15 Read Replicas
  per instance, within a single
  Region or 5 read replicas cross-
  region.

Instances | EC2 Management Con ☓   Create database - RDS Managem ☓   G Managed Services on AWS - Grea ☓   +

← → C 🔒 us-east-1.console.aws.amazon.com/rds/home?region=us-east-1#launch-dbinstance:gdb=false;s3-import=false

📁 Amazon   📁 Gym Equipment   📁 Trevors Computer   📄 Amazon Preparatio...   📄 System Design Inte...   Synology DiskStatio...   📅 Google Calendar   demaynas01   »   📁 Other bookmarks

Storage

Storage type Info

General Purpose SSD (gp2)
Baseline performance determined by volume size ▾

Allocated storage Info

20   GiB

The minimum value is 20 GiB and the maximum value is 6,144 GiB

Storage autoscaling Info
Provides dynamic scaling support for your database's storage based on your application's needs.

☑ Enable storage autoscaling

Enabling this feature will allow the storage to increase after the specified threshold is exceeded.

Maximum storage threshold Info

Charges will apply when your database autoscales to the specified threshold

1000   GiB

The minimum value is 22 GiB and the maximum value is 6,144 GiB

Connectivity Info   ⟳

Compute resource

Choose whether to set up a connection to a compute resource for this database. Setting up a connection will automatically change connectivity settings so that the compute resource can connect to this database.

🔘 **Don't connect to an EC2 compute resource**
Don't set up a connection to a compute resource for this database. You can manually set up a connection to a compute resource later.

⚪ **Connect to an EC2 compute resource**
Set up a connection to an EC2 compute resource for this database.

Network type Info

To use dual-stack mode, make sure that you associate an IPv6 CIDR block with a subnet in the VPC you specify.

🔘 **IPv4**
Your resources can communicate only over the IPv4 addressing protocol.

⚪ **Dual-stack mode**
Your resources can communicate over IPv4, IPv6, or both.

Virtual private cloud (VPC) Info

Choose the VPC. The VPC defines the virtual networking environment for this DB instance.

Default VPC (vpc-2a735750)
6 Subnets, 6 Availability Zones ▾

Only VPCs with a corresponding DB subnet group are listed.

ⓘ After a database is created, you can't change its VPC.

DB subnet group Info

Choose the DB subnet group. The DB subnet group defines which subnets and IP ranges the DB instance can use in the VPC that you selected.

default-vpc-2a735750
6 Subnets, 6 Availability Zones ▾

Public access Info

🔘 **Yes**
RDS assigns a public IP address to the database. Amazon EC2 instances and other resources outside of the VPC can connect to your database. Resources inside the VPC can also connect to the database. Choose one or more VPC security groups that specify which resources can connect to the database.

⚪ **No**
RDS doesn't assign a public IP address to the database. Only Amazon EC2 instances and other resources inside the VPC can connect to your database. Choose one or more VPC security groups that specify which resources can connect to the database.

VPC security group (firewall) Info

Choose one or more VPC security groups to allow access to your database. Make sure that the security group rules allow the appropriate incoming traffic.

🔘 **Choose existing**
Choose existing VPC security groups

⚪ **Create new**
Create new VPC security group

Existing VPC security groups

Choose one or more options ▾

caution-all-from-all ☓   default ☓

EC2 Management Console | Create database - RDS Managem | Managed Services on AWS - Grea | +

us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#SecurityGroup:groupId=sg-0ac3ef597cdb9b3a6

Amazon | Gym Equipment | Trevors Computer | Amazon Preparatio... | System Design Inte... | Synology DiskStatio... | Google Calendar | demaynas01 | » | Other bookmarks

aws | Services | Q Search [Alt+S] | N. Virginia ▾ | devops @ tdemay ▾

VPC | EC2 | S3 | Route 53 | WAF & Shield | CloudShell | Elastic Container Service | Elastic Container Registry | CloudFormation

