

Una formación de calidad
en ingeniería para el futuro

Centro de Convenciones Cartagena de Indias
15 al 18 de Septiembre de 2015

COMPUTACIÓN DE ALTO DESEMPEÑO PARA CÁLCULOS DE QUÍMICA MECANO-CUÁNTICA

Diego Armando Alvarado Escobar, Brandon Steven Ramírez Sierra, Luis Eduardo Sepúlveda Rodríguez, Nathalia Bibiana Duque Madrid, Christian Andrés Candela Uribe, Alberto Sánchez López

**Universidad del Quindío
Armenia, Colombia**

Resumen

El Grupo de Investigación en Redes, Información y Distribución - GRID en asocio con el Grupo de Investigación en Físicoquímica Ambiental y Computacional - GIFAC ambos de la Universidad del Quindío, identificaron la necesidad conjunta de formular un trabajo de grado en modalidad de investigación, cuyo objetivo principal se centra en realizar un acercamiento desde la Computación de Alto Desempeño (HPC por su sigla en inglés, High Performance Computing) aplicado a cálculos de química mecano cuántica.

El entorno problémico abordado en la investigación se enfoca en el grupo de investigación GIFAC, quienes mediante el uso del programa ORCA, realizan cálculos a nivel Ab-initio y DFT, de sistemas moleculares que demandan tiempo, a tal punto que la obtención de resultados se puede tardar días e incluso meses.

La cantidad de tiempo computacional empleado para la generación de resultados de valor es debido a diversos factores, entre ellos se destaca la complejidad de los algoritmos utilizados, las limitaciones físicas de la infraestructura computacional, además de la ineficiencia en las técnicas computacionales inherentes a la ejecución computacional misma. Sin importar la razón, la tardanza en la obtención de resultados es para el grupo de investigación GIFAC una situación desventajosa que retrasa el análisis de información y la respectiva toma de decisiones. La situación descrita representa una oportunidad para la realización de un trabajo interdisciplinar, mediante la aplicación de HPC, de tal forma que se propicie la reducción de los tiempos de procesamiento utilizados en este tipo de cálculos.

La expectativa de solución a la problemática se centra en la implementación de un clúster de procesamiento paralelo de bajo costo e implementado con tecnologías libres

y/o abiertas que permitan obtener el mayor aprovechamiento de los recursos hardware existentes al interior de los grupos de investigación partícipes del proyecto.

Palabras clave: HPC; cómputo paralelo; química cuántica

Abstract

The Investigation Group about Nets, Information and Distribution - IGNID in association with the Physicochemical Environment and Computational Investigation Group - PECIG, both from the University of Quindío, they identified the necessity of working together in a project in the investigation modality. The main objective is to focus on a HPC (High performance computing) approach applied to the mechano-quantum chemistry.

The main problem is focus on the investigation group PECIG, those using the ORCA software, perform calculations at the level Ab- initio and DFT, molecular systems that demands time, to the point that obtaining results can take days or even months.

The amount of computational time reserved to generate results of value is due to various factors, including itself emphasizes the complexity of the algorithms used, the physical limitations of computational Infrastructure, in addition to the inherent inefficiency in computational techniques in the same performance. No matter the reason, the delay in obtaining results is a disadvantage for investigation group PECIG because of the analysis of information and respective decision-making. This situation permits an opportunity for the realization of an interdisciplinary work, through the application of HPC, so it could exist a reduction of processing time used in these types of calculus.

The solution expected to the problem focuses on the implementation of a cluster of parallel processing low cost and Implemented with Free and open Technologies which could achieve the better use of existing hardware resources within investigation groups participating in this project.

