



STEP ON IT!

Rapid Event-Based AWS Cost Control With Step Functions

Damien Jones (he/him)

AWS Consultant @ Steamhaus

2025-04-02 DevOps Manchester x DTX Manchester











MY AWS ACCOUNT



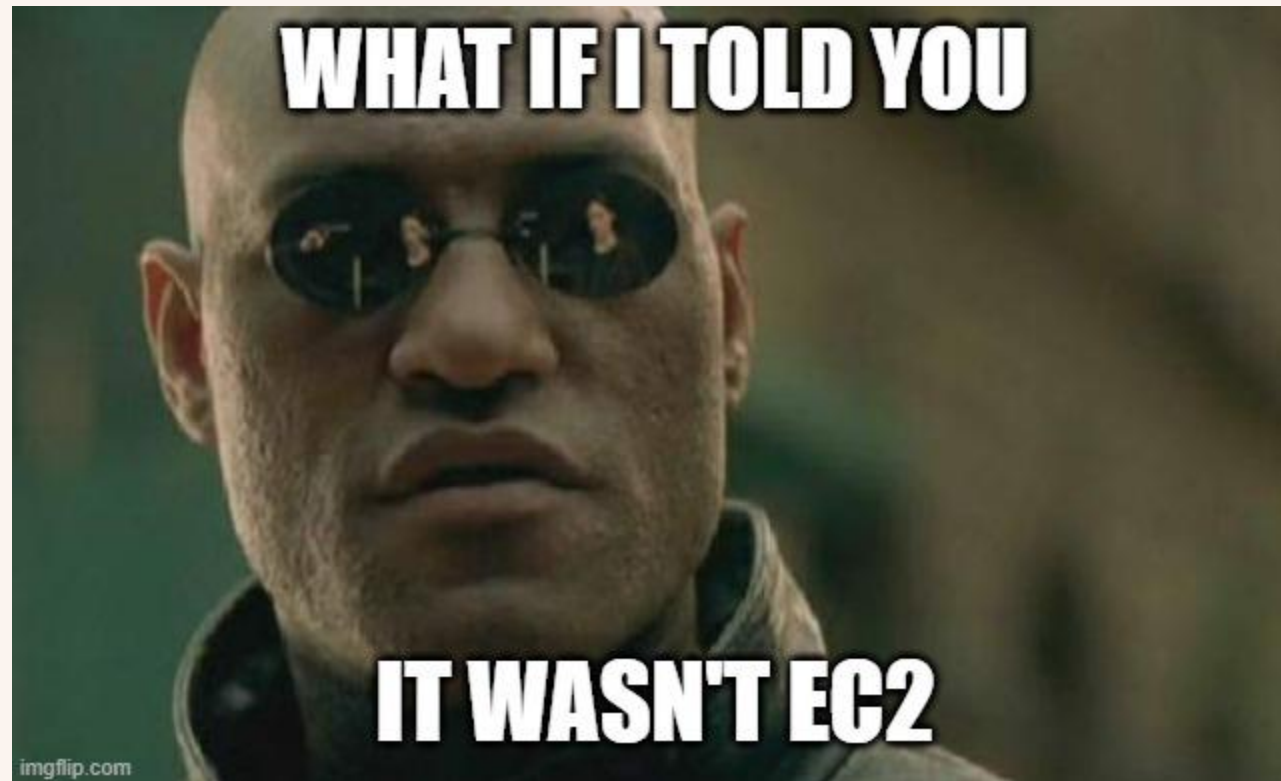
MY BILL



LOL ROFL LMAO ETC

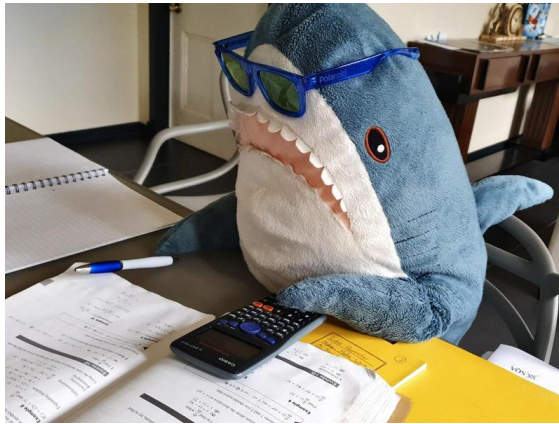


NOPE



AGENDA

WHY DID THIS
HAPPEN?



WHAT DID I DO?



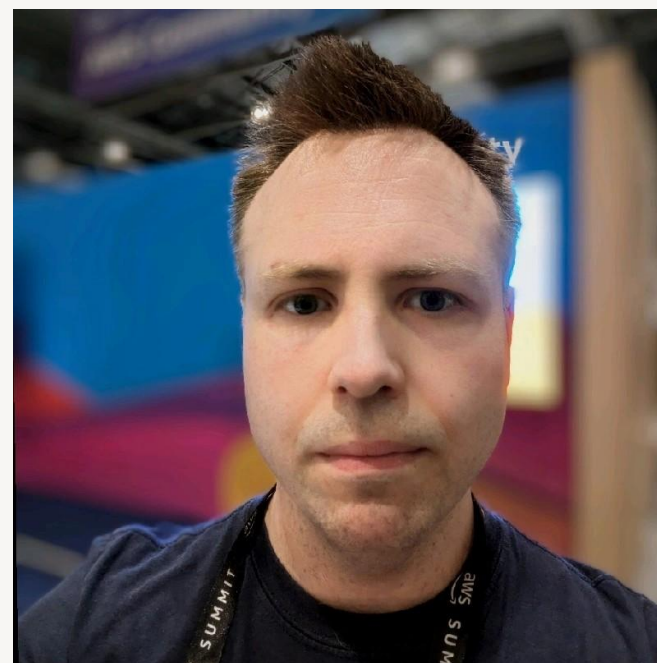
HOW DID I BUILD IT?





BUT FIRST

Hi.





