

Almacenamiento EMR - HDFS

- **HDFS (Hadoop Distributed File System)**

- Múltiples copias almacenadas en instancias del clúster para redundancia
- Archivos almacenados como bloques (tamaño por defecto de 128 MB)
- Los datos en HDFS se pierden cuando el clúster se termina (efímeros)
- Útil para el almacenamiento intermedio o cargas de trabajo con I/O aleatorio significativo
- Hadoop intenta procesar los datos donde están almacenados en HDFS