



HADOOP YARN Y ARQUITECTURA DE CLÚSTER

Profesores:

Miguel Angel Sánchez Hernández
Omar Mendoza González

Alumna:

Belem Anahi Mendieta Hernández

CONTENIDO

1. Introducción
2. ¿Qué es YARN?
3. Componentes principales
4. Flujo de ejecución
5. Arquitectura de clúster Hadoop
6. Ventajas y desafíos de YARN

De Hadoop 1 a Hadoop YARN

En Hadoop 1.x, el JobTracker y TaskTracker se encargaban de coordinar y ejecutar las tareas de MapReduce.

Limitación 1

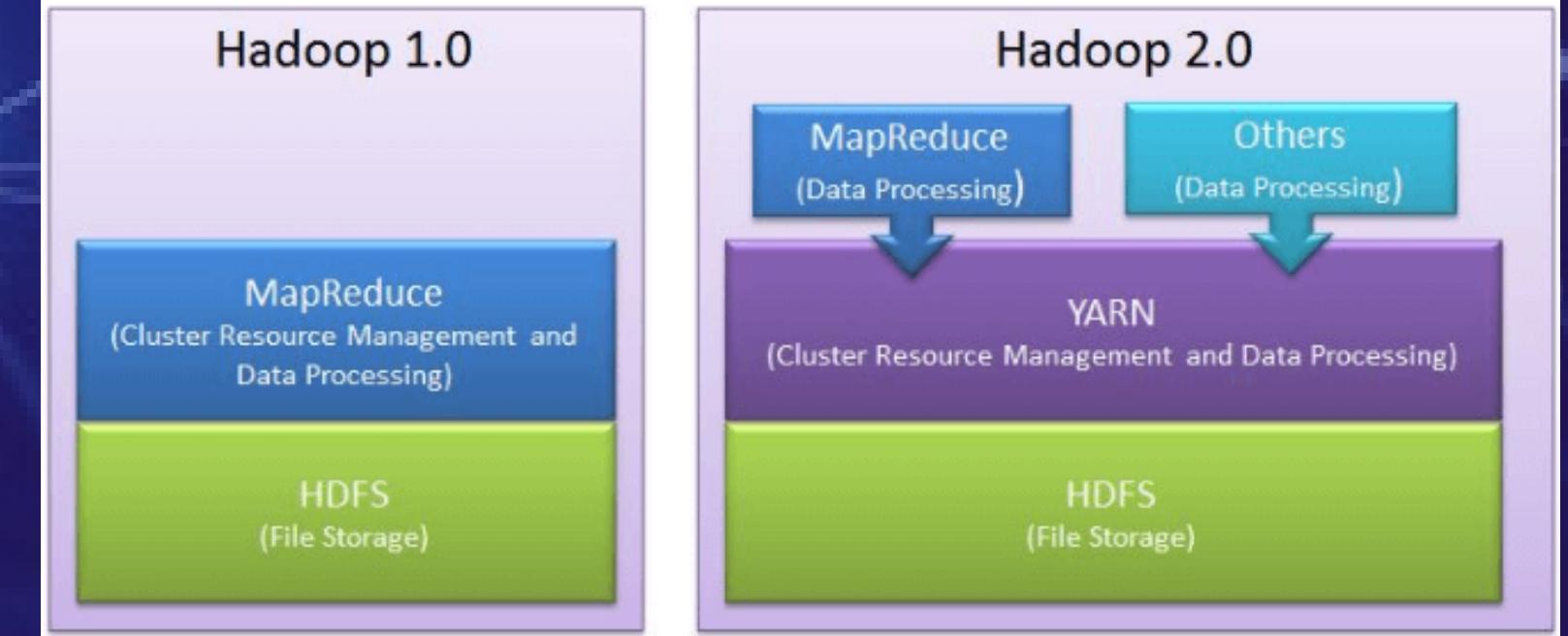
El JobTracker manejaba demasiadas funciones (planificación, monitoreo, fallos).

Limitación 2

No se podían ejecutar otros motores de procesamiento además de MapReduce.

Limitación 3

Escalaba mal en clústeres grandes.

