

## ACTIVIDAD DE CLASE

### HADOOP vs SPARK

Completa y responde según el [vídeo](#)

#### HADOOP

Nombra las tres partes	¿Cuál es su nombre en Hadoop?
1: Almacenamiento	HDFS
2: Reducción y procesado	MapReduce
3: Gestión de los Recursos	Hadoop Yarn

1. **¿Qué ocurre si uno de los bloques se corrompe? ¿Se pierden los datos? ¿Por qué?**

No, se hacen copias en otros nodos por replicación y por eso es tolerante a fallos

#### SPARK

2. **¿En qué se basa?** En los RDD, lo hace tolerante a fallos y es más rápido
- ¿Qué ocurre si se pierde un nodo?** Hace copias como Hadoop
  - ¿A partir de qué puede ser creado?** A partir de un archivo de texto, base de datos, hdfs de hadoop
  - ¿Es compatible con MapReduce?** Sí
  - ¿Qué más operaciones puede hacer?** Input, Separación, Orden
    - ¿Dónde hace estas operaciones?** En RAM
  - ¿Los usuarios son los que se encargan de gestionar los recursos?**  
No
3. **¿Qué es un DAG?** Grafos Acíclicos Dirigidos
4. **¿Cuál es la mejor característica que tiene Spark? ¿Gracias a qué ocurre esto?**  
Su velocidad, gracias a RDD y la computación de memoria
5. **¿Cuál es su mayor desventaja?**  
La cantidad de memoria RAM que necesita
- 

6. **¿Qué tienen Hadoop y Spark en común?**

Ambos son tolerantes a fallos, usan un procesamiento por lotes

7. **¿Cómo se trabaja en la actualidad en cuanto a los dos sistemas?**

Es común usar los dos, Hadoop para el almacenamiento y Spark en las capas de procesamiento.