Data Engineering assessment

Dear Candidate,

Thank you for applying to the position of Data Engineer at nPlan. We are thrilled to learn more about what you can do. For this reason, we would appreciate it if you could complete the following task. The purpose of this is for us to assess how you would go about creating a pipeline to process some data and perform some straightforward analysis upon it, using a streaming data source.

The data source you will be using is stock market streaming data from IEX Cloud. You will need to register an account, which only requires an email address (make sure you choose the "Start" plan). To access their streaming API, we will need to utilise their sandbox mode which contains artificial data and limitless access. To enable this, ensure that you have "Sandbox View" turned on in the console, and that the API token you are using starts with `Tpk_`. See the sandbox documentation for more details.

We would like you to produce a stream of the top 3 highest-priced stocks averaged over the last 1, 5, and 15 minutes, with stream elements being produced at least once per 10 seconds. We suggest you start by looking at the Quote endpoint, but you can use any endpoint or approach you prefer. Since there are many companies in the streaming data, we would recommend choosing a sample, for example use only 10 companies for this exercise. We would also like to see at least one other streaming analysis or output that you think might be interesting (using any IEX Cloud data you like).

Please use Python to implement your solution. We use Google Cloud tools such as Dataflow / Apache Beam, PubSub, and BigQuery, but you can implement your pipeline using whichever tools or frameworks you prefer.

Aside from your code and its requirements (if you configure any external services, please detail that as well), include a README describing what you have done and your reflections on the solution and approach. The next stage of the process is a technical call where we will go into more depth regarding your processes and thoughts during this exercise.

Please spend no more than four hours on the task. We ask that you return your submission within one week. If it will take longer, please let us know within that time.