High Frequency Finance
Coursework I

Set: 06 February 2025

Due: 04 March 2025 (11:59am)

This assignment is part of the formal assessment for module, and accounts for 15% of the overall mark. There are three parts to complete.

Part I: Compare SQL and NoSQL (worth 5%)

You are asked to write *up to* 500 words to compare SQL and NoSQL databases, especially in connection with the storage and processing of high-frequency finance data.

You can focus on some (e.g., one) of the key differences, if the word limit is not sufficient to make your point. This is preferable to a long(er) list without enough elaboration. A table, whereby key differences are highlighted, and some are discussed in detail within the word limit is also acceptable.

This part will be assessed attending factors such as contents, clarity and explanations.

Part II: Create a database for high-frequency data (worth 8%)

You are asked to design and implement a database (either SQL or NoSQL) that stores the data contained in the high-frequency tick-data from the London Stock Exchange. The data is shipped in three different files, available on KEATS, which are described in the documentation. As is typical of big data, there are many inconsistencies and anomalies in the dataset, so you may need to pre-process and clean the data before it is usable.

You are welcome to use any database management system you like (so you are not constrained to using MySQL or MongoDB we covered during the module). For this part, you will need to submit the list of steps/commands taken; each step/command must be accompanied by a screenshot of the corresponding output.

This part will be assessed attending factors such as correctness of the steps taken and successful completion of the task.

Part III: Establish a connection to the database in Python (worth 2%)

You are asked to write a short Python script that connects to the database you have just created and performs some operation on it (e.g., execute a query or count the number of cancelled orders). You will need to submit the software code, together with its output.

This part will be assessed attending factors such as correctness of the code and successful completion of the task.

Deliverables

The submission of the assignment must be done through KEATS in a single Jupyter notebook file, exported as a pdf file, that includes

- your name and student number;
- the 500 words of part I;
- the list of steps/commands each accompanied by a screenshot of their output for Part II;
- the code and the output of your piece of software for Part III.

Please refer to the Informatics Handbook for the regulations on Plagiarism and Late Submissions.