Data Intensive Computing - The Final Project

The final project consists of two steps:

1. Define your project, submit a one-page description, and obtain approval for your project proposal from the examiner **(deadline: Sep. 13)**.
2. Implement the approved project **(deadline: Oct. 18)**.

The project proposal should include the following sections:

- **Problem Description:** What problem will you be investigating?
- **Tools:** In your project, it is recommended that you read data from a big data storage system (e.g., HDFS, HBase, etc.), process the data using Spark or another big data processing framework (e.g., Dast, Flink, etc.), and present the results through a visualization tool or by storing them in a NoSQL database. These are just suggested directions, but you are also welcome to propose other tools and approaches if they better fit your project idea.
- **Data:** What data will you use, and how will you collect it?
- **Methodology and Algorithm:** What method(s) or algorithm(s) do you propose to use?

You can implement your code either in Jupyter Notebook or as a stand-alone application. The final submission should include a zip file containing your code and a short 2-page report detailing what you have done, the dataset used, your methodology, your results, and instructions on how to run the code.