



**BIG DATA**  
**ACADEMY**

# LABORATORIO GESTIÓN DE CLÚSTERS

FORMADOR: ALONSO MELGAREJO  
[alonsoraulmgs@gmail.com](mailto:alonsoraulmgs@gmail.com)

## LABORATORIO CONFIGURACIÓN DE DATABRICKS

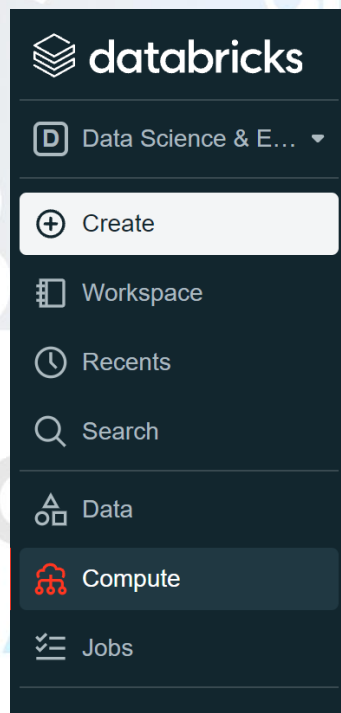
## 1. DESPLIEGUE DE CLÚSTER SPARK

**IMPORTANTE:** Durante cada clase, una hora antes de iniciar la clase, se enviará un correo para dejar encendiendo el clúster, ya que es un proceso que puede tomar entre 10 a 45 minutos.

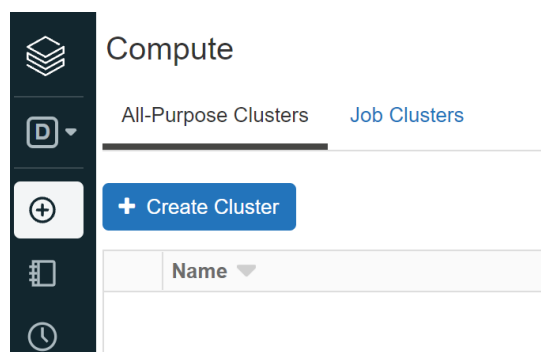
## 1.1. Entrar a la web de Databricks

<https://community.cloud.databricks.com/>

## 1.2. Seleccionar la opción "Compute"



## 1.3. Seleccionar la opción "Create Cluster"



- 1.4. Configuramos el tipo de clúster, seleccionaremos la versión de Scala y Spark más reciente y estable.

<b>Cluster Name</b>	BIG_DATA_ACADEMY
<b>Databricks Runtime Version</b>	Runtime: 8.3 (Scala 2.12, Spark 3.1.1)

Por defecto, en la edición gratuita Databricks te asignará un clúster de 15GB de RAM y 2 CPU.

Clic en “Create Cluster”

Create Cluster

New Cluster Cancel Create Cluster 0 Workers: 0 GB Memory, 0 Cores, 0 DBU  
1 Driver: 15.3 GB Memory, 2 Cores, 1 DBU

Cluster Name  
BIG\_DATA\_ACADEMY

Databricks Runtime Version   
Runtime: 8.3 (Scala 2.12, Spark 3.1.1)

**Note** Databricks Runtime 8.x and later use Delta Lake as the default table format. [Learn more](#)

Instance  
Free 15 GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For more configuration options, please upgrade your Databricks subscription.

Instances **Spark**

Availability Zone   
auto

- 1.5. El clúster comenzará a desplegarse, **este proceso puede tomar hasta 45 minutos.** Es recomendable hacerlo media hora antes de cada clase.

Clusters /

BIG\_DATA\_ACADEMY Edit Clone Restart Terminate Delete

**Configuration** **Notebooks** Libraries Event Log Spark UI Driver Logs Metrics Apps Spark Cluster UI - Master

Databricks Runtime Version  
8.3 (includes Apache Spark 3.1.1, Scala 2.12)

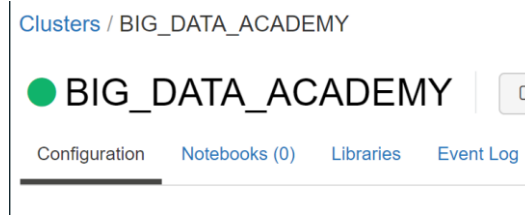
Driver Type  
Community Optimized 15.3 GB Memory, 2 Cores, 1 DBU

Instance  
Free 15 GB Memory: As a Community Edition user, your cluster will automatically terminate after an idle period of two hours. For more configuration options, please upgrade your Databricks subscription.