EC2 > Security Groups > sg-0ac3ef597cdb9b3a6 - caution-all-from-all

sg-0ac3ef597cdb9b3a6 - caution-all-from-all

Actions ▾

Details

Security group name	Security group ID	Description	VPC ID
caution-all-from-all	sg-0ac3ef597cdb9b3a6	For troubleshooting	vpc-2a735750

Owner	Inbound rules count	Outbound rules count	
032822161467	4 Permission entries	1 Permission entry	

Inbound rules | Outbound rules | Tags

ⓘ You can now check network connectivity with Reachability Analyzer | Run Reachability Analyzer | ✕

Inbound rules (4)

Filter security group rules

Manage tags | Edit inbound rules

< 1 >

	Name	▽	Security group rule ID	▽	Port range	▽	Source	▽	Description	▽
☐	–		sgr-0379c24dd50234c9f		0 - 65535		0.0.0.0/0		–	
☐	–		sgr-09900a9a9a314a92b		0 - 65535		::/0		–	
☐	–		sgr-06ba42c04768fc97d		0 - 65535		::/0		–	
☐	–		sgr-0681da9e3e71ad8b9		0 - 65535		0.0.0.0/0		–	

☰

MySQL ✕

Availability Zone Info

No preference ▾

MySQL is the most popular open source database in the world. MySQL on RDS offers the rich features of the MySQL community edition with the flexibility to easily scale compute resources or storage capacity for your database.

RDS Proxy

RDS Proxy is a fully managed, highly available database proxy that improves application scalability, resiliency, and security.

☐ Create an RDS Proxy Info

RDS automatically creates an IAM role and a Secrets Manager secret for the proxy. RDS Proxy has additional costs. For more information, see Amazon RDS Proxy pricing 🔗.

- Supports database size up to 64 TiB.
- Supports General Purpose, Memory Optimized, and Burstable Performance instance classes.

Certificate authority - *optional* Info

Using a server certificate provides an extra layer of security by validating that the connection is being made to an Amazon database. It does so by checking the server certificate that is automatically installed on all databases that you provision.

- Supports automated backup and point-in-time recovery.
- Supports up to 15 Read Replicas per instance, within a single Region or 5 read replicas cross-region.

rds-ca-2019 (default) ▾

If you don't select a certificate authority, RDS chooses one for you.

▶ Additional configuration

Database authentication

Database authentication options Info

🔘 Password authentication
Authenticates using database passwords.

⚪ Password and IAM database authentication
Authenticates using the database password and user credentials through AWS IAM users and roles.

⚪ Password and Kerberos authentication
Choose a directory in which you want to allow authorized users to authenticate with this DB instance using Kerberos Authentication.

Monitoring

Monitoring

☐ Enable Enhanced monitoring
Enabling Enhanced monitoring metrics are useful when you want to see how different processes or threads use the CPU.

▼ Additional configuration

Database options, backup turned off, backtrack turned off, maintenance, CloudWatch Logs, delete protection turned off.

Database options

Initial database name Info

[   ]

If you do not specify a database name, Amazon RDS does not create a database.

DB parameter group Info

default.mysql8.0 ▾

Option group Info

default:mysql-8-0 ▾

Backup

☐ Enable automated backups
Creates a point-in-time snapshot of your database.

Log exports

Select the log types to publish to Amazon CloudWatch Logs

☐ Audit log
☐ Error log
☐ General log
☐ Slow query log

IAM role

The following service-linked role is used for publishing logs to CloudWatch Logs.

← → ⟳ 🔒 us-east-1.console.aws.amazon.com/rds/home?region=us-east-1#launch-dbinstance:gdb=false;s3-import=false

📁 Amazon 📁 Gym Equipment 📁 Trevors Computer ⬜ Amazon Preparatio... ⬜ System Design Inte... Synology DiskStatio... Google Calendar demaynas01 » 📁 Other bookmarks

aws ⊞ Services 🔍 Search [Alt+S] ⟩_ 🔔 ? N. Virginia ▾ devops @ tdemay ▾

🔷 VPC 🔶 EC2 S3 🔶 Route 53 WAF & Shield CloudShell Elastic Container Service Elastic Container Registry CloudFormation

☰

▼ **Additional configuration**
Database options, backup turned off, backtrack turned off, maintenance, CloudWatch Logs, delete protection turned off.

