

PrimaryBid - Data Engineer Tech Assessment

The technical assessment for a data engineer consists of two parts. Details below:

Part 1

This part of the case study is designed to test out the candidate's core programming and data manipulation skills. Details of the case below:

With reference to the input file "CC Application Lifecycle.csv", please manipulate the data to return various Application stages as column names, and the time of stage completion, as values against each customer ID, as indicated in the work-in-progress output file "Application Lifecycle Output.csv". You need to do this data wrangling task using Python

Part 2

This part of the case study is designed to test out the candidate's ability to architect and design a technical solution. Details of the case below:

The data team now needs to build a pipeline to ingest data from a MongoDB database - including 5 collections) - into a Redshift data platform. The team is looking to include more data sources in the near future. How would you approach the problem and design the framework code in a way that is reusable and easy to maintain and change? Please include any supporting materials that you think are helpful including but not limited to the following:

1. A public Github repo
2. A high-level specification document
3. QA and testing considerations
4. Any other artefacts that you think are relevant

You won't need to go into a great level of detail in the coding. High-level class objects should do as long as you think that's sufficient to tell the story. When you return your work, please make sure to include the URL to the Github repo and any other artefacts in an accessible format (e.g. word, PDF etc.)