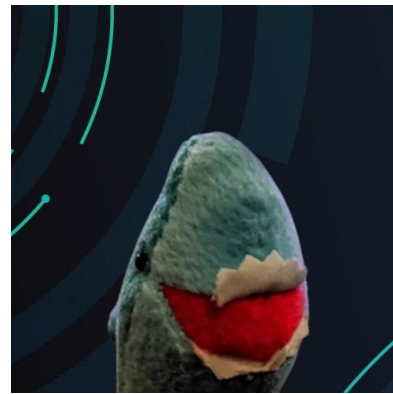
AWS Consultant

Steamhaus UK



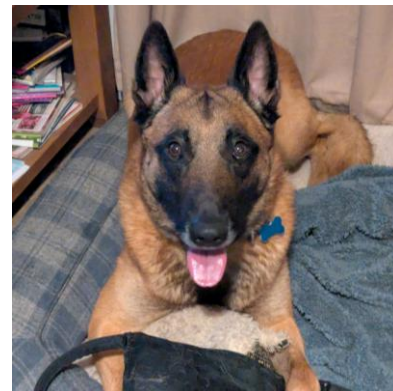
AWS Certified

x3



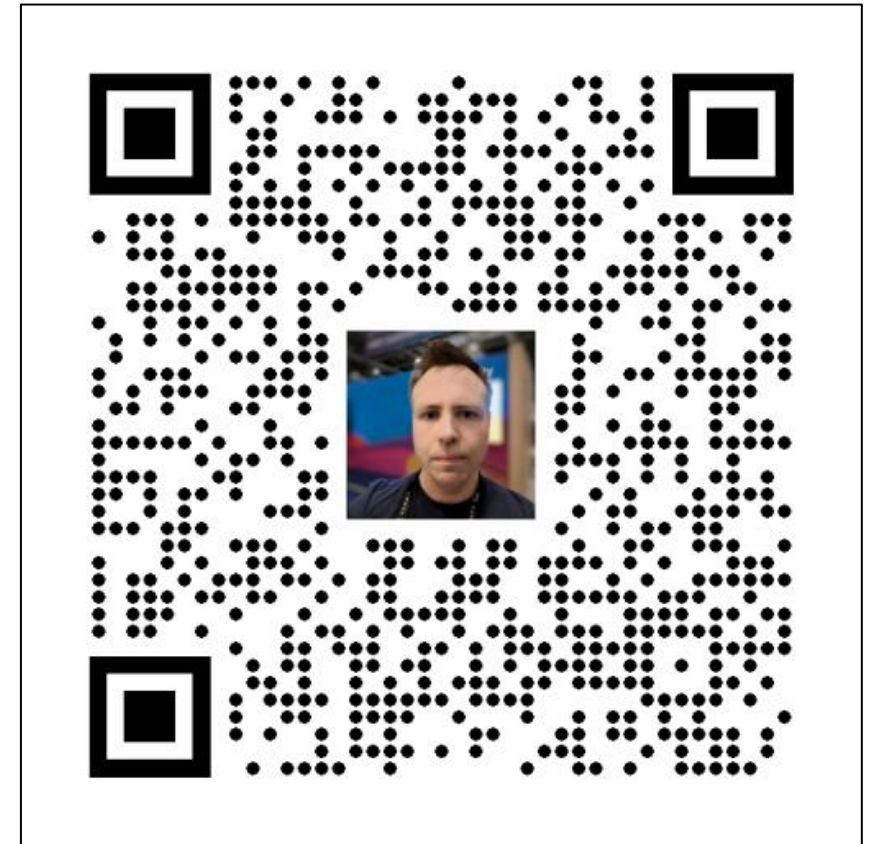
Content Creator

amazonwebshark.com



Dog Dad

to this noodle





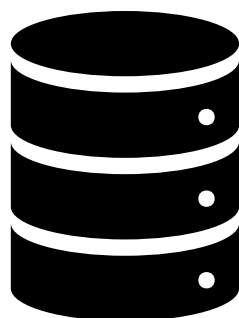
AWS User Group Leader

Lancashire UK

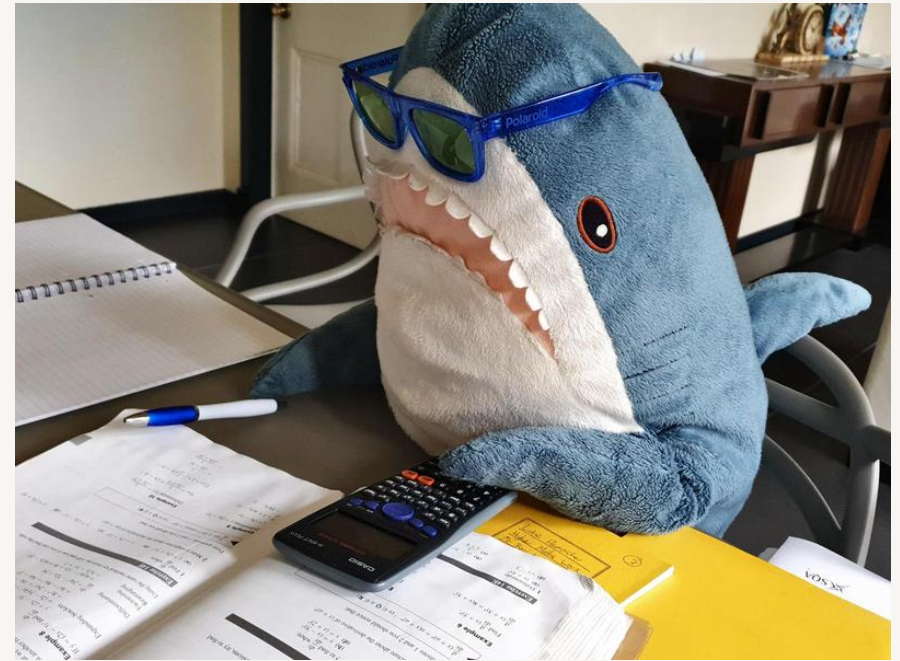


AWS Community Builder

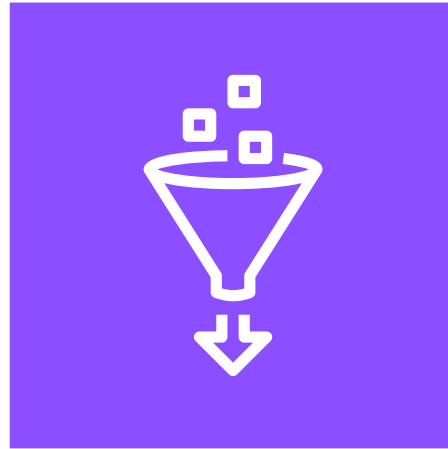
Data Category



WHY DID THIS
HAPPEN?



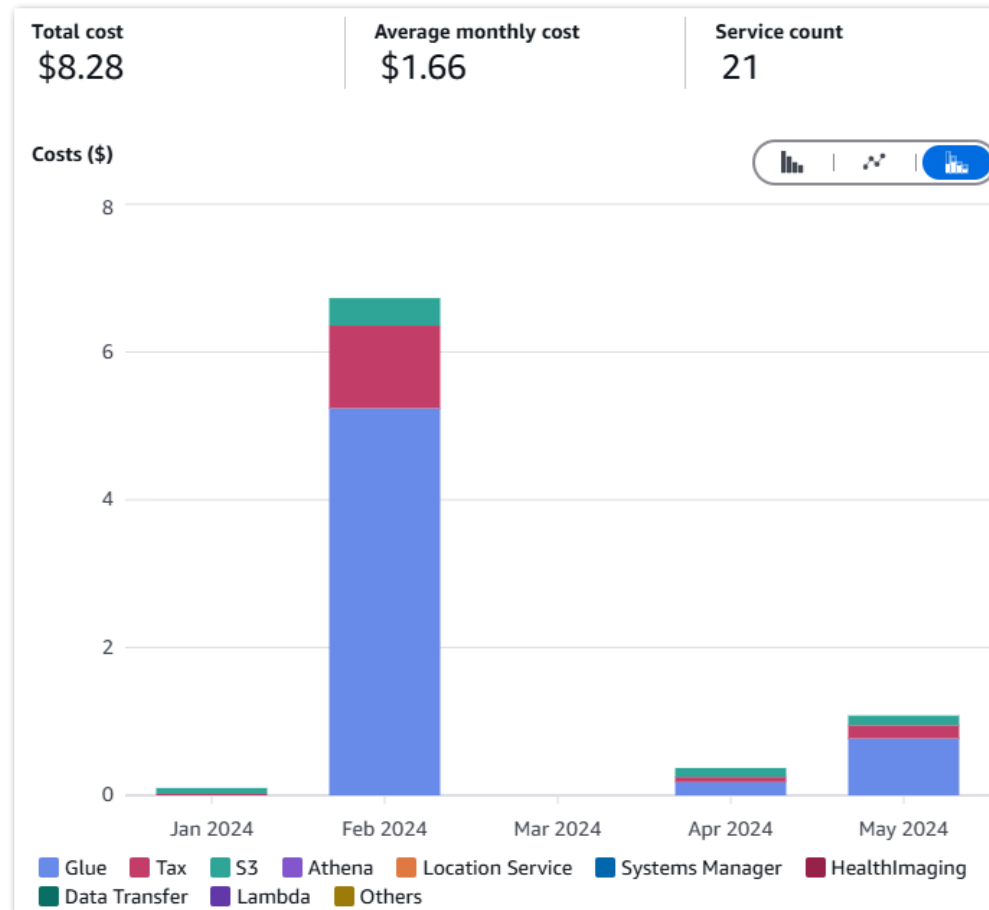
AWS GLUE



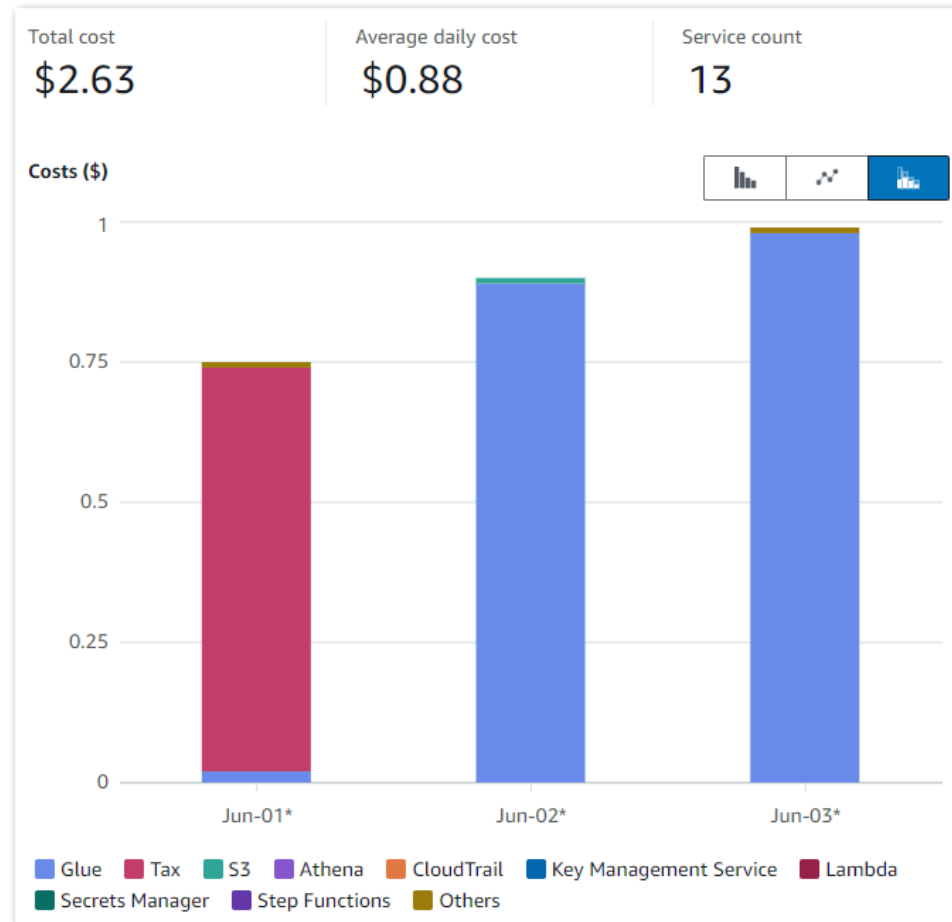
COSTS: 2024 JUNE



COSTS: 2024 H1



COSTS: 2024 JUNE



COST & USAGE: API OPERATION



COST & USAGE: API OPERATION - TAX



COST: API OPERATION – GLUE JOB RUN



COST: API OPERATION – GLUE INTERACTIVE SESSION



COST & USAGE: API OPERATIONS

API operation	API operation total	Jun-01*	Jun-02*	Jun-03*
Total costs	\$2.63	\$0.75	\$0.90	\$0.99
GlueInteractiveSession	\$1.50	-	\$0.54	\$0.97
No operation	\$0.72	\$0.72	-	-
Jobrun	\$0.33	-	\$0.33	-
CrawlerRun	\$0.06	\$0.02	\$0.02	\$0.02
PutObject	\$0.01	\$0.00	\$0.00	\$0.00
ListBucket	\$0.00	\$0.00	\$0.00	\$0.00
ReadACL	\$0.00	\$0.00	\$0.00	\$0.00

COST & USAGE: API OPERATIONS

API operation	API operation total	Jun-01*	Jun-02*	Jun-03*
GlueInteractiveSession	\$1.50	-	\$0.54	\$0.97
No operation	\$0.72	\$0.72	-	-
Jobrun	\$0.33	-	\$0.33	-



WHAT'S AN INTERACTIVE SESSION?



- AWS Glue

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations New

▼ Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

▼ Data Integration and ETL

Zero-ETL integrations New

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Interactive Sessions

Data classification tools

Sensitive data detection

Record Matching

Triggers

Workflows (orchestration)

Blueprints

Security configurations

Cost management New

▼ Legacy pages

Connections (legacy)

Jobs (legacy)

Workflows (legacy)

Welcome to AWS Glue

Get started by setting up your account and users, cataloging your data, and building ETL jobs to prepare data for analytics.

Prepare your account for AWS Glue

Admins: Grant access to AWS Glue and **set a default IAM role.**

Set up roles and users

Catalog and search for datasets

View your databases & tables and catalog data using Crawlers.

