UNIVERSIDAD EAFIT

INGENIERÍA DE SISTEMAS

ST0263 TÓPICOS ESPECIALES EN TELEMÁTICA, 2022-1

UNIDAD 3 BIG DATA

DOCUMENTACIÓN y EVIDENCIAS DE LAB 6

ESTE DOCUMENTO DEBE SER ENTREGADO POR BUZON DE ENTREGAS DE LA MATERIA

Además entregar por buzón la URL del repo github donde tiene todo de los LABs 5 y 6

NOMBRE DEL ALUMNO: Santiago Ignacio Vanegas Pérez

EMAIL: [svaneg11@eafit.edu.co](mailto:svaneg11@eafit.edu.co)

GITHUB DEL ALUMNO PARA LOS LABS <https://github.com/svaneg11/svaneg11_ST0263/tree/main/U3-Lab2>

Realizar las actividades relacionadas con el lab6 en: <https://github.com/st0263eafit/st0263-2261/tree/main/bigdata>

|  |
| --- |
| LAB6-EVIDENCIAS DE LA EJECUCIÓN DEL WORDCOUNG EN:   1. pyspark desde CLI interactivo en EMR 2. Pyspark desde CLI como archivo de entrada EMR 3. Pyspark desde jupyter notebooks en EMR 4. Ejecución del notebook jupyter: Data\_processing\_using\_PySpark.ipynb en EMR con datos en S3   párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc. |
|  |
| LAB6- Ejecución del notebook jupyter: Data\_processing\_using\_PySpark.ipynb en EMR con datos en S3  párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc. |
| LAB6-Evidencias de la gestión de tablas (creación y consultas) en Hive para los datos de la ONU en EMR  párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc. |
|  |
| LAB6-Evidencias del Wordcount en Hive con datos en S3 en ERM  párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc. |
| * **Ordernado por palabra:**      * **Ordenado por frecuencia de mayor a menor** |
| LAB6-Evidencias de la consulta de tablas Hive, desde SparkSQL en un jupyter notebook en EMR con datos en S3  párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc. |
|  |
| LAB6-evidencias de algo no considerado en las anteriores evidencias y requeridas en el LAB  párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc. |