Null Innovation

Technical Test

Problem statement

Design a simple ETL process – Extract Transform and Load

What is ETL?

https://www.youtube.com/watch?v=te1rduz601Q&ab_channel=edureka%21
https://www.youtube.com/watch?v=r4Hrp1fFpMs&ab_channel=GreatLearning
https://en.wikipedia.org/wiki/Extract,_transform,_load

First Part:

Extract (Reader)

- Read a file from a directory.
- It can have any text.
- Provide that data to transformation part.

Transformation

- Apply below transformation on data provided by Reader.
- Capitalize the contents of the Data.
- Provide transformed data to writer.

Load (Writer)

- Write data provided by transformer into another output file in a different directory

Second Part:

Extract (Reader)

- Read a file from a directory.
- It can have any text.
- Provide that data to transformation part.

Transformation

- Apply below transformation on data provided by Reader.
- Read all the unique words from the file. For each word calculate the word count
- For eg: for a file containing the content "Hello world, hello"
- In this case the summary would be
- hello -> 2
- world -> 1
- For simplicity cases should be ignored – Hello and hello mean the same word.
- Provide this transformed data to writer.

Load (Writer)

  • Write data provided by transformer into another output file in a different directory

Design Considerations:

  • It should be easy to add or remove transformations

  • Input source could change i.e. In future we could read contents from a SQL table

  • Output source could also change i.e – In future we could write contents to a SQL table

Hint :

We need three diff components. To read a file , to apply diff transformation on content of that file, to write a output.

  • Reader will read data from file or SQL and pass it to Transformer.

  • Transformer doesn't need to know from where this data came. It's job is to apply some transformation on data provided by Reader and pass it to Writer .

  • Writer will write data to file or DB, It doesn't need to know which kind of transformation applied on it or not. It's only job is to write given data at specified place.

Expectations:

  • You are required to use Python as your coding language

  • You are required to help create the software that works according to the sample inputs applied and generates output as explained in problem statement.

  • You should demonstrate the working software by building a console application or writing test program that exercises the sample inputs, for this purpose you may use Django test framework or any other testing framework.

  • Maturity of your solution will be judged on your clean coding, modularity, design analysis and design skills. Your coding skills will also be judged.

  • Please note that the problems below do not require any relational database or knowledge of underlying OS or Windowing platform.

  • Unit test cases will add extra value to your solution.

  • Upload your solution to your git hub repo and send link to careers@nullinnovation.com or janak@nullinnovation.com along with your resume.

https://nullinnovation.com/