



Proyecto de Innovación

Convocatoria 2020/2021

Nº de proyecto 80

Cluster de Raspberry Pi para prácticas de Programación Paralela y Big Data

Responsable del proyecto:  
Luis Fernando Llana Díaz

## 1. Objetivos propuestos

El responsable de este proyecto está impartiendo este curso las asignaturas de 'Programación Paralela' y 'Big Data' en la Facultad de Ciencias Matemáticas. Uno de los contenidos de ambas asignaturas es la programación usando Hadoop y Spark. Tanto Hadoop como Spark son entornos de programación de naturaleza distribuida. Normalmente, la docencia de estos entornos está limitada por no poder disponer de un clúster real para que los alumnos puedan ejecutar sus programas. Se ha intentado usar el propio laboratorio para usarlo como clúster, pero la experiencia no fue satisfactoria. Por ese motivo, las prácticas que deberían ejecutarse en entornos distribuidos, se ejecutan en los ordenadores personales simulando los entornos distribuidos.

Con este proyecto se pretende cubrir esa necesidad. Además al ser de bajo coste, se puede plantear la posibilidad de tener un conjunto de clústeres para cubrir la demanda de todos los alumnos de la asignatura.

Los objetivos específicos que se quiere conseguir con este proyecto son:

- Aumentar la motivación de los alumnos en estas asignaturas al poder ejecutar sus programas en entornos distribuidos reales.
- Mejorar la calidad docente de la asignatura puesto que los alumnos habrán podido practicar en entornos muy parecidos a los entornos reales de Big Data.
- Poder realizar prácticas en las que se puedan cambiar los parámetros del clúster para poder estudiar su efecto en el rendimiento de la aplicación.

## 2. Objetivos Alcanzados

La financiación que ha recibido el proyecto ha sido claramente insuficiente para la realización del mismo. Finalmente el proyecto ha sido posible gracias a la aportación de fondos por parte del Departamento de Sistemas Informáticos y Computación.

Actualmente el clúster está en funcionamiento y a disposición de los alumnos. Actualmente deben realizar una serie de prácticas en las que deben usar el clúster.

## 3. Metodología empleada en el proyecto

El proyecto se ha realizado siguiendo las tareas que se enumeraban en la petición del mismo:

1. Montaje del clúster.
2. Evaluación de la alimentación energética y necesidades de refrigeración del clúster.

3. Instalación de las herramientas Hadoop y Spark.
4. Diseño de experimentos para la evaluación del rendimiento.
5. Uso del clúster por grupos seleccionados de estudiantes.
6. Publicación de la experiencia.

## 4. Recursos humanos

La realización del clúster ha sido posible gracias a las aportaciones del equipo que formaba parte del proyecto. Gracias a la experiencia previa del profesor Alberto Núñez, se ha solventado satisfactoriamente el montaje físico del clúster.

El software necesario para su uso ha sido configurado con éxito con la ayuda del Profesor Carlos Gregorio.

La realización de las piezas impresas con tecnología 3D ha sido posible gracias a los estudiantes de doctorado Miguel Benito y Alfredo Ibias.

La labor del resto del equipo ha sido la validación y depuración del sistema en su conjunto: usabilidad, fiabilidad y seguridad.

## 5. Desarrollo de las actividades

Para el montaje físico del clúster se ha aprovechado la experiencia de proyectos anteriores en el que se realizó un clúster de Raspberry Pi3. Uno de los retos del proyecto ha sido la alimentación energética: Las nuevas Raspberry Pi4 necesitan una mayor alimentación y el uso de discos duros externos complica aún más esta tarea. Para solventar este problema se ha utilizado una fuente de alimentación de un PC convencional de sobremesa.

Para acoplar todos los elementos y facilitar la refrigeración se ha reutilizado la caja de un PC desusado de sobremesa. Para acoplar los discos duros y las Raspberry Pi a la caja se han diseñado unas serie de piezas que luego has sido realizadas con una impresora 3D.

Una vez acopladas las piezas y comprobado la estabilidad de la corriente eléctrica se ha instalado el software necesario. Además de la instalación prevista de Hadoop y Spark, se ha instalado la base de datos MongoDB. Esta última tarea ha sido algo más compleja de lo que puede parecer a simple vista debido a que el sistema operativo base de las Raspberry Pi es de 32-bits y las versiones recientes de MongoDB solo funcionan en sistemas operativos de 64 Bits.

Actualmente el clúster está siendo utilizado por alumnos de la Facultad de Matemáticas de las asignaturas de 'Programación Paralela' y 'Datos Masivos: Big Data'.