

Apache Hadoop and Spark Assignment

The goal of the project is to use the gdelt data to determine the suitability of a country for business investment in comparison to five(5) other countries. Students will use Apache Spark SQL to transform, perform exploratory analysis and find insights from the provided data. Students will then be required to present their insights and recommendation using charts and any necessary data visualization tools. **All your recommendations should be strongly supported by demonstratable and reproduceable analysis.**

This project should be completed using Apache Spark and any other supporting technology. Students should use the gdelt 2.0 Events dataset and any other supporting data.

Students are expected to submit the following:

1. A single notebook containing all code,report and visualizations.
2. Please be sure to include the sources of your data, it is not neccessary to include the actual files, but the data files that you use should be publicly accessible.
3. You are required to submit all notebooks to the course moodle assignment page.
4. The project will graded using the following rubric located at this link: [Grading Rubric](#)
5. Good luck and remember passion keeps motivation healthy