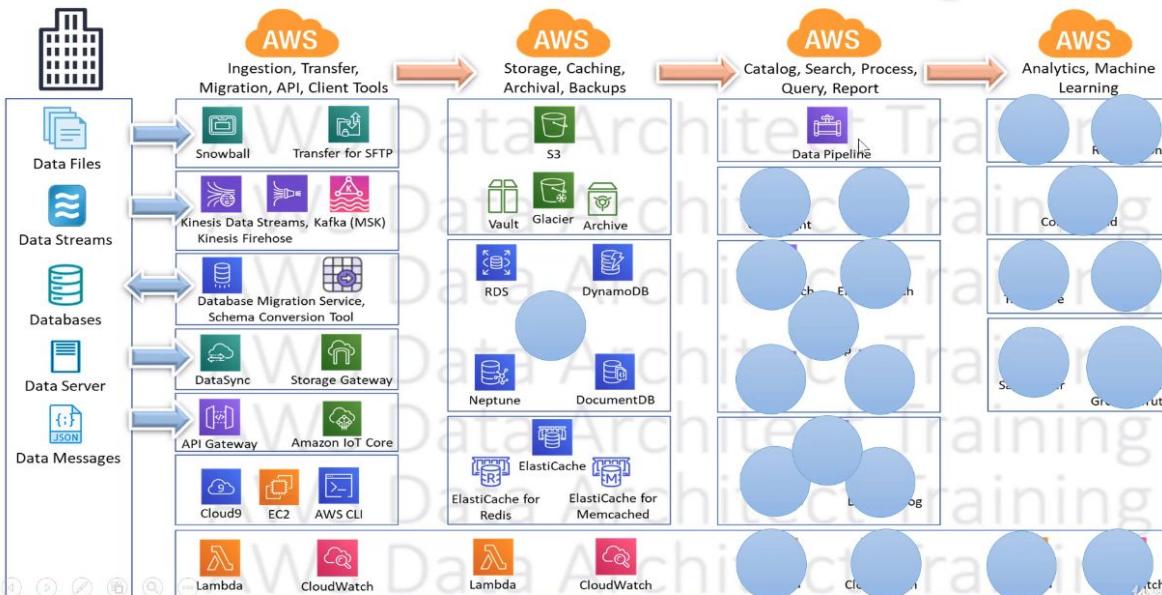


# sec\_21\_Data\_Pipeline

# AWS Data Architect Training



AWS Data Architect B x (115) Learn JavaScript - F x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content)

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing

AWS | Amazon Data Pipeline - X

<https://aws.amazon.com/datapipeline/>

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

Contact Sales Support English My Account

Sign in to the Console

# AWS Data Pipeline

Easily automate the movement and transformation of data.

[Get started with AWS Data Pipeline](#)[Product Details](#)[Pricing](#)[Developer Resources](#)[FAQs](#)[Documentation](#)[Big Data Blog](#)

AWS Data Pipeline is a web service that helps you reliably process and move data between different AWS compute and storage services, as well as on-premises data sources, at specified intervals. With AWS Data Pipeline, you can regularly access your data where it's stored, transform and process it at scale, and efficiently transfer the results to AWS services such as Amazon S3, Amazon RDS, Amazon DynamoDB, and Amazon EMR.

AWS Data Pipeline helps you easily create complex data processing workloads that are fault tolerant, repeatable, and highly available. You don't have to worry about ensuring resource availability, managing inter-task dependencies, retrying transient failures or timeouts in individual tasks, or creating a failure notification system. AWS Data Pipeline also allows you to move and process data that was previously locked up in on-premises data silos.



1.5x 2:14 / 10:44

Speaker settings

[Course content](#) [Overview](#) [Q&A](#) [Notes](#) [Announcements](#)

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing

https://aws.amazon.com/datapipeline/

**aws**

Contact Sales Support English My Account Sign In to the Console

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More



**Scalable**

AWS Data Pipeline makes it equally easy to dispatch work to one machine or many, in serial or parallel. With AWS Data Pipeline's flexible design, processing a million files is as easy as processing a single file.



**Low Cost**

AWS Data Pipeline is inexpensive to use and is billed at a low monthly rate. You can try it for free under the AWS Free Usage. [Learn more.](#)



**Transparent**

You have full control over the computational resources that execute your business logic, making it easy to enhance or debug your logic. Additionally, full execution logs are automatically delivered to Amazon S3, giving you a persistent, detailed record of what has happened in your pipeline.

Start Using AWS Data Pipeline

Get Started

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Leave a rating Your progress Share

177 AWS Data Pipeline - Architecture Diagram, Use Case, Pricing

What is AWS Data Pipeline? - AWS Data Pipeline Document History - AWS Data Pipeline

https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/DocHistory.html

AWS Data Pipeline Developer Guide (API Version 2012-10-29)

Documentation - This Guide Search

What is AWS Data Pipeline? AWS Data Pipeline Concepts Setting Up Getting Started with AWS Data Pipeline Working with Pipelines Controlling Access to Pipelines and Resources Tutorials Pipeline Expressions and Functions Pipeline Object Reference Working with Task Runner Troubleshooting Logging AWS Data Pipeline API Calls By Using AWS CloudTrail Limits AWS Data Pipeline Resources Document History

Document History

This documentation is associated with the 2012-10-29 version of AWS Data Pipeline.

Latest documentation update: 9 November 2018.

Change	Description	Release Date
Updated the lists of supported Amazon EC2 and Amazon EMR instances.	Updated the lists of supported Amazon EC2 and Amazon EMR instances. For more information, see <a href="#">Supported Instance Types for Pipeline Work Activities</a> .	9 November 2018
Updated the list of IDs of the HVM (Hardware Virtual Machine) AMIs used for the instances.	Updated the list of IDs of the HVM (Hardware Virtual Machine) AMIs used for the instances. For more information, see <a href="#">Syntax</a> and search for <code>ImageId</code> .	
Added configuration for attaching Amazon EBS volumes to cluster nodes, and for launching an Amazon EMR cluster into a private subnet.	Added configuration options to an <code>EMRcluster</code> object. You can use these options in pipelines that use Amazon EMR clusters.  Use the <code>coreEbsConfiguration</code> , <code>masterEbsConfiguration</code> , and <code>TaskEbsConfiguration</code> fields to configure the attachment of Amazon EBS volumes to core, master, and task nodes in the Amazon EMR cluster. For more information, see <a href="#">Attach EBS volumes to cluster nodes</a> .  Use the <code>emrManagedMasterSecurityGroupId</code> , <code>emrManagedSlaveSecurityGroupId</code> , and <code>ServiceAccessSecurityGroupId</code> fields to configure an Amazon EMR cluster in a private subnet. For more information, see <a href="#">Configure an Amazon EMR cluster in a private subnet</a> .  For more information about <code>EMRcluster</code> syntax, see <a href="#">EMRcluster</a> .	19 April 2018
Added the list of supported Amazon EC2 and Amazon EMR instances.	Added the list of instances that AWS Data Pipeline creates by default, if you do not specify an instance type in the pipeline definition. Added a list of supported Amazon EC2 and Amazon EMR instances. For more information, see <a href="#">Supported Instance Types for Pipeline Work Activities</a> .	22 March 2018

Have a question? Try the forums Did this page help you? Yes No Feedback

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Leave a rating Your progress Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing

What is AWS Data Pipeline? - AWS Data Pipeline

https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html

English Sign in to the Console

## AWS Data Pipeline

Developer Guide (API Version 2012-10-29)

Documentation - This Guide Search

What is AWS Data Pipeline? Related Services

Supported Instance Types for Pipeline Work Activities

AWS Data Pipeline Concepts

Setting Up

Getting Started with AWS Data Pipeline

Working with Pipelines

Controlling Access to Pipelines and Resources

Tutorials

Pipeline Expressions and Functions

Pipeline Object Reference

Working with Task Runner

Troubleshooting

AWS Documentation » AWS Data Pipeline » Developer Guide » What is AWS Data Pipeline?

## What is AWS Data Pipeline?

AWS Data Pipeline is a web service that you can use to automate the movement and transformation of data. With AWS Data Pipeline, you can define data-driven workflows, so that tasks can be dependent on the successful completion of previous tasks. You define the parameters of your data transformations and AWS Data Pipeline enforces the logic that you've set up.

The following components of AWS Data Pipeline work together to manage your data:

- A *pipeline definition* specifies the business logic of your data management. For more information, see [Pipeline Definition File Syntax](#).
- A *pipeline* schedules and runs tasks by creating Amazon EC2 instances to perform the defined work activities. You upload your pipeline definition to the pipeline, and then activate the pipeline. You can edit the pipeline definition for a running pipeline and activate the pipeline again for it to take effect. You can deactivate the pipeline, modify a data source, and then activate the pipeline again. When you are finished with your pipeline, you can delete it.
- Task Runner* polls for tasks and then performs those tasks. For example, Task Runner could copy log files to Amazon S3 and launch Amazon EMR clusters. Task Runner is installed and runs automatically on resources created by your pipeline definitions. You can write a custom task runner application, or you can use the Task Runner application that is provided by AWS Data Pipeline. For more information, see [Task Runners](#).

For example, you can use AWS Data Pipeline to archive your web server's logs to Amazon Simple Storage Service (Amazon S3) each day and then run a weekly Amazon EMR (Amazon EMR) cluster over those logs to generate traffic reports. AWS Data Pipeline schedules the daily tasks to copy data and the weekly task

Did this page help you? Yes No Feedback

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - (115) Learn JavaScript - F... | udemy.com

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Leave a rating Your progress Share

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing  
Automating Analytic Workflows

https://aws.amazon.com/blogs/big-data/automating-analytic-workflows-on-aws/

Contact Sales Support My Account Sign In to the Console

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

Blog Home Category Edition Follow Search Blogs

### AWS Data Pipeline reporting ETL workflow

Data Pipeline lets you define a dependent chain of data sources and destinations with an option to create data processing activities in a pipeline. You can schedule the tasks within the pipeline to perform various activities of data movement and processing. In addition to scheduling, you can also include failure and retry options in the data pipeline workflows.

