

Título de la tarea: Administración y monitorización de sistemas Hadoop.

Curso de Especialización y módulo: Inteligencia Artificial y Big Data - Big Data Aplicado

## ¿Qué contenidos o resultados de aprendizaje trabajaremos?

### Resultados de aprendizaje

- ✓ **RA4.** Realiza el seguimiento de la monitorización de un sistema, asegurando la fiabilidad y estabilidad de los servicios que se proveen.

### Contenidos

- 1.- Introducción a la administración de Hadoop.
- 2.- Interfaz de HDFS: Namenode UI.
- 3.- Interfaz de YARN: ResourceManager UI.
- 4.- Apache Ambari.
  - 4.1.- Ejemplo: parada de nodos worker
- 5.- Cloudera Manager.
- 6.- Ganglia.
- 7.- Guía práctica de Spark en Databricks

Siguiente »

# 1.- Descripción de la tarea.



## Consultas en Apache Spark

En esta práctica aprenderás a realizar consultas con Apache Spark sobre un clúster de Azure Databricks. Apache Spark es probablemente la herramienta Big Data más utilizada. En vez de instalar Spark en tu equipo local, usaremos Databricks. Necesitarás una cuenta "community" de Databricks. Si no tienes esta cuenta, es el momento de registrarte. Únicamente tendrás que indicar tu nombre y tu correo electrónico, pero no es necesario introducir ni teléfono ni ninguna información bancaria.



[Pexels](#) (Dominio público)

Algunas de las consultas serán idénticas a las de las tareas anteriores. De esta forma podrás comparar tus consultas con las que realizaste con otras herramientas como MRJob, Pig o Hive.

### ¿Qué te pedimos que hagas?

Descarga el enlace al fichero zip inferior, descomprímelo e importa el fichero "Tarea BDA04.dbc" en Databricks. Para ello entra en [Community Databricks](#); en el menú lateral pulsa sobre "Workspaces", en el panel lateral que se abre, pulsa sobre botón derecho del ratón y luego "import". Una vez importado abre el "notebook" "Tarea BDA04.dbc" en Databricks y sigue las instrucciones.

Una vez que hayas completado las consultas del libro, sube únicamente el fichero "dbc".

- [Tarea BDA04.zip \(Ventana nueva\)](#)

#### NOTA IMPORTANTE

El fichero entregado debe contener las celdas ejecutadas y con el comentario "TODO" eliminado. Los ejercicios son individuales. Pon tu nombre en la primera celda.

## 2.- Información de interés.



### Recursos necesarios y recomendaciones

#### Recursos necesarios

- ✓ Conexión a Internet para consultar la información o investigar sobre las cuestiones planteadas.
- ✓ Navegador web.
- ✓ Software para comprimir los archivos de la tarea.

#### Recomendaciones

- ✓ Antes de abordar la tarea:
  - Revisa la guía práctica y realiza paso a paso lo que allí se indica.
  - Si tienes problemas, escribe en el foro.
- ✓ No olvides subir el fichero "dbc" con las celdas ejecutadas.



### Indicaciones de entrega

Una vez realizada la tarea, el envío se realizará a través de la plataforma. El archivo se nombrará siguiendo las siguientes pautas:

**Apellido1\_Apellido2\_Nombre\_BDA04\_Tarea**

Asegúrate que el nombre no contenga la letra ñ, tildes ni caracteres especiales extraños. Así por ejemplo la alumna **Begoña Sánchez Mañas para la cuarta unidad del MP de BDA04**, debería nombrar esta tarea como...

**sanchez\_manas\_begona\_BDA04\_Tarea**

« Anterior   Siguiente »

### 3.- Evaluación de la tarea.

#### Criterios de evaluación implicados

**Criterios de evaluación RA4**

- ✓ a) Se han aplicado herramientas de monitorización eficiente de los recursos.
- ✓ b) Se han recogido métricas, procesamiento y visualización de los datos.
- ✓ c) Se han generado alertas para detectar un riesgo o mal funcionamiento.
- ✓ f) Se ha analizado la estabilidad de servicios.

#### ¿Cómo valoramos y puntuamos tu tarea?

Rúbrica de la tarea	
Pregunta 1	2 puntos
Pregunta 2	2 puntos
Pregunta 3	2 puntos
Pregunta 4	2 puntos
Pregunta 5	2 puntos
Redacción clara y correcta, sin errores ortográficos	Se resta 0,1 puntos por cada error ortográfico o expresiones incorrectas.

[« Anterior](#)