

PROYECTO EN CURSO

Universidad	Universidad Tecnológica de Pereira
Programa Académico	Ingeniería de Sistemas y Computación
Nombre del Semillero	Semillero de investigación JointDeveloper
Nombre del Grupo de Investigación	Grupo de Investigación Sirius
Línea de Investigación	Computación de Alto Desempeño (HPC)
Nombre del Tutor del Semillero	John Haiber Osorio Rios
Email Tutor	john@sirius.utp.edu.co
Título del Proyecto	Aplicaciones Web, una alternativa para la gestión de procesos dentro de una Arquitectura Distribuida.
Autores del Proyecto	Ramiro Andrés Barrios Valencia Alejandro Esteban Rendón Diosa
Ponente (1)	Alejandro Esteban Rendón Diosa
Documento de Identidad	CC.1088323558
Email	alejorendon@utp.edu.co
Teléfonos de Contacto	3310223 - 3153635626
Nivel de formación de los estudiantes ponentes (Semestre)	9no Semestre
MODALIDAD	PONENCIA Investigación en Curso (X) Investigación Terminada ()
Área de la investigación	<ul style="list-style-type: none"> ● Ciencias Naturales () ● Ingenierías y Tecnologías (X) ● Ciencias Médicas y de la Salud. () ● Ciencias Agrícolas () ● Ciencias Sociales () ● Humanidades () ● Artes, arquitectura y diseño ()

Aplicaciones Web, una alternativa para la gestión de procesos dentro de una Arquitectura Distribuida.

Ramiro Andrés Barrios Valencia¹. Alejandro Esteban Rendón Diosa²

Resumen

En la actualidad se considera que los los Sistemas Distribuidos son una gran herramienta en el área de Computación de Alto Desempeño debido a sus potenciales capacidades de tolerancia a fallos y escalabilidad, ya que permiten adecuarse a la gran demanda de procesamiento de datos (*e.g. Clusters, Grids*). Esta gran demanda implica una buena administración de este tipo de sistemas, por lo cual el tiempo es el recurso más importante del que se dispone.

Al estar construido a partir de componentes bastante similares en la mayoría de casos, los sistemas distribuidos realizan operaciones casi que repetitivas en algunas ocasiones. Estas operaciones (*procesos*) implican el uso de recursos que podrían hacerse con un menor esfuerzo pero conservando su utilidad.

Con el avance tecnológico se dispone de algunas herramientas/técnicas transversales a varias áreas del conocimiento, como lenguajes de programación, los cuales pueden mejorar el rendimiento en diferentes procesos de varias áreas del conocimiento.

Palabras claves: Computación Cluster, Automatización, Sistemas Distribuidos, Aplicación Web, Escalabilidad.

Introducción

Es innegable que el avance tecnológico llegue cada vez más rápido, con ésto también llega la dependencia en las labores y el diario vivir de los seres humanos. Desde la alarma del despertador, hasta los tableros digitales en las clases de las escuelas; la tecnología va involucrándose en nuestro diario vivir, teniendo como propósito principal apoyar la evolución de la sociedad. Un factor que juega un papel importante en toda la historia de modernización es el tiempo. El tiempo se ha convertido en uno de los recursos casi invaluable para los seres humanos, ya que en el medio cambiante actual el buen uso de éste garantiza éxito en las diferentes labores que se desempeñan.

Un concepto bastante conocido en la computación, en el campo de teoría de sistemas (también aplica en Sistemas Operativos y Sistemas Distribuidos) es el *proceso*. Un proceso es una instancia de un programa en ejecución [3], el cual requiere cierto tiempo y algunos recursos para que se ejecute como es lo esperado. Dicho lo anterior, se espera que los procesos en cualquier

¹ ramiro@sirius.utp.edu.co, Universidad Tecnológica de Pereira

² alejorendon@utp.edu.co, Universidad Tecnológica de Pereira

campo necesiten del menor tiempo para su ejecución pero que sigan generando la misma o mejor utilidad de lo que se espera.

Como propuesta, se busca gestionar procesos en una Arquitectura de Cómputo Distribuida mediante el uso de diferentes tecnologías y técnicas, que permitan obtener mejor rendimiento (Disminuir tiempo dedicado y/o recursos) al menor costo posible.

