



UNIVERSIDAD
COMPLUTENSE
MADRID

Proyecto de Innovación

Convocatoria 2018/2019

Proyecto nº 48

Adaptación y ejecución de aplicaciones MPI en entornos de simulación para
analizar el rendimiento de arquitecturas distribuidas.

Responsable del proyecto: Alberto Núñez Covarrubias

Facultad de Informática

Departamento de Sistemas Informáticos y Computación

1.- Objetivos propuestos en la presentación del proyecto

Una parte fundamental de la asignatura PSD es el desarrollo de aplicaciones distribuidas. En concreto, la programación de aplicaciones paralelas utilizando MPI tiene un peso considerable, tanto en la parte práctica, como en la parte teórica de la asignatura. Las aplicaciones MPI desarrolladas en las prácticas de PSD pueden, normalmente, ejecutarse en un solo ordenador utilizando los distintos núcleos del procesador, lo cual se conoce como modo “stand-alone”.

Sin embargo, limitar a los alumnos a ejecutar sus aplicaciones distribuidas en un único procesador complica en gran medida la labor docente, especialmente cuando se trata de aplicaciones MPI formadas por varios procesos que deben ejecutarse en los múltiples procesadores que proporciona un sistema distribuido. Primero, el alumno no presta atención a los aspectos relacionados con el rendimiento, los cuales se enfocan, generalmente, en explotar de la forma más eficiente posible los distintos recursos ofrecidos por el sistema distribuido, tales como CPU, memoria y almacenamiento. Segundo, es posible que el alumno no desarrolle una aplicación paralela de forma escalable ya que, al ejecutar la misma en un único ordenador, se pierde la visión del sistema distribuido, limitando la escalabilidad a los recursos de un único ordenador. Finalmente, el alumno puede experimentar una pérdida de motivación al no realizar y visualizar la ejecución distribuida de su aplicación.

El objetivo principal del proyecto consiste en **mitigar los problemas existentes para ejecutar las aplicaciones MPI, requeridas en las prácticas de PSD, en entornos distribuidos reales**. Para ello proponemos el uso del simulador SIMCAN, el cual ya ha sido adaptado en el pasado para su uso docente. Concretamente, se propone utilizar SIMCAN para reproducir las trazas de ejecución de aplicaciones MPI en diferentes arquitecturas distribuidas, de forma que se alivien, en la medida de lo posible, aquellos aspectos de interés que se tratan con dificultad en la ejecución “stand-alone”, en particular, la visualización del funcionamiento distribuido y su rendimiento. De esta forma, se utilizarán distintas arquitecturas distribuidas, modeladas con el simulador SIMCAN, para ejecutar las aplicaciones MPI estudiadas en la asignatura.

Seguidamente se exponen los objetivos específicos y las propuestas de valor asociadas a los mismos:

Objetivo 1: Mejorar el proceso de aprendizaje en los aspectos impartidos en la materia de sistemas distribuidos, consiguiendo un proceso didáctico más natural y fluido. Esto se conseguirá gracias a la posibilidad de ejecutar las prácticas de la asignatura en un amplio rango de arquitecturas distribuidas modeladas con SIMCAN.

Objetivo 2: Aumentar el interés de los alumnos por la asignatura. Esto se conseguirá al permitir que las aplicaciones MPI desarrolladas se puedan ejecutar en modelos que representen arquitecturas reales y de alta escalabilidad, como los centros de datos de un sistema *cloud*. Además, los alumnos podrán analizar cómo afecta la escalabilidad del sistema donde se ejecutan las aplicaciones, comparando la ejecución

de las prácticas en un solo ordenador con la ejecución en un sistema distribuido completo.