Keywords: HPC; parallel computing; quantum chemistry

1. Introducción

La Computación de Alto Desempeño se encuentra asociada generalmente con la supercomputación o el uso de supercomputadoras, por ende, son léxicos que tienen el mismo significado. El supercómputo hace uso de una técnica indispensable para que se genere un alto rendimiento, conocida como computación paralela o procesamiento paralelo, la cual, consiste en acelerar la ejecución de un programa mediante su división en fragmentos que pueden ser procesados simultáneamente, cada uno en un procesador, agrupando computadores independientes de forma que la imagen de éstos hacia el usuario final sea la de un único computador (Iberico Hidalgo, 2009; Reyes & Jiménez, 2003).

El grupo de investigación GIFAC realiza investigaciones en las cuales utiliza métodos Ab-initio y DFT. Este tipo de métodos requiere alta capacidad de recursos computacionales, especialmente memoria y procesamiento. A raíz de esta situación, el grupo de investigación GRID propuso la formulación del trabajo de grado aquí presentado, cuya realización beneficia al grupo de investigación GIFAC, mediante la aplicación de técnicas de Computación de Alto Desempeño que propiciarán la reducción en los tiempos de ejecución de las simulaciones computacionales.

El abordaje de la problemática se llevó a cabo haciendo uso de una técnica de supercómputo denominada clúster. Esta técnica, como la define el Dr. Thomas Sterling en la investigación de Hernández Vázquez (2005), es una clase de arquitectura de computador paralelo que se basa en unir máquinas independientes y cooperativas integradas por medio de redes de interconexión, para proveer un sistema coordinado capaz de procesar una carga.

En la realización de este proyecto, es importante destacar el trabajo interdisciplinar entre los grupos de investigación implicados, por medio de la cual se llega a una solución que beneficia la investigación en química mecano-cuántica computacional, dado que presenta una alternativa de mejora en el rendimiento del tiempo de ejecución de las simulaciones realizadas con el software ORCA, ampliamente utilizado por la comunidad científica.

2. Descripción del problema

El grupo de investigación GIFAC de la Universidad del Quindío realiza estudios moleculares con diferentes métodos de química mecano-cuántica, por medio del software ORCA. Dependiendo de las moléculas que se deseen simular, ORCA tarda lapsos de tiempo que van desde días hasta meses, para entregar resultados. La duración del proceso que realiza el software para obtener un resultado, se debe entre otros factores a la complejidad de los cálculos que ejecuta el algoritmo y a la eficiencia en las técnicas inherentes a la ejecución de procesos computacionales. Es precisamente este último aspecto, el que se evidencia una necesidad para establecer buenas prácticas en la ejecución de técnicas computacionales, que permitan entregar tiempos menores en la realización de las simulaciones de química para el grupo de investigación GIFAC.

3. Objetivos

3.1. General

Aplicar Computación de Alto Desempeño para cálculos de química mecano-cuántica realizados por el grupo de investigación GIFAC.

3.2. Específicos

- Identificar necesidades relacionadas en Computación de Alto Desempeño para la ejecución simulada de procesos moleculares en química mecano-cuántica, realizada por el grupo de investigación GIFAC.

- Determinar prácticas tecnológicas de Computación de Alto Desempeño que favorezcan la ejecución simulada de procesos moleculares en química mecano-cuántica según las necesidades identificadas en el grupo de investigación GIFAC.
- Especificar un procedimiento para la aplicación de prácticas de Computación de Alto Desempeño en la ejecución simulada de procesos moleculares en química mecano-cuántica en el grupo de investigación GIFAC.
- Realizar una implementación del procedimiento, aplicando prácticas de Computación de Alto Desempeño en la ejecución de cálculos mecano-cuánticos, desarrollados por el grupo de investigación GIFAC.
- Evaluar la implementación mediante la realización de cálculos mecano-cuánticos en la ejecución simulada de procesos moleculares para el grupo de investigación GIFAC

4. Justificación

Dentro de la química computacional se desarrollan cálculos mediante Ab-initio y DFT, que son un grupo de métodos en los que se trata de resolver la ecuación de Schrödinger para sistemas de varios átomos. El grupo de investigación GIFAC realiza investigaciones en las cuales utiliza dichos métodos. Este tipo de cálculos requieren alta capacidad de recursos computacionales, especialmente memoria y procesamiento. Por tanto, los estudios moleculares llevados a cabo por el grupo de investigación GIFAC, son ideales para la utilización de entornos con aplicación de Computación de Alto Desempeño (HPC) donde se requiere el uso de prácticas tecnológicas para mejorar tiempos de ejecución de los procesos computacionales (Aliseda Casado, 2011).

