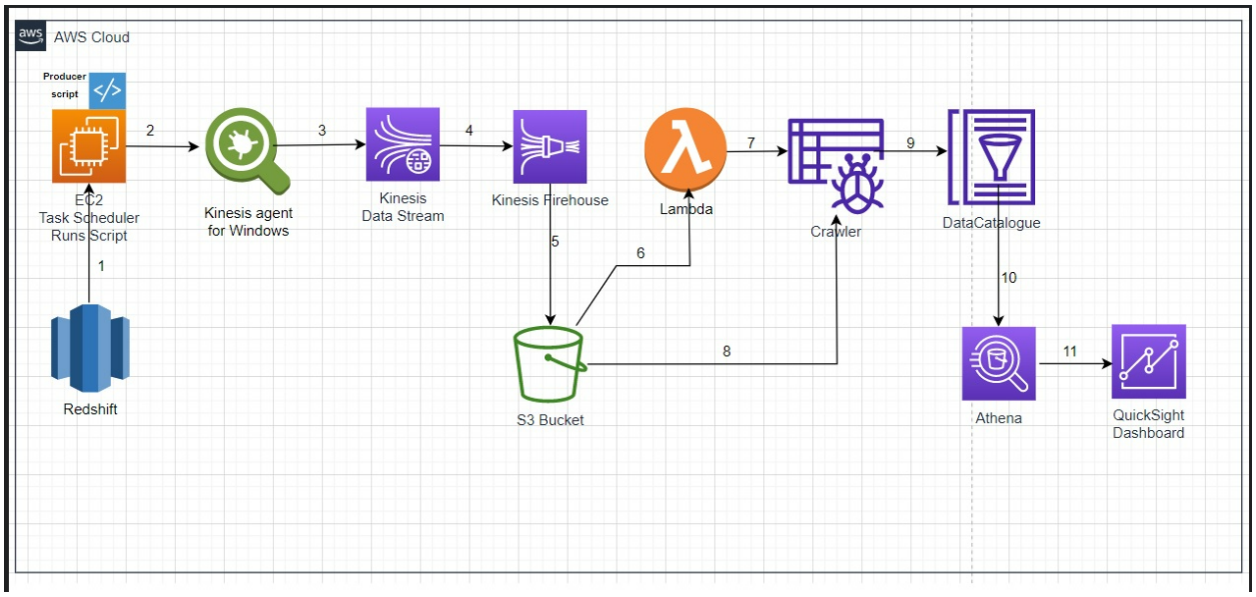
**LIVE DATA STREAM WITH KINESIS FROM REDSHIFT AND DELIVERY TO QUICKSIGHT DASHBOARD**

**Warning:** This project will pick up where Project 3 ended - Redshift table. To successfully complete this project you must have finished Project 3. This project is using EC2, Security Group, some IAM Roles, Redshift and S3 bucket that were created in Project 3.

**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, Ready-to-go Redshift cluster with database and imported table, downloaded files from folder 'Project 4' in GitHub link provided: https://github.com/Myself1214/Upwork.git

**Plan of Work (Pseudo Work)**:

1. Program a data producer and load it on EC2 through a shared path that is mounted from local to EC2. Put it on a schedule run through windows task manager
2. Create a role for producer to write data on Kinesis Data Stream and attach it to EC2
3. Create Kinesis Data Stream
4. Create Kinesis Firehose
5. Create a folder in your S3 raw bucket and name it 'from-firehouse'
6. Install and configure Kinesis Agent for Windows on your EC2 to stream data from EC2 server into S3 bucket in 'from-firehouse' folder. Run the agent
7. Create crawler to crawl streaming data from raw S3 bucket and create table metadata as a catalog
8. Create Lambda function to trigger crawler every time when new data is streamed and loaded in S3 raw bucket
9. Create Quicksight dashboard to be updated through direct athena queries

**Actual Steps:**

---

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

---

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