Yara Weening - 642287

Hadoop  23-6-2022

Assignment 2

**GitHub url with my source code:** <https://github.com/yaraweening/Hadoop>

For Hadoop a second assignment is made to use Pig for the data of a diplomacy dataset.

To complete the assignment the following steps took place:

**Step 1:** Because of the previous assignment the Oracle VM and Hortonbox are already installed. They need to be started first.

**Step 2:** In HDFS in the file view a new folder is created for the dataset. Services -> HDFS -> user -> maria\_dev -> diplomacy and the diplomacy csv files are uploaded.

Afbeelding met tafel

Automatisch gegenereerde beschrijving

**Step 3:** Now the dataset is saved in the folder some queries are written in Pig View. For the assignment the following url is used: <https://www.tutorialspoint.com/apache_pig/index.htm>.

**Assignment 1:**

* Make a alphabetic list from all locations from the orders.csv.
* Group by “location” with target “Holland”
* Count how many times Holland was the target from that location

Afbeelding met tekst

Automatisch gegenereerde beschrijving

Afbeelding met tekst

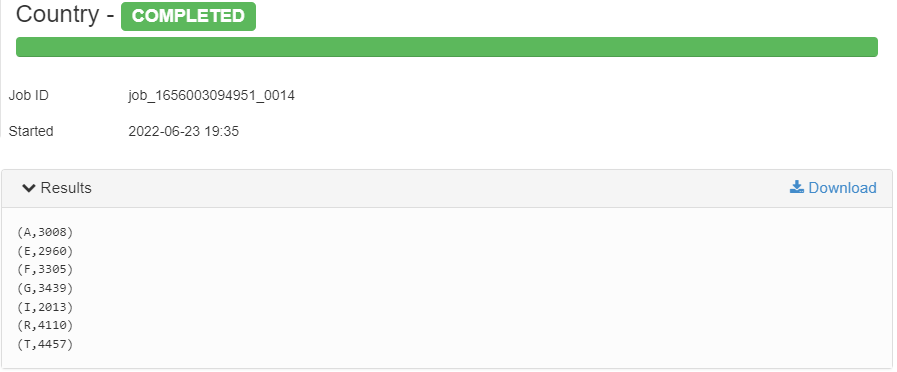
Automatisch gegenereerde beschrijving

**Assignment 2:**

* Make a list how many times every country won

Afbeelding met tekst

Automatisch gegenereerde beschrijving



**Assignment 3:**

* Make a list from the top 10 countries that had the most turns and won the game.
* Names had to be the full name and not the index.

Afbeelding met tekst

Automatisch gegenereerde beschrijving