En este sentido, el presente trabajo de grado es importante para el grupo de investigación GIFAC y el grupo de investigación GRID, dado que genera una dinámica de trabajo interdisciplinar, además de centrarse en el análisis de necesidades para llevar a cabo la ejecución simulada de procesos moleculares en química mecano-cuántica computacional.

Los beneficios proyectados tras la realización del proyecto son:

- Permitir una perspectiva clara acerca de la paralelización automática de algoritmos con tecnologías de Computación de Alto Desempeño (HPC), además, impactar futuras formulaciones de proyectos de investigación y trabajos de grado para estudiantes interesados en el tema.
- Reducir el tiempo empleado en los cálculos realizados sobre procesos moleculares, aprovechando potencialmente los recursos de la máquina.
- Favorecer procesos claves de análisis de resultados y toma de decisiones para acelerar las conclusiones sobre las prácticas desarrolladas por el grupo de investigación GIFAC con relación a la ejecución simulada de procesos moleculares en química mecano-cuántica.
- Mejorar la percepción del usuario sobre el comportamiento general del sistema utilizado para la ejecución de las simulaciones debido a la disminución de los tiempos de respuesta.

5. Marco Teórico

5.1. HPC (High Performance Computing)

La Computación de Alto Desempeño, ampliamente conocida como HPC (por su sigla en inglés, High Performance Computing) es cualquier técnica computacional que soluciona un problema grande de forma más rápida que usando posibles sistemas simples. Se apoya en tecnologías computacionales como clústeres, computación paralela, computación distribuida, computación Grid, procesadores de alto rendimiento, entre otros (Larrazábal, 2014).

5.2. Procesamiento paralelo o computación paralela

De acuerdo con Ibérico Hidalgo (2009), la computación paralela consiste en agrupar computadores independientes de forma que la imagen de éstos hacia el usuario final sea la de un único computador cuyas propiedades, en la medida de lo posible, es la suma de las prestaciones de los equipos que formen el sistema paralelo. Esta agrupación de computadores tiene como objetivo realizar una tarea o proceso computacionalmente intensivo, y para ello, la tarea necesita ser descompuesta en distintas subtareas independientes, debiendo redireccionar cada una de éstas hacia un computador asociado al proyecto, el cual tenga los recursos necesarios para la resolución de las mismas. A medida que las subtareas sean completadas por cada uno de los computadores, se centralizan en uno de ellos para formar la tarea inicial, obteniendo el resultado en un tiempo mucho menor si es que se realizará en un sólo computador (p. 26).

El procesamiento paralelo puede existir en muchos niveles software y hardware, la Figura 1 muestra las diferentes técnicas de supercómputo que pueden ser aplicadas para diferentes niveles de paralelismo:

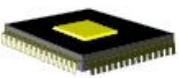
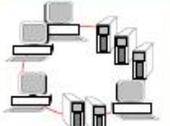
nivel		técnicas de implementación
estructuras paralelas	 paralelismo a nivel de procesador	- segmentación - división funcional - procesadores vectoriales
	 paralelismo en multiprocesadores	- memoria compartida - memoria distribuida
	 paralelismo en multicomputadores	- <i>clusters</i> - sistemas distribuidos

Figura 1. Tipos de paralelismo.
Fuente: (Instituto Politécnico Nacional, 2014).

