



Universidad Católica  
**San Pablo**

## **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](https://github.com/angel452/Ejercicios_MapReduce/tree/master)

# 2. Distributed Grep

- Crear la tarea (cargar el script y parametros)

Clonar paso s-08652091WENA7BT9AU4R

Información

Configuración de pasos

Tipo

☒ JAR personalizado

Agrega un paso que le permite escribir un script personalizado para procesar los datos utilizando el lenguaje de programación Java.

☐ Programa de transmisión

Agrega un paso que utiliza la entrada estándar para ejecutar scripts de asignación/reducción y enviar los resultados a la salida estándar.

☐ 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.

☐ Script de shell

Solucione los problemas que se presentan con el clúster.

Nombre

GrepTask

Ubicación de JAR

La ubicación de JAR puede ser una ruta en S3 o una clase de nombre completo en el classpath.

s3://aws-logs-709124348852-us-east-1/JAR/GrepJob.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.

GrepJob s3://aws-logs-709124348852-us-east-1/Input/ElAlquimista.txt s3://aws-logs-709124348852-us-east-1/output bolsillo

- Corriendo el programa

Pasos (5) <span>Información</span>									
Cada paso es una unidad de trabajo que contiene instrucciones para manipular los datos para su procesamiento por software instalado en el clúster.									
Pasos simultáneos: 1									
<div>Filtrar pasos por estado</div> <div>Buscar pasos</div>									
	ID de paso	Estado	Nombre	Archivos de registro	Hora de creación (UTC-05:00)	Hora de inicio (UTC-05:00)	Tiempo transcurrido		
<input type="checkbox"/>	s-08652091WENA7BT9AU4R	Running	GrepTask	No se han creado registros aún	7 de octubre de 2024, 02:04	7 de octubre de 2024, 02:04	46 segundos		
<input type="checkbox"/>	s-09187491LOD4YRMLN2H5	Failed	GrepTask	<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr</a>   <a href="#">stdout</a>	7 de octubre de 2024, 02:01	7 de octubre de 2024, 02:01	6 segundos		
<input type="checkbox"/>	s-04699272WW282Y4LO7KB	Failed	GrepTask	<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr</a>   <a href="#">stdout</a>	7 de octubre de 2024, 01:55	7 de octubre de 2024, 01:55	2 segundos		
<input type="checkbox"/>	s-037458813477L2D3VMRB	Failed	GrepTask	<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr</a>   <a href="#">stdout</a>	7 de octubre de 2024, 01:51	7 de octubre de 2024, 01:52	2 segundos		
<input type="checkbox"/>	s-10076403NVK1SUV44FY	Failed	GrepTask	<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr</a>   <a href="#">stdout</a>	7 de octubre de 2024, 01:50	7 de octubre de 2024, 01:50	2 segundos		

- Programa completado

Pasos (5) Información

Actualizar tabla

Cancelar pasos

Clonar paso

Agregar paso

Cada paso es una unidad de trabajo que contiene instrucciones para manipular los datos para su procesamiento por software instalado en el clúster.

Pasos simultáneos: 1

Filtrar pasos por estado

Buscar pasos

< 1 >

<input type="checkbox"/>	ID de paso	Estado	Nombre	Archivos de registro	Hora de creación (UTC-05:00)	Hora de inicio (UTC-05:00)	Tiempo transcurrido
<input type="checkbox"/>	s-08652091WENA78T9AU4R	Completed	GrepTask	No se han creado registros aún	7 de octubre de 2024, 02:04	7 de octubre de 2024, 02:04	38 segundos
<input type="checkbox"/>	s-09187491L0D4YRMLN2H5	Failed	GrepTask	<a href="#">controller</a> <a href="#">syslog</a> <a href="#">stderr</a> <a href="#">stdout</a>	7 de octubre de 2024, 02:01	7 de octubre de 2024, 02:01	6 segundos
<input type="checkbox"/>	s-04699272WW282Y4LO7KB	Failed	GrepTask	<a href="#">controller</a> <a href="#">syslog</a> <a href="#">stderr</a> <a href="#">stdout</a>	7 de octubre de 2024, 01:55	7 de octubre de 2024, 01:55	2 segundos
<input type="checkbox"/>	s-037450813477L2D3VMR8	Failed	GrepTask	<a href="#">controller</a> <a href="#">syslog</a> <a href="#">stderr</a> <a href="#">stdout</a>	7 de octubre de 2024, 01:51	7 de octubre de 2024, 01:52	2 segundos
<input type="checkbox"/>	s-10076403NVK15UV44FJY	Failed	GrepTask	<a href="#">controller</a> <a href="#">syslog</a> <a href="#">stderr</a> <a href="#">stdout</a>	7 de octubre de 2024, 01:50	7 de octubre de 2024, 01:50	2 segundos

