

## PROGRAMA PROFESIONAL

Ciencias de la Computación

**CURSO** 

Big Data

**TEMA** 

Ejercicios - Map Reduce

## Integrantes:

Loayza Huarachi Angel Josue

**SEMESTRE: VIII** 

**AÑO**: 2024

"Los alumnos declaran haber realizado el presente trabajo de acuerdo a las normas de la Universidad Católica San Pablo

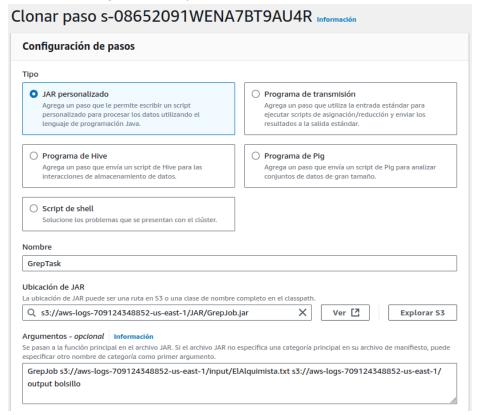
#### 1. Introducción

A continuación se muestran los resultados obtenidos de los ejercicios que pueden ser realizados por Map Reduce. Estos comprenden: Distributed Grep, Count of URL Access Frequency, Reverse Web-Link Graph, y Term-Vector per Host.

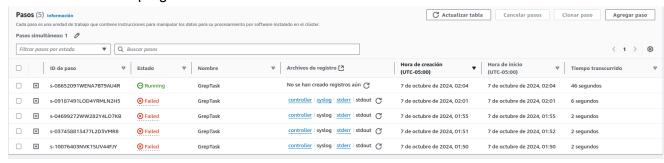
En el siguiente link a Github se encuentran todos los archivos realizados en el lenguaje de Java: https://github.com/angel452/Ejercicios MapReduce/tree/master

## 2. Distributed Grep

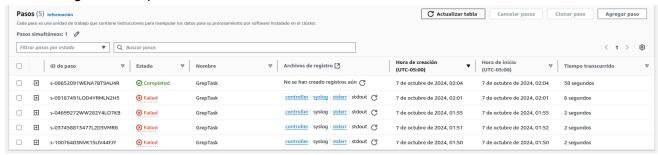
- Crear la tarea (cargar el script y parametros)



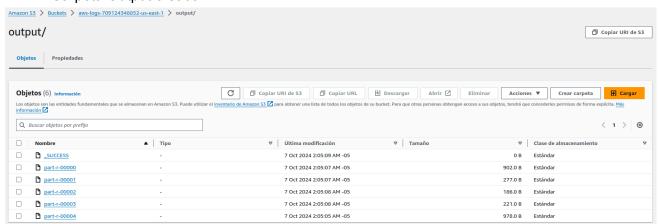
Corriendo el programa



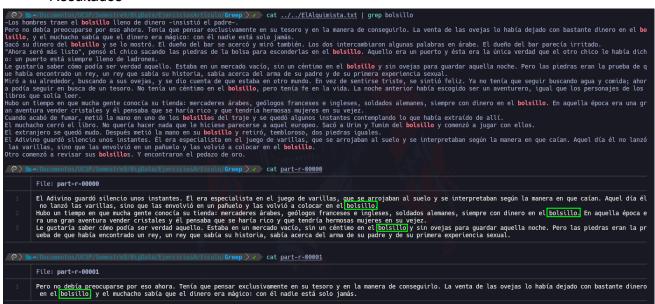
#### Programa completado



#### - Carpeta Output creada



#### - Resultados



C) — Mocumento // C//sestre/Albata/Ejercicoarticulo/Greep // cat ../../ElAlquimista.txt | grep holsillo
-Los hobbres traen el bolsillo lleno de dinero -insistió el padre-Pero no debía procuparse por eso abnar. Tenda que pensar exclusivamente en su tesoro y en la manera de conseguirlo. La venta de las ovejas lo había dejado con hastante dinero en el bo
lsillo, y el muchacho sabía que el dinero era mágico: con él nadie está solo jamás.

