COP 290
Design Practices

Project-1
Trading Simulator and Analyzer
(Subtask-1)

Introduction

The aim of this subtask is to get stocks data for a given stock symbol and a given amount of time and store them in a file.

This can be done in a lot of different file formats. Some of them being :

- **csv** – comma separated values file or just csv file stores tabular data in plain text, where each line (typically) represents one data record. With a very good balance between size and time to write in file, this is one of the most popular file formats to store data.

- **text –** plain text is an uncommon way of storing data. In terms of size and time to write in file, it is pretty similar to csv.

- **json –** json is a open standard file format that uses human-readable text to store and transmit data objects consisting of attribute-value pairs and arrays. This file format is created in a very small time but has a bigger size because of formatting.

- **parquet** – parquet is an open source, column-oriented data file format designed for efficient data storage and retrieval. It uses exceptionally low space but is very time consuming.

- **orc** – optimized row columnar or just orc file format provides a highly efficient way to store data. It uses lower time as well as sizes to store data in comparison with csv.

- **feather –** feather is a portable file format for storing Arrow tables or data frames. This file uses a lightweight binary format and is created very quickly.

- **pickle** – pickle uses a compact binary representation to store data. This file format takes the lowest time to write in the list. Size is similar to csv file format.

To benchmark these files, I used the **os** library for size and **time** library for time.

- os.path.getsize(path/to/file) gives the size of the file in bytes. I converted this to kilobytes.
- Time.time() is used to get the current time. I obtain the current time before and after writing in the file. The difference give the time in second and I convert it to 10^{-4} seconds.

These benchmarks differentiate these files in terms of time and size. The graph created shows a good comparison between these files.

Aditya Jha
2022CS11102