

# Trabajos prácticos de PGITIC (2020).

## *Clusters de alta disponibilidad*

### Trabajo 1

**Objetivo:** crear un *cluster* de alta disponibilidad basado en *Corosync* y *Pacemaker*, formado por tres o más máquinas virtuales

Usar *Vagrant* y *Ansible* para configurar y aprovisionar las máquinas virtuales y formar con ellas un *cluster* Linux de alta disponibilidad. A modo orientativo, el aprovisionamiento y configuración deben consistir en:

1. Crear tres o más máquinas virtuales Linux (con la misma versión de sistema operativo), cada una con su propia IP privada. Las máquinas deben llamarse *nodo1*, *nodo2*, *nodo3*...
2. Preparar SSH entre los nodos, de modo que desde el cualquier nodo podamos conectarnos con cualquier otro. Por ejemplo, que podamos conectarnos, desde otro nodo, con el nodo 3 mediante `sudo ssh root@nodo3`.
3. Instalar *nginx*, *libqb0*, *fence-agents*, *pacemaker* y *corosync*.
4. Configurar el cortafuegos para permitir el tráfico en alguno (o varios) de los puertos 5404 a 5406.
5. Configurar el *cluster Corosync*.
6. Iniciar y configurar *Pacemaker*.
7. Añadir recursos.
8. Probar el *cluster*.

Se recomienda consultar, al menos, las siguientes páginas web:

<https://wiki.debian.org/Debian-HA/ClustersFromScratch>

[https://clusterlabs.org/pacemaker/doc/en-US/Pacemaker/1.1/html/Clusters from Scratch/index.html](https://clusterlabs.org/pacemaker/doc/en-US/Pacemaker/1.1/html/Clusters_from_Scratch/index.html)

<https://www.digitalocean.com/community/tutorials/how-to-create-a-high-availability-setup-with-corosync-pacemaker-and-floating-ips-on-ubuntu-14-04>

Es fundamental que los pasos de aprovisionamiento y configuración estén perfectamente comentados en los archivos asociados a *Vagrant* y *Ansible* ya que este trabajo debe servir de base para que otros compañeros de la asignatura mejoren y/o amplíen las posibilidades de este primer *cluster*. Por ese motivo también es necesario que la entrega del trabajo se haga dentro del plazo establecido.

Todos los miembros del equipo de trabajo deben saber explicar el contenido de los archivos de aprovisionamiento y configuración.

## Trabajo 2

**Objetivo:** *hacer funcionar Stonith y Quorum en un cluster de alta disponibilidad, basado en Corosync y Pacemaker, formado por tres o más máquinas virtuales*

## Trabajo 3

**Objetivo:** *instalar Apache y DRBD (Distributed Replicated Block Device) en un cluster de alta disponibilidad, basado en Corosync y Pacemaker, formado por tres o más máquinas virtuales*

## Trabajo 4

**Objetivo:** probar el software de alta disponibilidad SafeKit con la arquitectura *mirror*

La información sobre este software es accesible a partir de:

<https://www.evidian.com/products/high-availability-software-for-application-clustering/>

La principal documentación es la “SafeKit User's Guide”.

Se recomienda obtener una clave gratuita de licencia de un mes. Ver apartado 2.1.4 de la “SafeKit User's Guide”.

Las principales partes del trabajo resultante han de ser:

- Crear y aprovisionar con Vagrant y Ansible los servidores Linux y redes que formarán el *cluster*.
- Configurar el *cluster* con la arquitectura *mirror* y replicación de archivos. A ser posible integrando en el *playbook* de Ansible los comandos necesarios (de los del apartado 9 de la “SafeKit User's Guide”). Como segunda opción usar la consola web de *SafeKit*.
- Comprobar el *cluster* (preferentemente de forma automatizada con Ansible y, si esto no se consigue, comprobarlo usando la consola web). Tener en cuenta la documentación (especialmente los apartados 4.1, 4.2 y 5) pero destacar la comprobación de:
  - Operación normal.
  - Replicación de archivos.
  - *Failover*.
  - *Failback* y reintegración.
  - *Split brain*.

Además de los archivos del proyecto Vagrant-Ansible debe elaborarse un documento (en formato Word, pdf o PowerPoint) explicando paso a paso las tareas realizadas y, en especial, cómo se ha comprobado que los módulos *mirror* ejecutados en el *cluster* presentan alta disponibilidad.

## Trabajo 5

### Objetivo: crear un *cluster* virtual CDH 5 con un HDFS de alta disponibilidad

Se trata de usar *Vagrant* y *Ansible* para configurar un *cluster* virtual CDH 5 sobre el que configurar y desplegar un HDFS (*Hadoop Distributed File System*) de alta disponibilidad.

Como se indica en la documentación de Cloudera, los tres pasos básicos son:

- Configurar el hardware para el HDFS de alta disponibilidad.
- Configurar el software para el HDFS de alta disponibilidad.
- Desplegar el HDFS de alta disponibilidad.

No se exige configurar MapReduce de alta disponibilidad ni otros componentes de CDH.

El trabajo práctico debe ir acompañado de un documento que explique: a) la organización del *cluster* (componentes hardware y software, red...); b) el proceso de creación del mismo; c) qué distingue al HDFS de alta disponibilidad del HDFS sin esta propiedad. Además, debe explicarse cómo se ha verificado que, en efecto, el HDFS implementado es tolerante a fallos.

Es fundamental consultar la documentación de Cloudera (por ejemplo, es imprescindible consultar [https://www.cloudera.com/documentation/enterprise/latest/topics/admin\\_ha.html](https://www.cloudera.com/documentation/enterprise/latest/topics/admin_ha.html) y/o <https://www.cloudera.com/documentation/cdh/5-0-x/CDH5-High-Availability-Guide/CDH5-High-Availability-Guide.html>). Pero, previamente, debe comprenderse cómo se instala CDH 5 en un *cluster*, cuál es la arquitectura del HDFS y cómo instalar un *cluster* ZooKeeper. Para ello, además de la documentación de Cloudera, puede ser útil la de Apache (por ejemplo, <http://zookeeper.apache.org/>, <http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HDFSHighAvailabilityWithNFS.html> y <https://archive.cloudera.com/cdh5/cdh/5/hadoop/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html>). Obviamente, los enlaces sugeridos no son los únicos que pueden ser útiles.

+++++

Trabajo	Quién lo realiza	Fecha límite de entrega
1	Los dos primeros equipos de trabajo que lo soliciten (un equipo de trabajo es un grupo o un grupo más un estudiante sin grupo). Cada equipo hará un trabajo independiente del otro.	7 de mayo
2	El primer grupo o estudiante sin grupo que lo solicite	3 de junio
3	El primer grupo o estudiante sin grupo que lo solicite	3 de junio
4	El primer grupo o estudiante sin grupo que lo solicite	9 de mayo
5	El primer grupo o estudiante sin grupo que lo solicite	9 de mayo