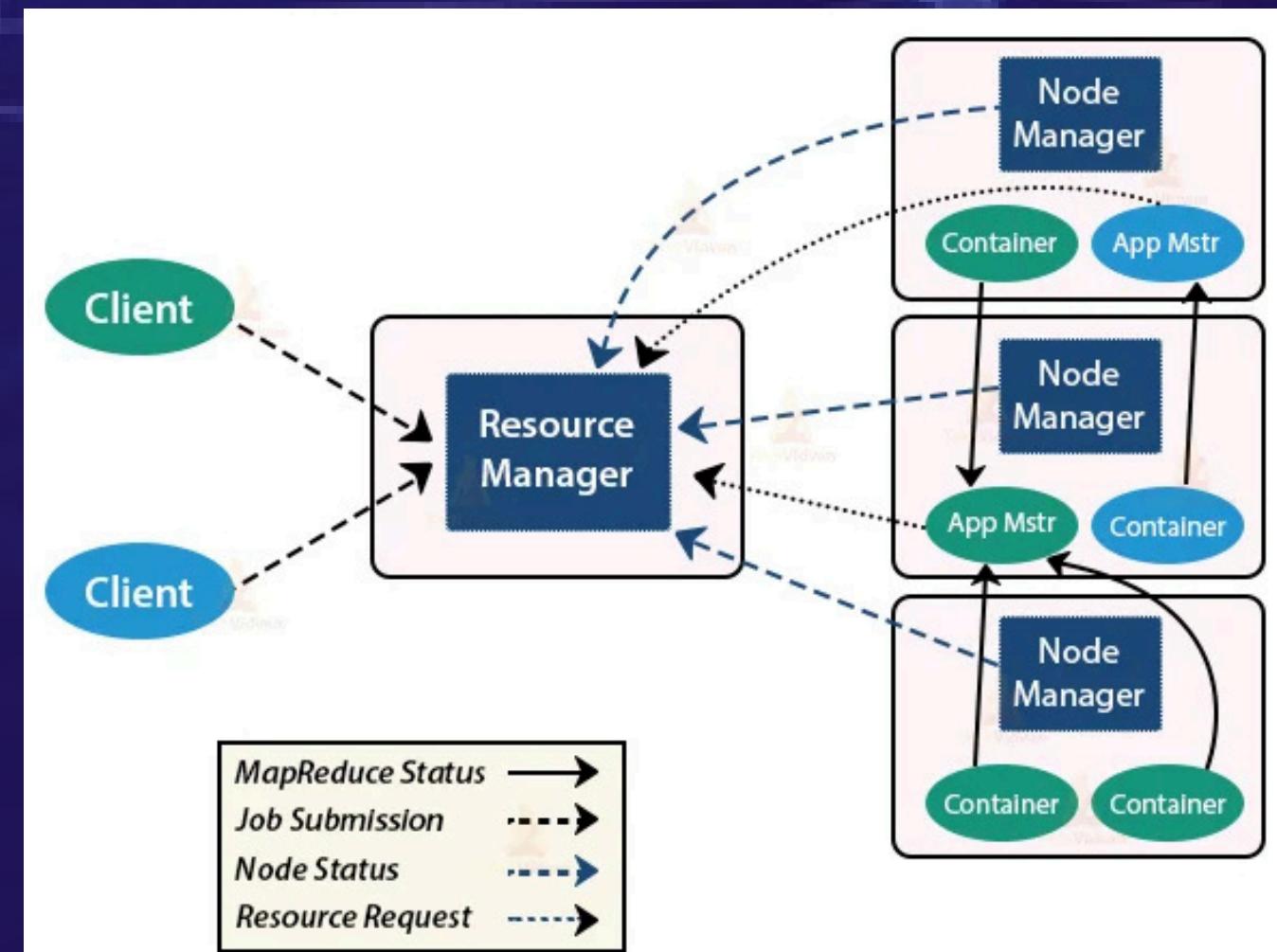


Para resolverlo, Hadoop 2 introdujo **YARN** (*Yet Another Resource Negotiator*).

¿Qué es YARN?

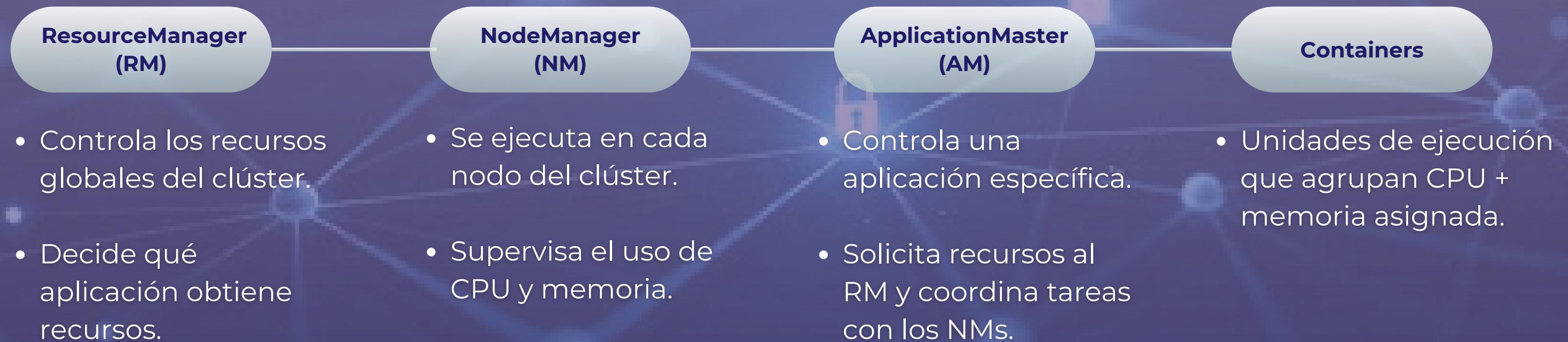
Es el sistema de gestión de recursos y programación de tareas de Hadoop. Separa la gestión de recursos del modelo de procesamiento, lo que permite ejecutar diferentes motores sobre el mismo clúster.

