Project

Abstract

The project is to solve a Big Data problem on the Cloud, including data collection, implementation, and performance analysis of an end-to-end application.

Project Software

Create a software repository on GitHub. Include also evaluation data sets and test cases.

Project Report

Include a README.md file in the repository describing the complete project. The following topics must be addressed:

1. Description of the problem.
2. Need for Big Data and Cloud.
3. Description of the data (Where does it come from? How was it acquired? What does it mean? What format is it? How big is it? 1 GB minimum).
4. Description of the application, programming model(s), platform and infrastructure.
5. Software design (architectural design, code baseline, dependencies…)
6. Usage (including screenshots that demonstrate how it works).
7. Performance evaluation (speed-up with different number of vCPUs and nodes, identified overheads, optimizations done…).
8. Advanced features (tools/models/platforms not explained in class, advanced functions, techniques to mitigate overheads, challenging implementation aspects…).
9. Conclusions (goals achieved, improvements suggested, lessons learnt, future work, interesting insights…).
10. References.

Submission

Attach the GitHub link to the assignment in Google Classroom.