5.3. Clúster o agrupamiento de nodos

La siguiente definición se presenta de una cita en la investigación de Rodríguez et al. (2009), la cual enmarca el lineamiento del presente proyecto, así:

“Un cluster es la variación de bajo precio de un multiprocesador masivamente paralelo (miles de procesadores, memoria distribuida, red de baja latencia), con las siguientes diferencias: cada nodo es una máquina quizás sin algo del hardware (monitor, teclado, mouse, etc.), el nodo podría ser SMP o PC. Los nodos se conectan por una red de bajo precio como Ethernet o ATM aunque en clústeres comerciales se pueden usar tecnologías de red propias. El interfaz de red no está muy acoplado al bus I/O. Todos los nodos tienen disco local. Cada nodo tiene un sistema operativo UNIX con una capa de software para soportar todas las características del clúster”.

Estos sistemas multicomputador pueden ser de tres tipos fundamentalmente: clústeres de Alto Desempeño (HP por su sigla en inglés, High Performance); clústeres de Alta Disponibilidad (HA por su sigla en inglés, High Availability); y por último, clústeres de Alta Confiabilidad (HR por su sigla en inglés, High Reliability).

6. Metodología

La metodología de este proyecto se sitúa en la Investigación Tecnológica en Ciencias de la Ingeniería, propuesta por Sonia Dalila Ríos en su libro Metodología de la Investigación Tecnológica, publicado en 1998; en la cual se propone iniciar recopilando información bibliográfica alrededor de los clúster de procesamiento. Continuando con un estudio para recolectar la información donde se identifican necesidades y oportunidades en la ejecución de cálculos científicos para química mecano-cuántica realizados por el grupo de investigación GIFAC. Posteriormente, se analizan y determinan características distintivas para diseñar el clúster de procesamiento que permita dar respuesta a las necesidades del grupo de investigación GIFAC, adaptando procesos y procedimientos de acuerdo a estándares y buenas prácticas computacionales. Al finalizar, se planea la implementación del clúster diseñado utilizando la infraestructura disponible en el grupo de investigación GIFAC.

7. Resultados

En el clúster de procesamiento implementado se realizaron ejecuciones simuladas para el análisis de un tipo de moléculas denominadas chalconas. El tipo de moléculas a simular se eligió de acuerdo con los estudios que realiza el grupo de investigación GIFAC. A continuación, se describen los resultados obtenidos en el análisis de dos moléculas tipo chalcona, con referencia al tiempo de ejecución de las simulaciones, utilizando diferente número de procesadores del clúster.

En las tablas 1 y 2 se pueden observar los tiempos de procesamiento (en minutos), de las ejecuciones simuladas de moléculas tipo chalconas, las cuales se presentan como

chalcona 1 y chalcona 2; de igual forma se relaciona la cantidad de núcleos utilizados en la ejecución de las simulaciones en el clúster. En la columna final de las tablas se muestra la ganancia de tiempo que se tiene al hacer uso de este clúster.

Para la ejecución simulada de la chalcona 1 (Tabla 1) se tiene como referencia una máquina que posee 4 núcleos de procesamiento. Para la chalcona 2 (Tabla 2) se ejecutó la simulación inicialmente haciendo uso del clúster compuesto por 3 máquinas, cada una con 4 núcleos, teniendo así un total de 12 núcleos de procesamiento.

Procesadores	Tiempo (minutos)	Ganancia de tiempo
4	434	0%
12	167	61,52%
18	174	59,90%
30	143	67,05%

Tabla 1. Diferencias de tiempo de ejecución Chalcona 1.

Procesadores	Tiempo (minutos)	Ganancia de tiempo
12	1030	0%
15	979	4,95%
30	725	29,61%

Tabla 2. Diferencias de tiempo de ejecución Chalcona 2.

Las figuras 2 y 3 muestran los resultados relacionados con los tiempos de ejecución obtenidos en la simulación de moléculas tipo Chalcona.

