

Laboratorio

Módulo 02

Entorno

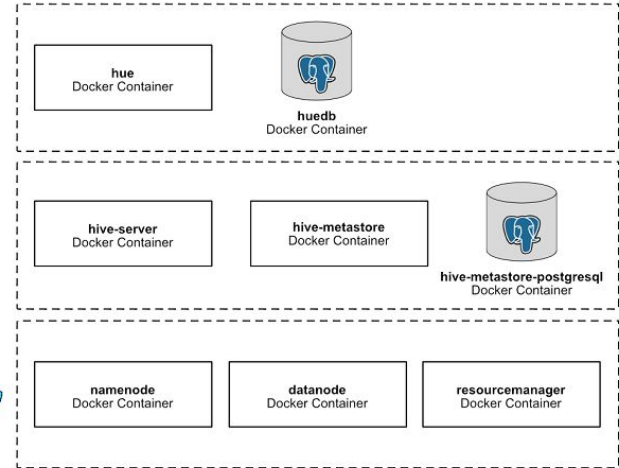
El instructor brindara un acceso para cada alumno a un entorno previamente configurado para realizar los ejercicios

En el caso que el alumno quisiera replicar este ambiente, debería descargar desde el repositorio:

[Docker Hadoop Hive Parquet](https://github.com/tech4242/docker-hadoop-hive-parquet.git)

Instrucciones para su configuración:

- `sudo apt install -y docker-compose`
- `git clone https://github.com/tech4242/docker-hadoop-hive-parquet.git`
- `cd docker-hadoop-hive-parquet/`
- `sudo docker-compose up`



Analizar Datos con Hive

Ingresar a Hue (ver imágenes)

http://<ip>:8888/hue

En la sección de archivos, cargar los archivos de la carpeta data y replicar la misma estructura de directorios en HDFS

En la sección de mis documentos, cargar el archivo clase-03.json y luego seleccionar el editor Hive.


Enlace sugerido para lectura:

Towardsdatascience : [Big-data-with-hadoop-hive-parquet-hue-and-docker](#)








Query. Explore. Repeat.

Iniciar sesión




Consulta ▾






Hive




Bases de datos (1) + ↻




Filter databases...





default





Archivos





Search data and saved documents...

Jobs  

Explorador de archivos

Buscar nombre de archivo



Acciones ▾

Eliminar permanentemente

Cargar

Nuevo ▾

[Inicio](#) / [user](#) / **instructor**

<input type="checkbox"/>	Nombre	Tamaño	Usuario	Grupo	Permisos	Fecha
<input type="checkbox"/>	 j		root	supergroup	drwxr-xr-x	June 21, 2020 02:53 PM
<input type="checkbox"/>	 .		instructor	instructor	drwxr-xr-x	June 21, 2020 03:14 PM

Mostrar





45 ▾


 de 0 elementos

Página




1








 de 1



Consulta ▾





< Hive

Bases de datos (1) + ↺

Filter databases...