Database options

Initial database name   Info

[   ]

If you do not specify a database name, Amazon RDS does not create a database.

DB parameter group   Info

[default.mysql8.0   ▼]

Option group   Info

[default:mysql-8-0   ▼]

Backup

☐ Enable automated backups
   Creates a point-in-time snapshot of your database

Log exports

Select the log types to publish to Amazon CloudWatch Logs
☐ Audit log
☐ Error log
☐ General log
☐ Slow query log

IAM role
The following service-linked role is used for publishing logs to CloudWatch Logs.

[RDS service-linked role   ]

Maintenance

Auto minor version upgrade   Info
☐ Enable auto minor version upgrade
   Enabling auto minor version upgrade will automatically upgrade to new minor versions as
   they are released. The automatic upgrades occur during the maintenance window for the
   database.

Maintenance window   Info
Select the period you want pending modifications or maintenance applied to the database by Amazon RDS.
○ Choose a window
⦿ No preference

Deletion protection

☐ Enable deletion protection
   Protects the database from being deleted accidentally. While this option is enabled, you can't delete the database.

Estimated monthly costs

The Amazon RDS Free Tier is available to you for 12 months. Each calendar month, the free tier will allow you to
use the Amazon RDS resources listed below for free:

- 750 hrs of Amazon RDS in a Single-AZ db.t2.micro, db.t3.micro or db.t4g.micro Instance.
- 20 GB of General Purpose Storage (SSD).
- 20 GB for automated backup storage and any user-initiated DB Snapshots.

Learn more about AWS Free Tier. 🔗

When your free usage expires or if your application use exceeds the free usage tiers, you simply pay standard,
pay-as-you-go service rates as described in the Amazon RDS Pricing page. 🔗

ⓘ You are responsible for ensuring that you have all of the necessary rights for any third-party products or
   services that you use with AWS services.

Cancel   [**Create database**]

MySQL   ✕

MySQL is the most popular open
source database in the world.
MySQL on RDS offers the rich
features of the MySQL community
edition with the flexibility to easily
scale compute resources or storage
capacity for your database.

- Supports database size up to 64
  TiB.

- Supports General Purpose,
  Memory Optimized, and
  Burstable Performance instance
  classes.

- Supports automated backup
  and point-in-time recovery.

- Supports up to 15 Read Replicas
  per instance, within a single
  Region or 5 read replicas cross-
  region.

us-east-1.console.aws.amazon.com/rds/home?region=us-east-1#database:id=databrew-rds;is-cluster=false

Amazon   Gym Equipment   Trevors Computer   Amazon Preparatio...   System Design Inte...   Synology DiskStatio...   Google Calendar   demaynas01   »   Other bookmarks

aws   Services   Q Search   [Alt+S]   N. Virginia ▾   devops @ tdemay ▾

VPC   EC2   S3   Route 53   WAF & Shield   CloudShell   Elastic Container Service   Elastic Container Registry   CloudFormation

RDS > Databases > databrew-rds

databrew-rds

Modify   Actions ▾

Summary

DB identifier	CPU	Status	Class
databrew-rds	▭ 3.61%	⊘ Available	db.t2.micro
Role	Current activity	Engine	Region & AZ
Instance	▭ 0 Connections	MySQL Community	us-east-1b

Connectivity & security   Monitoring   Logs & events   Configuration   Maintenance & backups   Tags

Connectivity & security

Endpoint & port

Endpoint
databrew-rds.c7hkuw409aei.us-east-1.rds.amazonaws.com

Port
3306

Networking

Availability Zone
us-east-1b

VPC
default-vpc (vpc-2a735750)