- Carpeta Output creada

Amazon S3 > Buckets > aws-logs-709124348852-us-east-1 > output/

output/

Copiar URI de S3

Objetos

Propiedades

Objetos (6) Información

Copiar URI de S3

Copiar URL

Descargar

Abrir

Eliminar

Acciones

Crear carpeta

Cargar

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [Inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

Buscar objetos por prefijo

< 1 >

<input type="checkbox"/>	Nombre	Tipo	Última modificación	Tamaño	Clase de almacenamiento
<input type="checkbox"/>	<a href="#">_SUCCESS</a>	-	7 Oct 2024 2:05:09 AM -05	0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00000</a>	-	7 Oct 2024 2:05:07 AM -05	902.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00001</a>	-	7 Oct 2024 2:05:07 AM -05	277.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00002</a>	-	7 Oct 2024 2:05:08 AM -05	186.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00003</a>	-	7 Oct 2024 2:05:08 AM -05	221.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00004</a>	-	7 Oct 2024 2:05:05 AM -05	978.0 B	Estándar

- Resultados

~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep > cat ../EIA/quimista.txt | grep bolsillo

-Los hombres traen el bolsillo lleno de dinero -insistió el padre-. Pero no debía preocuparse por eso ahora. Tenía que pensar exclusivamente en su tesoro y en la manera de conseguirlo. La venta de las ovejas lo había dejado con bastante dinero en el bolsillo, y el muchacho sabía que el dinero era mágico: con él nadie está solo jamás. Sacó su dinero del bolsillo y se lo mostró. El dueño del bar se acercó y miró también. Los dos intercambiaron algunas palabras en árabe. El dueño del bar parecía irritado. "Ahora seré más listo", pensó el chico sacando las piedras de la bolsa para esconderlas en el bolsillo. Aquello era un puerto y ésta era la única verdad que el otro chico le había dicho: un puerto está siempre lleno de ladrones. Le gustaría saber cómo podía ser verdad aquello. Estaba en un mercado vacío, 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 acerca del arma de su padre y de su primera experiencia sexual. Miró a su alrededor, buscando a sus ovejas, y se dio cuenta de que estaba en otro mundo. En vez de sentirse triste, se sintió feliz. Ya no tenía que seguir buscando agua y comida; ahora podía seguir en busca de un tesoro. No tenía un céntimo en el bolsillo, pero tenía fe en la vida. La noche anterior había escogido ser un aventurero, igual que los personajes de los libros que solía leer. Hubo un tiempo en que mucha gente conocía su tienda: mercaderes árabes, geólogos franceses e ingleses, soldados alemanes, siempre con dinero en el bolsillo. En aquella época era una gran aventura vender cristales y él pensaba que se haría rico y que tendría hermosas mujeres en su vejez. Cuando acabó de fumar, metió la mano en uno de los bolsillos del traje y se quedó algunos instantes contemplando lo que había extraído de allí. El muchacho cerró el libro. No quería hacer nada que le hiciese parecerse a aquel europeo. Sacó a Urim y Tumim del bolsillo y comenzó a jugar con ellos. El extranjero se quedó mudo. Después metió la mano en su bolsillo y retiró, temeroso, dos piedras iguales. El Adivino guardó silencio unos instantes. Él 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 envolvió en un pañuelo y las volvió a colocar en el bolsillo. Otro comenzó a revisar sus bolsillos. Y encontraron el pedazo de oro.

