Take-Home challenge

INTERNSHIP - Data/Analytics Engineer

Context

First of all, thank you for taking the time to apply at Astrafy and be willing to complete this technical take-home challenge. We are well aware that your time is precious and are grateful that you are willing to take the time to demonstrate to us your skills via this challenge.

The challenge is made up of a design challenge and a coding challenge. Once you submit your solution to those challenges, we will review those upfront and then set a meeting for you to go through your solution and explain the process that led you to this design and code architecture.

Generic Guidelines

There is not a perfect solution for each of those challenges and so is the case for the projects you would be working on at Astrafy. Problems can be tackled in many different ways and we are as much interested in the thought process that led you to those designs and code than in the final results ("the journey matters more than the destination").

Some guidelines we highly recommend you to follow:

- **Documentation is as important as the code**: we should be able to understand everything you did by reading the README, Docstrings in your code, eventual short tutorials (using Loom for instance), etc.
- **Do not reinvent the wheel**: some say that savvy developers excel at "copy-pasting". Chances are that a library or something similar to what you are trying to do has already been done. A developer's job is sometimes putting together different lego blocks and making it work all together in a seamless way. We therefore encourage you to look on the web and take inspiration from similar problems.
- **Automation is key**: We are all about automation. While the challenges do not put constraints on the devops/dataops side, we highly encourage you to:
  - to use CI/CD
  - to deploy eventual Cloud resources via IaC (using terraform)

©Astrafy 2023 Proprietary and Confidential. All Rights Reserved.

Submission

Once you are done with the challenge, please follow those guidelines:

1. For the design challenge, save your deliverables as pdf.
2. For the coding challenge, put your code on a public Github repository.
3. Reply to the email you received with this "Take-Home challenge" with the pdf from step 1 attached and a link to the Github repository (step 2).

Part 1: Design Challenge

a. Task

There is a conversation in the office between Astrafy colleagues about a new customer project. The customer has many different sources of data ranging from transactional databases (PostgreSQL, MySQL and MongoDB) to well known applications (SAP, Salesforce and SurveyMonkey). The customer wants to revamp fully its current analytical approach and has asked Astrafy to design a potential solution. The end goal is to serve BI dashboards for various departments as well as starting to develop ML models for recommendations and predictions.

The customer has the following requirements:

- Google Cloud as hosting platform
- Open-source technologies as much as possible
- High focus on GitOps and DataOps
- **OPTIONAL** - data mesh paradigm. Their data domains in that case would be **customers**, **products** and **maisons**.

b. Deliverables

- Lucidchart architecture for the solution (you can use lucidchart free tier)
- Small slide deck (maximum 5 slides of content) explaining your design, assumptions and motivations.

No code is expected for this task.

©Astrafy 2023 Proprietary and Confidential. All Rights Reserved.

Part 2: Coding Challenge

a. Task

Bitcoin Cash is a cryptocurrency that allows more bytes to be included in each block relative to its common ancestor Bitcoin.

There is a public dataset on BigQuery that contains the blockchain data in their entirety (dataset ID: bigquery-public-data.crypto_bitcoin_cash) with data pre-processed to be human-friendly and to support common use cases such as auditing, investigating, and researching the economic and financial properties of the system.

In a new Google Cloud project you are expected to:

- **Materialise a staging table** from the raw table "transactions" that only selects the last three months of data from this raw table
- **Materialise a data mart table** that gives the current balance for all addresses and exclude addresses that had at least one transaction on Coinbase.
- From the data mart, **display a graph** and plot the amount of transactions on the Y axis with the date on the X axis.

b. Deliverables

- Create a new Google Cloud project (**optional** - use Terraform)
- Use either scheduled queries, dbt cloud or dbt core to materialise the two tables aforementioned.
- A Google colab notebook using python to plot the chart of daily transactions. **Optional** - use a python class to encapsulate the logic.

Important: By selecting only the last three months of data, your queries will not cost anything and will enter in the free-tier offered by Google Cloud on any new projects.

©Astrafy 2023 Proprietary and Confidential. All Rights Reserved.