The following diagram is an example of a Data Pipeline reporting ETL pipeline that moves data from the replicated slave database on-premises to Amazon Redshift:

```
graph LR; subgraph OnPremises [On Premises]; D[Database] --> S[Server]; end; subgraph AWSCloud [AWS Cloud]; subgraph ExtractedFiles [Extracted Files Amazon S3]; E[Extracted Files] --> A[Amazon MR ETL Cluster]; end; subgraph TransformedData [Transformed Data Amazon S3]; T[Transformed Data] --> R[Amazon Redshift]; end; subgraph BIAnalytics [BI Analytics]; BI[BI Analytics] --> AHQ[Ad Hoc Queries]; end; E --> A; A --> T; T --> BI;
```

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Leave a rating Your progress Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing  
Automating Analytic Workflows

https://aws.amazon.com/blogs/big-data/automating-analytic-workflows-on-aws/

Contact Sales Support My Account Sign In to the Console

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More Q

Blog Home Category Edition Follow Search Blogs

The diagram illustrates a data pipeline architecture. It starts with a 'Database' and 'Server' on-premises, which extract data into 'Extracted Files Amazon S3'. These files are then processed by an 'Amazon EMR ETL Cluster', which outputs to 'Transformed Data Amazon S3'. Finally, this transformed data is loaded into 'Amazon Redshift'. From Redshift, the data can be queried via 'Ad Hoc Queries' or used for 'BI Analytics'.

The above pipeline performs the following activities:

- ShellCommandActivity – Incrementally extracts data from an on-premises data store to Amazon S3 using a custom script hosted on an on-premises server with Task Runner installed.
- EmrActivity – Launches a transient cluster that uses the extracted dataset as input, validates, and transforms it, and then outputs to an Amazon S3 bucket as illustrated in the blog post on ETL Processing Using AWS Data Pipeline and Amazon Elastic MapReduce.
- CopyActivity – Performs an Amazon Redshift COPY command on the transformed data and loads it into an Amazon Redshift table for analytics and reporting.

Data Pipeline checks the data for readiness. It also allows you to schedule and orchestrate the data movement while providing you with failure handling, SNS notifications in case of error, and built-in retry logic. You can also specify preconditions as decision logic. For example, a precondition can check whether data is present in an S3 bucket before a pipeline copy activity attempts to load it into Amazon Redshift.

Data Pipeline is useful for creating and managing periodic, batch-processing, data-driven workflows. It optimizes the data processing experience, especially as it relates to data on AWS. For on-demand or real-time processing, Data Pipeline is not an ideal choice.

Course content Overview Q&A Notes Announcements

Section 1: Introduction  
3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Leave a rating Your progress Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing  
ETL Processing Using AWS Data Pipeline

https://aws.amazon.com/blogs/big-data/etl-processing-using-aws-data-pipeline-and-amazon-elastic-mapreduce/

Contact Sales Support My Account Sign In to the Console

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

Blog Home Category Edition Follow

Search Blogs

EMR committer

## AWS Big Data Blog

### ETL Processing Using AWS Data Pipeline and Amazon Elastic MapReduce

by Manjeet Chayel | on 26 NOV 2014 | in Amazon EMR, AWS Big Data, AWS Data Pipeline | Permalink | Comments | Share

Manjeet Chayel is an AWS Solutions Architect

This blog post shows you how to build an ETL workflow that uses AWS Data Pipeline to schedule an Amazon Elastic MapReduce (Amazon EMR) cluster to clean and process web server logs stored in an Amazon Simple Storage Service (Amazon S3) bucket. AWS Data Pipeline is an ETL service that you can use to automate the movement and transformation of data. It launches an Amazon EMR cluster for each scheduled interval, submits jobs as steps to the cluster, and terminates the cluster after tasks have completed.

In this post, you'll create the following ETL workflow:



To create the workflow, we'll use the Pig and Hive examples discussed in the blog post "Ensuring Consistency when Using Amazon S3 and Amazon EMR." This ETL workflow pushes webserver logs to an Amazon S3 bucket, cleans and filters the

Resources

- Amazon Athena
- Amazon EMR
- AWS Glue
- Amazon DynamoDB
- Amazon Kinesis
- Amazon QuickSight
- Amazon Redshift

Follow

- Twitter
- Facebook
- LinkedIn
- Twitch

Email Updates

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com

Leave a rating Your progress Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing  
ETL Processing Using AWS Data

https://aws.amazon.com/blogs/big-data/etl-processing-using-aws-data-pipeline-and-amazon-elastic-mapreduce/

Contact Sales Support My Account Sign In to the Console

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

Blog Home Category Edition Follow Search Blogs

start date in the past.

You can create your workflow using the AWS Management console or use the AWS command line interface or API to automate the process of creating and managing pipelines. To learn more about the CLI and SDK see the following links:

- [AWS Command Line Interface \(CLI\)](#) — Provides commands for a broad set of AWS products, and is supported on Windows, Mac, and Linux/Unix.
- [AWS Software Development Kits \(SDK\)](#) — Lets you to build applications that create and manage pipelines using language-specific APIs.
- [Web Service API](#) — AWS provides a low-level interface that you can use to call the web service directly using JSON.

We will use the AWS CLI to create and activate our ETL pipeline. You can install and setup CLI using [these guidelines](#).

### Six-step Workflow

Step 1: Check if log files are available in the Amazon S3 bucket.

Step 2: Create an Amazon EMR cluster with EMRFS on it.

Step 3: Run emrfs sync to update metadata with contents of the Amazon S3 bucket.

Step 4: Submit a Pig job on Amazon EMR cluster as step.

Step 5: Reuse the same Amazon EMR cluster to launch a Hive job for generating reports.

Step 6: Clean EMRFS metadata that is older than two days (emrfs delete).

Let's look at how each step in our workflow can be represented as an AWS Data Pipeline object and how it translates to ISOM

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing Analyze data in Amazon DynamoDB

https://aws.amazon.com/blogs/big-data/analyze-data-in-amazon-dynamodb-using-amazon-sagemaker-for-real-time-prediction/ Contact Sales Support My Account Sign In to the Console

aws Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

Blog Home Category Edition Follow Search Blogs

## Solution architecture

The following diagram shows the overall architecture of the solution.

The diagram illustrates the architecture of the solution:

- DynamoDB** (represented by a blue sphere) is the primary data source.
- Data is **Exported** (Step 1) to an **S3 Bucket** (represented by a red cylinder).
- In the **S3 Bucket**, data is processed through **Conversion** (Step 2) and **Projection** (Step 2) to produce **JSON** and **CSV** files.
- The **JSON** and **CSV** files are stored in separate **S3 Buckets** (represented by orange cubes).
- The **S3 Buckets** serve as a **Data Source** for **SageMaker** (represented by a purple brain icon).
- SageMaker** performs **Renewal** (Step 3) of the **Model**.
- The **Model** provides **Real-time Prediction** (Step 4) to **Users** (represented by three people icons).
- Users can also perform **Ad Hoc Query** (Step 4) on the **Athena** service (represented by an orange hand icon).
- Users can **Update** or **Retrieve** data from **DynamoDB**.

The entire process is **Automated by Data Pipeline** (Step 5), indicated by a dashed line connecting the components.

The steps that data follows through the architecture are as follows:

Course content Overview Q&A Notes Announcements

Section 1: Introduction 3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing Analyze data in Amazon DynamoDB

<https://aws.amazon.com/blogs/big-data/analyze-data-in-amazon-dynamodb-using-amazon-sagemaker-for-real-time-prediction/>

Contact Sales Support My Account Sign In to the Console

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

Blog Home Category Edition Follow Search Blogs

The diagram illustrates the architecture for analyzing data in Amazon DynamoDB. It shows the following flow:

1. Data Pipeline regularly copies the full contents of a DynamoDB table as JSON into an S3 Bucket.
2. Exported JSON files are converted to comma-separated value (CSV) format to use as a data source for Amazon SageMaker.
3. Amazon SageMaker renews the model artifact and update the endpoint.
4. The converted CSV is available for ad hoc queries with Amazon Athena.
5. Data Pipeline controls this flow and repeats the cycle based on the schedule defined by customer requirements.

The architecture components include:

- DynamoDB
- S3 Buckets (JSON and CSV)
- SageMaker
- Athena
- Model
- Real-time Prediction
- Ad Hoc Query
- Update Retrieve

The diagram also shows user interaction at the top, with icons for leaving a rating, viewing progress, sharing, and signing in to the console.

The video player at the bottom shows the progress is at 8:57 / 10:44. The navigation bar below the video includes Course content, Overview, Q&A, Notes, and Announcements.

Section 1: Introduction  
3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851026#content

Leave a rating Your progress Share

177. AWS Data Pipeline - Architecture Diagram, Use-Case, Pricing

https://aws.amazon.com/datapipeline/pricing/

aws

Contact Sales Support English My Account Sign In to the Console

Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Explore More

Region: US West (Oregon)

	High Frequency	Low Frequency
Activities or preconditions running On AWS	\$1.00 per month	\$0.60 per month
Activities or preconditions running On-Premise	\$2.50 per month	\$1.50 per month
Inactive pipelines: \$1.00 per month		

Except as otherwise noted, our prices are exclusive of applicable taxes and duties, including VAT and applicable sales tax. For customers with a Japanese billing address, use of AWS is subject to Japanese Consumption Tax. [Learn more.](#)

For example, a pipeline that runs a daily job (a Low Frequency activity) on AWS to replicate an Amazon DynamoDB table to Amazon S3 would cost \$0.60 per month. If an Amazon EC2 activity was added to this same pipeline to produce a report based on the data in Amazon S3, the total cost of the pipeline would be \$1.20 per month (two activities X \$0.60 per activity per month). If the pipeline was changed to run every 6 hours, it would cost \$2.00 per month, because it would then consist of two High Frequency activities (at \$1.00 per month for each activity).

Activities or preconditions that are active for part of a month are pro-rated hourly. Your Amazon EC2, Amazon Elastic MapReduce, Amazon S3, Amazon RDS, Amazon DynamoDB, and Amazon SNS activities associated with AWS Data Pipeline are billed separately according to those services' normal prices.