default

Explorador de archivos

Buscar nombre de archivo

Acciones ▾

Eliminar permanentemente




Cargar

Nuevo ▾

Archivo

Directorio

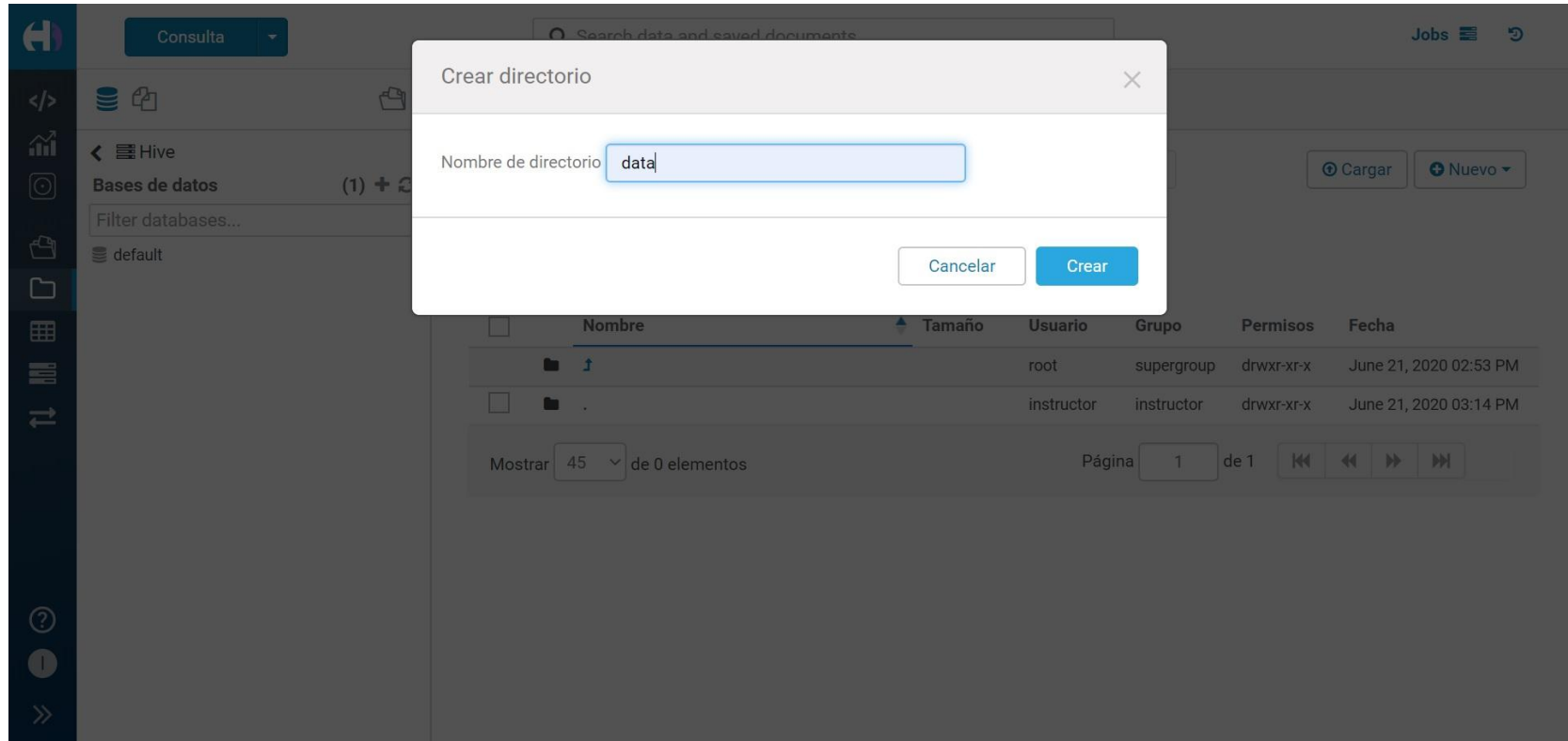
Inicio / user / instructor

<input type="checkbox"/>	Nombre	Tamaño	Usuario	Grupo	Permisos	Fecha
<input type="checkbox"/>	 		root	supergroup	drwxr-xr-x	June 21, 2020 02:53 PM
<input type="checkbox"/>	 .		instructor	instructor	drwxr-xr-x	June 21, 2020 03:14 PM

Mostrar 45 ▾ de 0 elementos

Página 1 de 1

⏮ ⏪ ⏩ ⏭



Consulta

Search data and saved documents

Jobs

Cargar a /user/instructor/data/bikeshare/stations

Seleccionar archivos o arrástrelos y suéltelos aquí

austin_bikeshare_stations.csv 5.7kB

Tablas (1) +

Filtrar...


access_log





	Nombre	Tamaño	Usuario	Grupo	Permisos	Fecha
<input type="checkbox"/>	↑		instructor	instructor	drwxr-xr-x	June 21, 2020 03:31 PM
<input type="checkbox"/>	.		instructor	instructor	drwxr-xr-x	June 21, 2020 03:51 PM


Mostrar 45 de 0 elementos


Página 1 de 1


Cargar Nuevo


 Mis documentos


 Papelera


Nombre	Descripción	Última mod
 Clase-02		29/06/202

 Copiar

 Mover a papelera

 Cambiar nombre de la carpeta

 Exportar

 Importar

The screenshot displays the Databricks workspace interface. On the left, a sidebar contains navigation icons for home, code editor, clusters, jobs, recent files, workspace, and help. The main area is divided into three panes:

- Left Pane:** Shows a table named 'flightdelays' with a filter bar.
- Center Pane:** A dropdown menu is open, showing options for 'Consulta' (top), 'Editor' (selected), and 'Programador'. The 'Editor' submenu is active, listing various languages and tools: Hive, Impala, Notebook, Pig, Java, Spark, MapReduce, Shell, Sqoop 1, Distcp, Scala, PySpark, R, and Spark Submit Jar.
- Right Pane:** Displays a code editor for a file named 'Clase-02'. The code is a Hive SQL query defining an external table 'flightdelays' with various integer and string columns. Below the code, a job execution log is visible, showing the upload of a file and the completion of a task.

At the top of the interface, there is a search bar labeled 'Search data and saved documents...' and a 'Jobs' link. The top right corner shows the file name 'Clase-02' and options to add a description or save the file.

¡Muchas gracias!

¡Sigamos trabajando!