~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep > cat part-r-00000

File: part-r-00000

1

El Adivino guardó silencio unos instantes. Él 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 envolvió en un pañuelo y las volvió a colocar en el bolsillo.

2

Hubo un tiempo en que mucha gente conocía su tienda: mercaderes árabes, geólogos franceses e ingleses, soldados alemanes, siempre con dinero en el bolsillo. En aquella época era una gran aventura vender cristales y él pensaba que se haría rico y que tendría hermosas mujeres en su vejez.

3

Le gustaría saber cómo podía ser verdad aquello. Estaba en un mercado vacío, 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 acerca del arma de su padre y de su primera experiencia sexual.

~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep > cat part-r-00001

File: part-r-00001

1

Pero no debía preocuparse por eso ahora. Tenía que pensar exclusivamente en su tesoro y en la manera de conseguirlo. La venta de las ovejas lo había dejado con bastante dinero en el bolsillo y el muchacho sabía que el dinero era mágico: con él nadie está solo jamás.

```

@> ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep >> cat ../../ElAlquimista.txt | grep bolsillo
-Los hombres traen el bolsillo lleno de dinero -insistió el padre-.
Pero no debía preocuparse por eso ahora. Tenía que pensar exclusivamente en su tesoro y en la manera de conseguirlo. La venta de las ovejas lo había dejado con bastante dinero en el bolsillo, y el muchacho sabía que el dinero era mágico: con él nadie está solo jamás.
Sacó su dinero del bolsillo y se lo mostró. El dueño del bar se acercó y miró también. Los dos intercambiaron algunas palabras en árabe. El dueño del bar parecía irritado.
"Ahora seré más listo", pensó el chico sacando las piedras de la bolsa para esconderlas en el bolsillo. Aquello era un puerto y ésta era la única verdad que el otro chico le había dicho: un puerto está siempre lleno de ladrones.
Le gustaría saber cómo podía ser verdad aquello. Estaba en un mercado vacío, 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 acerca del arma de su padre y de su primera experiencia sexual.
Miró a su alrededor, buscando a sus ovejas, y se dio cuenta de que estaba en otro mundo. En vez de sentirse triste, se sintió feliz. Ya no tenía que seguir buscando agua y comida; ahora podía seguir en busca de un tesoro. No tenía un céntimo en el bolsillo, pero tenía fe en la vida. La noche anterior había escogido ser un aventurero, igual que los personajes de los libros que solía leer.
Hubo un tiempo en que mucha gente conocía su tienda: mercaderes árabes, geólogos franceses e ingleses, soldados alemanes, siempre con dinero en el bolsillo. En aquella época era una gran aventura vender cristales y él pensaba que se haría rico y que tendría hermosas mujeres en su vejez.
Cuando acabó de fumar, metió la mano en uno de los bolsillos del traje y se quedó algunos instantes contemplando lo que había extraído de allí.
El muchacho cerró el libro. No quería hacer nada que le hiciese parecerse a aquel europeo. Sacó a Urim y Tumim del bolsillo y comenzó a jugar con ellos.
El extranjero se quedó mudo. Después metió la mano en su bolsillo y retiró, tembloroso, dos piedras iguales.
El Adivino guardó silencio unos instantes. Él 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 envolvió en un pañuelo y las volvió a colocar en el bolsillo.
Otro comenzó a revisar sus bolsillos. Y encontraron el pedazo de oro.

@> ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep >> cat part-r-00002
File: part-r-00002
1 El extranjero se quedó mudo. Después metió la mano en su bolsillo y retiró, tembloroso, dos piedras iguales.
2 Otro comenzó a revisar sus bolsillos. Y encontraron el pedazo de oro.

@> ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep >> cat part-r-00003
File: part-r-00003
1 -Los hombres traen el bolsillo lleno de dinero -insistió el padre-.
2 Cuando acabó de fumar, metió la mano en uno de los bolsillos del traje y se quedó algunos instantes contemplando lo que había extraído de allí.

