

# Almacenamiento EMR

- Uno de los aspectos clave de Amazon EMR es cómo maneja el almacenamiento de datos
- Hay varias opciones disponibles para almacenar y gestionar datos en un clúster de EMR:
  - **EBS (Elastic Block Store)**: EBS proporciona almacenamiento en bloque que puede ser utilizado por instancias EC2 dentro de un clúster de EMR
  - **HDFS (Hadoop Distributed File System)**: HDFS es el sistema de archivos distribuido utilizado por Hadoop para almacenar grandes volúmenes de datos en clústeres
  - **EMRFS (EMR File System)**: EMRFS es una implementación de Hadoop Compatible File System (HCFS) que permite a EMR acceder a Amazon S3 como si fuera un sistema de archivos Hadoop
  - **Discos Locales (Local Disks)**: Además de HDFS y S3, EMR puedes utilizar los discos locales de las instancias EC2
  - **Glue Data Catalog**: AWS Glue Data Catalog es un servicio que permite la gestión y catalogación de metadatos para los datos almacenados en S3 y otras fuentes