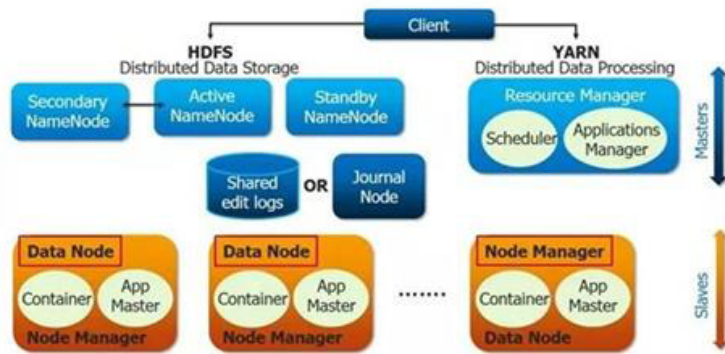


# Hadoop 3

Nueva versión de Hadoop, en la que se aplica un nuevo mecanismo llamado codificación de borrado para proporcionar tolerancia a fallos, ya no se usa la replicación de HDFS (se reduce la sobrecarga de almacenamiento anterior que era de un 200% a un 50%).

## Apache Hadoop 2.0 and YARN



### - Arquitectura

Se han realizado cambios en los componentes de la arquitectura de HDFS

- **Extensión del NameNode**

Nuevo protocolo denominado bloque jerárquico.

- **Extensión del Cliente**

Extensión de la lógica entrada/salida.

- + DFSStripedInputStream
- + DFSStripedOutputStream

- **Extensión del DataNode**

Ejecuta un ErasureCodingWorker adicional para recuperar bloques codificados por borrado fallido, lee los datos de los nodos de origen mediante subprocesos dedicados, codifica los datos y genera una salida.

- **Políticas de codificación de borrado HDFS**

Mezcla cómo codificar y decodificar un archivo.

- **Job History Server**

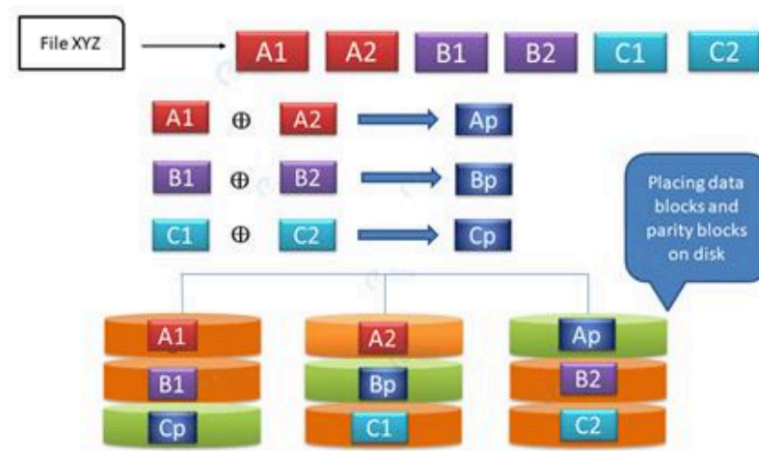
Proceso opcional cuya función es la de mantener un archivo de logs del job (hay un solo JHS por clúster)

## - Mejoras

- + YARN Timeline Services v.2.
- + Soporte para contenedores oportunistas y programación distribuida.
- + Soporte de más de dos NameNodes.
- + Cambios en los puertos predeterminados de varios servicios.
- + Gestión de daemon y task heap reelaborada.
- + Generalización del modelo de recursos de YARN.
- + Consistencia y almacenamiento en caché de metadatos para el cliente S3A.

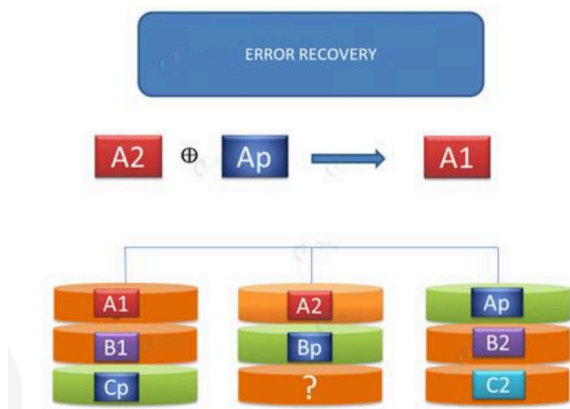
## - Funcionamiento

Se calcula el bloque de paridad y se almacena junto con los otros bloques en el disco.



Si por ejemplo, A1 se pierde.

Recuperamos A mediante el bloque de paridad Ap:



## - Ventajas

- + Soporte EC en línea
- + Baja sobrecarga
- + Recuperación bidireccional
- + Compatibilidad
- + Simplificación de operaciones