Planteamiento del problema

Algunas características que presenta una Arquitectura de Cómputo Distribuida son la *escalabilidad* y la *elasticidad*. La *escalabilidad* se refiere a la propiedad deseable de un software que indica su habilidad para extender el margen de operaciones sin perder calidad y se complementa con la *elasticidad*, la cual pretende atender a los cambios de demanda de una manera ágil y dinámica [1]. Para lograr que ambas características se complementen en un sistema, es necesario invertir tanto tiempo como infraestructura se quiera, es decir, entre más escale el sistema, más procesos se van a generar (procesos físicos y lógicos. Como configuraciones, manejo de infraestructura, etc).

De lo anterior se expone una de las desventajas en un sistema escalable: el tener que realizar procesos que implican tiempo y a veces soluciones manuales, en algunos casos innecesarios. Para el caso de estudio en este proyecto se muestran diferentes procesos de una Arquitectura Distribuida tipo Cluster.

Con el fin de aprovechar tecnologías modernas, como las tecnologías web que brindan diferentes características relacionadas con adaptarse a Sistemas Distribuidos, se consideran como una alternativa potencial para gestionar procesos en el tipo de arquitecturas ya mencionadas. Es por ello que la investigación busca responder a la hipótesis, ¿Es una aplicación Web una alternativa viable para la gestión de procesos dentro de una Arquitectura Distribuida?

Justificación

La administración de un Sistema en una Arquitectura de Cómputo Distribuida requiere satisfacer cierto tipo de condiciones para que haya una funcionalidad ideal, como lo son: Alta disponibilidad, soluciones de almacenamiento compartido, entre otras consideraciones [4]. Un ejemplo de esta clase de sistemas es el tipo Cluster.

Aunque los procesos en los sistemas descritos anteriormente lucen relativamente sencillos por poseer componentes teóricamente similares, en realidad se modifican por diferentes factores internos y externos; los cuales en ocasiones alteran el sistema y esto implica tener que actuar de manera algo diferente sobre los distintos componentes.

En este proyecto se busca generar medios para gestionar procesos (inicialmente con aplicaciones web), los cuales permitan optimizar tiempos y otros recursos dentro de la arquitectura ya mencionada. Además de contrastar con diferentes tecnologías/técnicas y obtener un grupo de

alternativas que ayuden a mejorar el desempeño en la administración de una Arquitectura de Cómputo Distribuida.

Objetivos

- **Objetivo General**

Determinar la efectividad de una aplicación web para la gestión de procesos dentro de un arquitectura distribuida.

- **Objetivos Específicos**

- Identificar procesos a ser gestionados
- Implementar prototipo funcional de una aplicación web para la gestión de procesos dentro de un arquitectura distribuida.
- Identificar otras técnicas/herramientas para gestionar procesos
- Comparar las diferentes técnicas/herramientas para gestionar procesos

Referente teórico

- **Sistema Distribuido:**

Un sistema distribuido es aquel en el que los componentes son computadores ubicados en red, comunican y coordinan sus acciones sólo mediante el paso de mensajes. [2].

Los sistemas distribuidos se caracterizan por su estructura: una distribución típica consistirá en un gran número de dispositivos interactivos donde cada uno ejecuta sus propios programas, los cuales se ven afectados por la recepción de mensajes, las actualizaciones de memoria compartida o los estados de otros dispositivos. Los sistemas de computación distribuidos van desde sistemas simples en donde un solo cliente habla con un solo servidor, hasta grandes redes amorfas como la Internet. [8]

- **Computación Cluster:**

- **Historia:**

La primera inspiración para la computación Clúster fue desarrollada en los 60's por IBM como una alternativa de enlazar grandes mainframes para proveer un forma más efectiva en costos de paralelismo comercial.

Sin embargo, la computación en Cluster no ganaría momentum hasta que convergieron tres importantes avances en los 80's: microprocesadores de alto desempeño, redes de alta velocidad, y herramientas estandarizadas para la computación distribuida de alto desempeño. Una posible cuarta tendencia fue el incremento en las necesidades de poder computacional de las aplicaciones científicas y comerciales que compitiera con los tradicionales supercomputadores de alto costo y baja accesibilidad.

