

# Instructions

## Overview

### ¡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:** 10% (Equivale al 10% de la nota final del curso)
- **Puntaje:** 12 puntos

## Instrumento de evaluación

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

## ¿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)**

CSV File

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



**mascotas (1)**

CSV File

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

## Review Criteria

### 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”