Go to the Data Catalog

Move and transform data Updated

Use Zero-ETL integrations to replicate data in near real-time, or ETL jobs to transform data in visual, notebook, or code interface.

Go to Zero-ETL integrations

Go to ETL jobs

Resources and tutorials

Getting started with AWS Glue: [Documentation](#)[AWS Training](#)

Glue in 5 Minutes Videos: [Authoring](#), [GenAI](#), [Monitoring](#), [Orchestration](#)

[Using connectors and connections](#)

[AWS Glue Documentation home](#)

Examples: [AWS Glue blog posts](#)[AWS Glue on GitHub](#)

What's new in Glue

Announcing support of AWS Glue Data Catalog views with AWS Glue 5.0

Mar 14, 2025

AWS Glue now available in two new regions

Mar 12, 2025

AWS Glue expands connectivity to 14 native connectors for applications

Jan 31, 2025

View more what's new in Glue

Data integration and management

Monitor & debug ETL jobs and track usage

Go to job run monitoring

Connect to your data stores

Go to connections

Orchestrate jobs to build data pipelines

Go to workflows

Manage and protect data

Create a job to evaluate data quality

Create a job to detect sensitive data

Visually clean and prepare a dataset for analysis

Go to AWS Glue DataBrew

Administration

Go to data catalog settings

Go to security configurations

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

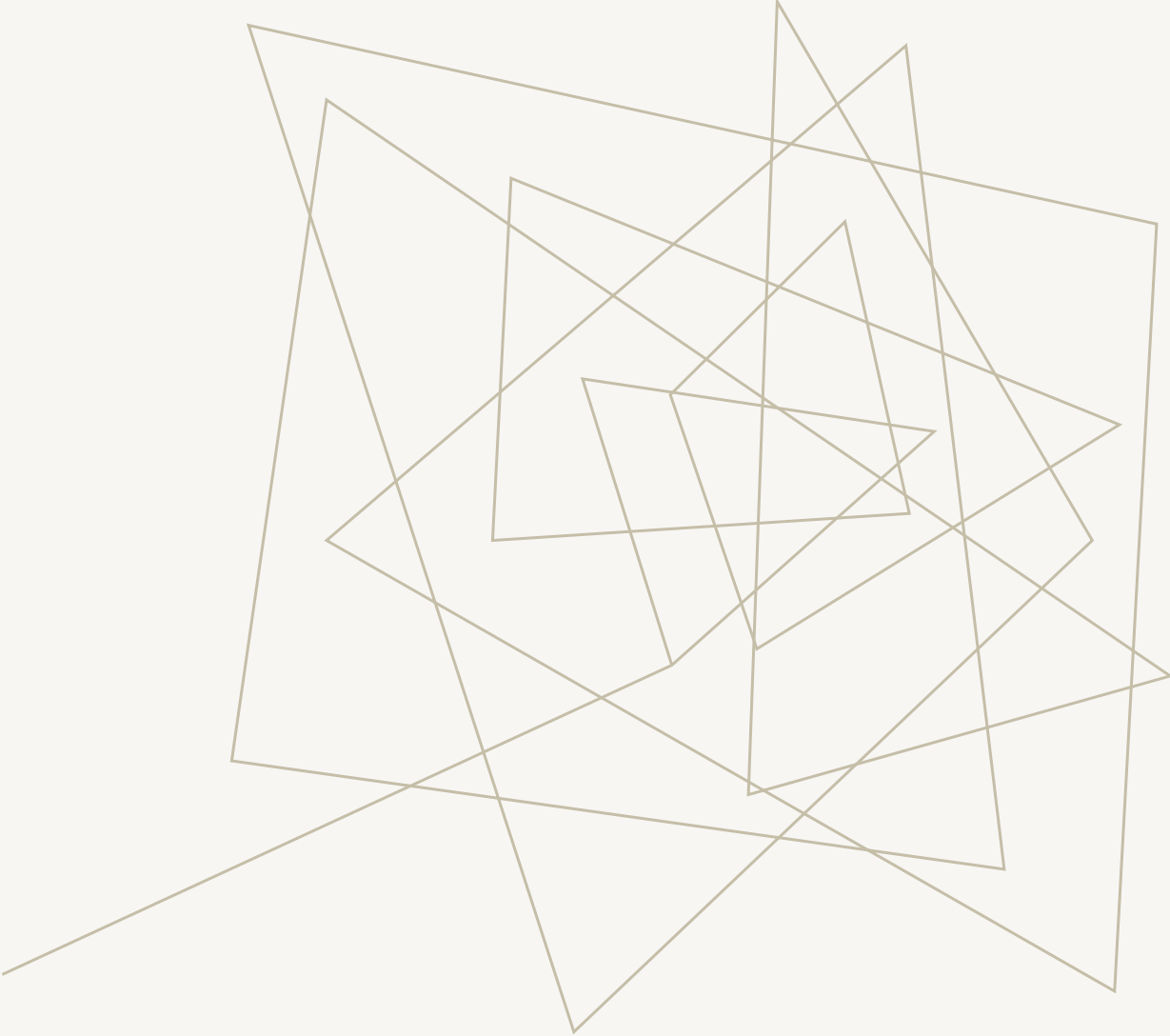
Cookie preferences

AWS Glue | eu-west-1

OBS 31.0.2 - Profile...

14:51

23/03/2025



COST DRIVER #1: OPEN SESSIONS

COST DRIVER: ACTIVE SESSION

Data preview (200) [Info](#) READY i ↺ End session Previewing 20 of 20 fields

⚙

id	post_date	post_date_gmt	post_title
----	-----------	---------------	------------

COST DRIVER: ACTIVE SESSION

Data preview (200) [Info](#) READY [i](#) [↺](#) **End session** [Previewing 20 of 20 fields](#)

[⚙️](#)

id	post_date	post_date_gmt	post_title
----	-----------	---------------	------------



