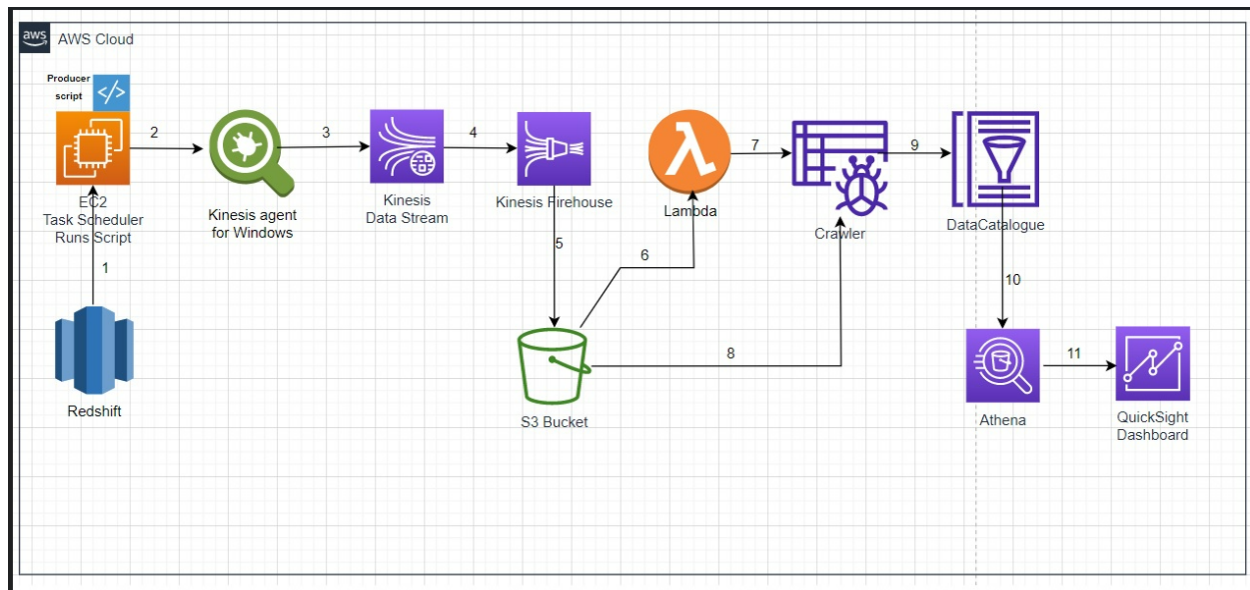


LIVE DATA STREAM WITH KINESIS FROM REDSHIFT AND DELIVERY TO QUICKSIGHT DASHBOARD

Purpose: Design an aws cloud based data pipeline that will read a table from redshift database, simulate streaming data delivery into S3 bucket, from where crawlers will read the data and create metadata table called - catalog, and finally quicksight dashboard will pick up streaming data from data catalog though direct athena queries when an update is performed.

Take a look at the following Diagram to get a high level glimpse into the expected work awaiting us:



Prerequisites: AWS Account, Installed SQL Workbench, PySpark, Python, Basic Network Configuration, json, SQL, RDC, Downloaded files from folder 'Project 4' in GitHub link provided: <https://github.com/Myself1214/Upwork.git>

Plan of Work (Pseudo Work):

1.

Actual Steps:

Create role (kinesis role) and attach it to ec2. This will allow ec2 to put records on kinesis data stream

To install kinesis agent for ms windows, go to:

<https://s3-us-west-2.amazonaws.com/kinesis-agent-windows/downloads/index.html>

Then follow instructions for 'Install using PowerShell'

Once installation is done, open notepad and create Kinesis Agent for Windows configuration file

```
Start-Service -Name AWSKinesisTap
```

```
Stop-Service -Name AWSKinesisTap
```