Subnet group
default-vpc-2a735750

Subnets
subnet-a8eef986
subnet-63a2fe5d
subnet-2e6a6249
subnet-740dd539
subnet-13f4e14f
subnet-4aec2e44

Network type
IPv4

Security

VPC security groups
default (sg-e7437fb3)
⊘ Active
caution-all-from-all (sg-0ac3ef597cdb9b3a6)
⊘ Active

Publicly accessible
Yes

Certificate authority   Info
rds-ca-2019

Certificate authority date
August 22, 2024, 13:08 (UTC-04:00)

DB instance certificate expiration date
August 22, 2024, 13:08 (UTC-04:00)

Security group rules (4)

Filter by Security group rules   < 1 > ⚙

Security group ▲	Type ▽	Rule ▽
caution-all-from-all (sg-0ac3ef597cdb9b3a6)	CIDR/IP - Inbound	0.0.0.0/0
caution-all-from-all (sg-0ac3ef597cdb9b3a6)	CIDR/IP - Outbound	0.0.0.0/0
default (sg-e7437fb3)	EC2 Security Group - Inbound	sg-e7437fb3
default (sg-e7437fb3)	CIDR/IP - Outbound	0.0.0.0/0

Replication (1)

Filter by Replication   < 1 > ⚙

DB identifier ▲	Role ▽	Region & AZ ▽	Replication source ▽	Replication state ▽	Lag ▽
databrew-rds	Instance	us-east-1b	-	-	-

Proxies (0)

Actions ▾   Create proxy

Filter by proxies   < 1 > ⚙

EC2 Management Console × | Database Details - RDS Managem × | G Managed Services on AWS - Grea × | +

us-east-1.console.aws.amazon.com/rds/home?region=us-east-1#database:id=databrew-rds;is-cluster=false;tab=configuration

Amazon | Gym Equipment | Trevors Computer | Amazon Preparatio... | System Design Inte... | Synology DiskStatio... | Google Calendar | demaynas01 | » | Other bookmarks

aws ::: Services | Q Search | [Alt+S] | | N. Virginia ▾ | devops @ tdemay ▾

VPC | EC2 | S3 | Route 53 | WAF & Shield | CloudShell | Elastic Container Service | Elastic Container Registry | CloudFormation

RDS > Databases > databrew-rds

databrew-rds

[Modify] [Actions ▾]

Summary

DB identifier	CPU	Status	Class
databrew-rds	▭ 3.61%	⊘ Available	db.t2.micro
Role	**Current activity**	**Engine**	**Region & AZ**
Instance	▭ 0 Connections	MySQL Community	us-east-1b

Connectivity & security | Monitoring | Logs & events | **Configuration** | Maintenance & backups | Tags

Instance

Configuration

DB instance ID
databrew-rds

Engine version
8.0.32

DB name
-

License model
General Public License

Option groups
default:mysql-8-0 ⊘ In sync

Amazon Resource Name (ARN)
▭ arn:aws:rds:us-east-1:032822161467:db:databrew-rds

Resource ID
db-BXTF3K4HVS7PS6TT5FAROAN2AM

Created time
June 22, 2023, 17:22 (UTC-04:00)

DB instance parameter group
default.mysql8.0 ⊘ In sync

Deletion protection
Disabled

Instance class

Instance class
db.t2.micro

vCPU
1

RAM
1 GB

Availability

Master username
root

Master password
****

IAM DB authentication
Not enabled

Multi-AZ
No

Secondary Zone
-

Storage

Encryption
Not enabled

Storage type
General Purpose SSD (gp2)

Storage
20 GiB

Provisioned IOPS
-

Storage throughput
-

Storage autoscaling
Enabled

Maximum storage threshold
1000 GiB

Performance Insights

Performance Insights enabled
Turned off

Recommendations (0)

[Dismiss] [Schedule] [Apply now]