Manage Your Resources Sign In to the Console

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content

Leave a rating Your progress Share

178. Lab: AWS Data Pipeline - Create source and destination data repositories

RDS - AWS Console

https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstances:

AWS Services Resource Groups Oregon Support

Amazon RDS Instances

Instances (1)

DB instance MySQL available 0.29% 0 Connections none db.r4.xlarge

Restore from S3 Create database

Filter instances

DB instance Engine Status CPU Current activity Maintenance Class

mysqlinstance MySQL available 0.29% 0 Connections none db.r4.xlarge

Subnet groups Parameter groups Option groups

Events Event subscriptions

Recommendations

Feedback English (US) 1.5x 0:10 / 4:09

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

This screenshot shows a dual-browser setup. The top browser window is a Chromium browser displaying the AWS Data Architect Bootcamp course on Udemy. The visible content is a lab titled '178. Lab: AWS Data Pipeline - Create source and destination data repositories'. Below the title, it says 'RDS - AWS Console' and shows the URL 'https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstances:'. The main content area is the 'Amazon RDS Instances' section, showing one instance named 'mysqlinstance' which is MySQL, available, and using the 'db.r4.xlarge' class. The bottom part of the Udemy page shows course navigation links like 'Course content', 'Overview', 'Q&A', 'Notes', and 'Announcements'. The bottom of the browser has standard controls for volume, brightness, and search. The bottom of the entire screen shows a media player bar with playback controls (play/pause, volume, etc.) and a timestamp '0:10 / 4:09'. The bottom-most part of the image shows the footer of the Udemy page with copyright information and links to 'Privacy Policy' and 'Terms & Conditions'.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content

Leave a rating Your progress Share

178. Lab: AWS Data Pipeline - Create source and destination data repositories

RDS - AWS Console

https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstances:

Amazon RDS Instances

Instances (1)

Filter instances

Create database

Name	Status	CPU	Current activity	Maintenance	Class	VPC	Multi-AZ	Replication role
QL	available	0.29%	0 Connections	none	db.r4.xlarge	vpc [REDACTED]	No	

Subnet groups Parameter groups Option groups Events Event subscriptions Recommendations

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot displays a dual-pane interface. The left pane is a web browser showing the Amazon RDS Instances page, where a single instance named 'QL' is listed as available. The right pane is a video player showing a lecture titled 'Section 1: Introduction' at 0:55 of 4:09. The browser's address bar shows the URL for the RDS home page. The video player has standard controls for play, volume, and progress.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content

Leave a rating Your progress Share

178. Lab: AWS Data Pipeline - Create source and destination data repositories

RDS - AWS Console

https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstance:id=mysqlinstance

Amazon RDS

Instances

Clusters

Performance Insights

Snapshots

Automated backups

Reserved instances

Subnet groups

Parameter groups

Option groups

Events

Event subscriptions

Recommendations

RDS > Instances > mysqlinstance

mysqlinstance

Summary

Engine: MySQL 5.6.40 DB instance class: db.r4.xlarge DB instance status: available Pending maintenance: none

CloudWatch (15) Legend: mysqlinstance

CPU Utilization (Percent)

DB Connections (Count)

Free Storage Space (MB)

Freeable Memory (MB)

Write IOPS (Count/Second)

Read IOPS (Count/Second)

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

This screenshot shows the AWS RDS MySQL instance details page for 'mysqlinstance'. The instance is running MySQL 5.6.40 and has been assigned a db.r4.xlarge class. It is currently available and has no pending maintenance. The 'CloudWatch' monitoring section displays four charts: 'CPU Utilization (Percent)', 'DB Connections (Count)', 'Free Storage Space (MB)', and 'Freeable Memory (MB)'. The 'CPU Utilization' chart shows a sharp drop from approximately 5% to near 0% between 12:30 and 13:00 on 11/18. The 'DB Connections' chart shows a count of 1 connection. The 'Free Storage Space' chart shows 20,000 MB available. The 'Freeable Memory' chart shows 1,500 MB available. Below the monitoring section, there are links for 'Course content', 'Overview', 'Q&A', 'Notes', and 'Announcements'.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content

Leave a rating Your progress Share

178. Lab: AWS Data Pipeline - Create source and destination data repositories

RDS - AWS Console

https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstance:id=mysqlinstance

Amazon RDS

Services Resource Groups

Oregon Support

Instances

Clusters

Performance Insights

Snapshots

Automated backups

Reserved instances

Subnet groups

Parameter groups

Option groups

Events

Event subscriptions

Recommendations

Connect

Endpoint: mysqlinstance [REDACTED] us-west-2.rds.amazonaws.com Port: 3306 Publicly accessible: Yes

Security group rules (2)

Security group	Type	Rule
rds-launch-wizard (sg-00000000)	CIDR/IP - Inbound	[REDACTED] 32
rds-launch-wizard (sg-00000000)	CIDR/IP - Outbound	0.0.0.0/0

Details

Configurations	Security and network	Instance and IOPS	Maintenance details
ARN: arn:aws:rds:us-west-2:000000000000:mysqlinstance	Availability zone: us-west-2a	Instance Class: db.r4.xlarge	Auto minor version upgrade: Yes
Engine: MySQL 5.6.40	VPC: [REDACTED]	Storage Type: General Purpose (SSD)	Maintenance window: tue:10:53-tue:11:23 UTC (GMT)
License Model: General Public License	Subnet group: def	Storage: 20 GiB	Pending Modifications: None

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bc x (115) Learn JavaScript - F- x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content)

## 178. Lab: AWS Data Pipeline - Create source and destination data repositories

MySQL Workbench

File Edit View Database Tools Scripting Help

# Welcome to MySQL Workbench

MySQL Workbench is the official graphical user interface (GUI) tool for MySQL. It allows you to design, create and browse your database schemas, work with database objects and insert data as well as design and run SQL queries to work with stored data. You can also migrate schemas and data from other database vendors to your MySQL database.

[Browse Documentation >](#)[Read the Blog >](#)[Discuss on the Forums >](#)MySQL Connections [⊕](#) [⊗](#)

MySQL Workbench could not detect any MySQL server running.

This means that MySQL is not installed or is not running.

[Rescan servers](#)

1.5x 1:42 / 4:09

[Course content](#)[Overview](#)[Q&A](#)[Notes](#)[Announcements](#)

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bc x (115) Learn JavaScript - F- x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content)

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share

## 178. Lab: AWS Data Pipeline - Create source and destination data repositories

MySQL Workbench

File Edit View Database Tools Scripting Help

- Connect to Database Ctrl+U
- Manage Connections...
- Reverse Engineer... Ctrl+R
- Schema Transfer Wizard...
- Migration Wizard...
- Edit Type Mappings for Generic Migration...

# Welcome to MySQL Workbench

MySQL Workbench is the official graphical user interface (GUI) tool for MySQL. It allows you to design, create and browse your database schemas, work with database objects and insert data as well as design and run SQL queries to work with stored data. You can also migrate schemas and data from other database vendors to your MySQL database.

[Browse Documentation >](#)[Read the Blog >](#)[Discuss on the Forums >](#)MySQL Connections + ⑧

MySQL Workbench could not detect any MySQL server running.

This means that MySQL is not installed or is not running.

[Rescan servers](#)

1.5x 1:47 / 4:09

[Course content](#) [Overview](#) [Q&A](#) [Notes](#) [Announcements](#)

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

## AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

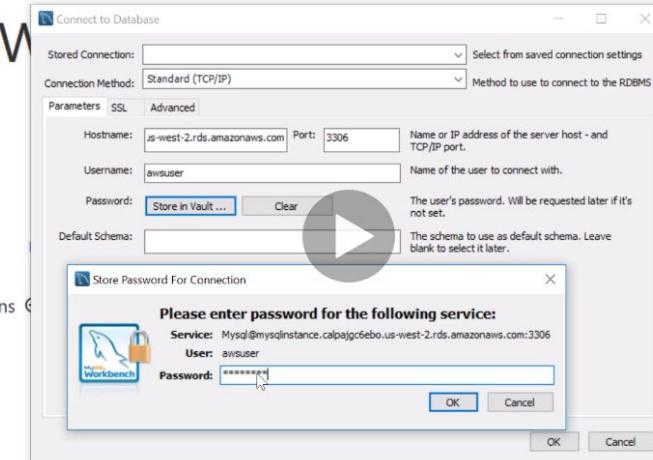
Your progress

Share

## 178. Lab: AWS Data Pipeline - Create source and destination data repositories

MySQL Workbench

File Edit View Database Tools Scripting Help



MySQL Connections

Filter connections

1.5x 1:55 / 4:09

Course content Overview Q&amp;A Notes Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect B... x (115) Learn JavaScript - Fi... x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content)

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share

⋮

## 178. Lab: AWS Data Pipeline - Create source and destination data repositories

MySQL Workbench  
Mysql@mysqlinstance [ ]  
File Edit View Query Database Server Tools Scripting Help

Navigator

MANAGEMENT  
Server Status  
Client Connections  
Users and Privileges  
Status and System Variables  
Data Export  
Data Import/Restore

INSTANCE

Startup / Shutdown  
Server Logs  
Options File

PERFORMANCE

Dashboard  
Performance Reports  
Performance Schema Setup

Administration Schemas

Information

No object selected



▶ 1.5x 2:06 / 4:09 ⏷



Course content Overview Q&amp;A Notes Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - (115) Learn JavaScript - F... | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content

u Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

178. Lab: AWS Data Pipeline - Create source and destination data repositories

MySQL Workbench

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

- innodb
- mytestdb
- sys

Query 1

1.

< >

Administration Schemas Information

No object selected

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content

Leave a rating Your progress Share

178. Lab: AWS Data Pipeline - Create source and destination data repositories

S3 Management Console

https://console.aws.amazon.com/s3/buckets/sid-testbucket-01/?region=us-west-2&tab=overview

aws Services Resource Groups

Amazon S3 > sid-testbucket-01

Overview Properties Permissions Management

Type a prefix and press Enter to search. Press ESC to clear.

Upload Create folder Download Actions

Last modified Nov 18, 2018 6:14:03 PM GMT+0530 166.0 B Storage class Standard

Name CityData.csv

Viewing 1 to 1

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows a dual-pane interface. The left pane displays a Udemy course page for 'AWS Data Architect Bootcamp'. The right pane shows the AWS S3 console for the 'sid-testbucket-01' bucket. In the S3 console, there is one object named 'CityData.csv' with a size of 166.0 B and a storage class of Standard. The file was last modified on Nov 18, 2018, at 6:14:03 PM GMT+0530. The browser's address bar shows the URL for the S3 console. At the bottom, a video player indicates the video is at 2:49 of 4:09, with a 1.5x speed setting.