COST DRIVER #2: MULTIPLE SESSIONS

COST DRIVER: MULTIPLE SESSIONS

ONE TAB

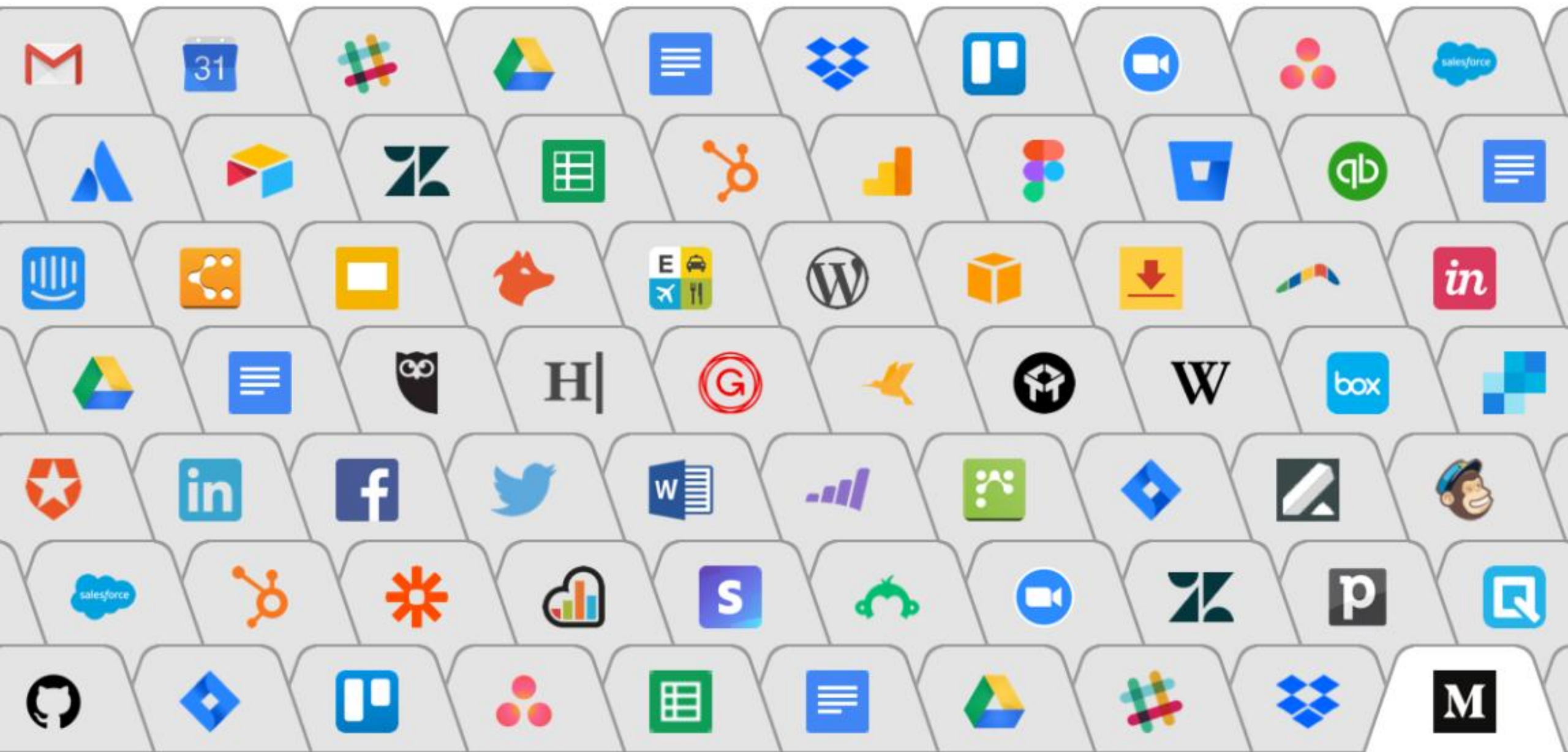


ONE SESSION



ONE BILLABLE
RESOURCE



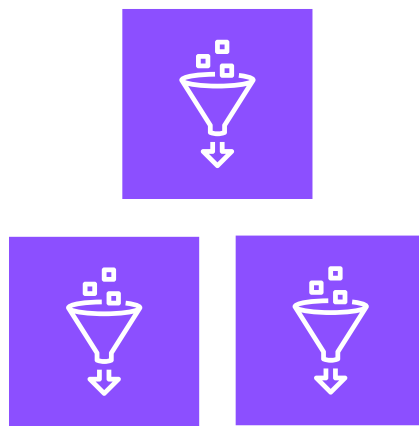


COST DRIVER: MULTIPLE SESSIONS

MANY TABS



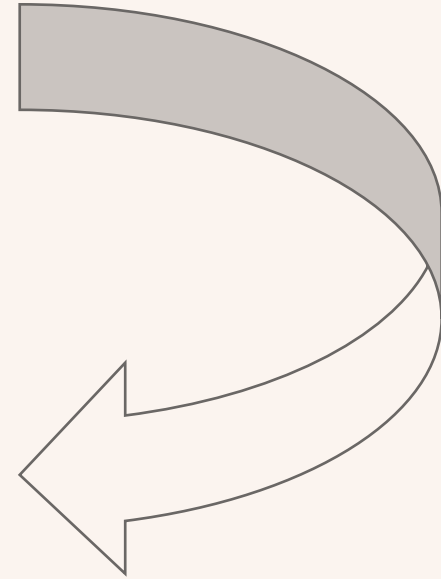
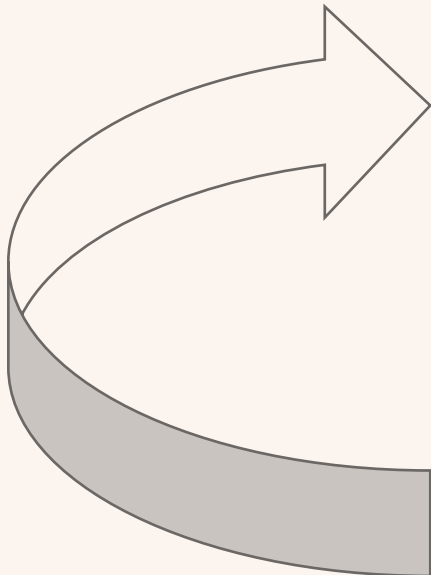
MANY SESSIONS



MANY BILLABLE
RESOURCES!



THE CAUSE



WHAT DID I DO?





OPTION ONE:

PARAMETER ADJUSTMENT

Change the timeout value manually.

CreateSession

PDF

Creates a new session.

Request Syntax

```
{
  "Command": {
    "Name": "string",
    "PythonVersion": "string"
  },
  "Connections": {
    "Connections": [ "string" ]
  },
  "DefaultArguments": {
    "string" : "string"
  },
  "Description": "string",
  "GlueVersion": "string",
  "Id": "string",
  "IdleTimeout": number,
  "MaxCapacity": number,
```

IdleTimeout

The number of minutes when idle before session times out. Default for Spark ETL jobs is value of Timeout. Consult the documentation for other job types.

Type: Integer

Valid Range: Minimum value of 1.

Required: No



OPTION TWO:

DENY PERMISSIONS

Stop anyone from starting a session.

PERMISSION BLOCKING

Start a data preview session

IAM role

To start a data preview session, choose an IAM role for this job. Changing the role will end an existing data preview session.

[Create IAM role.](#) 

Developing and testing AWS Glue job scripts locally

↓ PDF

↓ RSS

☐ Focus mode

► On this page

When you develop and test your AWS Glue for Spark job scripts, there are multiple available options:

- AWS Glue Studio console
 - Visual editor
 - Script editor
 - AWS Glue Studio notebook
- Interactive sessions
 - Jupyter notebook
- Docker image
 - Local development
 - Remote development
- AWS Glue Studio ETL library
 - Local development

Developing and testing AWS Glue job scripts locally

[PDF](#)

[RSS](#)

☐ Focus mode

► On this page

When you develop and test your AWS Glue for Spark job scripts, there are multiple available options:

- AWS Glue Studio console
 - Visual editor
 - Script editor
 - AWS Glue Studio notebook
- Interactive sessions
 - Jupyter notebook
- Docker image
 - Local development
 - Remote development
- AWS Glue Studio ETL library
 - Local development



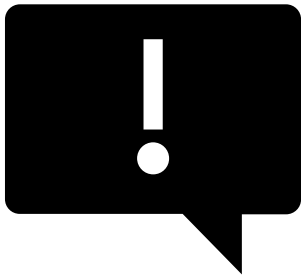
OPTION THREE:

EVENT-BASED AUTOMATION

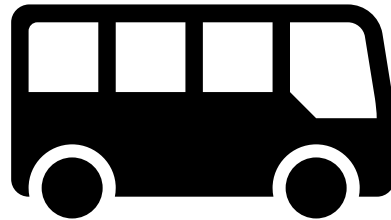
Automate all the things.

