

CARTA DE ACEPTACIÓN DE PROYECTO TERMINAL

TÍTULO

Esquemas de distribución de trabajo para un sistemas ERP en la nube.

AUTORES

110300083

Cindy Guadalupe Canul Canul

110300083@ucaribe.edu.mx

110300097

Cristian Ariel Kumul Uc

110300097@ucaribe.edu.mx

110300079

Jonathan Peraza Feliciano

110300079@ucaribe.edu.mx

ASESOR

Joel Antonio Trejo Sánchez


RESUMEN


El cómputo en la nube es una tecnología prometedora como plataforma del futuro, permite el acceso a recursos de hardware de manera remota. Dentro del funcionamiento de los servicios es necesario optimizar los recursos en el centro de datos y para esto se utilizan los esquemas de calendarización. El problema incrementa con el número de tareas, pues es un problema NP-difícil, debido a las distintas características de cada host.

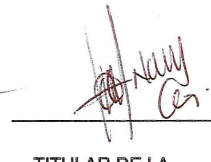
En este proyecto se estudiará y se hará una comparación de varios esquemas de calendarización para proponer una mejora en la optimización de recursos de acuerdo a las peticiones de un sistema ERP en la nube.

FIRMAS




AUTORES


ASESOR


TITULAR DE LA
ASIGNATURA

Resumen

El cómputo en la nube es una tecnología prometedora como plataforma del futuro, permite el acceso a recursos de *hardware* de manera remota. Dentro del funcionamiento de los servicios es necesario optimizar los recursos en el centro de datos y para esto se utilizan los esquemas de calendarización. La complejidad de la administración de los recursos y la calendarización incrementa con el número de tareas, pues es un problema NP-difícil debido a las distintas características de cada *host* y lo heterogéneas que son las tareas. En este proyecto se estudiaron y compararon cuatro esquemas de calendarización para proponer una mejora en la optimización de recursos en un sistema *ERP* en la nube, para ello se utilizaron dos métricas; costo de procesamiento y tiempo de ejecución de las tareas.

Abstract

Cloud Computing is a promising technology as a platform of the future, allows access to hardware resources remotely. The main goal on a cloud environment is optimize the resources on the Data Center, for this scheduling algorithms are used.

The complexness of resources administration and scheduling increases with the number of tasks, is a NP-hard problem due to the heterogeneous nature of the tasks and resources. In this project they were studied and compared fourth scheduling schemes to propose an improvement in resources in an ERP system in the cloud. Two metrics were used; processing cost and time of execution of tasks.

Índice general

| | |
|--|-----------|
| Resumen | 3 |
| Abstract | 4 |
| Introducción | 10 |
| Antecedentes y Planteamiento del problema | 10 |
| Propuesta de solución y Justificación | 11 |
| Objetivos | 13 |
| Objetivo general | 13 |
| Objetivos específicos | 13 |
| Hipótesis | 14 |
| 1. Marco contextual | 15 |
| Introducción | 15 |
| 1.1. Situación Actual | 16 |
| 1.2. Estado del Arte | 17 |
| 1.3. Marco teórico | 18 |
| Centro de Datos | 18 |
| Esquemas de Distribución de Trabajos | 19 |
| Balanceo de Cargas | 20 |
| Migración | 21 |
| Sistemas ERP | 21 |
| 1.4. Marco Metodológico | 21 |
| 2. Desarrollo | 22 |
| Introducción | 23 |
| 2.1. Aplicación del marco metodológico y de actividades de experimentación | 23 |

| | |
|---|-----------|
| 2.1.1. Simulación del Centro de Datos en la Nube | 23 |
| 2.1.2. Implementación de los Algoritmos | 25 |
| Configuración de Elementos del Datacenter y DatacenterBrokers para Algoritmos de Calendarización | 29 |
| DatacenterBrokers para Algoritmos de Calendarización | 31 |
| 3. Resultados | 36 |
| Introducción | 36 |
| 3.1. Tiempo de ejecución y costo de procesamiento | 37 |
| Conclusiones | 44 |
| Glosario | 47 |
| A | 47 |
| C | 47 |
| E | 47 |
| F | 47 |
| P | 48 |
| Anexos | 49 |

Índice de figuras

| | | |
|-----|--|----|
| 1. | Propuesta de arquitectura en la nube, Fuente: Elaboración propia. | 12 |
| 2. | Esquema general del Cómputo en la Nube, Fuente: Elaboración propia. | 19 |
| 3. | Estilo de Trabajo de <i>CloudSim</i> , Fuente: Chatterjee et al. . . . | 23 |
| 4. | Esquema de trabajo de algoritmos de calendarización, Fuente: Elaboración propia. | 25 |
| 5. | Arquitectura FCFS para un entorno en la nube. Fuente: Elaboración propia. | 26 |
| 6. | Arquitectura de <i>Min-Min</i> para un entorno en la nube. Fuente: Elaboración propia. | 27 |
| 7. | Arquitectura de <i>Max-Min</i> para un entorno en la nube. Fuente: Elaboración propia. | 28 |
| 8. | Características del <i>datacenter</i> . Fuente: Elaboración propia . . | 29 |
| 9. | Características del <i>Host</i> . Fuente: Elaboración propia | 29 |
| 10. | Características de las máquinas virtuales. Fuente: Elaboración propia | 30 |
| 11. | Características de las Tareas (<i>cloudlets</i>). Fuente: Elaboración propia | 31 |
| 12. | <i>FCFS Broker</i> . Fuente: Elaboración propia | 32 |
| 13. | <i>Min-Min Broker</i> . Fuente: Elaboración propia | 33 |
| 14. | <i>Max-min Broker</i> . Fuente: Elaboración propia | 34 |
| 15. | <i>Algoritmos de ordenamientos</i> . Fuente: Elaboración propia . . . | 35 |
| 15. | Promedio tiempo de ejecución con tareas 100-500, Fuente: Elaboración propia. | 39 |

