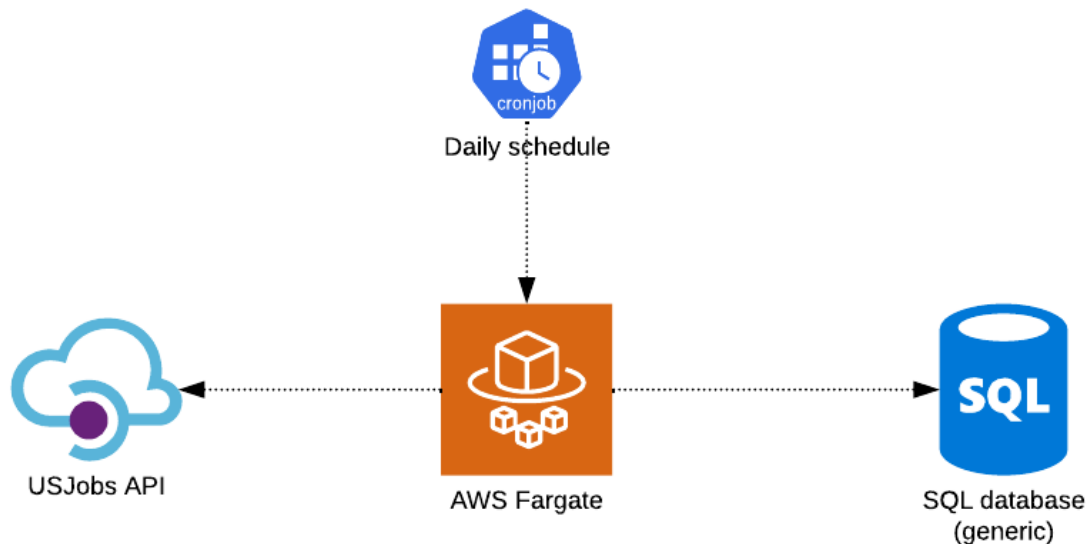


USJobs ETL Process

Solution Design Document

1. Process Description



1. The function is triggered automatically once a day based on cron schedule:

```
{  
  "Schedule": "cron(0 7 * * ? *)"
```

2. Function will execute the DDL queries if the database doesn't exist.

3. Repetitive or secure parameters will be stored on AWS SSM or in current code in params.txt file.

4. Function will call USJobs rest API for CodeList to populate lookup tables metadata. There is no information about the frequency of updates on lookup tables. But it is considered a low change table in the current process.

5. If there is any cloud storage available batch extract after processing in code will be stored in S3 bucket. In the current process data is stored in json format in memory and insert into the database will be in batch.

6. Based on the target source database different capabilities might be available. For example SnowFlake has a Copy Into capability which imports data in batch all at once in a fast and efficient process.

7. Dimjob data will be populated based on Upsert on machedID or positionID. A flag can be also added for deleted records scenarios.

8. In a better ELT process stage tables will be created to dump data into stage tables and transform will happen in DataWarehouse. For example SnowFlake can take json as a variant and the json process can be handled on SnowFlake.

Logical data model US Jobs

sepideh.hashemi | March 27, 2022