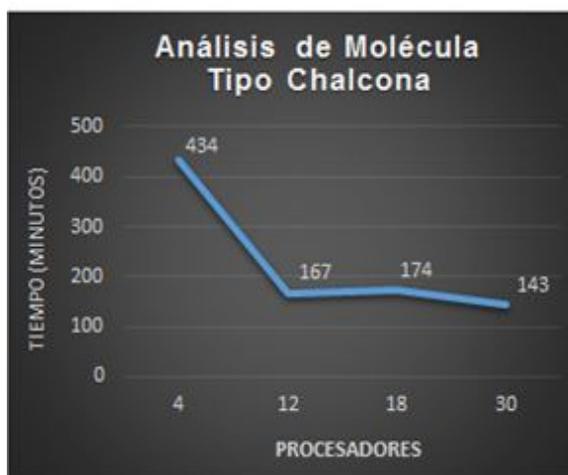


Figura 2. Tiempos de ejecución Chalcona 1

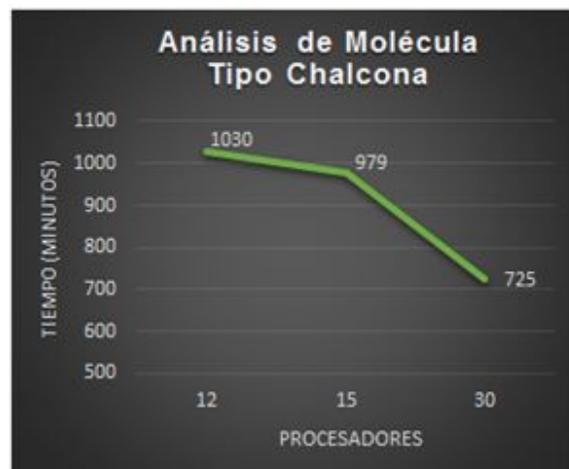


Figura 3. Tiempos de ejecución Chalcona 2

Los tiempos de ejecución alcanzados con la molécula Chalcona 1, evidencian una disminución significativa del 67,05%; pasando de 7 horas con 14 minutos a 2 horas con 23 minutos realizando ejecuciones en la máquina de referencia y el clúster con 30 núcleos respectivamente. De igual forma, para la molécula Chalcona 2, utilizando los mismos elementos tecnológicos, los tiempos de ejecución alcanzados también evidencian una disminución correspondiente al 29,61%; pasando de 17 horas con 10 minutos a 12 horas con 05 minutos.

La Figura 4 presenta una evidencia de la implementación del clúster en el cual se realizó la ejecución simulada de las moléculas tipo chalcona. A través de la herramienta Htop, los monitores de las máquinas muestran el procesamiento que cada uno de los núcleos está realizando. Dicho proceso fue realizado en el laboratorio de cómputo en la Facultad de Ingeniería de la Universidad del Quindío.