Objetivo 3: Fomentar el trabajo autónomo de los alumnos, lo cual tendrá un impacto positivo sobre la efectividad en la enseñanza semi-presencial de sistemas distribuidos. Puesto que el simulador SIMCAN puede ejecutarse en un ordenador convencional, no será obligatorio contar con los ordenadores de los laboratorios para desarrollar las prácticas de la asignatura. Con ello se pretende facilitar el desarrollo de las mismas a los alumnos que compaginan sus estudios con la vida laboral. De esta forma se evitarían, en la medida de lo posible, abandonos de la asignatura.

2.- Objetivos alcanzados

Este proyecto tiene como principal objetivo **brindar a los alumnos la oportunidad de ejecutar aplicaciones paralelas en entornos altamente distribuidos**, tales como los que ofrecen las plataformas de *cloud computing* actuales.

El objetivo de este proyecto se ha alcanzado de forma satisfactoria. Para ello, se ha utilizado la plataforma de simulación SIMCAN, con la cual se han analizado las trazas de distintas aplicaciones MPI previamente ejecutadas en entornos reales. En concreto, el alumno Bryan Vaca desarrolló el TFG titulado “Estudio del rendimiento y la escalabilidad de aplicaciones MPI en entornos distribuidos utilizando SIMCAN” enfocado, principalmente, a desarrollar un sistema que cumpliera los objetivos definidos en este proyecto.

Seguidamente, se detallan los objetivos específicos alcanzados en la realización de este proyecto.

Objetivo 1: Mejorar el proceso de aprendizaje en los aspectos impartidos en la materia de sistemas distribuidos, consiguiendo un proceso didáctico más natural y fluido.

El sistema desarrollado permite realizar las prácticas de la asignatura de forma fluida. Cabe destacar que el proceso de generar una traza con la ejecución de la aplicación MPI bajo estudio, transformarla a un formato legible por SIMCAN y ejecutar la misma en el propio simulador, puede llegar a requerir un tiempo y esfuerzo considerable. Sin embargo, con la herramienta desarrollada, este proceso se realiza de forma automática con una sola aplicación, permitiendo al alumno focalizar su esfuerzo en el análisis de la aplicación. Por ello, podemos concluir que el proceso de aprendizaje ha mejorado considerablemente.

Objetivo 2: Aumentar el interés de los alumnos por la asignatura.

Uno de los efectos esperados con la realización de este proyecto es obtener un aumento de interés en la asignatura. Sin embargo, hasta el próximo curso no será posible comprobar si se ha logrado este objetivo, ya que para ello es necesario realizar las encuestas a los estudiantes, donde no sólo comentarán su grado de interés por la asignatura, sino que realizarán comentarios sobre los aspectos a mejorar de la misma.

Objetivo 3: Fomentar el trabajo autónomo de los alumnos, lo cual tendrá un impacto positivo sobre la efectividad en la enseñanza semi-presencial de sistemas distribuidos.

Con la herramienta desarrollada, los alumnos serán capaces de realizar las prácticas de la asignatura de forma semi-presencial. Aunque es deseable que los alumnos asistan a un número mínimo de clases, donde se explican los objetivos y aspectos más relevantes de cada práctica, no dependerán de los recursos proporcionados por los laboratorios para su desarrollo. De esta forma, podrán realizar las prácticas en cualquier ordenador personal fuera de los horarios establecidos en la asignatura.

3.- Metodología empleada en el proyecto

Este proyecto está formado por 7 tareas, las cuales se detallan a continuación:

- 1) Implementar una biblioteca que genere la traza de ejecución de aplicaciones MPI cuando éstas son ejecutadas en sistemas reales.
- 2) Desarrollar un módulo en el simulador SIMCAN que reproduzca las trazas de las aplicaciones MPI generadas en la tarea anterior.
- 3) Definir y modelar diferentes arquitecturas de sistemas distribuidos empleando diferentes niveles de escalabilidad, tanto horizontal como vertical.
- 4) Generar las trazas de ejecución de las aplicaciones MPI estudiadas en la asignatura.
- 5) Reproducir en el simulador las trazas de las aplicaciones MPI sobre las diferentes arquitecturas modeladas.
- 6) Evaluar y comparar los resultados obtenidos en cada arquitectura.
- 7) Diseñar y planificar los indicadores necesarios que permitan llevar a cabo los procedimientos de evaluación.