```

```

@> ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep >> cat ../../ElAlquimista.txt | grep bolsillo
-Los hombres traen el bolsillo lleno de dinero -insistió el padre-.
Pero no debía preocuparse por eso ahora. Tenía que pensar exclusivamente en su tesoro y en la manera de conseguirlo. La venta de las ovejas lo había dejado con bastante dinero en el bolsillo, y el muchacho sabía que el dinero era mágico: con él nadie está solo jamás.
Sacó su dinero del bolsillo y se lo mostró. El dueño del bar se acercó y miró también. Los dos intercambiaron algunas palabras en árabe. El dueño del bar parecía irritado.
"Ahora seré más listo", pensó el chico sacando las piedras de la bolsa para esconderlas en el bolsillo. Aquello era un puerto y ésta era la única verdad que el otro chico le había dicho: un puerto está siempre lleno de ladrones.
Le gustaría saber cómo podía ser verdad aquello. Estaba en un mercado vacío, 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 acerca del arma de su padre y de su primera experiencia sexual.
Miró a su alrededor, buscando a sus ovejas, y se dio cuenta de que estaba en otro mundo. En vez de sentirse triste, se sintió feliz. Ya no tenía que seguir buscando agua y comida; ahora podía seguir en busca de un tesoro. No tenía un céntimo en el bolsillo, pero tenía fe en la vida. La noche anterior había escogido ser un aventurero, igual que los personajes de los libros que solía leer.
Hubo un tiempo en que mucha gente conocía su tienda: mercaderes árabes, geólogos franceses e ingleses, soldados alemanes, siempre con dinero en el bolsillo. En aquella época era una gran aventura vender cristales y él pensaba que se haría rico y que tendría hermosas mujeres en su vejez.
Cuando acabó de fumar, metió la mano en uno de los bolsillos del traje y se quedó algunos instantes contemplando lo que había extraído de allí.
El muchacho cerró el libro. No quería hacer nada que le hiciese parecerse a aquel europeo. Sacó a Urim y Tumim del bolsillo y comenzó a jugar con ellos.
El extranjero se quedó mudo. Después metió la mano en su bolsillo y retiró, tembloroso, dos piedras iguales.
El Adivino guardó silencio unos instantes. Él 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 envolvió en un pañuelo y las volvió a colocar en el bolsillo.
Otro comenzó a revisar sus bolsillos. Y encontraron el pedazo de oro.

@> ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/Greep >> cat part-r-00004
File: part-r-00004
1 "Ahora seré más listo", pensó el chico sacando las piedras de la bolsa para esconderlas en el bolsillo. Aquello era un puerto y ésta era la única verdad que el otro chico le había dicho: un puerto está siempre lleno de ladrones.
2 El muchacho cerró el libro. No quería hacer nada que le hiciese parecerse a aquel europeo. Sacó a Urim y Tumim del bolsillo y comenzó a jugar con ellos.
3 Miró a su alrededor, buscando a sus ovejas, y se dio cuenta de que estaba en otro mundo. En vez de sentirse triste, se sintió feliz. Ya no tenía que seguir buscando agua y comida; ahora podía seguir en busca de un tesoro. No tenía un céntimo en el bolsillo, pero tenía fe en la vida. La noche anterior había escogido ser un aventurero, igual que los personajes de los libros que solía leer.
4 Sacó su dinero del bolsillo y se lo mostró. El dueño del bar se acercó y miró también. Los dos intercambiaron algunas palabras en árabe. El dueño del bar parecía irritado.

```

### 3. Count of URL Access Frequency

- Creación de la tarea (cargar programa y parámetros)

#### Configuración de pasos

##### Tipo

☒ JAR personalizado

Agrega un paso que le permite escribir un script personalizado para procesar los datos utilizando el lenguaje de programación Java.

☐ Programa de transmisión

Agrega un paso que utiliza la entrada estándar para ejecutar scripts de asignación/reducción y enviar los resultados a la salida estándar.

☐ 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.

☐ 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.

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.

UrlCountJob s3://aws-logs-709124348852-us-east-1/Input/url-text.txt s3://aws-logs-709124348852-us-east-1/output

- Tarea completada

Pasos (6) Información

Actualizar tabla

Cancelar pasos

Clonar paso

Agregar paso

Cada paso es una unidad de trabajo que contiene instrucciones para manipular los datos para su procesamiento por software instalado en el clúster.