Instances **Spark** JDBC/ODBC Permissions

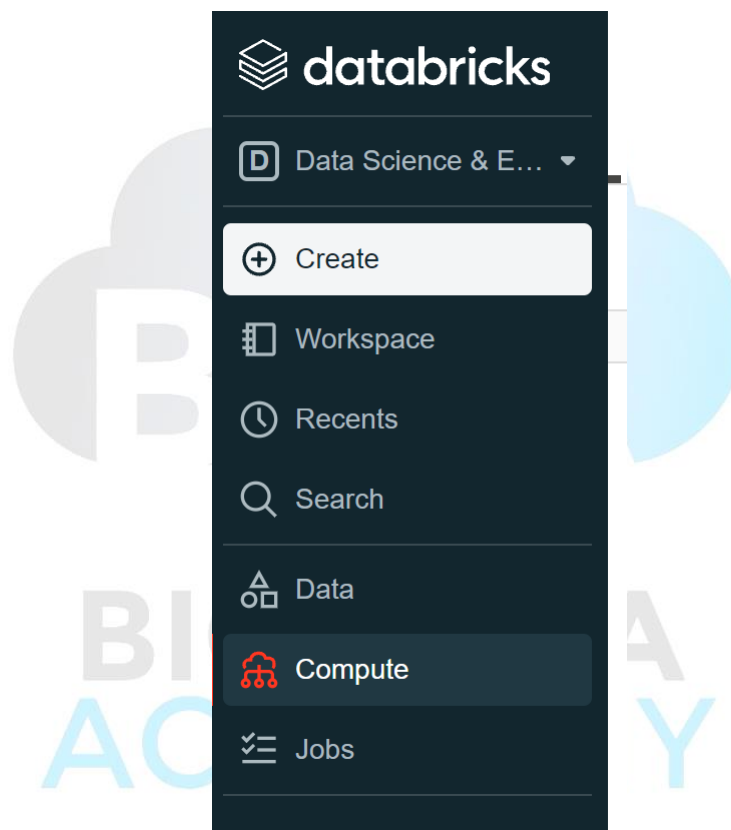
Availability Zone   
us-west-2a

Una vez esté desplegado, lo veremos de la siguiente manera:

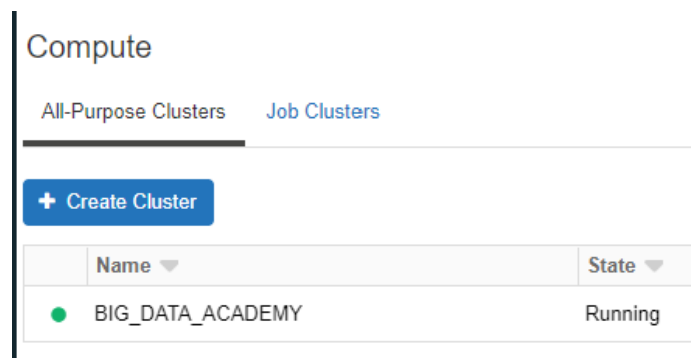


## 2. LISTAR CLÚSTER DISPONIBLES

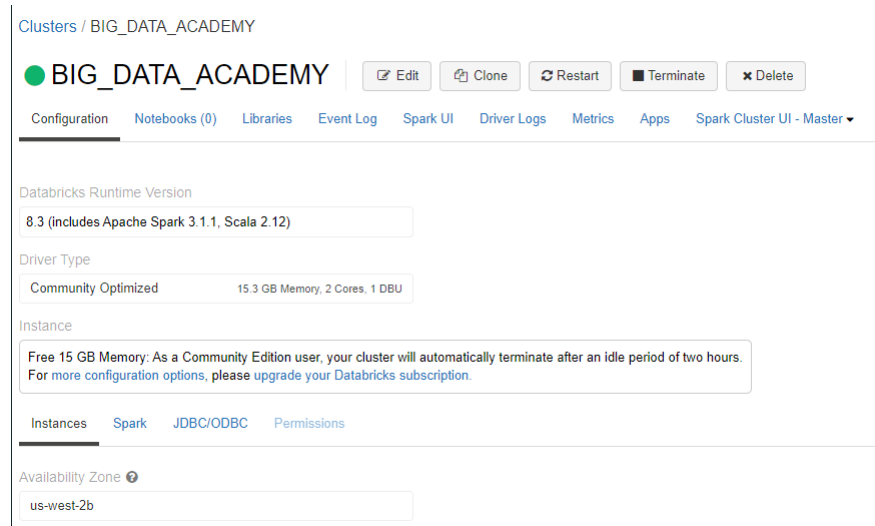
2.1. Podemos ver los clúster disponibles en la opción "Compute"



2.2. Se listarán los clúster desplegados



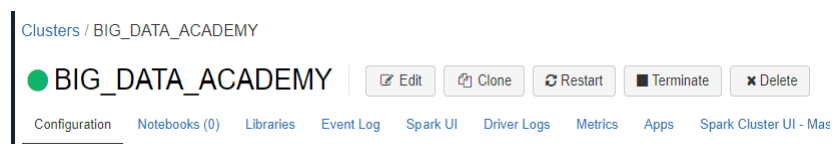
- 2.3. Si hacemos clic sobre el nombre del clúster, veremos los detalles y las opciones del clúster



**IMPORTANTE: Durante cada clase, una hora antes de iniciar la clase, se enviará un correo para dejar encendiendo el clúster, ya que es un proceso que puede tomar entre 10 a 45 minutos.**

### 3. DETENER CLÚSTER

- 3.1. Por defecto el clúster estará encendido por 2 horas, luego de eso se apagará de manera automática.
- 3.2. Si queremos detenerlo manualmente, deberemos seleccionar la opción “Terminate”



### 4. ELIMINAR CLÚSTER

- 4.1. Si queremos eliminar el clúster, deberemos seleccionar la opción “Delete”