- 1 Coordina cómo se usan CPU, memoria y red entre aplicaciones.
- 2 Permite que varios usuarios compartan el mismo clúster de forma segura.
- 3 Hace que Hadoop sea una plataforma multitarea y multipropósito.

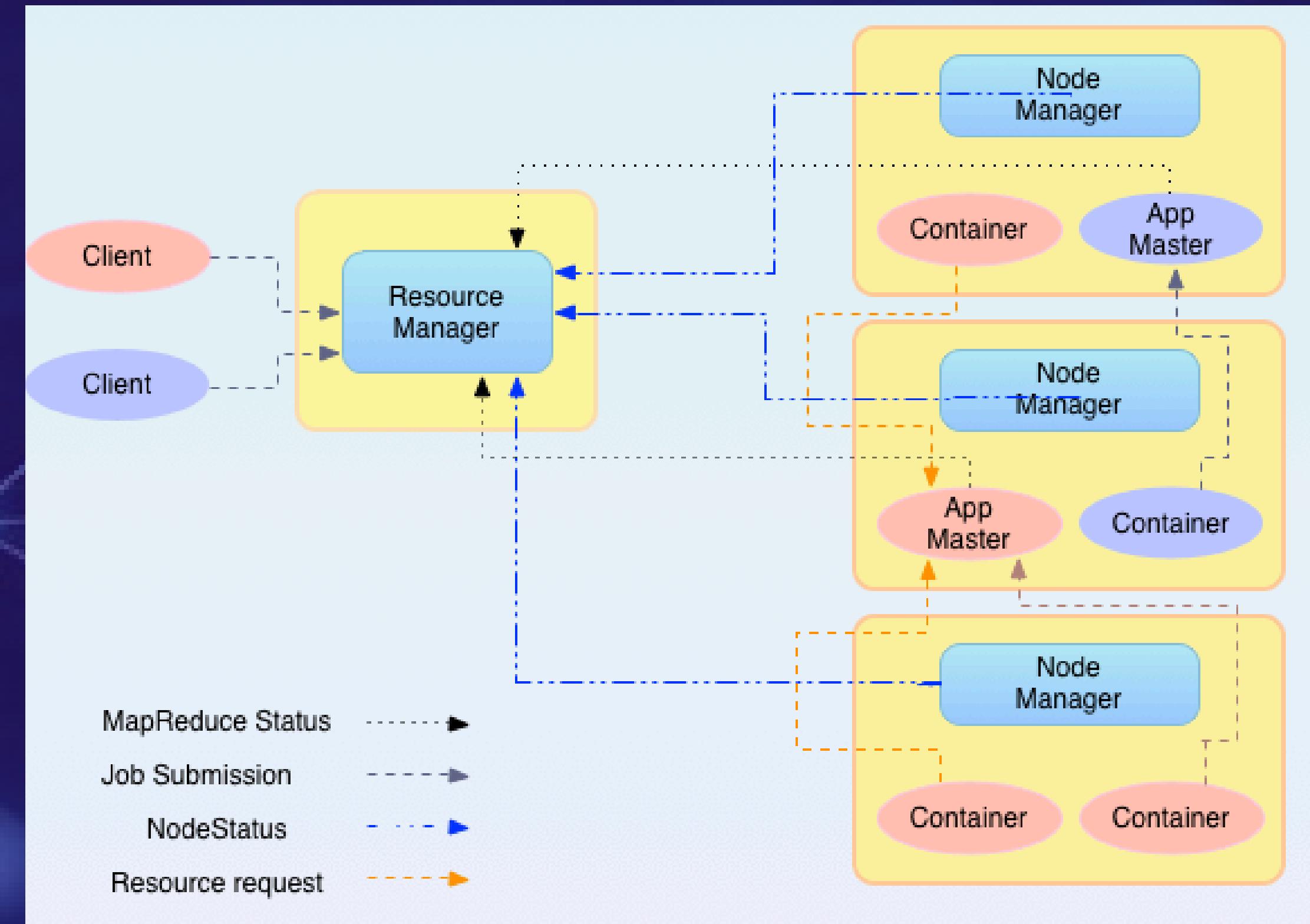


Componentes principales

Arquitectura YARN



Flujo de ejecución



Arquitectura de clúster Hadoop

Nodo de nombres



Almacena los metadatos de los archivos, es el único punto de control.