EVENT-BASED AUTOMATION

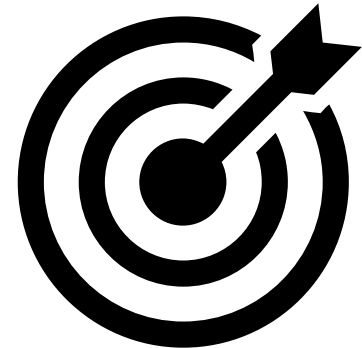
SOURCE



BUS



TARGET



EVENT-BASED AUTOMATION

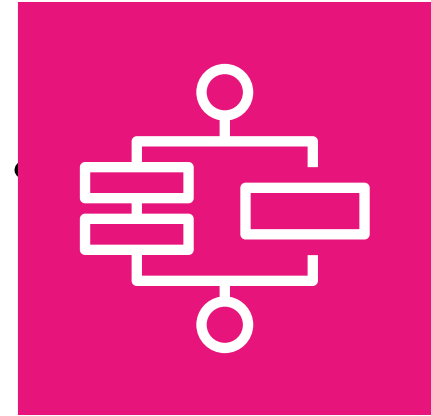
AWS GLUE



AMAZON
EVENTBRIDGE



AWS STEP
FUNCTIONS

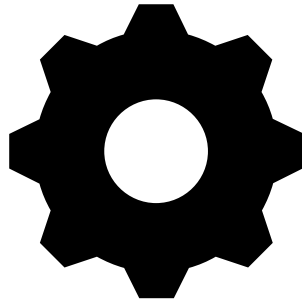


EVENT-BASED AUTOMATION

FAST



RESILIENT



SCALABILITY





EVENT-BASED & EVENT-DRIVEN

JAMES EASTHAM

@SERVERLESSJAMES

Developer Advocate @ Datadog

Ex AWS



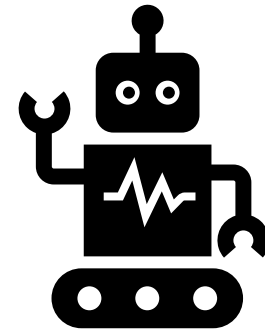
EVENT-DRIVEN SYSTEM

BUSINESS FOCUSED



EVENT-BASED SYSTEM

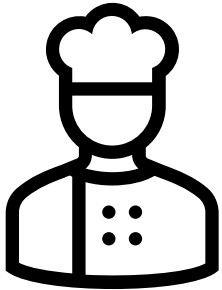
TECHNICAL FOCUSED



EVENT-DRIVEN SYSTEMS

BUSINESS EVENTS

PIZZA COOKED



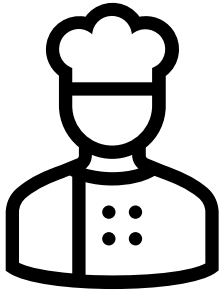
REVIEW LEFT



EVENT-DRIVEN SYSTEMS

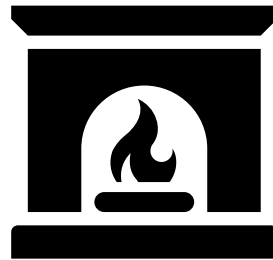
BUSINESS EVENTS

PIZZA COOKED



+

BIGGER OVEN



=

MORE PIZZAS

EVENT-DRIVEN SYSTEMS

BUSINESS EVENTS

REVIEW LEFT

CHEAPER REVIEW
PLATFORM



FEEDBACK COSTS
REDUCED

EVENT-BASED SYSTEMS

TECHNICAL EVENTS

SUCCESSFUL LOGIN



RELATIONAL
DATABASE WRITE

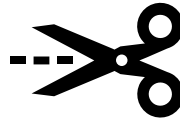


EVENT-BASED SYSTEMS

TECHNICAL EVENTS

LOGIN SUCCESSFUL

REMOVE LOGIN
REQUIREMENT



EASIER TO REVIEW

EVENT-BASED SYSTEMS

TECHNICAL EVENTS

RELATIONAL
DATABASE WRITE

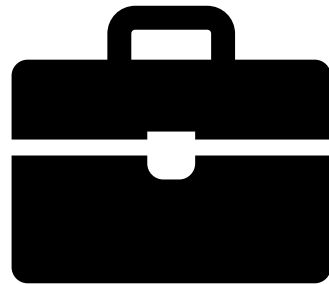
REARCHITECT
DATABASE



LESS DEVELOPER
OVERHEAD

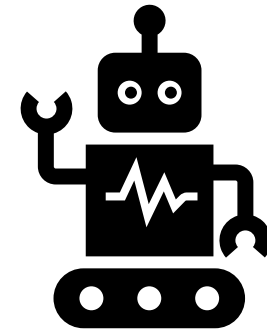
EVENT-DRIVEN SYSTEM

LANGUAGE OF
BUSINESS



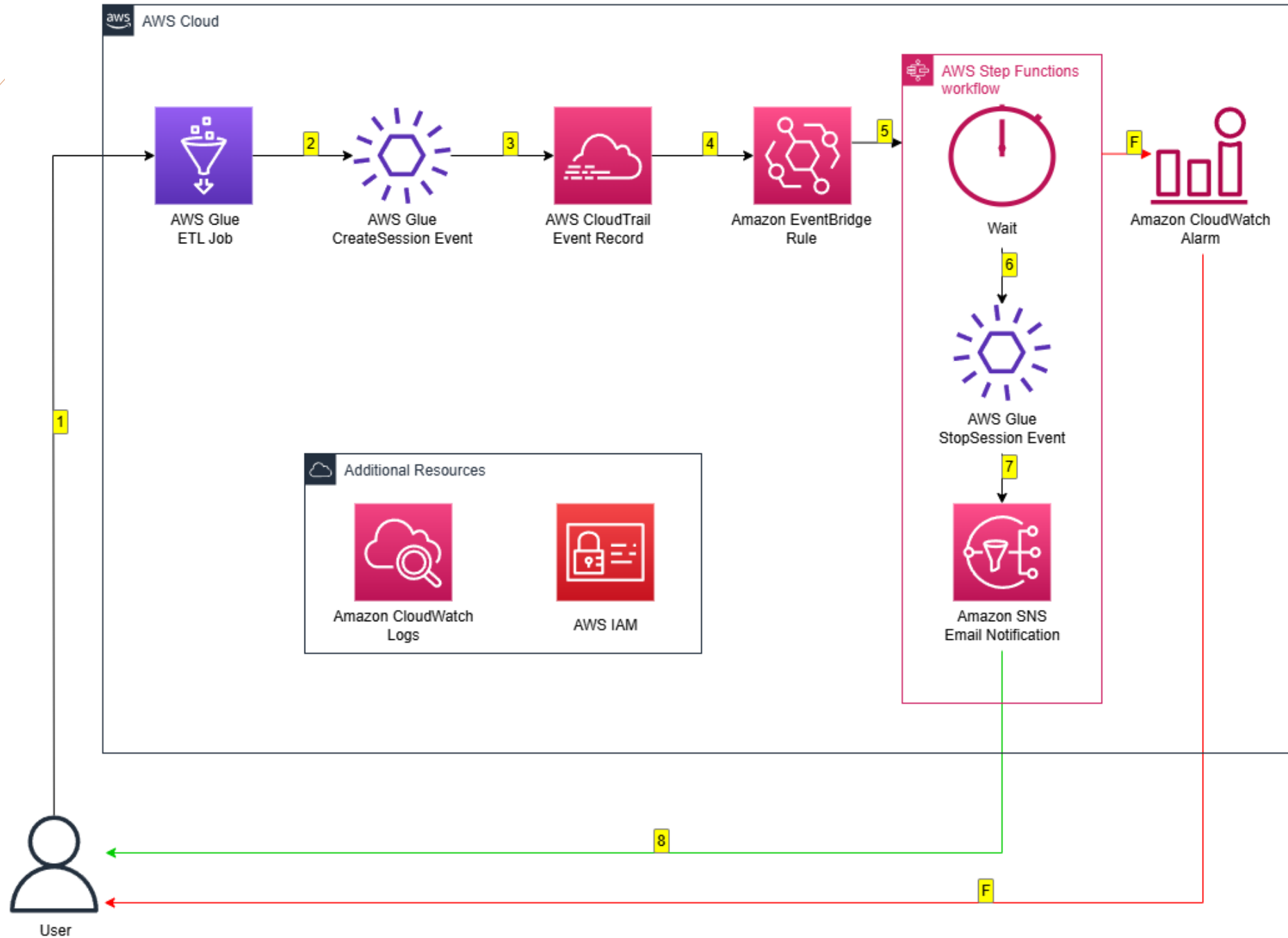
EVENT-BASED SYSTEM

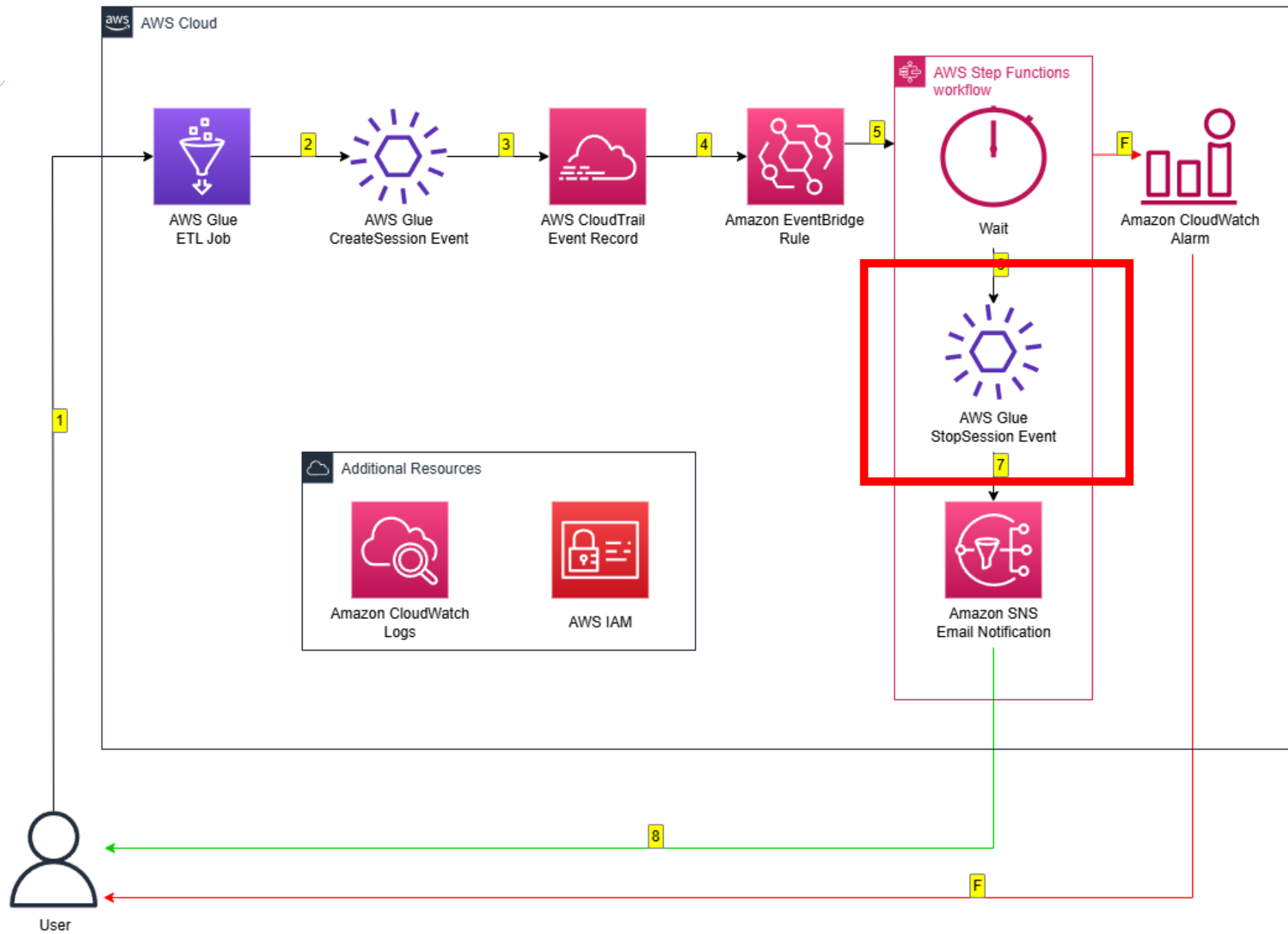
LANGUAGE OF
TECHNOLOGY



HOW DID I
CREATE IT?







