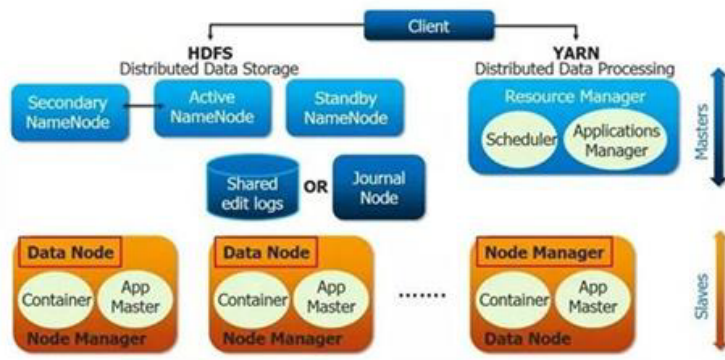


# Yarn

Incluida en Hadoop2, se encarga de asignar los recursos del sistema a las diversas aplicaciones que se ejecutan en un clúster Hadoop y programar tareas para que se ejecuten en diferentes nodos.

## Apache Hadoop 2.0 and YARN



### - Componentes

- **Resource Manager**

Es el responsable de la asignación de recursos y de optimizar la utilización del clúster.

Compuesto por dos componentes principales:

- Scheduler

Asigna los recursos a las aplicaciones en ejecución según su disponibilidad y sus políticas de planificación. Sólo se enfoca en la asignación de recursos.

- Application Manager

Gestiona el ciclo de vida de las aplicaciones que se ejecuten en el clúster. Cada aplicación tiene su propio Application Manager que negocia con el Resource Manager para obtener recursos y coordina la ejecución de las tareas.

- **Node Manager**

- + Se ocupa de los nodos individuales de Hadoop y de gestionar los trabajos de usuario y el flujo de trabajo en el nodo dado.
- + Se conecta con el RM y le envía HeartBeats con el estado de los nodos.
- + Administra los contenedores de aplicaciones asignados por el RM y que posteriormente son solicitados por el Aplicación Master.
- + Supervisa el uso de los recursos de los contenedores individuales.
- + Realiza la gestión de registros.
- + Se encarga de destruir el contenido cuando el RM lo indica.

- **Application Master**

- + Negocia los recursos apropiados con el RM y monitoriza su estado y progreso (de los recursos), también se encarga de gestionar la ejecución de todas las tareas en las que puede dividirse su aplicación.
- + Gestiona los fallos producidos.
- + Responsable de los contenedores de recursos adecuados con el RM para posteriormente rastrear su estado y monitorear el progreso.
- + Envía HeartBeats al RM para confirmar su estado y actualizar el registro de recursos.

- **Container**

- + Conjunto de recursos físicos dentro de un solo nodo(RAM, CPU, discos...)
- + Estos contenedores son administrados por un contexto de lanzamiento de contenedor(conjunto de configuraciones y parámetros que especifican cómo se debe iniciar un contenedor), que es su ciclo de vida.
- + Otorga derechos a una aplicación para utilizar una cantidad específica de recursos en un host específico.

- **Job History Server**

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

- **Envío de solicitud Yarn**

1. Envío de trabajo.
2. Obtener ID de la aplicación.
3. Contexto de envío de la solicitud.
4. Contenedor de inicio de lanzamiento e inicio de aplicación.
5. Asignación de recursos.
6. Contenedor.
7. Lanzamiento.
8. Ejecución.

- **Flujo de trabajo de la aplicación Hadoop Yarn**

1. El cliente envía una solicitud.
2. Resource Manager asigna un contenedor para iniciar el Application Manager.
3. Application Manager se conecta con Resource Manager.
4. Application Manager solicita los contenedores de Resource Manager.
5. El Application Master notifica al Node Manager que inicie sus contenedores.
6. El código de la aplicación se ejecuta en el contenedor.
7. El cliente se pone en contacto con Application Master/ Resource Manager para supervisar el estado de la aplicación.
8. Application Manager cancela el registro con Resource Manager.