

Data Engineering Skills Assessment

v1.1 April 2024

Privileged and Confidential

Document version control

Version	Last Modified on	Last Modified By
1.0	22 nd June 2022	Lal Hussain
1.1	11 th April 2024	Luke Lee

Contents

Privileged and Confidential	1
Document version control.....	2
Contents	3
Introduction.....	4
Background case study	4
Task Overview	4

Introduction

The purpose of this test is to assess your level of Data Engineering experience. You have 3 hours to work through the objectives. You may not complete everything, but please focus on the quality of each objective.

It is important to write clean and well-documented code so that others can continue where you left off. Where you have incomplete features, try to include notes about your intended direction so that any co-worker can easily understand how to continue.

It is recommended that you reserve the last few minutes of time to write code docs for your teammates.

Background case study

Alchemy inc. sells desirable used electronic devices. They wish to analyse sales data and monitor their revenue/units sold by customer and by month.

Task Overview

- Create a GitHub repo
- Devise and document a database schema to hold the data, to be applied via a script
- Write a script for loading the CSV files into an S3 datalake
- Write a script for importing the data into a data warehouse for
- Write queries to answer the questions Alchemy inc is asking
- Readme.md documenting
 - how to run your scripts
 - how the scripts would be scheduled and run on AWS
- Share a link to your repo

Notes on completing these tasks;

- There is no right or wrong way to do this. We're more interested in the process and methodology.
- Make sure that your code is executable.
- All input storage and output should be in UTF-8.
- You might find it easier to work with a subset of data when developing.
- Reporting needs to be in USD\$.

You have 3 hours to complete this assessment.

Don't forget that you'll be handing over unfinished work to a co-worker.

Good luck!