Nodos de datos



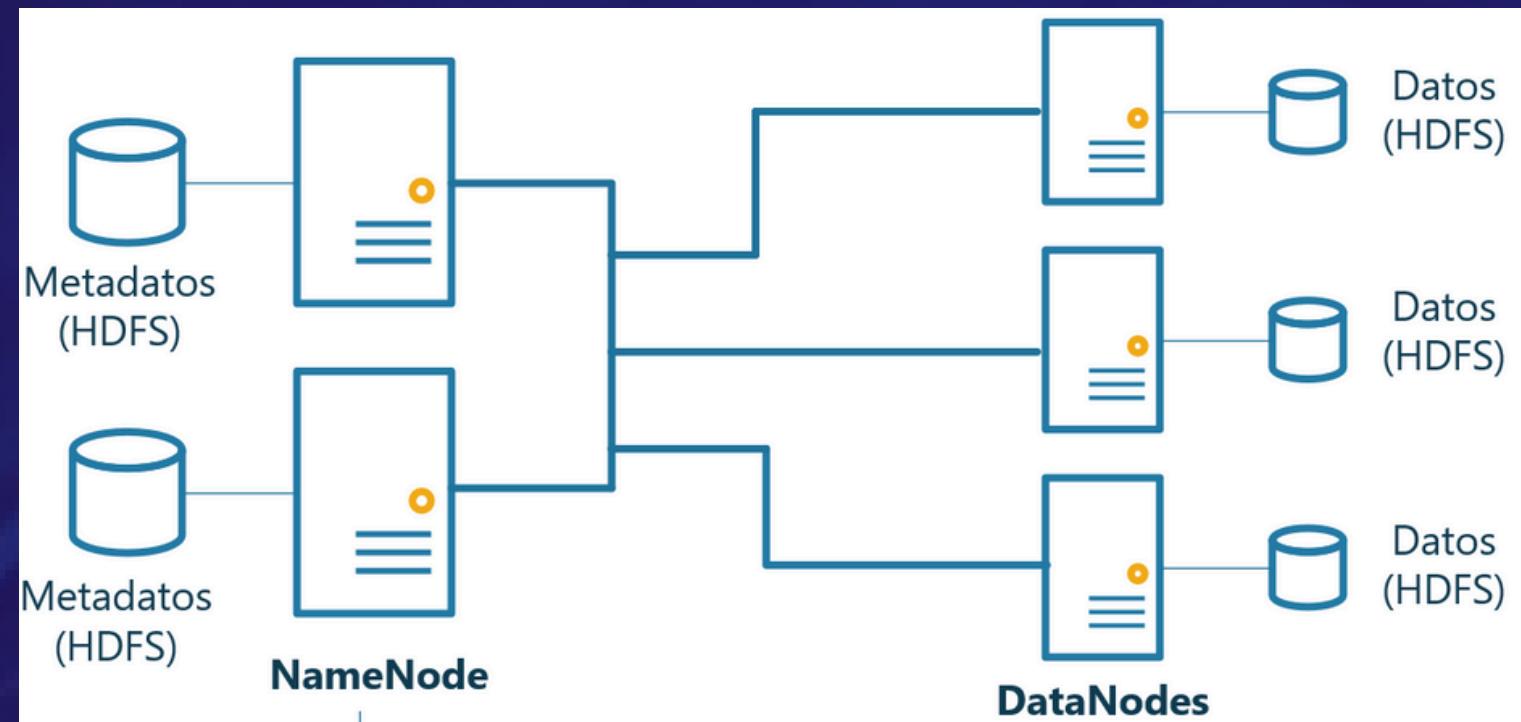
Servidores donde se almacena la información en forma de bloques

Red Interna



Comunica los componentes para compartir datos y recursos.

Ventajas y desafíos de YARN



Ventajas	Desafíos
Soporta múltiples motores	Configuración avanzada
Uso eficiente de recursos	Curva de aprendizaje alta
Escalabilidad horizontal	Requiere monitoreo constante
Mejor utilización de clústeres compartidos	Complejidad en la administración