

# UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA

ESCUELA PROFESIONAL DE CIENCIA DE LA COMPUTACIÓN  
BIG DATA



---

Motor de Búsqueda

---

*Presentado por*

Juan Manuel Soto Begazo

*Docente*

Alvaro Henry Mamani Aliaga



## 1. Config

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	<a href="#">hive-logs.hql</a>	hql	6 Jun 2024 10:54:29 AM -05	520.0 B	Estándar
<input type="checkbox"/>	<a href="#">hive-wordcount.hql</a>	hql	6 Jun 2024 9:36:51 AM -05	529.0 B	Estándar
<input type="checkbox"/>	<a href="#">wordcount_save.hql</a>	hql	6 Jun 2024 9:57:31 AM -05	554.0 B	Estándar

Figura 1: Programas de Hive subidos al bucket de S3

## 2. Steps Corridos






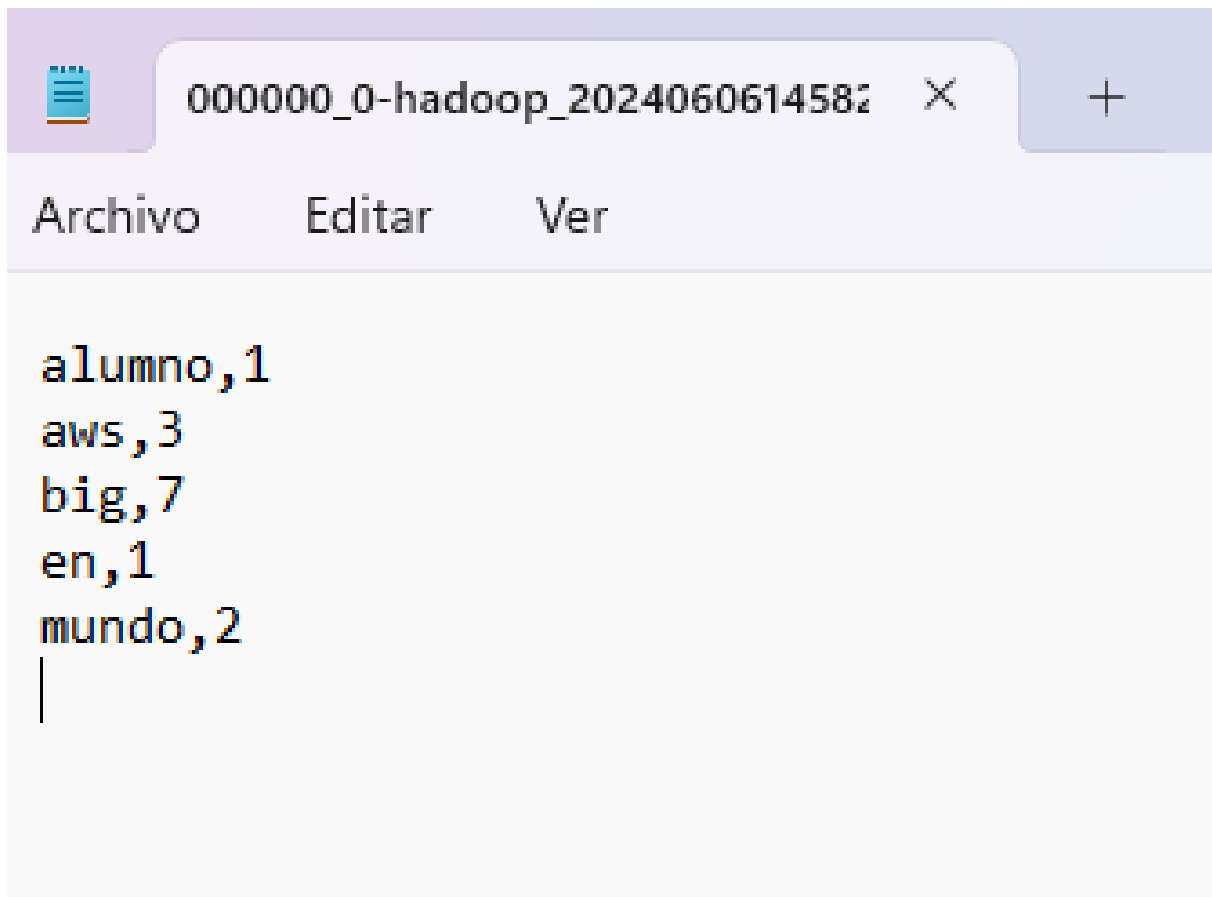
<input type="checkbox"/>	ID de paso ▼	Estado ▼	Nombre ▼	Archivos de registro 	Hora de creación (UT)
<input type="checkbox"/>	<a href="#">s-0387826WUXNR6J0P9JT</a>	<span>Failed</span>	Logs	No se han creado registros aún 	6 de junio de 2024, 11
<input type="checkbox"/>	<a href="#">s-09696843LV13LHAFI7GQ</a>	<span>Completed</span>	WordCount	<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr</a>   <a href="#">stdout</a> 	6 de junio de 2024, 0
<input type="checkbox"/>	<a href="#">s-058730617RCFBQWU5UPI</a>	<span>Completed</span>	WordCount	<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr</a>   <a href="#">stdout</a> 	6 de junio de 2024, 0
<input type="checkbox"/>	<a href="#">s-07394621HFC2BWLCE9GA</a>	<span>Failed</span>	WordCount	<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr</a>   <a href="#">stdout</a> 	6 de junio de 2024, 0

Figura 2: Steps

## 3. WordCount

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	<a href="#">_SUCCESS</a>	-	6 Jun 2024 9:58:41 AM -05	0 B	Estándar
<input type="checkbox"/>	<a href="#">000000_0-hadoop_20240606145820_006d9e17-d3ea-44a9-bda8-77a13f00372a-1</a>	-	6 Jun 2024 9:58:40 AM -05	34.0 B	Estándar
<input type="checkbox"/>	<a href="#">000001_0-hadoop_20240606145820_006d9e17-d3ea-44a9-bda8-77a13f00372a-1</a>	-	6 Jun 2024 9:58:41 AM -05	82.0 B	Estándar

Figura 3: Resultado de Word Count



```
alumno,1
aws,3
big,7
en,1
mundo,2
|
```

Figura 4: Resultado del Word Count descargado de AWS

4. Logs
5. Average
6. Rank