Dema.ai Data Engineer Hiring Test

The purpose of this hiring test is to get a better understanding of your technical skills and focus. We expect you to spend a few hours on it (not more than 5 hours) and let us know how much time it took to achieve where you end up.

Theme

At Dema, e-commerce platforms and tooling are at the core and we will build one of the functionalities around them. We will build a data ingestion pipeline that stores raw data, then transform it by combining order and inventory information.

Preface

- Create a data ingestion pipeline that ingests the two datasets, stores the raw data, then combines the data and persists it with an appropriate schema of your choice.
- We will be using two datasets:
    - *Orders Dataset*: Contains orders done on the e-commerce platform. (Download Here)
    - *Inventory Dataset*: Contains inventory information, including products description and their available quantities. (Download Here)

Requirements

- Your solution should be able to execute locally without external infrastructure dependencies and all needed infrastructure should be set up by infrastructure as code.
- All intermediate steps in your pipeline should be persisted.
- The resulting schema should be designed so that it scales for thousands of e-commerce merchants and be suitable for analytics queries.
- Your solution will be judged partly on your choice of technologies and tools so use proven industry standards that are known to scale and that make your pipeline easy to maintain.

Bonus Points

If you want to go further with this, these will be nice to haves:

- Add appropriate data validations in your pipeline.
- Write three report queries that aggregate relevant data from your small data lake.

Tips

- Feel free to use any open source tools to achieve the functionality, as long as you can motivate your choices.

Submission

Please submit your completed test as a GitHub repository.

- The repository should include a README file with instructions for running the application, and any other information you think is relevant.
- Please also include a summary on what you would have done given more time.

Please feel free to reach out at linn@dema.ai if you have any questions or concerns. Good luck and we look forward to seeing your solution!