La sincronización de las distintas tareas que forman el proyecto fue llevada a cabo por el responsable del mismo. Tanto las tareas realizadas por cada miembro del proyecto, así como el responsable de cada tarea, se detallan en la sección 5.-Desarrollo de las actividades.

Cabe destacar el esfuerzo realizado para sincronizar cada una de las tareas del proyecto con el TFG realizado por el alumno Bryan Vaca. Además, fueron necesarias varias reuniones de coordinación con los responsables de cada tarea.

Tras la finalización del proyecto, el responsable del mismo realizó un informe con la información generada en cada tarea.

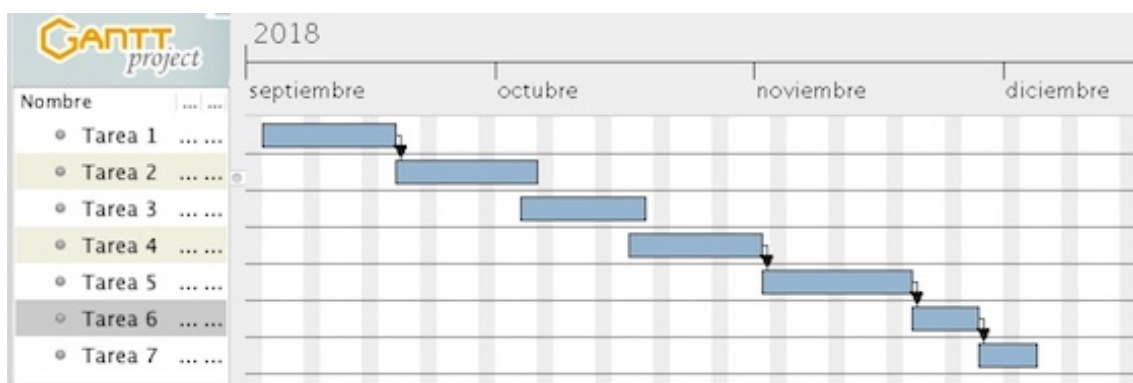
4.- Recursos humanos

El grupo de trabajo está formado, mayoritariamente, por personal de la Universidad Complutense de Madrid (UCM), aunque también existen miembros del mismo pertenecientes a la Universidad Castilla-La Mancha (UCLM). Además, incluye a Bryan Vaca, un alumno que aprobó la asignatura PSD el curso 17/18. Con ello se pretende aportar una visión adicional, desde el punto de vista de los estudiantes, para poder realizar con mayor precisión los métodos propuestos.

De los miembros PDI que forman este proyecto, 3 profesores han impartido asignaturas relacionadas con sistemas distribuidos. Concretamente, de la UCM, los profesores Simon Pickin y Alberto Núñez han impartido la asignatura PSD durante los últimos seis años académicos; esta asignatura fue puesta en marcha en el curso académico 2012/2013 por el profesor Alberto Núñez. Además, la profesora M. Emilia Cambronerero ha impartido, durante los últimos 8 años, la asignatura Sistemas Distribuidos del grado en Ingeniería Informática de la UCLM. El resto de los profesores del grupo de trabajo cuenta con amplia experiencia en asignaturas de programación, lo cual entra dentro del ámbito de la asignatura PSD. Por otro lado, el grupo de trabajo también cuenta con personal PAS de la UCM. Concretamente, Ana Bartolomé, que actualmente trabaja como Técnico de Laboratorio en la facultad de Informática y Pablo Cerro, el cual fue antiguo alumno de la asignatura PSD y cuenta con el grado en Ingeniería de Computadores. Así, queda demostrada tanto la idoneidad como la experiencia docente del grupo de trabajo en el ámbito en el cual se enmarca este proyecto.

