

Al igual que en el caso de MapReduce, el ejercicio elegido para hacer la prueba de ejecución en paralelo es el de las películas, por la dimensión de sus datos.

Ejecución normal: Al ejecutar el programa en Spark localmente desde Ubuntu se tarda 3:17:86 en obtener los datos. Como se puede ver es mucho más eficiente que haciéndolo con MapReduce.

Ejecución paralelizada: Ocurre lo mismo que en el caso de MapReduce, aumentar el número de nodos sobre los que corre el programa únicamente empeora su rendimiento. El número de datos no es lo suficientemente grande para que resulte beneficioso para el programa paralelizarlo en varios nodos.