Pasos simultáneos: 1

Filtrar pasos por estado

Buscar pasos

< 1 >

<input type="checkbox"/>		ID de paso	Estado	Nombre	Archivos de registro	Hora de (UTC-05)
<input type="checkbox"/>		s-06438641B9OMADSGV365	Completed	URL-Task	<a href="#">controller</a>   <a href="#">syslog</a>   stderr   stdout	7 de oct
<input type="checkbox"/>		s-09939251YPD4IWZZ1MWW	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>		s-0378395D0PZNIYWVVIW	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>		s-02401852IMN3D7WWHVON	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>		s-0270507VWL66BR8RX55	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>		s-004494427W6W6H1WAZ43	Cancelled	GrepTask	No se han creado registros aún	7 de oct

- Carpeta Output creada

Objetos (6) Información

🔄

Copiar URI de S3

Copiar URL

⬇️ Descargar

Abrir

Eliminar

Acciones ▾

Crear carpeta

Cargar

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [Inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

🔍

Buscar objetos por prefijo

< 1 > ⚙️

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	<a href="#">_SUCCESS</a>	-	7 Oct 2024 11:37:06 AM -05	0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00000</a>	-	7 Oct 2024 11:37:05 AM -05	27.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00001</a>	-	7 Oct 2024 11:37:05 AM -05	27.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00002</a>	-	7 Oct 2024 11:37:02 AM -05	27.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00003</a>	-	7 Oct 2024 11:37:02 AM -05	0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00004</a>	-	7 Oct 2024 11:37:00 AM -05	0 B	Estándar

- Resultados

📁 ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/UrlCount > ✓ cat url-text.txt

File: url-text.txt

1 2024-10-01 http://example.com/page1

2 2024-10-01 http://example.com/page2

3 2024-10-01 http://example.com/page1

4 2024-10-02 http://example.com/page3

5 2024-10-02 http://example.com/page1

6 2024-10-03 http://example.com/page2

7 2024-10-03 http://example.com/page3

8 2024-10-03 http://example.com/page1

📁 ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/UrlCount > ✓ cat part-r-00000

File: part-r-00000

1 http://example.com/page1 4

📁 ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/UrlCount > ✓ cat part-r-00001

File: part-r-00001

1 http://example.com/page2 2

📁 ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/UrlCount > ✓ cat part-r-00002

File: part-r-00002

1 http://example.com/page3 2

📁 ~/Documentos/UCSP/Semestre9/BigData/EjerciciosArticulo/UrlCount > ✓ cat part-r-00003

File: part-r-00003 <EMPTY>

```
~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / UrlCount > cat url-text.txt
File: url-text.txt
1 2024-10-01 http://example.com/page1
2 2024-10-01 http://example.com/page2
3 2024-10-01 http://example.com/page1
4 2024-10-02 http://example.com/page3
5 2024-10-02 http://example.com/page1
6 2024-10-03 http://example.com/page2
7 2024-10-03 http://example.com/page3
8 2024-10-03 http://example.com/page1

~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / UrlCount > cat part-r-00004
File: part-r-00004 <EMPTY>
```

#### 4. Reverse Web-Link Graph

- Configuración de la tarea (programas y parámetros)

##### Configuración de pasos

###### Tipo

☒ JAR personalizado

Agrega un paso que le permite escribir un script personalizado para procesar los datos utilizando el lenguaje de programación Java.

☐ Programa de transmisión

Agrega un paso que utiliza la entrada estándar para ejecutar scripts de asignación/reducción y enviar los resultados a la salida estándar.

☐ 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.

☐ Script de shell

Solucione los problemas que se presentan con el clúster.

###### Nombre

WebLink-Task

###### Ubicación de JAR

La ubicación de JAR puede ser una ruta en S3 o una clase de nombre completo en el classpath.

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.