5.- Desarrollo de las actividades

La línea temporal de cada una de las tareas que forman este proyecto se muestra en el siguiente diagrama de Gantt.



Asimismo, los intervalos temporales de cada tarea se describen a continuación:

- Tarea 1: 03/09/2018 - 18/09/2018
- Tarea 2: 19/09/2018 – 05/10/2018
- Tarea 3: 04/10/2018 – 18/10/2018
- Tarea 4: 17/10/2018 – 01/11/2018
- Tarea 5: 02/11/2018 – 19/11/2018
- Tarea 6: 20/11/2018 – 27/11/2018
- Tarea 7: 28/11/2018 – 04/12/2018

Las tareas en las que ha participado cada miembro del equipo de trabajo se detallan en el siguiente cuadro, donde 'C' indica el coordinador de cada tarea y 'X' la participación de un miembro en una tarea.

Miembro\Tarea	1	2	3	4	5	6	7
Alberto Núñez	C	C	C	X	X	X	C
Simon Pickin	X	X		C	X		
Víctor Lavín				X	X		X
Luis Llana	X	X	X		C	X	
Ana Bartolomé	X	X					
Pablo Cerro	X		X	X	X		
Emilia Cambronerero	X			X			X
Manuel Núñez						C	X
Bryan Vaca	X	X				X	X
Miguel Gómez-Zamalloa						C	X

La descripción detallada de cada tarea se presenta a continuación:

La tarea 1 consiste en implementar una biblioteca que genere la traza de ejecución de aplicaciones MPI. Esto se consigue, básicamente, capturando las llamadas a las funciones más relevantes de MPI, tales como MPI_Send, MPI_Recv o MPI_Bcast. De

esta forma, se escribe en un fichero de texto cada una de las operaciones invocadas por la aplicación, así como el instante de tiempo en el cual fueron invocadas. Como resultado, obtenemos un fichero con la lista de las funciones MPI invocadas, así como sus parámetros, ya que es relevante contar con información tal como los identificadores de los procesos involucrados o la cantidad de datos enviada.

En la tarea 2 se desarrolla un módulo para el simulador SIMCAN. El objetivo de este módulo consiste en la reproducción de las trazas generadas en la tarea anterior. De esta forma, se simula el comportamiento de aplicaciones MPI en diferentes sistemas distribuidos. Para ello, en la tarea 3 se definen y modelan distintas arquitecturas distribuidas empleando diferentes niveles de escalabilidad, tanto horizontal, incrementando el número de máquinas que forman el sistema, como vertical, mejorando las características hardware de cada máquina.

En la tarea 4 se seleccionan varias aplicaciones MPI para ser analizadas en diferentes arquitecturas. Generalmente, serán aplicaciones MPI estudiadas en clase. Además, utilizando la biblioteca desarrollada en el paso 1, se generan las trazas correspondientes de su ejecución.

En la tarea 5 se simulan las aplicaciones MPI que fueron seleccionadas en la tarea 4, sobre las distintas arquitecturas modeladas en la tarea 3. Para ello, se hace uso del simulador SIMCAN y del módulo desarrollado en la tarea 2. En la tarea 6 se evalúan los resultados obtenidos en las simulaciones realizadas en la tarea anterior.

Finalmente, en la tarea 7 se definen y llevan a cabo los indicadores para evaluar el impacto obtenido durante el desarrollo de este proyecto. Para ello, se ha contado con un equipo de trabajo heterogéneo, lo cual complementa y enriquece los conocimientos necesarios para cumplir con éxito los objetivos establecidos. Además de contar con una amplia experiencia en la materia de sistemas distribuidos y programación, se suma la participación de un alumno que ha cursado, y aprobado, la asignatura en la cual se centra este proyecto, pudiendo aportar una perspectiva diferente para llevar a cabo esta tarea.