AWS Data Architect Bc x (115) Learn JavaScript - F1 x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851032#content)

## 178. Lab: AWS Data Pipeline - Create source and destination data repositories

CityData - Excel

File Home Insert Draw Page Layout Formulas Data Review View Add-ins Team Tell me what you want to do

A1 A B C D E F G H I J K L M N O P Q

ID	Country	State	City	Amount
1	USA	NewYork	NewYork	101.4
2	England	London	London	102.4
3	India	Maharash	Mumbai	103.5
4	Singapore	Singapore	Singapore	104.6

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25



CityData

Ready

Course content Overview Q&amp;A Notes Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Leave a rating Your progress Share

179. Lab: AWS Data Pipeline - Create a new data pipeline

RDS AWS Console AWS Management Console

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

aws Services Resource Groups

### AWS services

- Data Pipeline
- Database Migration Service
- AWS AppSync
- AWS Cost Explorer
- AWS Glue
- AWS Migration Hub
- Amazon Comprehend
- Storage Gateway
- Database
- RDS
- DynamoDB
- ElastiCache
- Elastic Transcoder
- Kinesis Video Streams
- MediaConvert
- MediaLive
- MediaPackage
- MediaStore
- MediaTailor
- Application Integration
- Step Functions
- Amazon MQ
- Simple Notification Service
- Simple Queue Service
- SWF

### Helpful tips

- Manage your costs
- Create an organization

### Explore AWS

- Machine Learning with Amazon SageMaker
- Amazon Relational Database Service (RDS)
- AWS Fargate Runs Containers for You

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows a browser window with the following details:

- Address Bar:** https://us-west-2.console.aws.amazon.com/console/home?region=us-west-2
- Tab Bar:** AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content (active tab), (115) Learn JavaScript - F... (inactive tab), + (new tab).
- Header:** Udemy logo, AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools, Leave a rating, Your progress, Share, Help icon.
- Main Content:**
  - AWS services sidebar:** Data Pipeline, Database Migration Service, AWS AppSync, AWS Cost Explorer, AWS Glue, AWS Migration Hub, Amazon Comprehend, Storage Gateway, Database (RDS, DynamoDB, ElastiCache), Elastic Transcoder, Kinesis Video Streams, MediaConvert, MediaLive, MediaPackage, MediaStore, MediaTailor, Application Integration (Step Functions, Amazon MQ, Simple Notification Service, Simple Queue Service, SWF).
  - Helpful tips:** Manage your costs (Monitor your AWS costs, usage, and reservations using AWS Budgets. Start now), Create an organization (Use AWS Organizations for policy-based management of multiple AWS accounts. Start now).
  - Explore AWS:** Machine Learning with Amazon SageMaker (The fastest way to build, train, and deploy machine learning models. Learn more), Amazon Relational Database Service (RDS) (RDS manages and scales your database for you. RDS supports Aurora, MySQL, PostgreSQL, MariaDB, Oracle, and SQL Server. Learn more), AWS Fargate Runs Containers for You (AWS Fargate works with Amazon ECS to run and scale your containers. Learn more).
- Bottom Navigation:** Course content, Overview, Q&A, Notes, Announcements.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

(115) Learn JavaScript - Full Course

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content



AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share



170. Let's AWS Data Pipeline - Create a Data Pipeline

RDS - AWS Console

AWS Data Pipeline Management

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#CreatePipelinesPlace:

Source  Build using a template

Choose...

Choose...

Getting Started

Getting Started using ShellCommandActivity

AWS Command Line Interface (CLI) Templates

Run AWS CLI command

DynamoDB Templates

Export DynamoDB table to S3

Import DynamoDB backup data from S3

Elastic MapReduce (EMR) Templates

Run job on an Elastic MapReduce cluster

RDS Templates

Full copy of RDS MySQL table to S3

Incremental copy of RDS MySQL table to S3

Load S3 data into RDS MySQL table

Redshift Templates

Full copy of RDS MySQL table to Redshift

Incremental copy of RDS MySQL table to Redshift

Load data from S3 into Redshift

Copy execution logs to S3 [More](#)

Schedule

You can run your pipeline once or specify a schedule

Run

DynamoDB Templates

Export DynamoDB table to S3

Run every

Starting

Ending

Pipeline Configuration

Logging

S3 location for logs

s3://

Security/Access

IAM roles

 Default

Custom

IAM Roles let you control permissions for AWS

Data Pipeline and your EC2 applications. [More](#)

Course content

Overview

Q&amp;A

Notes

Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

179. Lab: AWS Data Pipeline - Create a new data pipeline AWS Data Pipeline Management Security Groups | VPC Management

https://us-west-2.console.aws.amazon.com/vpc/home?region=us-west-2#SecurityGroups:sort=groupId

VPC Dashboard Services Resource Groups

Create security group Actions

Filter by VPC: Select a VPC

Virtual Private Cloud Your VPCs Subnets Route Tables Internet Gateways Egress Only Internet Gateways DHCP Options Sets Elastic IPs Endpoints Endpoint Services NAT Gateways Peering Connections Security Network ACLs Security Groups

1 to 2 of 2

Name	Group ID	Group Name	VPC ID	Type	Description
sg-[REDACTED]	sg-[REDACTED]	rds-launch-wizard	vpc-[REDACTED]	EC2-VPC	Created from the ...
		default	vpc-[REDACTED]	EC2-VPC	default VPC secur...

Security Group: sg-06d7c672c0ece0428

Description Inbound Rules Outbound Rules Tags

Group ID: sg-[REDACTED] Group Name: rds-launch-wizard  
VPC ID: vpc-[REDACTED] Description: Created from the RDS Management Console:  
Owner: [REDACTED] 2018/11/18 06:55:35  
Outbound rule count: 1 Inbound rule count: 2

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows the AWS VPC Security Groups console. On the left, there's a sidebar with various network-related services like VPC Dashboard, Internet Gateways, and Security Groups. The main area shows a table of security groups with columns for Name, Group ID, Group Name, VPC ID, Type, and Description. One group, 'rds-launch-wizard', is selected. Below the table, there's a detailed view of the selected security group, including its group ID, VPC ID, owner, and rule counts. At the bottom, there are tabs for Description, Inbound Rules, Outbound Rules, and Tags, along with some basic information about the group.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

AWS Data Pipeline Management Edit inbound rules | VPC Manage +

https://us-west-2.console.aws.amazon.com/vpc/home?region=us-west-2#ModifyInboundSecurityGroupRules;g

aws Services Resource Groups

Security Groups > Edit inbound rules

## Edit inbound rules

Inbound rules control the incoming traffic that's allowed to reach the instance.

Type	Protocol	Port Range	Source	Description
MySQL/Aurora	TCP	3306	Custom	e.g. SSH for Admin Desktop
All traffic	All	All	Anywhere	e.g. SSH for Admin Desktop

Add Rule

NOTE: Any edits made on existing rules will result in the edited rule being deleted and a new rule created with the new details. This will cause traffic that depends on that rule to be dropped for a very brief period of time until the new rule can be created.

\* Required

Cancel Save rules

Feedback English (US)

© 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp

(115) Learn JavaScript - Full Course

+

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content)

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share



## 179. Lab: AWS Data Pipeline - Create a new data pipeline

MySQL Workbench

File Edit View Query Database Server Tools Scripting Help



Navigator

SCHEMAS

Filter objects

- innodb
- mytestdb
  - Tables
  - Views
  - Stored Procedures
  - Functions
- sys

10

Query 1

1 insert into mytestdb.citydata (id, country, state, city, amount) values (?, ?, ?, ?, ?);

2

3 CREATE TABLE mytestdb.citydata

4 (

5 id VARCHAR(25) NOT NULL,

6 country VARCHAR(25) NOT NULL,

7 state VARCHAR(25) NOT NULL,

8 city VARCHAR(25) NOT NULL,

9 amount VARCHAR(25) NOT NULL,

10 );



Administration Schemas

Information

No object selected

Object Info Session

Query Completed

Course content

Overview

Q&amp;A

Notes

Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Leave a rating Your progress Share

179. Lab: AWS Data Pipeline - Create a new data pipeline

AWS Data Pipeline Management

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#CreatePipelinesPlace:

Parameters

RDS MySQL password: .....  
RDS MySQL security group(s) (optional): sg-.....  
RDS MySQL username: awsuser  
Input S3 file path: s3://slid-testbucket-01/CityData.csv  
Insert SQL query: insert into mytestdb.citydata (id, country, state, city, amount) values  
RDS MySQL table name: mytestdb.citydata  
Create table SQL query (optional): CREATE TABLE mytestdb.citydata  
RDS Instance ID: mysqlinstance

Schedule

You can run your pipeline once or specify a schedule. More

Run:  on pipeline activation  on a schedule  
Run every: 1 day(s)  
Starting:  on pipeline activation  2018-11-18 14:20 UTC (Current time is 14:21 UTC)  
Ending:  never  after 1 occurrence(s)  2018-11-19 14:20 UTC (Current time is 14:21 UTC)

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating Your progress Share

179. Lab: AWS Data Pipeline - Create a new data pipeline RDS - AWS Console AWS Data Pipeline Management

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#CreatePipelinesPlace:

Schedule

You can run your pipeline once or specify a schedule. More

Run  on pipeline activation  on a schedule

Pipeline Configuration

Logging  Enabled  Disabled Copy execution logs to S3. More

Security/Access

IAM roles  Default  Custom IAM Roles let you control permissions for AWS Data Pipeline and your EC2 applications. More

Tags

Add up to 10 tags to your pipeline. These tags will be applied to the pipeline as well as any resources created by the pipeline. A tag consists of a case-sensitive key-value pair. Learn more

Key Value (Optional)

Add key to create

This pipeline launches an Amazon EC2 instance (t1.micro) in your account on every scheduled execution of the pipeline. Normal service charges for this resource will apply in addition to charges for other AWS services used by this pipeline.