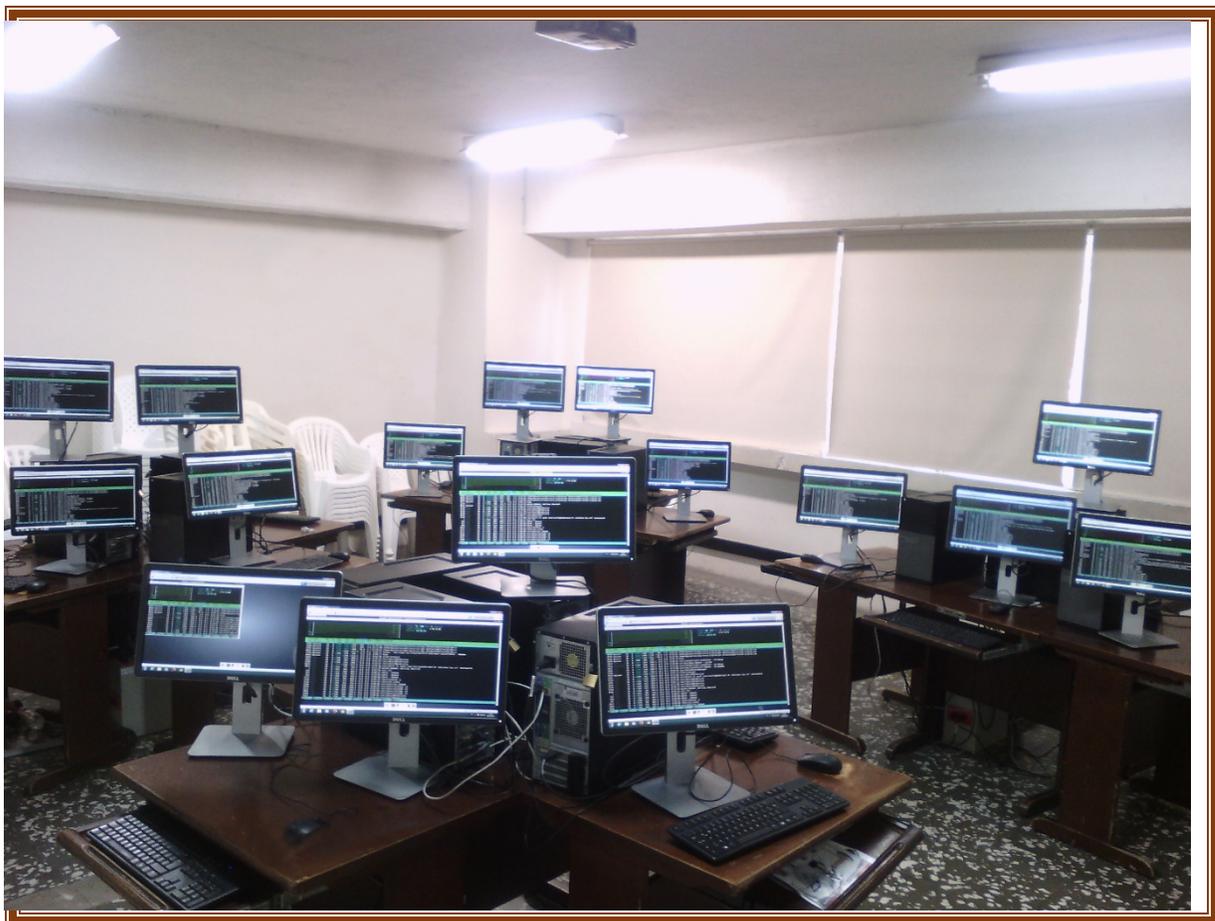


Figura 4. Implementación del Clúster HP en la Facultad de Ingeniería de la Universidad del Quindío.

Debido a los resultados presentados, fue posible implantar prácticas en la utilización de tecnologías de computación de alto desempeño, que permitieron mejorar la ejecución de simulaciones en química mecano-cuántica realizadas por el grupo de investigación GIFAC.

Otro resultado para destacar es el trabajo interdisciplinar realizado entre los grupos de investigación GRID y GIFAC, logrando en este caso particular el fortalecimiento del trabajo en equipo y sentando las bases para la construcción de redes internas de cooperación para la investigación en la Universidad del Quindío.

8. Conclusión

Con el desarrollo de esta investigación se puede demostrar la utilidad que tiene la computación paralela, teniendo como resultado la disponibilidad continua de un clúster utilizado para reducir el tiempo en las ejecuciones simuladas de procesos moleculares, realizadas por el grupo de investigación GIFAC. El clúster permite aprovechar al máximo los recursos de las máquinas permitiendo obtener un mayor rendimiento y llevar a una toma de decisiones más oportuna. Entre los beneficios que se pueden obtener al realizar este proyecto están:

- *Investigación:* Permitir una perspectiva más clara acerca de la paralelización automática de algoritmos con tecnologías de alto rendimiento, además, abrir el camino para futuras formulaciones de proyectos de investigación y trabajos de grado para estudiantes interesados en el tema.
- *Ejecución:* Reducir significativamente el tiempo en los cálculos moleculares realizados, aprovechando potencialmente los recursos de la máquina.
- *Gestión:* Posibilitar un mayor enfoque en los procesos claves de análisis de resultados y toma de decisiones para acelerar las conclusiones sobre las prácticas desarrolladas.
- *Guía de lineamientos:* Determinar prácticas tecnológicas cuando se requiera de cálculos científicos de procesamiento paralelo, considerando opciones diferentes a una ejecución secuencial.
- *Administración:* Decidir la cantidad de recurso que se desea usar en la computadora y el proceso como requiere realizar las ejecuciones, dando la posibilidad de efectuar actividades diferentes en la máquina utilizada.
- *Interdisciplinariedad:* Trabajar en conjunto los grupos de investigación GRID y GIFAC ambos de la Universidad del Quindío, los cuales tienen dentro de sus líneas de investigación temáticas de informática, computación y estudios en química molecular.

Para finalizar hay que tener claro que la computación de alto desempeño es imprescindible para la investigación que requiera cálculo científico computacional.

9. Referencias

- Universidad Carlos III de Madrid. (2011). Diseño de un Simulador de Dinámica Molecular basado en CORBA. Aliseda Casado, J. Consultado el 16 de Febrero de 2015 en <http://e-archivo.uc3m.es/handle/10016/11822>
- Pontificia Universidad Católica del Perú. (2009). Tesis: Administrador de proyectos de Grid computing que hacen uso de la capacidad de cómputo ociosa de laboratorios informáticos. Iberico Hidalgo, M. A. Consultado el 30 de Enero de 2015 en <http://repositorio.pucp.edu.pe/index/handle/123456789/38605?show=full>
- Universidad de Carabobo. (2014). Introducción a la Computación de Alto Rendimiento: Computación Paralela, Computación Distribuida, Computación Grid y Más. Larrazábal, G. Consultado el 27 de Enero de 2015 en <http://alfa.facyt.uc.edu.ve/~glarraz/clase2.ppt>

- Universidad Nacional Autónoma de México. (2003, Julio). Procesamiento Paralelo en Redes Linux Utilizando MPI. Reyes, V. F., & Jiménez, J. A. Consultado el 16 de Diciembre de 2014 en <http://beta.redes-linux.com/manuales/cluster/mpi-spanish.pdf>
- Universidad Politécnica de Madrid. (2005). El Procesamiento Distribuido y su aplicación al Tratamiento de Imágenes. Hernández, V. M. Consultado el 02 de Febrero de 2015 en <http://www.elai.upm.es/webantigua/spain/Investiga/GCII/personal/mhernandez/PFC/Proyecto.pdf>
- Instituto Politécnico Nacional. (2014). Fundamentos de la Computación. Consultado el 17 de Diciembre de 2014 en http://www.sites.upiicsa.ipn.mx/polilibros/portal/polilibros/p_terminados/PolilibroFC/Unidad IV/Unidad%20IV_4.htm

Sobre Los Autores

- **Diego Armando Alvarado Escobar.** Estudiante de Ingeniería de Sistemas y Computación. alvarado_escobar_diego@hotmail.com
- **Brandon Steven Ramírez Sierra.** Estudiante de Ingeniería de Sistemas y Computación. brandonsteven23@hotmail.com
- **Luis Eduardo Sepúlveda Rodríguez.** Ingeniero de Sistemas y Computación, Magister en Software Libre y Sistemas Operativos. Docente de planta e Investigador del grupo de investigación GRID. lesepulveda@uniquindio.edu.co
- **Nathalia Bibiana Duque Madrid.** Ingeniera de Sistemas y Computación, Magister en Ingeniería de Sistemas y Computación. Docente e Investigadora del grupo de investigación GRID. nbduque@uniquindio.edu.co
- **Christian Andrés Candela Uribe.** Ingeniero de Sistemas y Computación, Magister en Comercio Electrónico. Docente de planta e Investigador grupo del grupo de investigación GRID. christiancandela@uniquindio.edu.co
- **Alberto Sánchez López.** Químico, Magister en Química. Docente e Investigador grupo de investigación GIFAC. asanchezl@uniquindio.edu.co

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2015 Asociación Colombiana de Facultades de Ingeniería (ACOFI)