[Q Filter by recommendations] | ‹ 1 › ⚙

Recommendation type ▲	Recommendation	▽	Recommendation time	▽	Next maintenanc
	No recommendations found				

```
Windows PowerShell                                        ×    +   ∨                                   —   □   ×

PS C:\Users\demay\OneDrive\Desktop\courses\gl-cloud\projects\managed services\DataBrew\data-transformation-of-
rds-using-glue-databrew> mysql -uroot -p -h databrew-rds.c7hkuw409aei.us-east-1.rds.amazonaws.com
Enter password: ********
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 19
Server version: 8.0.32 Source distribution

Copyright (c) 2000, 2023, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> source DB_OBJECT_DDL.sql
Query OK, 1 row affected (0.01 sec)

Query OK, 0 rows affected (0.04 sec)

Query OK, 0 rows affected, 1 warning (0.06 sec)

Query OK, 0 rows affected, 1 warning (0.04 sec)

mysql> source SCHOOL_TBL__INSERT.sql
Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

mysql> source STUDENT_TBL_INSERT.sql
Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)

Query OK, 1 row affected (0.01 sec)
```
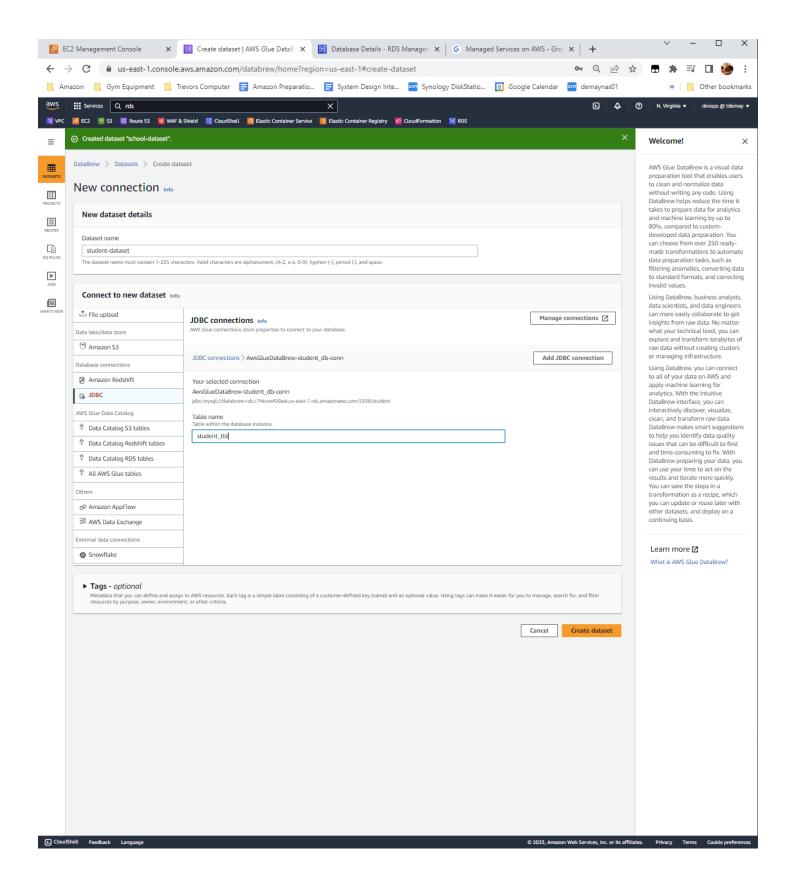
```
mysql> use student ;
Database changed
mysql> alter table study_details_tbl add column marks int DEFAULT NULL;
Query OK, 0 rows affected (0.03 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> source STUDY_DETAIL_TBL_INSERT.sql
```

