**Instrucciones**

**Sumario**

**¡Bienvenido(a)!**

Te invitamos a realizar una actividad para que pongas en práctica lo aprendido en el curso.

**Características de la evaluación**

* **Tipo de actividad:** grupal
* **Tipo de evaluación:** Sumativa (con calificación)
* **Ponderación:** 12% (Equivale al 12% de la nota final del curso)
* **Puntaje:** 12 puntos

**Instrumento de evaluación**

**Descarga el instrumento de evaluación y revísalo antes de realizar la actividad.**

**[Pauta de evaluación t1 .pdf](https://d3c33hcgiwev3.cloudfront.net/s6og_rpqStmnojVRXrGUEA_7e63033a9d144588941066957fe719f1_Pauta-de-evaluacion-t1-.pdf?Expires=1685491200&Signature=JAobCYg0LL4LWOsZiiAys~Hs1QxrrRl2pkWXn9doW9biV1XVq85sxlvUuCZjWsIMQjfesO4C6r~~5JTEf4J3GxbPz7~SggHmly9XyWeIs4azskw5~Kks2wNgo2oNAKNF-ufdykvFKEvNpyf1hJYontwaxZDaKntYRkAuH7kWl8E_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)**

[PDF File](https://d3c33hcgiwev3.cloudfront.net/s6og_rpqStmnojVRXrGUEA_7e63033a9d144588941066957fe719f1_Pauta-de-evaluacion-t1-.pdf?Expires=1685491200&Signature=JAobCYg0LL4LWOsZiiAys~Hs1QxrrRl2pkWXn9doW9biV1XVq85sxlvUuCZjWsIMQjfesO4C6r~~5JTEf4J3GxbPz7~SggHmly9XyWeIs4azskw5~Kks2wNgo2oNAKNF-ufdykvFKEvNpyf1hJYontwaxZDaKntYRkAuH7kWl8E_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**¿Cómo realizar la tarea?**

**1**. Para realizar la tarea, debes haber revisado los siguientes materiales:

**Clases**

* Map Reduce
* Algoritmos con Map Reduce
* Tutorial : Introducción a Databricks
* Tutorial: Más sobre Apache Spark

**Lecturas**

* Mining of Massive Datasets

**2**.Descarga y revisa los **dos datasets** que usarás en la tarea:

* **personas.csv** (tiene los campos id\_persona, nombre\_persona, direccion, edad)

**[personas (1)](https://d3c33hcgiwev3.cloudfront.net/jXlVR5J3SCyUIH1Y6MtDlQ_4b44c7f09cc845b98ada71a551f8e3f1_personas-1-.csv?Expires=1685491200&Signature=LcWiJ-6IH~jTKMkiCH7JcoPpcKsg8GPVn2gS-~UOpdYWkP4iyvAeu-vKzHPz5VS8aj9ABuEVY20QHdXLw~X9CNvdTJ4GyM2sbI6PZP29eL5I2KdLOMq~RdNDZFCF7W1Rxvymw1J3WV0EO6wXnW3FhmI--ZauiADB69FSS3-8KZU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)**

[CSV File](https://d3c33hcgiwev3.cloudfront.net/jXlVR5J3SCyUIH1Y6MtDlQ_4b44c7f09cc845b98ada71a551f8e3f1_personas-1-.csv?Expires=1685491200&Signature=LcWiJ-6IH~jTKMkiCH7JcoPpcKsg8GPVn2gS-~UOpdYWkP4iyvAeu-vKzHPz5VS8aj9ABuEVY20QHdXLw~X9CNvdTJ4GyM2sbI6PZP29eL5I2KdLOMq~RdNDZFCF7W1Rxvymw1J3WV0EO6wXnW3FhmI--ZauiADB69FSS3-8KZU_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

* **mascotas.csv (**tiene los campos id\_mascota, nombre\_mascota, tipo\_mascota, id\_dueño)

**[mascotas (1)](https://d3c33hcgiwev3.cloudfront.net/PEYAyBI3RZm3M6kd5N-QvA_09f4aa1da12041a291081a50f78f6ff1_mascotas-1-.csv?Expires=1685491200&Signature=FfIqDTKBt5zLIGXy0ikJHPnSM8oMWgHewwFsnqE9RRselFleFvcVse8EiRAHTYSJQfKcpDyX0qOfVEP6lnqumPTRUqXNGizmyDGbRVBleD8hMu~JRaIEcLdn~oJRMUnljmiJHKwqAo5d~U24tbV17dAhlqu3xyDWWsfo~-ttt2w_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)**

[CSV File](https://d3c33hcgiwev3.cloudfront.net/PEYAyBI3RZm3M6kd5N-QvA_09f4aa1da12041a291081a50f78f6ff1_mascotas-1-.csv?Expires=1685491200&Signature=FfIqDTKBt5zLIGXy0ikJHPnSM8oMWgHewwFsnqE9RRselFleFvcVse8EiRAHTYSJQfKcpDyX0qOfVEP6lnqumPTRUqXNGizmyDGbRVBleD8hMu~JRaIEcLdn~oJRMUnljmiJHKwqAo5d~U24tbV17dAhlqu3xyDWWsfo~-ttt2w_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A" \t "_blank)

**3**. La idea de esta actividad es que ocupes los operadores básicos de Apache Spark para simular consultas SQL.

**Criterios de revisión**

**Aspectos formales**

Considera los aspectos formales que se describen a continuación:

* El formato de entrega es un **pdf que contenga los programas.**
* El archivo debe tener el nombre “Tarea 1{Nombre1 Apellido1}{Nombre2 Apellido2}.pdf”

### \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Pregunta 1

**Preguntas**

**Esta actividad requiere que utilices los operadores básicos de Apache Spark ('map', 'reduceByKey', 'filter', 'join', etc.) para simular consultas SQL.**

1. Utiliza pyspark para cargar cada archivo como una "tabla", tal como lo hicimos en el tutorial. Recuerda que los archivos están disponibles en las instrucciones.

2. A partir de ahí, debes realizar las siguientes tareas:

* **Parte I:** Crea un programa que devuelve cada nombre de mascota y su tipo, junto con el nombre de su dueño. Es decir, tu programa debe entregar, para cada mascota en tus datos, el nombre de la mascota, el tipo de mascota (por ejemplo, perro, gato, etc.) y el nombre del dueño de la mascota.
* **Parte 2:** Crea un programa que devuelve, para cada número identificador y nombre de persona, el número de mascotas que posee. En otras palabras, para cada persona en tus datos, tu programa debería mostrar el ID de la persona, su nombre, y el número total de mascotas que esa persona tiene.

3. Queda prohibido el uso de Spark SQL y Data Frames en tu programa.

**Para enviar tus respuestas**

* Presiona el botón para cargar archivo. Se abrirá una nueva ventana que permite arrastrar o adjuntar el archivo.
* Comprueba que el archivo arrastrado es el correcto, ya que el documento quedará guardado en la plataforma.

*/FileStore/tables/mascotas.csv*

*/FileStore/tables/personas.csv*