LinkGraphJob s3://aws-logs-709124348852-us-east-1/Input/webLink.txt s3://aws-logs-709124348852-us-east-1/output

- Tarea Completada

Pasos (7) Información

Actualizar tabla

Cancelar pasos

Clonar paso

Agregar paso

Cada paso es una unidad de trabajo que contiene instrucciones para manipular los datos para su procesamiento por software instalado en el clúster.

Pasos simultáneos: 1

Filtrar pasos por estado

Buscar pasos

< 1 >

	ID de paso	Estado	Nombre	Archivos de registro	Hora de (UTC-05)
<input type="checkbox"/>	s-086646922TJL6RI35QU	Completed	WebLink-Task	No se han creado registros aún	7 de oct
<input type="checkbox"/>	s-06438641B9QMADSGV365	Completed	URL-Task	<a href="#">controller</a>   <a href="#">syslog</a>   stderr   stdout	7 de oct
<input type="checkbox"/>	s-09939251YPD4IWZZ1MWW	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>	s-0378395D0PZNIYVVVIW	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>	s-02401852IMN3D7WWHVON	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>	s-0270507VWL66BR8RXS5	Cancelled	GrepTask	No se han creado registros aún	7 de oct
<input type="checkbox"/>	s-004494427W6W6H1WAZ43	Cancelled	GrepTask	No se han creado registros aún	7 de oct

- Carpeta output creada

Objetos (6) Información

Copiar URI de S3

Copiar URL

Descargar

Abrir

Eliminar

Acciones

Crear carpeta

Cargar

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

Buscar objetos por prefijo

< 1 >

	Nombre	Tipo	Última modificación	Tamaño	Clase de almacenamiento
<input type="checkbox"/>	<a href="#">_SUCCESS</a>	-	7 Oct 2024 12:09:26 PM -05	0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00000</a>	-	7 Oct 2024 12:09:25 PM -05	50.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00001</a>	-	7 Oct 2024 12:09:25 PM -05	0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00002</a>	-	7 Oct 2024 12:09:26 PM -05	50.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00003</a>	-	7 Oct 2024 12:09:26 PM -05	50.0 B	Estándar
<input type="checkbox"/>	<a href="#">part-r-00004</a>	-	7 Oct 2024 12:09:23 PM -05	102.0 B	Estándar



## - Resultados

```
~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / Web-Link > cat webLink.txt
File: webLink.txt
1 http://example.com/pageA http://example.com/pageB http://example.com/pageC
2 http://example.com/pageB http://example.com/pageD
3 http://example.com/pageA http://example.com/pageD
4 http://example.com/pageC http://example.com/pageE http://example.com/pageD

~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / Web-Link > cat part-r-00000
File: part-r-00000
1 http://example.com/pageE http://example.com/pageC

~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / Web-Link > cat part-r-00001
File: part-r-00001 <EMPTY>

~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / Web-Link > cat part-r-00002
File: part-r-00002
1 http://example.com/pageB http://example.com/pageA

~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / Web-Link > cat part-r-00003
File: part-r-00003
1 http://example.com/pageC http://example.com/pageA

~ / Documentos / UCSP / Semestre9 / BigData / EjerciciosArticulo / Web-Link > cat part-r-00004
File: part-r-00004
1 http://example.com/pageD http://example.com/pageC, http://example.com/pageA, http://example.com/pageB
```

## 5. Term-Vector per Host

- Configuración de la tarea (programa y parámetros)

### Configuración de pasos

Tipo

☒ **JAR personalizado**  
Agrega un paso que le permite escribir un script personalizado para procesar los datos utilizando el lenguaje de programación Java.

☐ **Programa de transmisión**  
Agrega un paso que utiliza la entrada estándar para ejecutar scripts de asignación/reducción y enviar los resultados a la salida estándar.

☐ **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.

☐ **Script de shell**  
Solucione los problemas que se presentan con el clúster.

Nombre

TermVector-Task

Ubicación de JAR

La ubicación de JAR puede ser una ruta en S3 o una clase de nombre completo en el classpath.

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.

TermVectorJob

 s3://aws-logs-709124348852-us-east-1/input/TermVector.txt s3://aws-logs-709124348852-us-east-1/output