Cancel Edit in Architect Activate

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

1.5x 8:13 / 10:51

AWS Data Architect Bc x (115) Learn JavaScript - F... x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content)

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share

## 179. Lab: AWS Data Pipeline - Create a new data pipeline

AWS Data Pipeline Management x Security Groups | VPC Manager x +

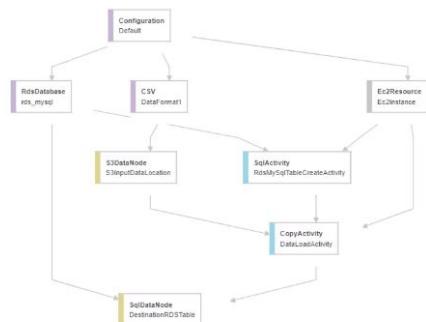
<https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ArchitectPlace:pipelineId=df-10464712QM4HMVLZLEBO>

Services

Resource Groups

Data Pipeline &gt; List Pipelines &gt; Architect: MyFirstDataPipeline (df-10464712QM4HMVLZLEBO) [Pending]

Add Save Activate Export View/Edit tags



Activities

DataNodes

Schedules

Resources

Preconditions

Others

Parameters

myRDSPassword

myEc2RdsSecurityGrps (optional)

myRDSUsername

myInputS3Loc

myRDSTableInsertSql

myRDSTableName

myRDSCreateTableSql (optional)

myRDSInstanceId

mySqlInstance



## Errors/Warnings

Object: Default

WARNING: 'pipelineLogUrl' is missing. It is recommended to set this value on Default object for better troubleshooting.



Course content Overview Q&amp;A Notes Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

179. Lab: AWS Data Pipeline - Create a new data pipeline

AWS Data Pipeline Management Security Groups VPC Manager

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ArchitectPlace:pipelineId=df-10464712QM4HMVLZLEBO

aws Services Resource Groups

Data Pipeline List Pipelines Architect: MyFirstDataPipeline (df-10464712QM4HMVLZLEBO) [Pending]

Add Save Activate Export View/Edit tags

Configuration Default

RdsDatabase (rds\_mydb)

CSV DataFormat

Ec2Resource EchoNone

S3DataNode (S3InputDataLocation)

SqlActivity (RdsSqlTableCreateActivity)

CopyActivity (DataLoadActivity)

SqlDataNode (DestinationRDSTable)

Activities DataNodes Schedules Resources Preconditions Others Parameters

myRDSPassword .....  
myEc2RdsSecurityGrps (optional) sg-[REDACTED]  
myRDSUsername awsuser  
myInputS3Loc s3://sid-testbucket-01/CityData.csv  
myRDSTableInsertSql insert into mytestdb.citydata (id, country, state, city, amount) values  
myRDSTableName mytestdb.citydata  
myRDSCreateTableSql (optional) CREATE TABLE mytestdb.citydata  
myRDSInstanceId mysqlinstance

Errors/Warnings

Object: Default  
WARNING: 'pipelineLogUrl' is missing. It is recommended to set this value on Default object for better troubleshooting.

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp

(115) Learn JavaScript - Full Course

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content



AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share

⋮

## 179. Lab: AWS Data Pipeline - Create a new data pipeline

EC2 Management Console

AWS Data Pipeline Management

+

Maximum Retries

Min Instance Count

On Fall

On Late Action

On Success

Region

Resource Role

Role

Run As User

Schedule Type

Security Group Ids

Security Groups

Spot Bid Price

Action On Task Failure:

Subnet Id

Use On Demand On Last Attempt

Worker Group

Terminate After:

Add an optional field...

Activities  
DataNodes  
Schedules  
Resources

Ec2Instance

Name:

Type:

Instance Type:

Action On Task Failure:

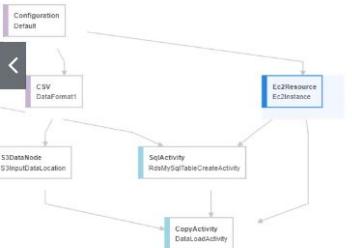
Security Groups:

Terminate After:

Preconditions

Others

Parameters



## Errors/Warnings

Object: Ec2Instance



Course content Overview Q&amp;A Notes Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

AWS Data Pipeline Management Subnets | VPC Management Con +

https://us-west-2.console.aws.amazon.com/vpc/home?region=us-west-2#subnets:sort=SubnetId

aws Services Resource Groups

VPC Dashboard

Create subnet Actions

Filter by VPC: Select a VPC

Virtual Private Cloud Your VPCs Subnets Route Tables Internet Gateways Egress Only Internet Gateways DHCP Options Sets Elastic IPs Endpoints Endpoint Services NAT Gateways Peering Connections Security Network ACLs Security Groups

Feedback English (US)

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows the AWS VPC Dashboard with three subnets listed:

Name	Subnet ID	State	VPC	IPv4 CIDR	Available IPv4	IPv6 CIDR	Availability Zone
	subnet-00	available		251	-	us-west-2c	
	subnet-07	available		250	-	us-west-2a	
	subnet-0c	available		251	-	us-west-2b	

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851034#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating Your progress Share

179. Lab: AWS Data Pipeline - Create a new data pipeline

AWS Data Pipeline Management Subnets | VPC Management Con +

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ArchitectPlace:pipelineId=df-10464712QM4HMLZLEBO

aws Services Resource Groups

Data Pipeline List Pipelines Architect: MyFirstDataPipeline (df-10464712QM4HMLZLEBO) [Pending]

Add Save Activate Export View/Edit tags

Configuration Detail

CSV DataFormat

S3DataNode StepInputDataLocation

SeqActivity ResultSqlTableCreateActivity

Ec2Resource Ec2Instance

CopyActivity DataLoadActivity

Activities DataNodes Schedules Resources

Ec2Instance

Name: Ec2Instance  
Type: Ec2Resource  
Subnet Id: subnet-00  
Instance Type: t1.micro  
Action On Task Failure: terminate  
Terminate After: 2 Hour(s)

Preconditions Others Parameters

Errors/Warnings

Object: Default  
WARNING: pipelineLogUrl is missing. It is recommended to set this value on Default object for better troubleshooting.

Feedback English (US)

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - (115) Learn JavaScript - F... | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

u Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management Subnets | VPC Management Con...

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ArchitectPlace:pipelineId=df-10464712QM4HMVLZLEBO

aws Services Resource Groups

Data Pipeline List Pipelines Architect: MyFirstDataPipeline (df-10464712QM4HMVLZLEBO) [Pending]

Add Save Activate Export View/Edit tags

Configuration Details

Activities DataNodes Schedules Resources Preconditions Others

CSV DataFormat1

Ec2Resource ECommerce

S3DataNode C3DeploymentLocation

SqlActivity RDSMySqlTableCreateActivity

CopyActivity DataLoadActivity

Pipeline objects were saved, but there are validation warnings. You can still activate the pipeline.

Cancel Activate

File0428

Bucket-01/CityData.csv

RDS Table (Sensu)

myRDSTableName mytestdb.citydata

myRDSCreateTableSql (optional) CREATE TABLE mytestdb.citydata

myRDSDatabaseId mysqlinstance

Errors/Warnings

Object: Default

WARNING: pipelineLogUri is missing. It is recommended to set this value on Default object for better troubleshooting.

Feedback English (US) 1.5x 0:06 / 4:32

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management Subnets | VPC Management Con +

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ExecutionDetailsPlace:pipelineId=df-10464712QM4HMVL...

aws Services Resource Groups

Data Pipeline List Pipelines Execution Details: MyFirstDataPipeline (df-10464712QM4HMVLZLEBO) Data Pipeline Help

Edit Pipeline Rerun Cancel Mark Finished Your pipeline is now active. Objects will be visible on pipeline execution.

Show all components in any state with Schedule Interval between 2018-11-04 14:36:56 UTC and 2018-11-18 14:36:56 UTC Apply

Filter: Activities Filter Instances ... 2 instances (all loaded)

Component Name	Schedule Interval (UTC)	Type	Status	Execution Start (UTC)	Execution End (UTC)	Attempt
DataLoadActivity	2018-11-18 14:31:53 – 14:31:53	CopyActivity	WAITING_ON_DEPENDENCIES	2018-11-18 14:31:57	–	–
RdsMySQLTableCreateActivity	2018-11-18 14:31:53 – 14:31:53	SqActivity	WAITING_FOR_RUNNER	2018-11-18 14:31:56	–	1 of 3 >

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management Subnet | VPC Management Con +

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ExecutionDetailsPlace:pipelineId=df-10464712QM4HMVL...

aws Services Resource Groups

Data Pipeline List Pipelines Execution Details: MyFirstDataPipeline (df-10464712QM4HMVLZLEBO) Data Pipeline Help

Edit Pipeline Rerun Cancel Mark Finished Your pipeline is now active. Objects will be visible on pipeline execution.

Show all components in any state with Schedule Interval between 2018-11-04 14:36:56 UTC and 2018-11-18 14:36:56 UTC Apply

Filter: Activities Filter Instances ... 2 instances (all loaded)

Component Name	Schedule Interval (UTC)	Type	Status	Execution Start (UTC)	Execution End (UTC)	Attempt
DataLoadActivity	2018-11-18 14:31:53 – 14:31:53	CopyActivity	WAITING_ON_DEPENDENCIES	2018-11-18 14:31:57	–	–
RdsMySQLTableCreateActivity	2018-11-18 14:31:53 – 14:31:53	SqActivity	WAITING_FOR_RUNNER	2018-11-18 14:31:56	–	1 of 3 >

Dependencies Attempts

Runs on resource Ec2Instance

RdsMySQLTableCreateActivity waiting on... Ec2Instance

Name Ec2Instance
Type Ec2Resource
Schedule start time 2018-11-18 14:31:53
Execution start time 2018-11-18 14:31:57
Status CREATING
Attempt 1 of 3 Latest Attempt (1) Details
Logs Logs not available

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

1.5x 1:08 / 4:32

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - Fi x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management EC2 Management Console

https://us-west-2.console.aws.amazon.com/ec2/v2/home?region=us-west-2#instances:sort=instanceId

aws Services Resource Groups

EC2 Dashboard Events Tags Reports Limits

INSTANCES Instances Launch Templates Spot Requests Reserved Instances Dedicated Hosts Scheduled Instances Capacity Reservations

IMAGES AMIs Bundle Tasks

ELASTIC BLOCK STORE Volumes Snapshots Lifecycle Manager

NETWORK &

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

Name	Instance ID	Instance Type	Availability Zone	Instance State	Status Checks	Alarm Status	Public DNS (IPv4)	IPv4 Put
L...	i-...	t1.micro	us-west-2c	running	Initializing	None	None	None

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management | EC2 Management Console

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ExecutionDetailsPlace:pipelineId=df-10464712QM4HMVL...

aws Services Resource Groups

Data Pipeline List Pipelines Execution Details: MyFirstDataPipeline (df-10464712QM4HMVLZLEBO) Data Pipeline Help

Edit Pipeline Rerun Cancel Mark Finished

Show all components in any state with Schedule Interval between 2018-11-04 14:36:56 UTC and 2018-11-18 14:36:56 UTC Apply

Filter: Activities Filter Instances ... 2 instances (all loaded)

Component Name	Schedule Interval (UTC)	Type	Status	Execution Start (UTC)	Execution End (UTC)	Attempt
DataLoadActivity	2018-11-18 14:31:53 – 2018-11-18 14:31:53	CopyActivity	RUNNING	2018-11-18 14:31:57	–	1 of 3
RdsMySQLTableCreateActivity	2018-11-18 14:31:53 – 2018-11-18 14:31:53	SqActivity	FINISHED	2018-11-18 14:31:56	2018-11-18 14:34:11	1 of 3

Dependencies Attempts

Attempt	Status	StartTime	EndTime	Logs	Error Message	Details
1	FINISHED	2018-11-18 14:33:59	2018-11-18 14:34:10	Logs not available		More...