Create an Amazon RDS connection

Create datasets

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

Amazon   Gym Equipment   Trevors Computer   Amazon Preparatio...   System Design Inte...   Synology DiskStatio...   Google Calendar   demaynas01   Other bookmarks

aws   Services   Q rds   N. Virginia ▾   devops @ tdemay ▾

VPC   EC2   S3   Route 53   WAF & Shield   CloudShell   Elastic Container Service   Elastic Container Registry   CloudFormation   RDS

✓ Created Glue connection "AwsGlueDataBrew-student_db-conn".   ✕

DataBrew > Datasets > Create dataset

New connection Info

New dataset details

Dataset name

school-dataset

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

Connect to new dataset Info

↥ File upload

Data lake/data store

🗄 Amazon S3

Database connections

🔲 Amazon Redshift

🗄 JDBC

AWS Glue Data Catalog

⚲ Data Catalog S3 tables

⚲ Data Catalog Redshift tables

⚲ Data Catalog RDS tables

⚲ All AWS Glue tables

Others

⚭ Amazon AppFlow

⬚ AWS Data Exchange

External data connections

❄ Snowflake

JDBC connections Info
AWS Glue connections store properties to connect to your database.

Manage connections ⧉

JDBC connections > AwsGlueDataBrew-student_db-conn   Add JDBC connection

Your selected connection
AwsGlueDataBrew-student_db-conn
jdbc:mysql://databrew-rds.c7hkuw409aei.us-east-1.rds.amazonaws.com:3306/student

Table name
Table within the database instance.

school_tbl

▶ **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.

Cancel   Create dataset

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

Amazon   Gym Equipment   Trevors Computer   Amazon Preparatio...   System Design Inte...   Synology DiskStatio...   Google Calendar   demaynas01   »   Other bookmarks

aws   ::: Services   Q rds   ×

VPC   EC2   S3   Route 53   WAF & Shield   CloudShell   Elastic Container Service   Elastic Container Registry   CloudFormation   RDS

⊘ Created dataset "school-dataset".   ×

Welcome!   ×

DataBrew > Datasets > Create dataset

New connection Info

AWS Glue DataBrew is a visual data preparation tool that enables users to clean and normalize data without writing any code. Using DataBrew helps reduce the time it takes to prepare data for analytics and machine learning by up to 80%, compared to custom-developed data preparation. You can choose from over 250 ready-made transformations to automate data preparation tasks, such as filtering anomalies, converting data to standard formats, and correcting invalid values.

New dataset details

Dataset name

| student-dataset |

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

Connect to new dataset Info

File upload
Data lake/data store
Amazon S3
Database connections
Amazon Redshift
JDBC
AWS Glue Data Catalog
Data Catalog S3 tables
Data Catalog Redshift tables
Data Catalog RDS tables
All AWS Glue tables
Others
Amazon AppFlow
AWS Data Exchange
External data connections
Snowflake

JDBC connections Info   Manage connections ↗

AWS Glue connections store properties to connect to your database.

JDBC connections > AwsGlueDataBrew-student_db-conn   Add JDBC connection

Your selected connection

AwsGlueDataBrew-student_db-conn

jdbc:mysql://databrew-rds.c7hkuw409aei.us-east-1.rds.amazonaws.com:3306/student

Table name

Table within the database instance.

| student_tbl| |

Using DataBrew, business analysts, data scientists, and data engineers can more easily collaborate to get insights from raw data. No matter what your technical level, you can explore and transform terabytes of raw data without creating clusters or managing infrastructure.

Using DataBrew, you can connect to all of your data on AWS and apply machine learning for analytics. With the intuitive DataBrew interface, you can interactively discover, visualize, clean, and transform raw data. DataBrew makes smart suggestions to help you identify data quality issues that can be difficult to find and time-consuming to fix. With DataBrew preparing your data, you can use your time to act on the results and iterate more quickly. You can save the steps in a transformation as a recipe, which you can update or reuse later with other datasets, and deploy on a continuing basis.

Learn more ↗

What is AWS Glue DataBrew?

▶ **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.