Los avances recientes en estas áreas y la disponibilidad de componentes baratos y comunes han hecho que la computación en Cluster o computadores en red sea ampliamente aceptadas y fácilmente construida. Y consecuentemente, han surgido como la principal tendencia en plataformas paralelas y distribuidas para alto desempeño, alta tasa de transferencia y alta disponibilidad.

- **Infraestructura:**

Un Cluster es un tipo de sistema computacional paralelo y distribuido, que consiste en una colección de computadores individuales trabajando unidos como un único recurso integrado [5][6]. La arquitectura típica de un Clúster se muestra en la figura TAL. Los principales componentes de un Clúster incluyen múltiples computadores simples (PCs, estaciones de trabajo, etc), sistemas operativos, interconexión de alto desempeño, middleware (capa de interfaz), ambientes de programación paralela y aplicaciones.

Para la construcción de un Clúster se debe tener claros tres elementos: tecnologías de interconexión y software de comunicación, Imagen Unica de Sistema (SSI - Single System Image) y un sistema de administración de recursos. Los Clústers necesitan incorporar tecnologías de interconexión rápidas para soportar altas tasas de transferencia y baja latencia para la comunicación entre procesadores de los diferentes nodos. Tecnologías lentas de interconexión siempre han sido un factor de embotellamiento en la computación Clúster. Seleccionar la red de interconexión depende de varios factores, como compatibilidad entre el hardware del clúster y el sistema operativo, precio y desempeño.

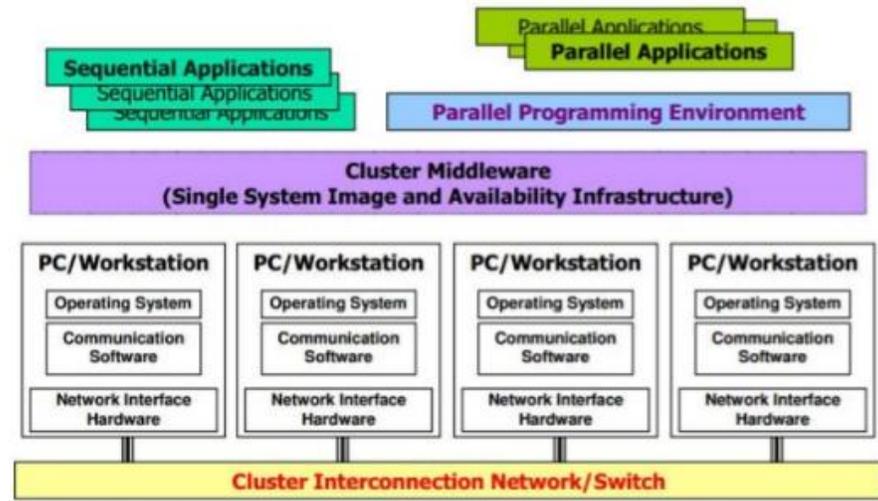


Figura 1: Cluster architecture. R. Buyya [5]

- **Escalabilidad**

La escalabilidad es una propiedad deseable de un sistema, lo que indica su capacidad para manejar cantidades crecientes de trabajo de una manera agradable o su capacidad para mejorar el rendimiento cuando se agregan recursos adicionales

(típicamente hardware). Un sistema cuyo desempeño mejora después de añadir hardware, proporcionalmente a la capacidad añadida, se dice que es un Sistema escalable. De manera similar, se dice que un algoritmo escala si es adecuadamente eficiente y práctico cuando se aplica a situaciones grandes (por ejemplo, un gran conjunto de datos de entrada o un gran número de nodos participantes en el caso de un sistema distribuido). Si el algoritmo no funciona cuando los recursos aumentan, entonces no es escalable [7].

