

PR_02.1

Instalación de Apache Ambari

Apache Ambari es una herramienta **open source** desarrollada originalmente por **Hortonworks** (dentro del ecosistema Apache) que sirve para **instalar, configurar, gestionar y monitorear clústeres de Hadoop y sus componentes**.

En palabras simples, **Ambari es un panel de control para Hadoop**.

Funciones principales de Apache Ambari

1. Instalación y configuración automatizada

- Facilita la instalación de los diferentes componentes de Hadoop (HDFS, YARN, Hive, HBase, Spark, etc.) en múltiples nodos.
- Permite configurar servicios mediante una interfaz web en lugar de editar manualmente archivos XML de configuración.

2. Gestión centralizada del clúster

- Permite **iniciar, detener, reiniciar o reconfigurar servicios** desde una consola central.
- Ofrece gestión de usuarios y roles.

3. Monitoreo y alertas

- Muestra métricas de rendimiento, uso de CPU, memoria, disco, y estado de los servicios.
 - Integra alertas automáticas cuando un servicio falla o el rendimiento cae.
-

Arquitectura básica

- **Ambari Server:** componente central que gestiona todo el clúster.
 - **Ambari Agent:** corre en cada nodo del clúster y ejecuta las órdenes que recibe del servidor.
 - **Ambari Web UI:** interfaz gráfica accesible vía navegador para administración y monitoreo.
-

Historia y situación actual

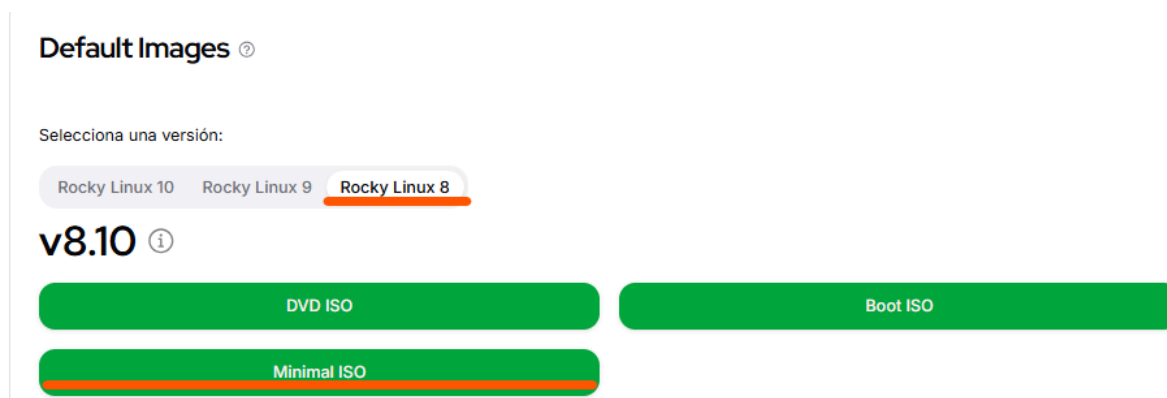
- Fue **desarrollado y mantenido por Hortonworks** como parte de *Hortonworks Data Platform (HDP)*.
- Tras la fusión de Hortonworks con Cloudera (2019), **Ambari dejó de desarrollarse activamente** dentro del ecosistema Cloudera.
- Cloudera, en su nueva plataforma **Cloudera Data Platform (CDP)**, **ya no usa Ambari**, sino **Cloudera Manager** para la administración de clústeres.
- Ambari en estos momentos no depende de Cloudera sino de la Apache Foundation.

Sistema operativo base

Dado que CentOS 7 llega a fin de soporte en junio de 2024, las mejores alternativas actualmente son:

Rocky Linux 8.x / AlmaLinux 8.x

Nosotros usaremos Rocky Linux: <https://rockylinux.org/es-ES/download>



Instalación de Rocky Linux

- Vamos a intentar instalar un *clúster* de Hadoop utilizando máquina virtuales. Será una actividad en grupo (4 alumnos de cada fila por grupo). Crearemos un *clúster* por cada fila de mesas en el aula. es decir, constará de 4 equipos.
- Para ello descargaremos e instalaremos **Rocky Linux**, con VirtualBox en cada uno de nuestros equipos (como indica la imagen de arriba, la versión **Minimal**). **Rocky Linux** está basado en Red Hat, por lo que algunos comandos y configuraciones pueden cambiar respecto a Ubuntu.

- A cada máquina virtual les pondremos 8 GB de memoria RAM y 150 GB de disco duro.
- Le pondremos como nombre de usuario `alumno` y contraseña `alumno`
- Para nombrar las máquinas utilizaremos la siguiente convención:
Comenzando siempre desde la ventana del aula
1ª fila `nodo11` , `nodo12` , `nodo13` , `nodo14`
2ª fila `nodo21` , `nodo22` , `nodo23` , `nodo24`
...y así sucesivamente
- Uno de los ordenadores ha de hacer de `master` (servidor -Ambari Server-) y los demás de `workers` (nodos -Ambari Agents-) Escoger como `master` mejor uno de los del centro, el terminado en 2 o 3.
- Como hicimos en la práctica con Ubuntu, tendréis que ponerles IP's estáticas a los equipos. Seguid la misma convención que ya vimos allí: IP: X.X.X.**201** en el equipo con Windows será IP: X.X.X.**101** en la máquina virtual, X.X.X.**202** → X.X.X.**102**, y así sucesivamente.

Instalación Apache Ambari

Intentad instalar el `clúster` siguiendo las instrucciones de los siguientes enlaces:

Fijaros muy bien en los comandos y archivos que hay que utilizar, pues algunos de ellos habréis de adaptarlos a vuestra configuración.

▼ Descargas e instrucciones iniciales

<https://ambari.apache.org/docs/3.0.0/quick-start/download>

⚙ Configuraciones de los equipos:

<https://ambari.apache.org/docs/3.0.0/quick-start/environment-setup/bare-metal-kvm-setup>

✓ Instalaciones finales:

<https://ambari.apache.org/docs/3.0.0/quick-start/installation-guide>

- En este apartado utilizaremos MySQL como base de datos (no PostgreSQL)