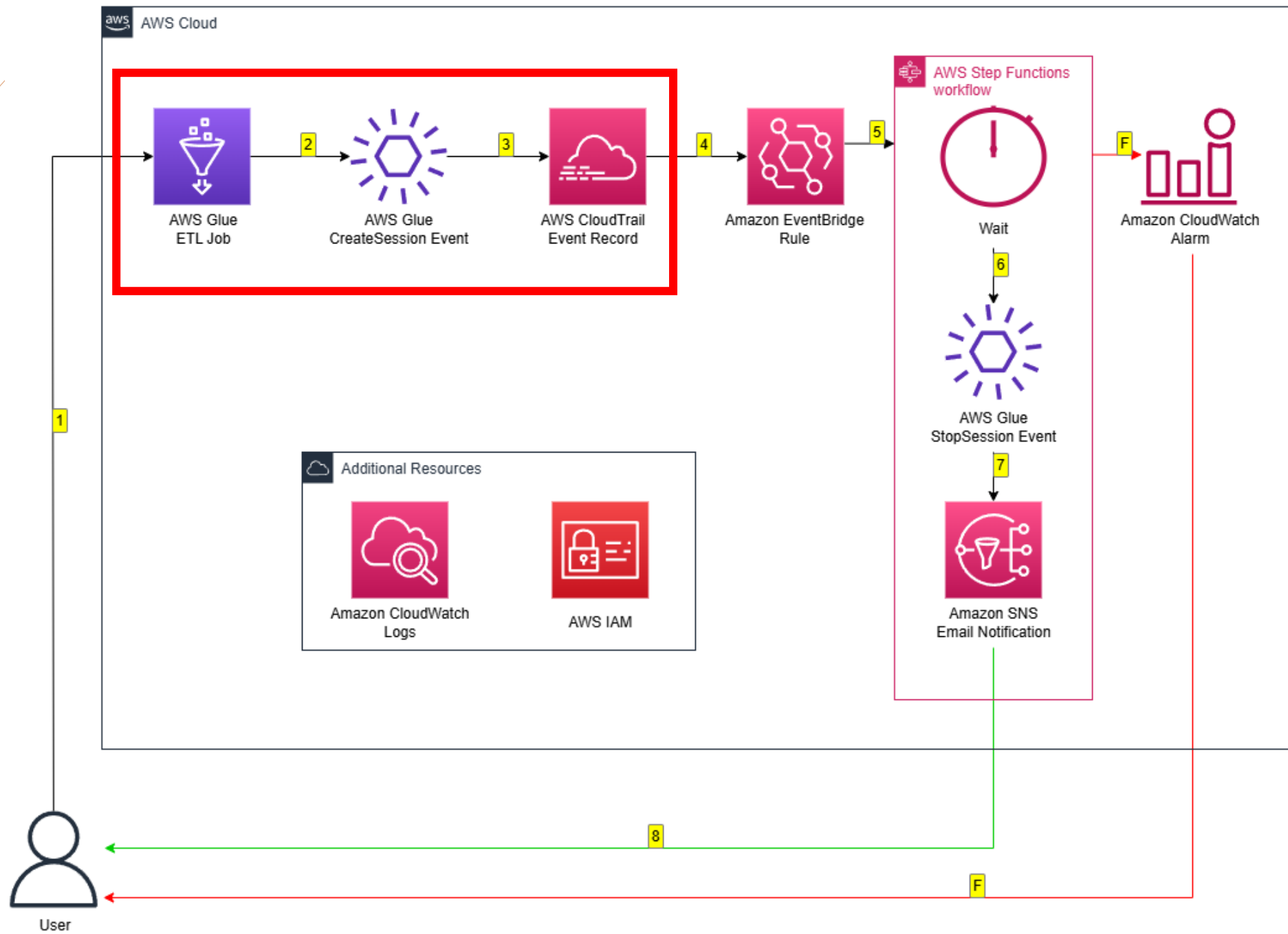
StopSession

PDF

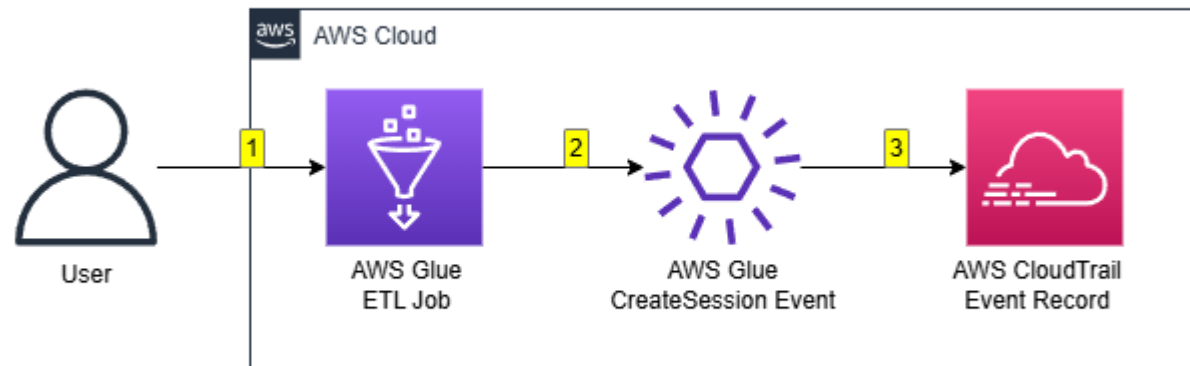
Stops the session.

Request Syntax

```
{  
  "Id": "string",  
  "RequestOrigin": "string"  
}
```



ARCHITECTURE DIAGRAM 1-3

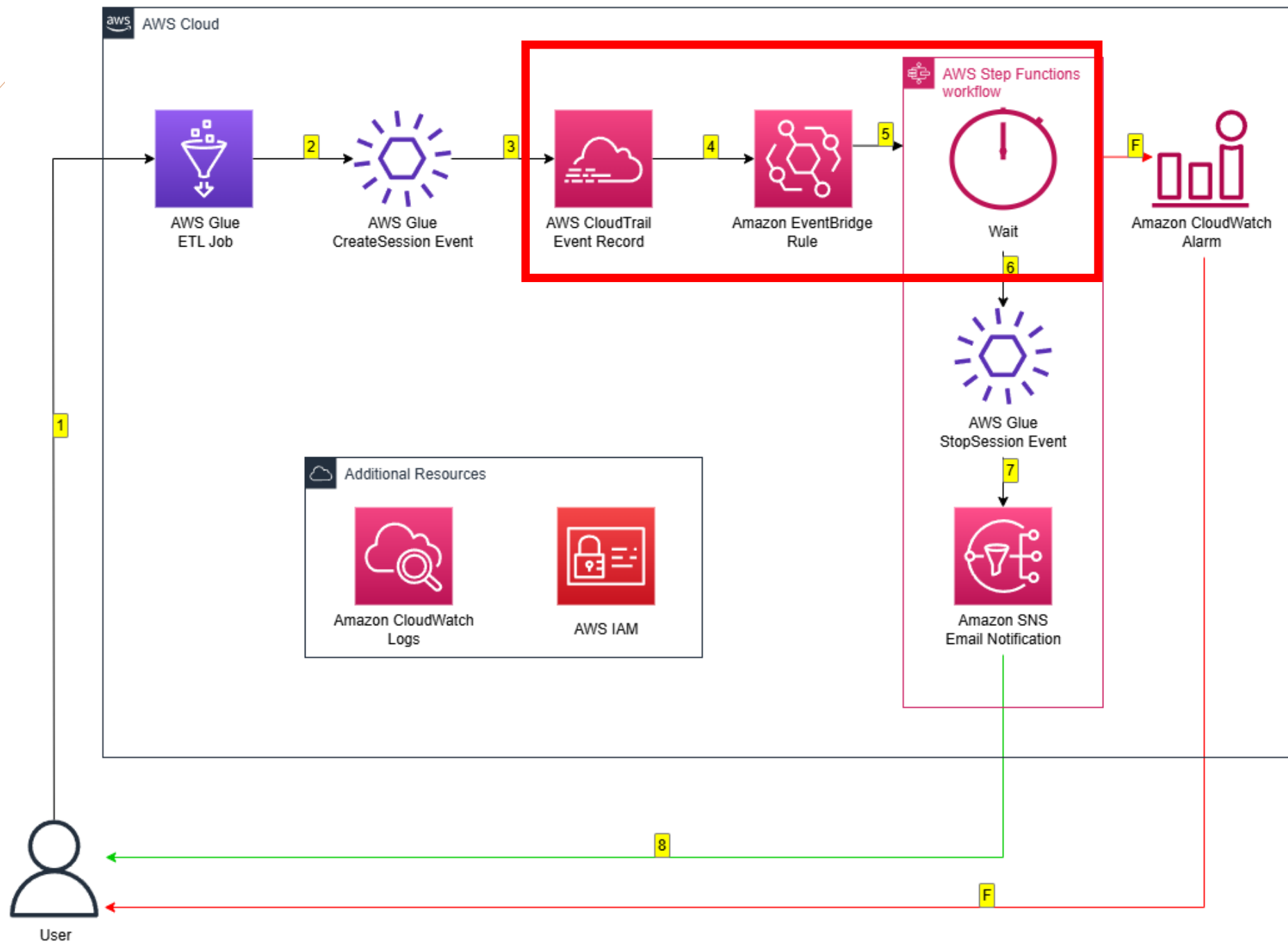


GLUE CREATESSESSION EVENT

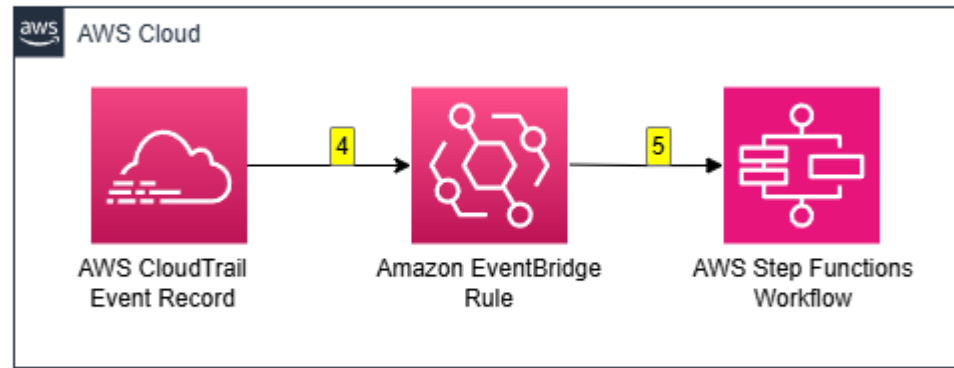
```
{
  "eventSource": "glue.amazonaws.com",
  "eventName": "CreateSession",
  "requestParameters": {
    "id": "glue-studio-datapreview-e09f88a9-4d7f-4e64-95f2-e435fbd1963a",
    "command": {
      "name": "glueetl",
      "pythonVersion": "3"
    },
    "idleTimeout": 30,
    "maxCapacity": 2,
    "glueVersion": "4.0",
    "requestOrigin": "GlueStudioDataPreview"
  },
}
```

GLUE CREATESSESSION EVENT

```
{
  "eventSource": "glue.amazonaws.com",
  "eventName": "CreateSession",
  "requestParameters": {
    "id": "glue-studio-datapreview-e09f88a9-4d7f-4e64-95f2-e435fbd1963a",
    "command": {
      "name": "glueetl",
      "pythonVersion": "3"
    },
    "idleTimeout": 30,
    "maxCapacity": 2,
    "glueVersion": "4.0",
    "requestOrigin": "GlueStudioDataPreview"
  },
}
```



ARCHITECTURE DIAGRAM 3-5



EVENTBRIDGE RULE – EVENT PATTERN

Event pattern [Info](#)

```
1 {  
2   "source": ["aws.glue"],  
3   "detail-type": ["AWS API Call via CloudTrail"],  
4   "detail": {  
5     "eventSource": ["glue.amazonaws.com"],  
6     "eventName": ["CreateSession"]  
7   }  
8 }
```

EVENTBRIDGE RULE – EVENT PATTERN

Targets		
Details	Target Name	Type
▼	GlueSession-WaitAndStop ↗	Step Functions state machine
Input to target:		Input transformer View transformer

EVENTBRIDGE RULE – INPUT TRANSFORMER

Input transformer ×

Input path

```
1 {  
2   "id": "$.detail.requestParameters.id"  
3 }
```

Input template

```
1 {  
2   "session_id": "<id>"  
3 }
```

Input transformer



Input path

```
1 {  
2   "id": "$.detail.requestParameters.id"  
3 }
```



```
{  
  "eventSource": "glue.amazonaws.com",  
  "eventName": "CreateSession",  
  "requestParameters": {  
    "id": "glue-studio-datapreview-e09f88a9-4d7f-4e64-95f2-e435fbd1963a",  
    "command": {  
      "name": "glueetl",  
      "pythonVersion": "3"  
    },  
    "idleTimeout": 30,  
    "maxCapacity": 2,  
    "glueVersion": "4.0",  
    "requestOrigin": "GlueStudioDataPreview"  
  },  
}
```