- **Gestión de Procesos**

- **Automatización**

La automatización de los procesos es la sustitución de tareas tradicionalmente manuales por las mismas realizadas de manera automática por máquinas, robots o cualquier otro tipo de automatismo. La automatización tiene ventajas muy evidentes en los procesos industriales. Se mejora en costes, en servicio y en calidad. El trabajo es más rápido y no necesita de una cantidad determinada de operarios, que antes eran necesarios. Además se producen menos problemas de calidad por realizarse el trabajo de una manera más uniforme debido a las especificaciones dadas al automatismo. Otras ventajas se obtienen de la automatización son el aumento de producción, menor gasto energético, mayor seguridad para los trabajadores.

Metodología, enfoque y tipo de investigación

Basados en los puntos anteriores se considera que éste es un Proyecto de investigación; ya que se pretende que por medio de la obtención de nuevos conocimientos a través de la investigación, experiencia y observación activa se produzca un resultado o entregable después de la culminación (El entregable esperado es la herramienta/técnica, la cual permita alcanzar los objetivos propuestos).

En esta investigación se utiliza un enfoque cuantitativo. El objetivo de una investigación cuantitativa es el de adquirir conocimientos fundamentales y la elección del modelo más adecuado que permita conocer la realidad de una manera más imparcial, ya que se recogen y analizan los datos a través de los conceptos y variables. Se elige este tipo de investigación ya que para apoyar las conclusiones y los resultados que se esperan es necesario comparar y estudiar una serie de variables, las cuales indican si la hipótesis propuesta para la investigación es válida.

Resultados esperados

- Construir una Aplicación web funcional que cumpla con los requisitos establecidos en la investigación
- A partir de la presente investigación, se espera explorar otras posibles alternativas para la gestión de procesos dentro de una Arquitectura distribuida (Además del expuesto inicialmente)
- Identificar procesos los cuales al ser gestionados de una manera correcta pueden producir mejoras en el rendimiento del sistema tratado.

- Con una comparativa entre las diferentes tecnologías/técnicas investigadas, se espera poder realizar una clasificación sobre el uso y la importancia de cada una.

Conclusiones

- A pesar de que hay una técnica propuesta como base de la investigación, es importante plantear una serie de tecnologías alternas para poder realizar comparaciones y así hallar la alternativa más viable.
- La gestión de procesos podría implicar una serie de consecuencias no estimadas si no se selecciona correctamente la manera de abordar el problema.
- No todos los procesos en una Arquitectura tipo Cluster se pueden gestionar según lo planteado en el proyecto. Por esta razón es requerido una clasificación de éstos, ya que podrían alterar el funcionamiento del sistema, agregando carga de procesamiento innecesario.

Impactos

Como se ha mencionado en apartados anteriores, existen diferentes factores que llevan a que los diferentes procesos impliquen el uso de recursos que a veces son innecesarios ya que se podrían completar con el menor costo y generando la misma o mejor utilidad. Por esta razón se considera que al cumplir la serie de objetivos planteados se puede lograr un buen impacto económico en el área de aplicación.

Además, se considera que el impacto económico en el área de aplicación es bastante importante, ya que los recursos aprovechados del resultado de la buena gestión pueden ser utilizados en otros. Lo cual beneficiará tanto al sistema como a los usuarios del mismo.

Bibliografía

- [1] F. Godino González, "Diseño de una arquitectura escalable y de alta disponibilidad para un sistema middleware", Doctorado, Universidad Politécnica de Madrid, 2013.
- [2] G. Coulouris, J. Dollimore, T. Kindberg and G. Blair, DISTRIBUTED SYSTEMS Concepts and Design, 5th ed. Addison-Wesley, 2017, p. 1.
- [3] A. Silberschatz, G. Gagne and P. Baer Galvin, Operating System Concepts, 9th ed. p. Capítulo 3.
- [4] J. Carr, "INTRODUCTION TO LINUX CLUSTERING", 2016.
- [5] R. Buyya (ed.). High Performance Cluster Computing: Architectures and Systems. Prentice Hall, 1 edition, 1999.
- [6] G. F. Pfister. In Search of Clusters. Prentice Hall, 2 edition, 1998.
- [7] D. Agrawal, A. J. Elmore, S. Das and A. El Abbadi, "Database Scalability, Elasticity, and Autonomy in the Cloud", p. 5, 2017.
- [8] J. Aspnes, "Introduction", in Notes on Theory of Distributed Systems, 1st ed., Yale University, 2016.