The effect of renewable energy development on carbon emissions

Aknur Kassym, Alex Ma, Aibar Talip

**Description:**

Hello! We are delighted to see your interest in our project! The purpose of this application is to examine the renewable energy development trends and patterns throughout different countries of the world over the period from 2000 to 2016. We intend to employ advanced big data analytics tools such as Hadoop and Spark to process and analyze the amount of resources being invested into the renewable energy sector, its effect on the energy generation volumes and global CO2 emissions, aiming to gain meaningful insights. Our analysis encompasses the identification of trends, patterns, and factors contributing to the renewable energy development and carbon emissions fluctuations.

**Folder structure:**

Although a dedicated folder for data ingestion and data cleaning could be observed, both of them have been copied and used in the analytics code (included in the folder ana\_code). However, the code for profiling should be accessed separately (in the profiling\_code folder), since it was not used in the analytics code.

**The access to the datasets:**

The access has been given to as17321\_nyu\_edu, lj2330\_nyu\_edu and adm209\_nyu\_edu. All three datasets *fossil\_co2.csv*, *re\_generation.csv*, *investment\_data.csv* can be accessed in ak8827 user’s folder called “project”.

**Running and testing the codes:**

After downloading / accessing the datasets, we advise to download the codes for analysis, which can be found in the ana\_code folder. It is important to know that since every member of the group worked on its part of the project independently on their own local machine, the paths to the dataset are different and should be updated accordingly.

To run the program, use the command:

spark-shell --deploy-mode client -i <name of the scala file>

Thank you!