EVENTBRIDGE RULE – INPUT TRANSFORMER

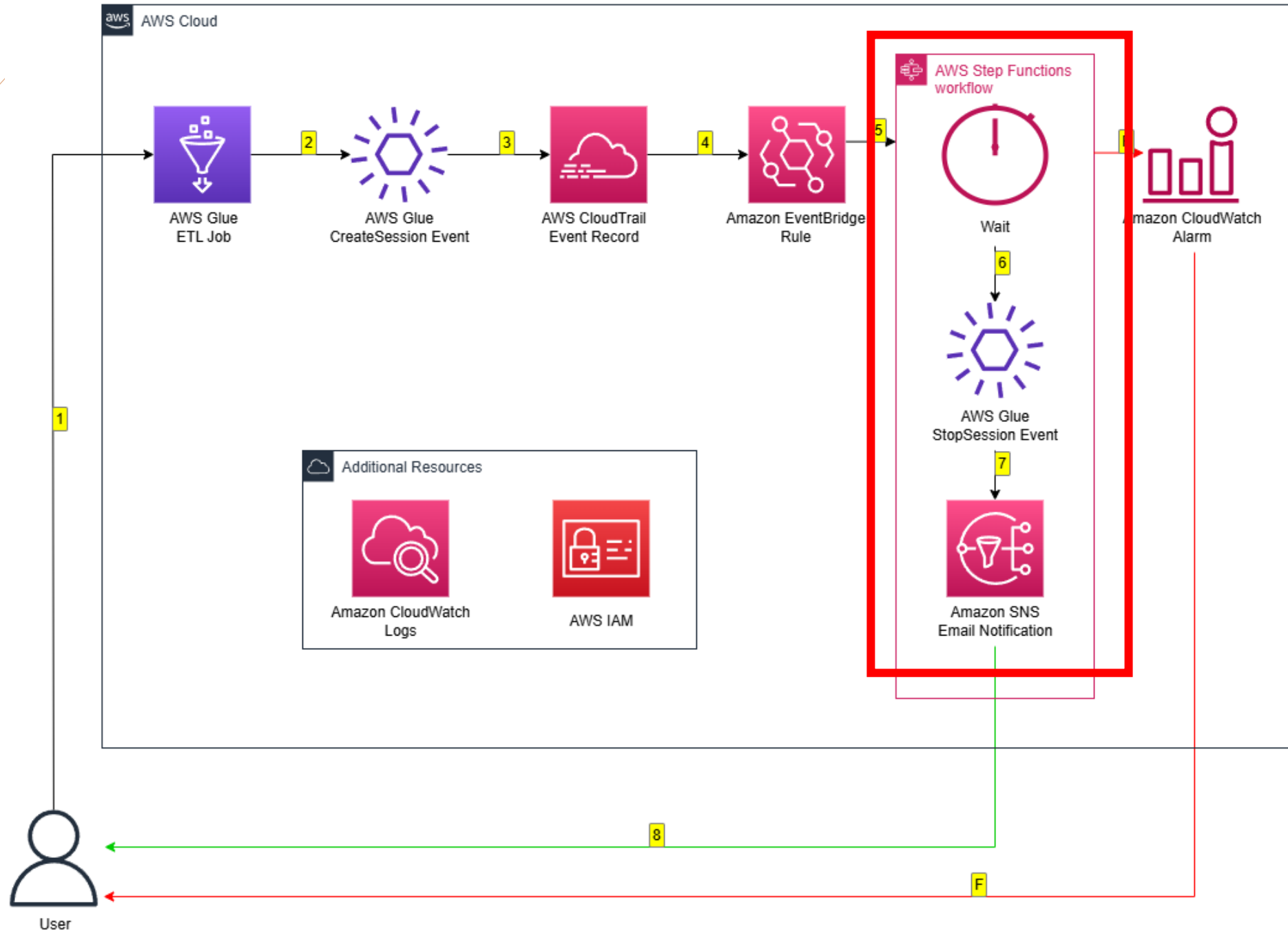
Input transformer ×

Input path

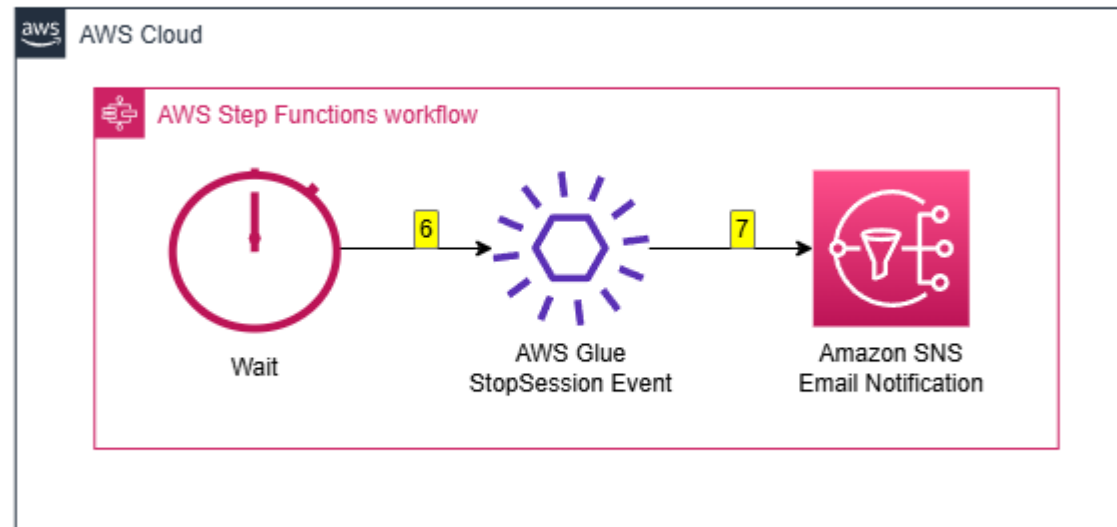
```
1 {  
2   "id": "$.detail.requestParameters.id"  
3 }
```

Input template

```
1 {  
2   "session_id": "<id>"  
3 }
```



ARCHITECTURE DIAGRAM STEP F



WAIT



```
"Wait": {  
  "Type": "Wait",  
  "Seconds": 30,  
  "Next": "StopGlueSession"  
}
```



```
"StopGlueSession": {  
  "Type": "Task",  
  "Resource": "arn:aws:states:::aws-sdk:glue:stopSession",  
  "Parameters": {  
    "Id.$": "$.session_id"  
  },  
  "Next": "SNS Publish",  
  "Retry": [  
    {  
      "ErrorEquals": [  
        "States.ALL"  
      ],  
      "IntervalSeconds": 10,  
      "MaxAttempts": 3  
    }  
  ]  
}
```

```
"StopGlueSession": {
  "Type": "Task",
  "Resource": "arn:aws:states:::aws-sdk:glue:stopSession",
  "Parameters": {
    "Id.$": "$.session_id"
  },
  "Next": "SNS Publish",
  "Retry": [
    {
      "ErrorEquals": [
        "States.ALL"
      ],
      "IntervalSeconds": 10,
      "MaxAttempts": 3
    }
  ]
}
```

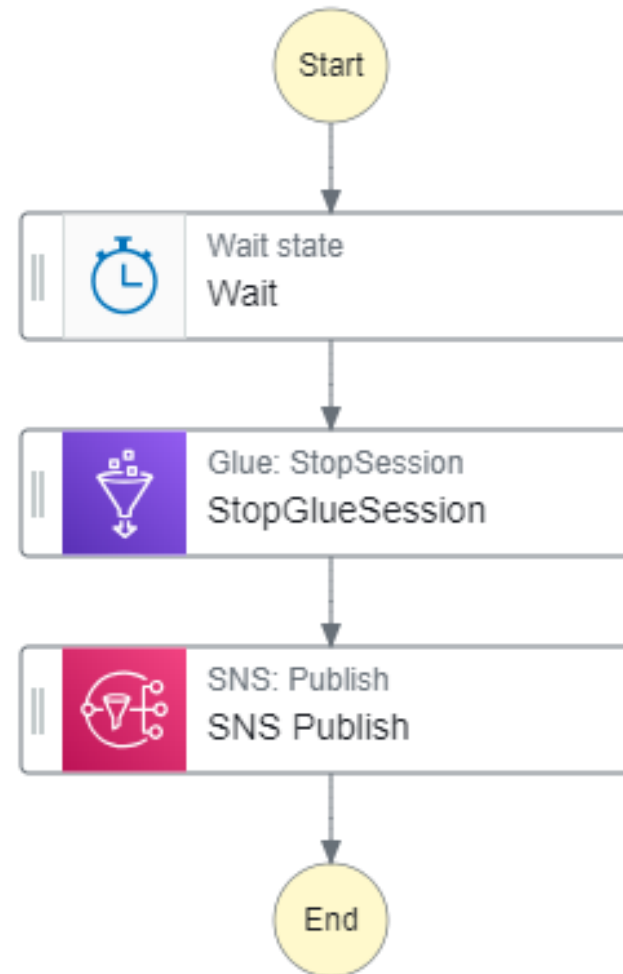


```
"StopGlueSession": {
  "Type": "Task",
  "Resource": "arn:aws:states:::aws-sdk:glue:stopSession",
  "Parameters": {
    "Id.$": "$.session_id"
  },
  "Next": "SNS Publish",
  "Retry": [
    {
      "ErrorEquals": [
        "States.ALL"
      ],
      "IntervalSeconds": 10,
      "MaxAttempts": 3
    }
  ]
}
```

SNS

```
"SNS Publish": {  
  "Type": "Task",  
  "Resource": "arn:aws:states:::sns:publish",  
  "Parameters": {  
    "TopicArn": "arn:aws:sns:REDACTED:REDACTED:success-stepfunction",  
    "Message.$": "States.Format('Hi! AWS Step Functions has stopped this Glue session for you: {}', $)"  
  },  
  "End": true  
}
```

STEP FUNCTIONS STUDIO





DEMO: GLUE

Jobs - AWS Glue Studio

eu-west-1

aws

Search

[Alt+S]

AWS Glue

> Jobs

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Zero-ETL integrations

New

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Zero-ETL integrations

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Interactive Sessions

Data classification tools

Sensitive data detection

Record Matching

Triggers

Workflows (orchestration)

Blueprints

Security configurations

Cost management

New

Legacy pages

Connections (legacy)

Jobs (legacy)

AWS Glue Studio

Info

Create job

Info

Author in a visual interface focused on data flow.