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

1.5x 1:37 / 4:32

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management | EC2 Management Console

https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ExecutionDetailsPlace:pipelineId=df-10464712QM4HMVLZLEBO...

aws Services Resource Groups

Data Pipeline List Pipelines Execution Details: MyFirstDataPipeline (df-10464712QM4HMVLZLEBO) Data Pipeline Help

Edit Pipeline Rerun Cancel Mark Finished

Show all components in any state with Schedule Interval between 2018-11-04 14:36:56 UTC and 2018-11-18 14:36:56 UTC Apply

Filter: Activities Filter Instances ... 2 instances (all loaded)

Component Name	Schedule Interval (UTC)	Type	Status	Execution Start (UTC)	Execution End (UTC)	Attempt
DataLoadActivity	2018-11-18 14:31:53 – 2018-11-18 14:31:53	CopyActivity	FINISHED	2018-11-18 14:31:57	2018-11-18 14:34:15	1 of 3
RdsMySQLTableCreateActivity	2018-11-18 14:31:53 – 2018-11-18 14:31:53	SqActivity	FINISHED	2018-11-18 14:31:56	2018-11-18 14:34:11	1 of 3

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The image shows a screenshot of a Linux desktop environment with a browser window open. The browser has two tabs: 'AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy' and '(115) Learn JavaScript - F... | AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com'. The main content area displays the AWS Data Pipeline Management console, specifically the 'Execution Details' for a pipeline named 'MyFirstDataPipeline'. The table lists two activities: 'DataLoadActivity' and 'RdsMySQLTableCreateActivity', both of which are shown as FINISHED. Below the table, there are navigation arrows for more instances. At the bottom of the browser window, there's a video player interface showing a progress bar at 2:02 / 4:32 and a 1.5x speed button. The browser's address bar shows the URL https://console.aws.amazon.com/datapipeline/home?region=us-west-2#ExecutionDetailsPlace:pipelineId=df-10464712QM4HMVLZLEBO... The desktop taskbar at the top includes icons for various applications like a terminal, file manager, and system monitor.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - (115) Learn JavaScript - F... | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

MySQL Workbench

Mysql@mysqlinstance.calpaj... File Edit View Query Database Server Tools Scripting Help

Navigator Schemas Filter objects

innodb mytestdb Tables citydata Columns id country state city amount

Indexes Foreign Keys Views Stored Procedures Functions

sys Administration Schemas

Information Schema: mytestdb

Result Grid Filter Rows Export Wrap Cell Content

citydata 17 x Read Only

Object Info Session

Query Completed 1.5x 2:25 / 4:32

Course content Overview Q&A Notes Announcements

Section 1: Introduction 3 / 3 | 13min

The screenshot shows the MySQL Workbench interface. In the top-left corner, there's a browser tab for 'AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium'. Below it, another tab says '(115) Learn JavaScript - F... | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content'. The main window is titled '180. Lab: AWS Data Pipeline - Execute data pipeline'. It displays a MySQL session for 'Mysql@mysqlinstance.calpaj...'. The 'Schemas' pane shows 'innodb' and 'mytestdb'. Under 'mytestdb', the 'Tables' section is expanded, showing a table named 'citydata' with columns: id, country, state, city, and amount. The 'Result Grid' pane shows the following data:

ID	Country	State	City	Amount
1	USA	NewYork	NewYork	101.4
2	England	London	London	102.4
3	India	Maharashtra	Mumbai	103.5
4	Singapore	Singapore	Singapore	104.6

The bottom status bar indicates 'Query Completed' with a progress bar at 2:25 / 4:32.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - (115) Learn JavaScript - F... | (115) Learn JavaScript - F... | +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management EC2 Management Console

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

aws Services Resource Groups Oregon Support

Data Pipeline List Pipelines DataPipeline Help

Create new pipeline Actions

Filter: All Filter pipelines... 1 pipelines (all loaded)

Pipeline ID	Name	Schedule State	Health Status	Creation Time (UTC)
df-10464712QM4HMVLZLEBO	MyFirstDataPipeline	SCHEDULED Runs every 100 years	No completed executions	2018-11-18 14:28:04

Schedule information

Activity	Type	Health Status	Last Completed Run (UTC)	Active Run (UTC)	Next Run (UTC)
DataLoadActivity	CopyActivity	HEALTHY	2018-11-18 14:31:53	2018-11-18 14:31:56	-
RdsMySQLTableCreateActivity	SqActivity	HEALTHY	2018-11-18 14:31:53	2018-11-18 14:31:56	-

Tags View all / Edit

Edit pipeline View execution details

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows a browser window with the Udemy 'AWS Data Architect Bootcamp' course page. The main content area displays the AWS Data Pipeline Management console. A single pipeline, 'MyFirstDataPipeline', is listed. The pipeline is scheduled to run every 100 years. It contains two activities: 'DataLoadActivity' (CopyActivity, HEALTHY) and 'RdsMySQLTableCreateActivity' (SqActivity, HEALTHY). Both activities have completed runs at 2018-11-18 14:31:53 and 2018-11-18 14:31:56 respectively. The pipeline has no completed executions. The browser title bar shows 'AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium'.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - F x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management EC2 Management Console

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

aws Services Resource Groups

Data Pipeline List Pipelines DataPipeline Help

Create new pipeline Actions

Filter: All Filter

Pipeline ID: dt-10464712QM4H

Edit Clone Delete Export Activate Deactivate

1 pipelines (all loaded)

Name	Schedule State	Health Status	Creation Time (UTC)
MyFirstDataPipeline	SCHEDULED Runs every 100 years	No completed executions	2018-11-18 14:28:04

Schedule information

Start 2018-11-18 14:31:53 (UTC)  
End 2018-11-18 14:31:53 (UTC)  
Period Runs every 100 years

Tags

View all / Edit

Activity Type Health Status Last Completed Run (UTC) Active Run (UTC) Next Run (UTC)

DataLoadActivity	CopyActivity	HEALTHY	2018-11-18 14:31:53	2018-11-18 14:31:56	-
RdsMySQLTableCreateActivity	SqActivity	HEALTHY	2018-11-18 14:31:53	2018-11-18 14:31:56	-

Edit pipeline View execution details

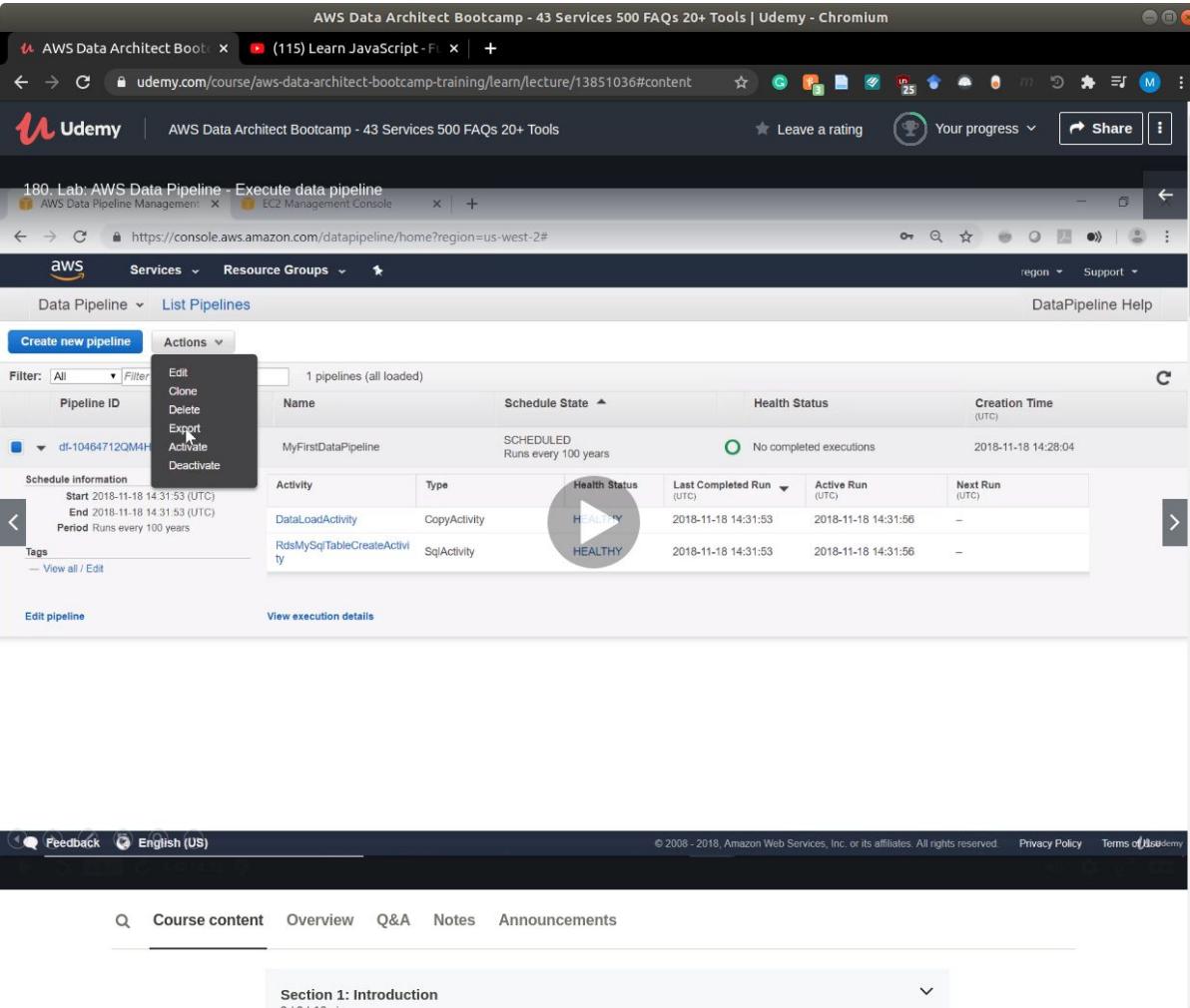
Feedback English (US)

© 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min



AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bc (115) Learn JavaScript - F1 +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851036#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

180. Lab: AWS Data Pipeline - Execute data pipeline

AWS Data Pipeline Management EC2 Management Console

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

aws Services Resource Groups

Data Pipeline List Pipelines

Create new pipeline Actions

Filter: All Filter pipelines...

