Kovrr Home Assignment

HN Analyzer
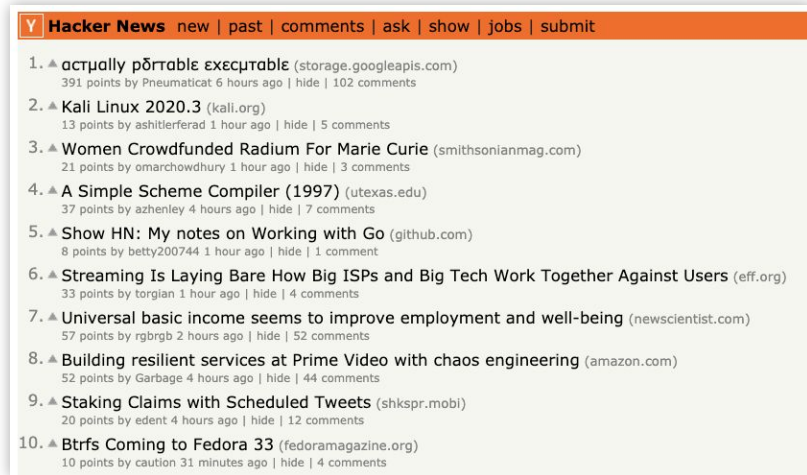
Data Engineer

Modelling Team

Introduction

Hacker News is a tech news feed that allows users to post interesting articles and rank them.

The feed is **ordered by rank** that is defined by a **ranking** algorithm which combines an article's up-vote count, submission time and gravity.

Attached to the exercise is a json file with 3 months of posts downloaded from HN feed

The Task

Build a data pipeline program that **loads** the data given, **analyzes** it and generates a **summary report** of the findings.

The report should consist the following information:

1. All posts published in a hourly granularity with the following fields aggregated in the following way:
   - mean and median score
   - sum of words in all titles
   - max id
   - sum of descendants
   This part should be given in a .csv format.

2. Create a view of the data that will help to learn as much as possible about users activity in HNews in whatever way you see fit.
   This part can be done in whatever format you see fit.

3. Explain what would you do in order to make the two tasks above to run daily on the real data from https://github.com/HackerNews/API
   This part can be done as part of the README file.

Please Pay Attention to:

- **The project should be submitted in a working state**

- Code readability, structure, and Python conventions, documentation

- Errors should be handled gracefully

- Logging

- The assignment should include a simple readme file with installation instructions (+ part 3 of the task)

- Feel free to use python libraries and open APIs as you see fit

* Attached is the json file with hacker news stories from the past several months.

KOVRR
Cyber Risk Modeling

Good Luck!