Test task for Data Platform Engineer

- Background
- Data and questions

Background

This task should not take more than few hours to complete by someone who has good working knowledge of any data processing system. It may take slightly more if used system is new and some learning is required. In any case you should not spend more than 8 hours on this task.

Writing solid production-grade solution that will work in various conditions can be very demanding task. It's OK to cut corners in test task and point out shortcomings of offered solutions or assumptions made.

Data and questions

For this task we are using pre-generated datasets for DBT from jaffle-shop-classic (https://github.com/dbt-labs/jaffle-shop-classic/tree/main/seeds Can't find link).

Do the following:

- Register 14-day trial Pipedrive account (https://www.pipedrive.com).
- Write a small program to download dataset from Github and create or update deals in Pipedrive account through the API (https://developers.pipedrive.com/docs/api/v1)
  - Deals should have a Title of "LastName FirstName"
  - Apply simple transformation logic on each row (like multiply "value" field by 2).
    - Be creative!
  - If "value" in Pipedrive account for the deal is different from calculated new value than it should be updated in Pipedrive API.
- Think of it as it is regular daily job to keep dataset from S3 and deals in your Pipedrive account in sync.

Requirements:

- Can use any of these languages - Python/Go/NodeJS/Scala
- Don't store any secret in publicly available places (ie Github)

Additional questions:

- Think about the performance and be prepared to discuss on it.
  - What happens if there will be 100K, 1M deals to synchronise.
- How easy is it to make source pluggable (ie take file from HTTP endpoint) and destination (ie to Postgres) and how to make this data synchronisation task more generic.
- Can this be containerised and be triggered to run daily from CLI?
  - Maybe something else can be used to automate the daily process?