Visual ETL

Author using an interactive code notebook.

Notebook

Author code with a script editor.

Script editor

▶ Example jobs

Info

Create example job

Your jobs (6)

Info

Filter jobs by property

Job name

▼

WordPress_Gold_statisticspagespostsjoin_botodelete

WordPress_Gold_statisticspagespostsjoin_mod

WordPress_Gold_statisticspagespostsjoin

wordpressapi_silver

bronze-statistics_pages_crawler

bronze-statistics_pages

Type

Glue ETL

Glue ETL

Glue ETL

Python shell

Glue ETL

Glue ETL

Created by

Script

Script

Visual

Script

Visual

Visual

Last modified

28/10/2024, 22:29:26

28/10/2024, 17:51:15

23/08/2024, 11:04:05

09/08/2024, 13:30:58

12/03/2024, 15:30:57

12/03/2024, 13:09:19

AWS Glue version

4.0

4.0

4.0

4.0

4.0

4.0

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

15:07

23/03/2025



DEMO: STEP FUNCTION

Step Functions <

- State machines
- Activities
- ▼ Developer resources
 - Online learning workshop
 - Local Development
 - Data flow simulator
 - Feature spotlight
 - Documentation

Join our feedback panel

GlueSession-WaitAndStop

Edit Actions Start execution

Details

Arn

arn:aws:states:eu-west-1:973122011240:stateMachine:GlueSession-WaitAndStop

IAM role ARN

arn:aws:iam::973122011240:role/service-role/StepFunctions-GlueSession-WaitAndStop-role-b8l7sy5xc

Type

Standard

Status

Active

Creation date

Jun 5, 2024, 19:31:49 (UTC+01:00)

X-Ray tracing

Disabled

Executions Monitoring Logging Definition Aliases Versions Tags

Executions (0/0)

View details Stop execution Redrive Start execution

Filter executions by property or value All Last 15 months Local timezone 0 matches

Name Status Start Time (local) End Time (local) Duration Info Version Alias

No executions

Start execution

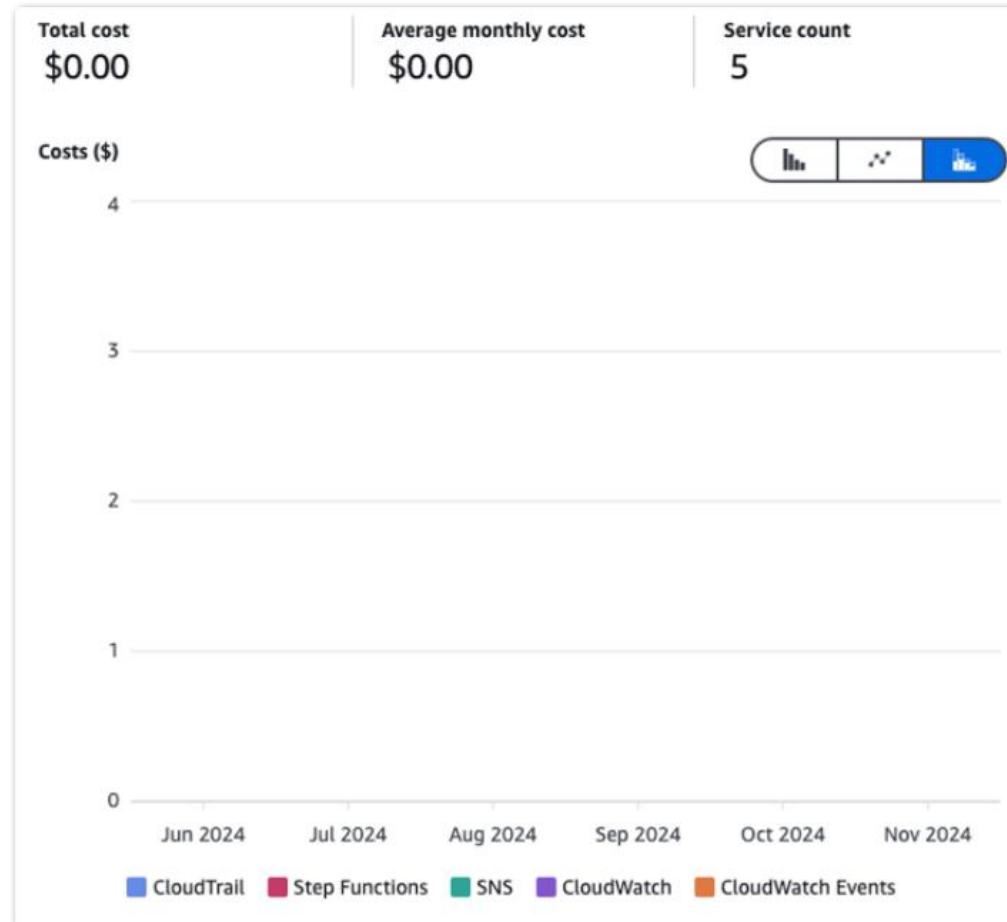


OUTCOMES

GLUE INTERACTIVE SESSIONS 2024 H2



ARCHITECTURE COSTS 2024 H2

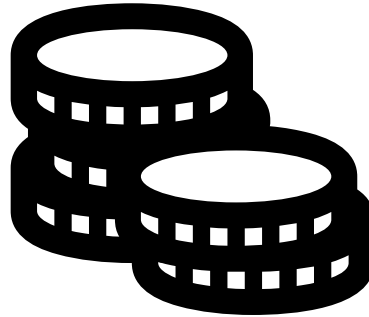


LIMITATIONS

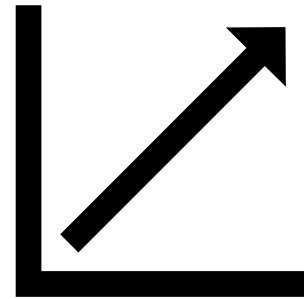
REACTIVE SOLUTION



STILL SOME GLUE
COST

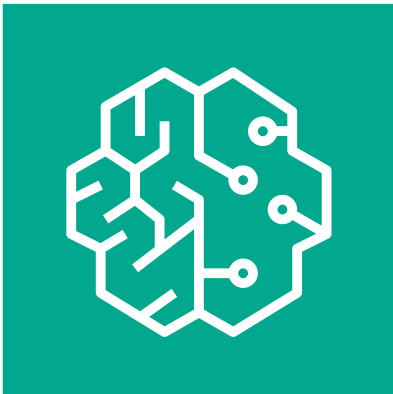


POTENTIAL
WORKFLOW COST

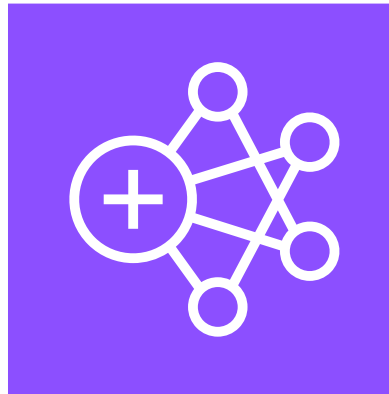


OTHER APPLICATIONS: STOP OTHER SERVICES

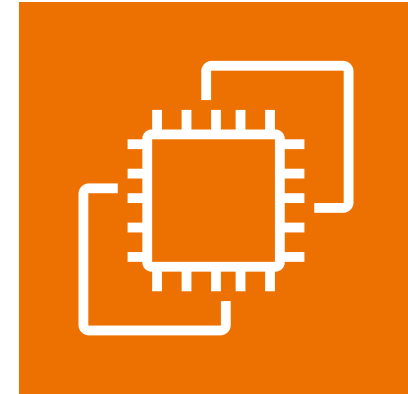
AMAZON
SAGEMAKER



AMAZON EMR



AMAZON EC2

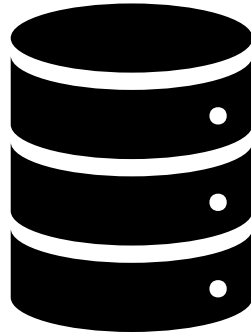


OTHER APPLICATIONS: PERFORM OTHER ACTIONS

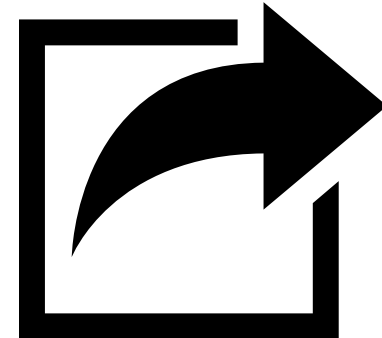
EMAIL & SLACK
NOTIFICATIONS



BESPOKE LOGGING



TRIGGER SEPARATE
WORKFLOW

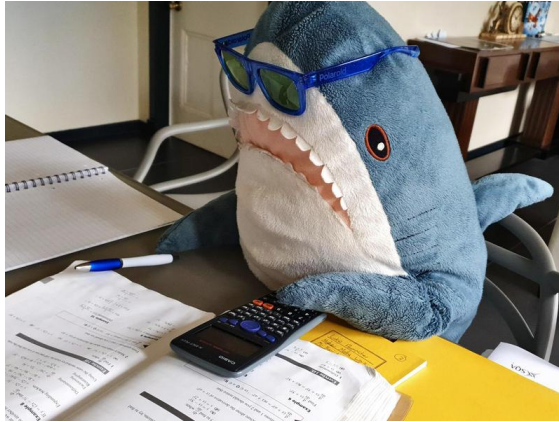




SUMMARY

AGENDA

FIND THE CAUSE



CONSIDER YOUR
OPTIONS



BUILD A SOLUTION





Me, Session Resources

THANK YOU



AWS User Group Lancashire