

UNIDAD 2: MODELO DE MAPEO Y REDUCCIÓN

PRÁCTICA: RESILIENT DISTRIBUTED DATASET

Blanca Vázquez

Febrero 2020

Aprender a usar la abstracción principal de datos de Spark (RDD) a través de la generación de transformaciones y acciones.

- Ejercicio 1: procesamiento de lenguaje natural para el análisis de sentimientos usando Apache Spark

1. Crear clúster
2. Cargar datos
3. Crear notebook usando el mismo menú de 'data' y adjuntamos clúster
4. Codificamos
5. Guardamos la tabla de resultados
6. Exportamos el código en un archivo IPython Notebook
7. Terminamos la ejecución del clúster