Cancel   Create dataset

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

aws   Services   Q rds   ×   N. Virginia ▾   devops @ tdemay ▾

VPC   EC2   S3   Route 53   WAF & Shield   CloudShell   Elastic Container Service   Elastic Container Registry   CloudFormation   RDS

⊘ Created dataset "student-dataset".   ×

Welcome!   ×

DataBrew > Datasets > Create dataset

New connection Info

New dataset details

Dataset name

study-detail-dataset

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

Connect to new dataset Info

Connection options	
⬆ File upload	
Data lake/data store	
⬡ Amazon S3	
Database connections	
▤ Amazon Redshift	
▤ JDBC	
AWS Glue Data Catalog	
⬡ Data Catalog S3 tables	
⬡ Data Catalog Redshift tables	
⬡ Data Catalog RDS tables	
⬡ All AWS Glue tables	
Others	
⬡ Amazon AppFlow	
⬡ AWS Data Exchange	
External data connections	
❄ Snowflake	

JDBC connections Info   Manage connections ⬈

AWS Glue connections store properties to connect to your database.

JDBC connections > AwsGlueDataBrew-student_db-conn   Add JDBC connection

Your selected connection

AwsGlueDataBrew-student_db-conn

jdbc:mysql://databrew-rds.c7hkuw409aei.us-east-1.rds.amazonaws.com:3306/student

Table name

Table within the database instance.

study_details_tbl

▸ **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.

Cancel   Create dataset

Create a project using the datasets

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

Amazon   Gym Equipment   Trevors Computer   Amazon Preparatio...   System Design Inte...   Synology DiskStatio...   Google Calendar   demaynas01   »   Other bookmarks

aws   ⠿ Services   🔍 rds   ✕   N. Virginia ▾   devops @ tdemay ▾

VPC   EC2   S3   Route 53   WAF & Shield   CloudShell   Elastic Container Service   Elastic Container Registry   CloudFormation   RDS

✓ Created dataset "study-detail-dataset".   ✕   ⓘ

DataBrew > Projects > Create project

Create project Info

Project details

Project name

my-rds-proj

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

Recipe details Info

Data cleaning steps in DataBrew are stored as a recipe. A recipe is connected to a project by default. An existing recipe with no associated project could also be applied to a project.

Attached recipe   Recipe name

Create new recipe ▼   my-rds-proj-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 ▼
◉	study-detail-dataset	Database table	Database connection	a minute ago June 22, 2023, 5:36:53 pm
○	student-dataset	Database table	Database connection	2 minutes ago June 22, 2023, 5:36:19 pm
○	school-dataset	Database table	Database connection	2 minutes ago June 22, 2023, 5:35:36 pm

▶ **Sampling** - *optional*
Select the type and size of your sample

▶ **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.

AWSGlueDataBrewServiceRole-student   ▼   ⟳

By clicking "Create project" you are authorizing DataBrew to add required permissions to access all the datasets in this project to the selected service 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

EC2 Management Console   × | Project workspace | AWS Glue Da   × | Database Details - RDS Managem   × | G Managed Services on AWS - Grea   ×   +

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

Amazon   Gym Equipment   Trevors Computer   Amazon Preparatio...   System Design Inte...   Synology DiskStatio...   Google Calendar   demaynas01   »   Other bookmarks

aws   Services   Q rds

VPC   EC2   S3   Route 53   WAF & Shield   CloudShell   Elastic Container Service   Elastic Container Registry   CloudFormation   RDS

Created project "my-rds-proj".

Project failed to load   ×

VPC S3 gateway endpoint validation failed for SubnetId: subnet-740dd539. VPC: vpc-2a735750. Reason: Could not find S3 gateway endpoint or NAT gateway for subnetId: subnet-740dd539 in Vpc vpc-2a735750 (Service: AWSGlueDataBrewJobExecutor; Status Code: 400; Error Code: InvalidInputException; Request ID: bc913b82-d52b-42dc-b70c-c17e9afd8106; Proxy: null)

Edit dataset   Ok