Albata seré más listo, pensó el chico sacando las piedras de la bolsa para esconderlas en el bolsillo. Apuello era un puerto y ésta era la onica verdad que el otro chico le había dich
leg ustaría saber como podía ser verdad aquello. Estaba en un nercado vacio, sin un céntimo en el bolsillo. Apuello era un puerto y ésta era la onica verdad que el otro chico le había dich
leg ustaría saber como podía ser verdad aquello. Estaba en un nercado vacio, sin un céntimo en el bolsillo y sin ovejas para guardar aquella noche. Pero las piedras eran la prueba de que había encontrado un rey, un rey que sabía su historia, sabía sacrez del arma de su padre y de su prinera experiencia secual.

Mirá a su alreddor, buscando a sus ovejas, y se dio cuenta de que estaba en otro numdo. En vez de sentirse triste, se sintió fellz. Ya no tenía que seguir buscando aqua y comáa; ahor a podía seguir en busca de un tesoro. No tenía un céntumo en el bolsillo, pero tenía fe en la vida. La noche anterior había escogidos er un aventurero, igual que los personajes de los libros que solia leer.

Hibbo un tiempo en que estaleg y el pomaba que se haría rico y que teorir a herbosa superes en su vejo:

Cuando acabó de fumar, estió la mano en un desir prico y que teorir a herbosa superes en su vejo:

Cuando acabó de fumar, estió la mano en su bissillo y retiró, tembloroso, dos piedras iguales.

El Advino quardó silencio unos instantes. El era especialista en el juego de varillas, que se arrojaban al suelo y se interpretaban según la manera en que caían. Aquel día él no lanzó las varillas, sino que las envolviós en un p

Coloration (NCC) Security (Notation (Notation

## 3. Count of URL Access Frequency

- Creación de la tarea (cargar programa y parámetros)

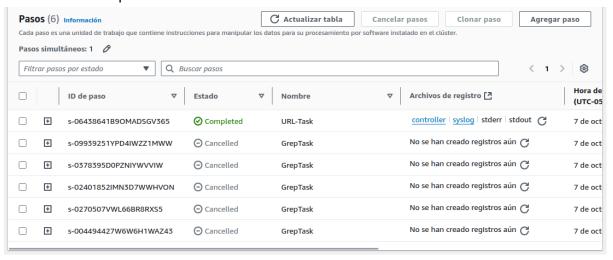
#### Configuración de pasos Tipo JAR personalizado Programa de transmisión Agrega un paso que le permite escribir un script Agrega un paso que utiliza la entrada estándar para personalizado para procesar los datos utilizando el ejecutar scripts de asignación/reducción y enviar los lenguaje de programación Java. resultados a la salida estándar. O Programa de Hive O Programa de Pig Agrega un paso que envía un script de Hive para las Agrega un paso que envía un script de Pig para analizar interacciones de almacenamiento de datos. conjuntos de datos de gran tamaño. Script de shell Solucione los problemas que se presentan con el clúster. Nombre **URL-Task** Ubicación de JAR La ubicación de JAR puede ser una ruta en S3 o una clase de nombre completo en el classpath. Q s3://aws-logs-709124348852-us-east-1/JAR/UrlCountJob.jar Ver 🖸 Explorar S3

Argumentos - opcional Información

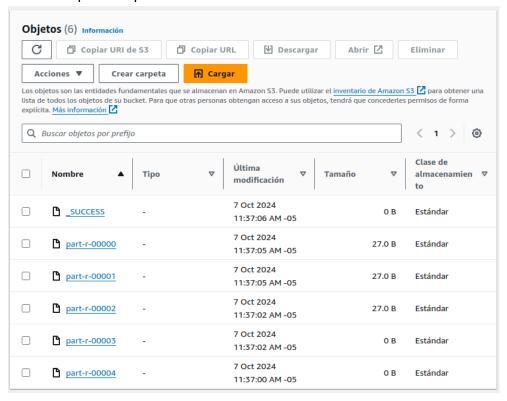
Se pasan a la función principal en el archivo JAR. Si el archivo JAR no especifica una categoría principal en su archivo de manifiesto, puede especificar otro nombre de categoría como primer argumento.

