ETL Code Challenge

You have to write an AWS Glue Job(s) in order to fulfill the following scenario:

At ABC Company, each year every employee is given a percentage based increment to their salaries based on the departments they belong to. You have to write a script(s) to read from flat_data.csv and store into employees and department DB tables in the schema below. Furthermore you need to read tables from the database, calculate the updated salaries and write them back. Please note that you will need to create these tables in the PostgreSQL DDL and provide your code.

The database contains the following schemas:

- **employee:** id :: UUID, first_name::Text, last_name::Text, salary::numeric, department_id::UUID
- **department:** id::UUID, name::Text, salary_increment::numeric

The salary_increment column in **department** contains the percentage value for calculating the salary. The output of the process should be the following table

- **updated_salaries:** employee_id, updated_salary

All the tables must be created by script and have the necessary key relationships between them.

For doing this task you have two options.

- Use your own AWS account for the development.
- Setup AWS Glue Env on your local machine https://docs.aws.amazon.com/glue/latest/dg/aws-glue-programming-etl-libraries.html

Completion

When you are finished (or out of time) you will push up the application to a git repo. Then please notify Ali Usman via email (ali@pumpjackdataworks.com) with the subject Pumpjack ETL Code Challenge and the link to the github repo.

Ali is available for any questions you may have via email at ali@pumpjackdataworks.com.