Pipeline ID Name Schedule information Tags

dt-10464712QM4HMLZLEBO MyFirstDataActivity DataLoadActivity RdsMySqlTable

Start 2018-11-18 14:31:53 (UTC)  
End 2018-11-18 14:31:53 (UTC)  
Period Runs every 100 years

View all / Edit

Edit pipeline View execution

**Pipeline Definition (JSON Format)**

```
{ "objects": [ { "password": "#{myRDPSPassword}", "name": "rds_mysql", "jdbcProperties": "allowMultiQueries=true", "id": "rds_mysql", "type": "RdsDatabase", "rdsInstanceId": "#{myRDSInstanceId}", "username": "#{myRDSUsername}" }, { "output": { "ref": "DestinationRDSTable" }, "input": { "ref": "S3InputDataLocation" } }, { "dependsOn": { "ref": "RdsMySqlTableCreateActivity" } } ] }
```

Creation Time (UTC)

Run	Next Run (UTC)
11-18 14:31:56	-
11-18 14:31:56	-

Close

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - Fi x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851038#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

181. Lab: AWS Data Pipeline - Delete data pipeline

AWS Data Pipeline Management EC2 Management Console

https://us-west-2.console.aws.amazon.com/ec2/v2/home?region=us-west-2#instances:sort=instanceId

aws Services Resource Groups

EC2 Dashboard Events Tags Reports Limits

INSTANCES Instances Launch Templates Spot Requests Reserved Instances Dedicated Hosts Scheduled Instances Capacity Reservations

IMAGES AMIs Bundle Tasks

ELASTIC BLOCK STORE Volumes Snapshots Lifecycle Manager

NETWORK &

Course content Overview Q&A Notes Announcements

Section 1: Introduction 3 / 3 | 13min

The screenshot shows the AWS EC2 Management Console interface. On the left, there's a sidebar with navigation links for EC2 Dashboard, Events, Tags, Reports, Limits, Instances (selected), Launch Templates, Spot Requests, Reserved Instances, Dedicated Hosts, Scheduled Instances, Capacity, Reservations, IMAGES, AMIs, and Bundle Tasks. Below that is another section for ELASTIC BLOCK STORE with Volumes, Snapshots, and Lifecycle Manager. At the bottom, there's a NETWORK & section. The main content area displays a table of instances. The table has columns: Name, Instance ID, Instance Type, Availability Zone, Instance State, Status Checks, Alarm Status, Public DNS (IPv4), and IPv4 Put. There is one row visible with the following data: Name (redacted), Instance ID (redacted), i1.micro, us-west-2c, terminated, None, and a yellow exclamation mark icon. The status bar at the bottom shows the video player controls: 1.5x, 0:25 / 2:42, and a refresh icon.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851038#content

181. Lab: AWS Data Pipeline - Delete data pipeline

AWS Data Pipeline Management

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

aws Services Resource Groups

Data Pipeline List Pipelines

Create new pipeline

Actions ▾

- Edit
- Clone
- Delete**
- Export
- Activate
- Deactivate

Filter: All Filter

Pipeline ID

Name	Schedule State	Health Status	Creation Time (UTC)
MyFirstDataPipeline	FINISHED Runs every 100 years	HEALTHY	2018-11-18 14:28:04

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows the AWS Data Pipeline Management interface. A single pipeline named "MyFirstDataPipeline" is listed in the table. The "Actions" dropdown menu is open, and the "Delete" option is highlighted with a red box. The pipeline details show it is FINISHED and runs every 100 years, with a current status of HEALTHY and a creation time of 2018-11-18 14:28:04.

Name	Schedule State	Health Status	Creation Time (UTC)
MyFirstDataPipeline	FINISHED Runs every 100 years	HEALTHY	2018-11-18 14:28:04

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - Fi x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851038#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating Your progress Share

181. Lab: AWS Data Pipeline - Delete data pipeline

vpc | VPC Management Console RDS - AWS Console

https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstances:

aws Services Resource Groups

Amazon RDS Instances

Instances (1)

DB instance Engine MySQL

mysqlinstance

Instance actions ▾

Restore from S3 Create database

Modify Stop Reboot Current activity Maintenance

Delete Create read replica Create Aurora read replica Promote read replica

Take snapshot Restore to point in time Migrate snapshot

Filter instances

Subnet groups Parameter groups Option groups Events Event subscriptions Recommendations 2

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows a dual-pane interface. The left pane is a video player for a Udemy course lecture, and the right pane is the AWS RDS (Amazon Relational Database Service) console. In the video player, the title is '181. Lab: AWS Data Pipeline - Delete data pipeline'. The video progress bar is at 1:24 / 2:42. The AWS RDS console shows one MySQL DB instance named 'mysqlinstance'. A context menu is open over this instance, with the 'Delete' option highlighted. Other options in the menu include Modify, Stop, Reboot, Create database, Restore from S3, Create read replica, Create Aurora read replica, Promote read replica, Take snapshot, Restore to point in time, and Migrate snapshot. The browser's address bar shows the URL https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstances:. The overall context is a lab exercise for deleting a data pipeline in AWS.

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

(115) Learn JavaScript - Full Course | Udemy - Chromium

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851038#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851038#content)

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools

Leave a rating

Your progress

Share



## 181. Lab: AWS Data Pipeline - Delete data pipeline

vpc | VPC Management Console

RDS - AWS Console

<https://us-west-2.console.aws.amazon.com/rds/home?region=us-west-2#dbinstances>

aws

Services

Resource Groups

Amazon RDS

Dashboard

Instances

Clusters

Performance Insights

Snapshots

Automated backups

Reserved Instances

Subnet groups

Parameter groups

Option groups

Events

Event subscriptions

Recommendations

## Delete mysqlinstance instance?



Are you sure you want to Delete the mysqlinstance DB Instance?

 Create final snapshot?

Determines whether a final DB Snapshot is created before the DB instance is deleted.

 Retain automated backups?

Determines whether retaining automated backups for 7 days after deletion.

 I acknowledge that upon instance deletion, automated backups, including system snapshots and point-in-time recovery, will no longer be available.

To confirm deletion, type delete me into the field

delete me

We strongly recommend taking a final snapshot before instance deletion since after your instance is deleted, automated backups will no longer be available.

Cancel

Delete

1.5x 1:37 / 2:42



Course content

Overview

Q&amp;A

Notes

Announcements

## Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp x (115) Learn JavaScript - Fi x +

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851038#content

Udemy AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools Leave a rating Your progress Share

181. Lab: AWS Data Pipeline - Delete data pipeline

vpc | VPC Management Console

https://us-west-2.console.aws.amazon.com/vpc/home?region=us-west-2#vpcs:sort=VpId

Services Resource Groups

VPC Dashboard

Create VPC Actions

Delete VPC

Filter by tag: Keyword

1 to 1 of 1

Name	State	IPv4 CIDR	IPv6 CIDR	DHCP options set	Route table
04d0	available	-	-	dopt-0b7a5e72	-

Virtual Private Cloud

Your VPCs Subnets Route Tables Internet Gateways Egress Only Internet Gateways DHCP Options Sets Elastic IPs Endpoints Endpoint Services NAT Gateways Peering Connections Security Network ACLs Security Groups

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

The screenshot shows the AWS VPC Management Console interface. On the left, there's a sidebar with various navigation links like Virtual Private Cloud, Your VPCs, Subnets, etc. The main area displays a table of VPCs. A single row is selected, showing details such as Name (04d0), State (available), IPv4 CIDR, IPv6 CIDR, DHCP options set (dopt-0b7a5e72), and Route table. A context menu is open over this row, with 'Delete VPC' as the top item. A large, semi-transparent play button is centered over the table. At the bottom, there's a navigation bar with Course content, Overview, Q&A, Notes, and Announcements, and a video player showing a timestamp of 1:56 / 2:42.

AWS Data Architect Bc x (115) Learn JavaScript - Fi x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851040#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851040#content)

## 182. AWS Data Pipeline - FAQs

## Q: What is AWS Data Pipeline?



AWS Data Pipeline is a web service that makes it easy to schedule regular data movement and data processing activities in the AWS cloud. AWS Data Pipeline integrates with on-premise and cloud-based storage systems to allow developers to use their data when they need it, where they want it, and in the required format. AWS Data Pipeline allows you to quickly define a dependent chain of data sources, destinations, and predefined or custom data processing activities called a pipeline. Based on a schedule you define, your pipeline regularly performs processing activities such as distributed data copy, SQL transforms, MapReduce applications, or custom scripts against destinations such as Amazon S3, Amazon RDS, or Amazon DynamoDB. By executing the scheduling, retry, and failure logic for these workflows as a highly scalable and fully managed service, Data Pipeline ensures that your pipelines are robust and highly available.

## Q: What can I do with AWS Data Pipeline?



Using AWS Data Pipeline, you can quickly and easily provision pipelines that remove the development and maintenance effort required to manage your daily data operations, letting you focus on generating insights from that data. Simply specify the data sources, schedule, and processing activities required for your data pipeline. AWS Data Pipeline handles running and monitoring your processing activities on a highly reliable, fault-tolerant infrastructure. Additionally, to further ease your development process, AWS Data Pipeline provides built-in activities for common actions such as copying data between Amazon S3 and Amazon RDS, or running a query against Amazon S3 log data.

|| 1.5x 0:13 / 6:54 ⏪



AWS Data Architect Bc x (115) Learn JavaScript - F1 x +

[udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851040#content](https://udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851040#content)

## 182 AWS Data Pipeline - FAQs

## Q: What is a pipeline?

A pipeline is the AWS Data Pipeline resource that contains the definition of the dependent chain of data sources, destinations, and predefined or custom data processing activities required to execute your business logic.