 $\label{lem:unlcount} \begin{tabular}{ll} Url Count Job s3://aws-logs-709124348852-us-east-1/input/url-text.txt s3://aws-logs-7091243488-0000$ 

Tarea completada

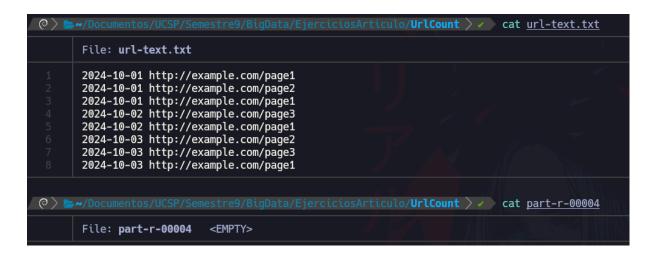


- Carpeta Output creada



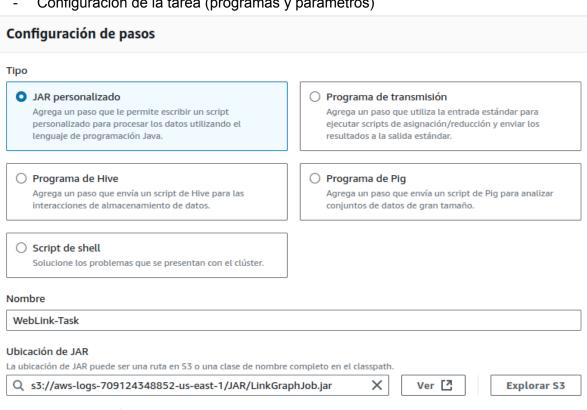
Resultados





#### 4. Reverse Web-Link Graph

Configuracion de la tarea (programas y parámetros)

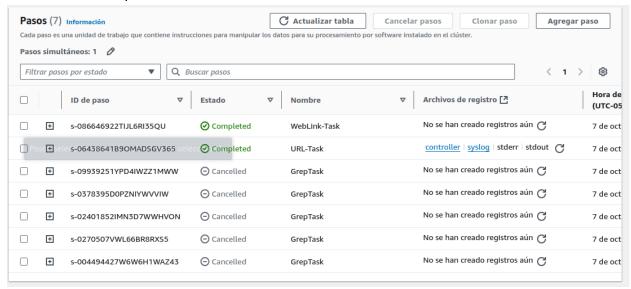


#### Argumentos - opcional Información

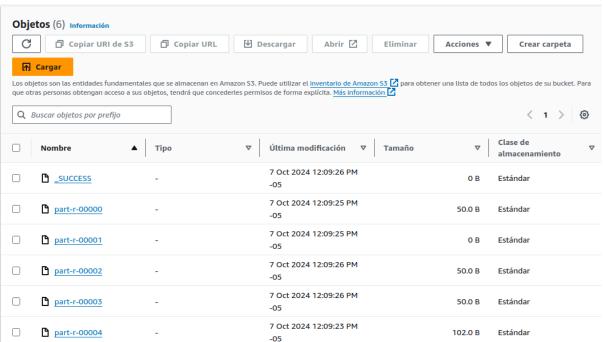
Se pasan a la función principal en el archivo JAR. Si el archivo JAR no especifica una categoría principal en su archivo de manifiesto, puede especificar otro nombre de categoría como primer argumento.

LinkGraphJob s3://aws-logs-709124348852-us-east-1/input/webLink.txt s3://aws-logs-709124348852-us-east-1/ ouput

### - Tarea Completada



### - Carpeta output creada



#### Resultados



# 5. Term-Vector per Host

- Configuración de la tarea (programa y parámetros)

ро	
<ul> <li>JAR personalizado         Agrega un paso que le permite escribir un script personalizado para procesar los datos utilizando el lenguaje de programación Java.     </li> </ul>	<ul> <li>Programa de transmisión</li> <li>Agrega un paso que utiliza la entrada estándar para ejecutar scripts de asignación/reducción y enviar los resultados a la salida estándar.</li> </ul>
Programa de Hive     Agrega un paso que envía un script de Hive para las interacciones de almacenamiento de datos.	Programa de Pig  Agrega un paso que envía un script de Pig para analizar conjuntos de datos de gran tamaño.
O Script de shell Solucione los problemas que se presentan con el clúster.	
ombre	
TermVector-Task	
bicación de JAR a ubicación de JAR puede ser una ruta en S3 o una clase de nombre	e completo en el classpath.
Q s3://aws-logs-709124348852-us-east-1/JAR/TermVe	ectorJob.jar X Ver 🖸 Explorar S3
rgumentos - <i>opcional</i> Información e pasan a la función principal en el archivo JAR. Si el archivo JAR no	o especifica una categoría principal en su archivo de manifiesto, puec