## Q: What is a data node?

A data node is a representation of your business data. For example, a data node can reference a specific Amazon S3 path. AWS Data Pipeline supports an expression language that makes it easy to reference data which is generated on a regular basis. For example, you could specify that your Amazon S3 data format is s3://example-bucket/my-logs/logdata-#{scheduledStartTime('YYYY-MM-dd-HH')}.tgz.

## Q: What is an activity?

An activity is an action that AWS Data Pipeline initiates on your behalf as part of a pipeline. Example activities are EMR or Hive jobs, copies, SQL queries, or command-line scripts.

## Q: What is a precondition?

A precondition is a readiness check that can be optionally associated with a data source or activity. If a data source has a precondition check, then that check must complete successfully before any activities consuming the data source are launched. If an activity has a precondition, then the precondition check must complete successfully before the activity is run. This can be useful if you are running an activity that is expensive to compute, and should not run until specific criteria are met.

## Q: What is a schedule?

Schedules define when your pipeline activities run and the frequency with which the service expects your data to be available. All schedules must have a start date and a frequency; for example, every day starting Jan 1, 2013, at 3pm. Schedules can optionally have an end date, after which time the AWS Data Pipeline service does not execute any activities. When you associate a schedule with an activity, the activity runs on it. When you associate a schedule with a data source, you are telling the AWS Data Pipeline service that you expect the data to be updated on that schedule. For example, if you define an Amazon S3 data source with an hourly schedule, the

service expects that the data source contains new files every hour.

1.5x 1:48 / 6:54



AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851040#content

Leave a rating Your progress Share

182. AWS Data Pipeline - FAQs

**Q: Does Data Pipeline supply any standard Activities?**

Yes, AWS Data Pipeline provides built-in support for the following activities:

- **CopyActivity:** This activity can copy data between Amazon S3 and JDBC data sources, or run a SQL query and copy its output into Amazon S3.
- **HiveActivity:** This activity allows you to execute Hive queries easily.
- **EMRActivity:** This activity allows you to run arbitrary Amazon EMR jobs.
- **ShellCommandActivity:** This activity allows you to run arbitrary Linux shell commands or programs.

**Q: Does AWS Data Pipeline supply any standard preconditions?**

Yes, AWS Data Pipeline provides built-in support for the following preconditions:

- **DynamoDBDataExists:** This precondition checks for the existence of data inside a DynamoDB table.
- **DynamoDBTableExists:** This precondition checks for the existence of a DynamoDB table.
- **S3KeyExists:** This precondition checks for the existence of a specific AmazonS3 path.
- **S3PrefixExists:** This precondition checks for at least one file existing within a specific path.
- **ShellCommandPrecondition:** This precondition runs an arbitrary script on your resources and checks that the script succeeds.

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | Udemy - Chromium

AWS Data Architect Bootcamp - 43 Services 500 FAQs 20+ Tools | udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851040#content

Leave a rating Your progress Share

**Q: Can I supply my own custom activities?**

Yes, you can use the ShellCommandActivity to run arbitrary Activity logic.

**Q: Can I supply my own custom preconditions?**

Yes, you can use the ShellCommandPrecondition to run arbitrary precondition logic.

**Q: Can you define multiple schedules for different activities in the same pipeline?**

Yes, simply define multiple schedule objects in your pipeline definition file and associate the desired schedule to the correct activity via its schedule field. This allows you to define a pipeline in which, for example, log files are stored in Amazon S3 each hour to drive generation of an aggregate report one time per day.

**Q: What happens if an activity fails?**

An activity fails if all of its activity attempts return with a failed state. By default, an activity retries three times before entering a hard failure state. You can increase the number of automatic retries to 10; however, the system does not allow indefinite retries. After an activity exhausts its attempts, it triggers any configured onFailure alarm and will not try to run again unless you manually issue a rerun command via the CLI, API, or console button.

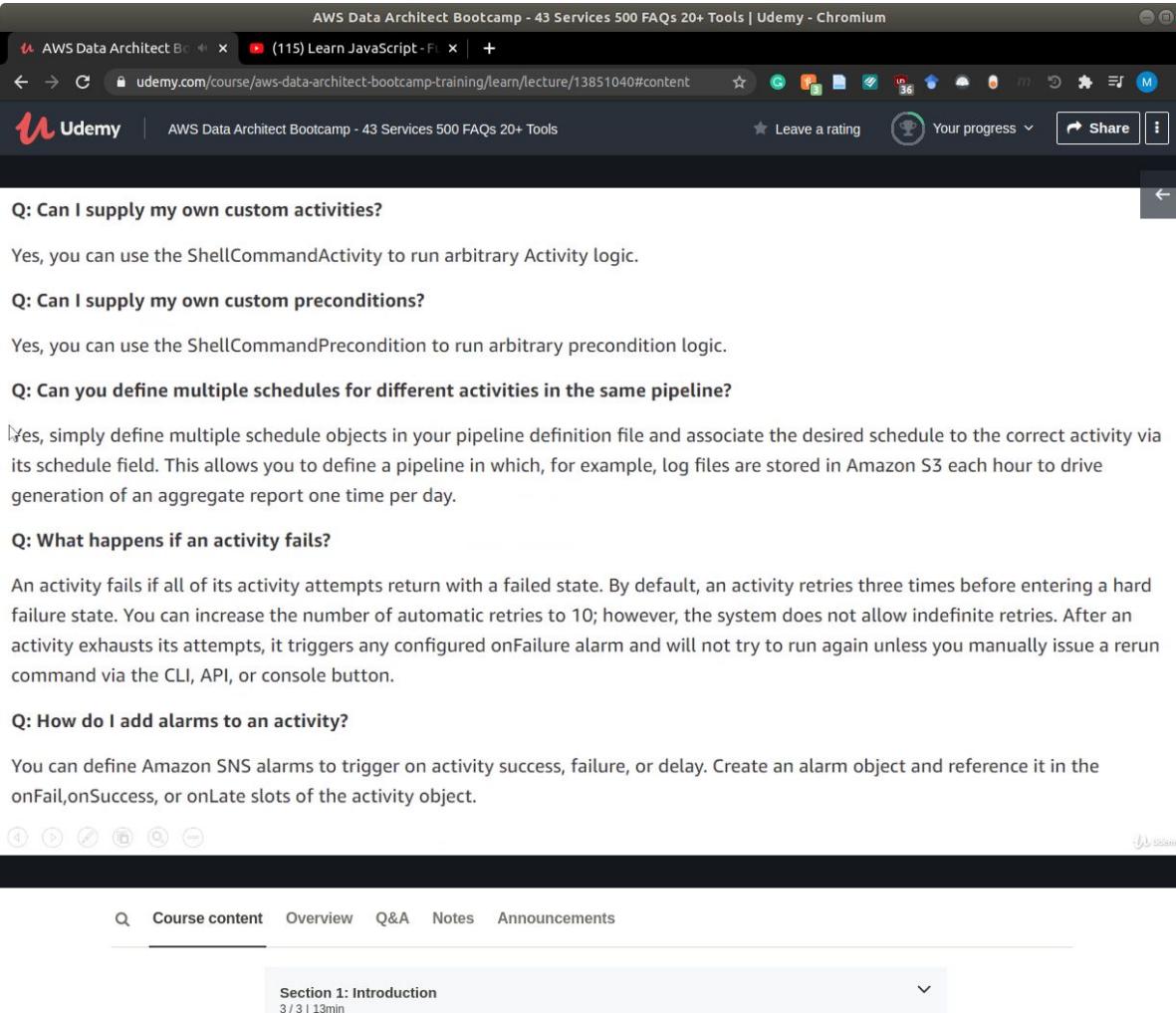
**Q: How do I add alarms to an activity?**

You can define Amazon SNS alarms to trigger on activity success, failure, or delay. Create an alarm object and reference it in the onFail, onSuccess, or onLate slots of the activity object.

Course content Overview Q&A Notes Announcements

Section 1: Introduction

3 / 3 | 13min



AWS Data Architect Bc (115) Learn JavaScript - F...

udemy.com/course/aws-data-architect-bootcamp-training/learn/lecture/13851040#content

**Q: On what resources are activities run?**

AWS Data Pipeline activities are run on compute resources that you own. There are two types of compute resources: AWS Data Pipeline-managed and self-managed. AWS Data Pipeline-managed resources are Amazon EMR clusters or Amazon EC2 instances that the AWS Data Pipeline service launches only when they're needed. Resources that you manage are longer running and can be any resource capable of running the AWS Data Pipeline Java-based Task Runner (on-premise hardware, a customer-managed Amazon EC2 instance, etc.).

**Q: Will AWS Data Pipeline provision and terminate AWS Data Pipeline-managed compute resources for me?**

Yes, compute resources will be provisioned when the first activity for a scheduled time that uses those resources is ready to run and those instances will be terminated when the final activity that uses the resources has completed successfully or failed.

**Q: Can multiple compute resources be used on the same pipeline?**

Yes, simply define multiple cluster objects in your definition file and associate the cluster to use for each activity via its runsOn field. This allows pipelines to combine AWS and on-premise resources, or to use a mix of instance types for their activities – for example, you may want to use a t1.micro to execute a quick script cheaply, but later on the pipeline may have an Amazon EMR job that requires the power of a cluster of larger instances.

**Q: Can I execute activities on on-premise resources, or AWS resources that I manage?**

Yes. To enable running activities using on-premise resources, AWS Data Pipeline supplies a Task Runner package that can be installed on your on-premise hosts. This package continuously polls the AWS Data Pipeline service for work to perform. When it's time to run a particular activity on your on-premise resources, for example, executing a DB stored procedure or a database dump, AWS Data Pipeline will issue the appropriate command to the Task Runner. In order to ensure that your pipeline activities are highly available, you can optionally assign multiple Task Runners to poll for a given job. This way, if one Task Runner becomes unavailable, the others will simply pick up its work.