| | | |
|-----|--|----|
| 16. | Promedio costo de procesamiento con tareas 100-500, Fuente: Elaboración propia. | 40 |
| 17. | Tiempo ejecución 50 muestras <i>FCFS</i> , Fuente: Elaboración propia | 41 |
| 18. | Tiempo ejecución 50 muestras <i>Max-Min</i> , Fuente: Elaboración propia | 42 |
| 19. | Tiempo ejecución 50 muestras <i>Min-Min</i> , Fuente: Elaboración propia | 43 |

Índice de cuadros

| | | |
|----|--|----|
| 1. | Configuración <i>Datacenter</i> , Fuente: Elaboración propia. | 37 |
| 2. | Configuración de <i>Host</i> , Fuente: Elaboración propia. | 37 |
| 3. | <i>Virtual Machine</i> , Fuente: Elaboración propia. | 38 |
| 4. | Configuración <i>Cloudlet</i> , Fuente: Elaboración propia. | 38 |
| 5. | Desviación estándar del tiempo de ejecución, Fuente: Elabo- ración propia | 43 |

Introducción

Antecedentes y Planteamiento del problema

El cómputo en la nube es una tecnología que ha abarcado gran parte de los negocios para dar soporte a ellos. Ésta tecnología provee a los negocios una ventaja competitiva en los costos de recursos como se ve en los negocios tradicionales, además de la versatilidad que provee al ajustarse a la necesidad de las empresas (Srinivasan, 2014).

La tecnología en la nube ha desarrollado una infraestructura fuerte después del surgimiento del cómputo distribuido (Chen and Deng, 2009). Para obtener las ventajas de dicha tecnología los usuarios simplemente necesitarán conectarse a internet y de esta manera tendrán el acceso al procesamiento de manera remota (Aranganathan and Mehata, 2011). Sin embargo, para aprovechar el máximo potencial de éstos recursos, es necesario tener en consideración algunas variables, ya que en un entorno en la nube existe un comportamiento dinámico de los recursos a manera que se les provea a los usuarios el servicio (Shimpy and Sidhu, 2014). Una de las prácticas con mayor importancia en la nube es la calendarización, ya que tiene como objetivo administrar las tareas del centro de datos para optimizar los recursos del mismo. De esta manera la eficiencia de la carga de trabajo en la nube aumenta (Shimpy and Sidhu, 2014). En general, el objetivo de la calendarización en la nube es utilizar los recursos de manera apropiada, mientras la carga de trabajo es distribuida uniformemente para mejorar los tiempos de ejecución (Shimpy and Sidhu, 2014). Debido a la atención que se tiene en la tecnología en la nube, los centros de datos han tomado un papel muy importante para los servicios empresariales (Shimpy and Sidhu, 2014). Un centro de datos está compuesto por miles de servidores virtuales ejecutándose en una instancia de tiempo alojando muchas tareas, al mismo tiempo el centro de datos recibe miles de peticiones a esas tareas. Es aquí en donde la

programación de trabajos tiene un rol muy importante para el cómputo en la nube, ya que influye en el rendimiento del mismo (Srinivasan, 2014).

El problema de la calendarización pertenece a los algoritmos NP-Difícil, lo cual tiene un amplio rango de soluciones posibles y se toma mucho más tiempo de encontrar una respuesta óptima, ya que no existe un método para resolver estas incógnitas. Sin embargo, es posible estar cerca de la mejor solución contemplando algunos entornos (Shimpy and Sidhu, 2014).

Propuesta de solución y Justificación

Los centros de datos son una parte esencial en la tecnología en la nube, están conformados por varios servidores virtuales que alojan varios trabajos, ejecutándose por estancias de tiempo y a la vez recibiendo peticiones a esos trabajos (Shimpy and Sidhu, 2014). Para que el cómputo en la nube pueda rendir correctamente, se necesita una buena programación de trabajos o calendarización. Este problema tiene un gran número de soluciones posibles, toma mucho tiempo en encontrar una respuesta óptima y por eso está dentro de los algoritmos NP-Difícil. No existe un método para resolver lo anterior, sin embargo es posible obtener una solución cercana contemplando algunos entornos (Shimpy and Sidhu, 2014). En esta investigación se propone una mejora a un algoritmo de calendarización para el caso de estudio de un sistema *ERP* en la nube, teniendo en cuenta sus posibles escenarios y peticiones heterogéneas. De esta manera se podrá contar con un esquema de distribución de trabajos que satisfaga las necesidades de un sistema *ERP*, mejorando el tiempo de respuesta y aprovechando los recursos. Con beneficios a los proveedores de servicios (*SaaS*) de sistemas *ERP*.

En la figura (1) se puede observar la arquitectura del centro de datos que se implementará en la investigación, el cual está compuesto de los siguientes elementos:

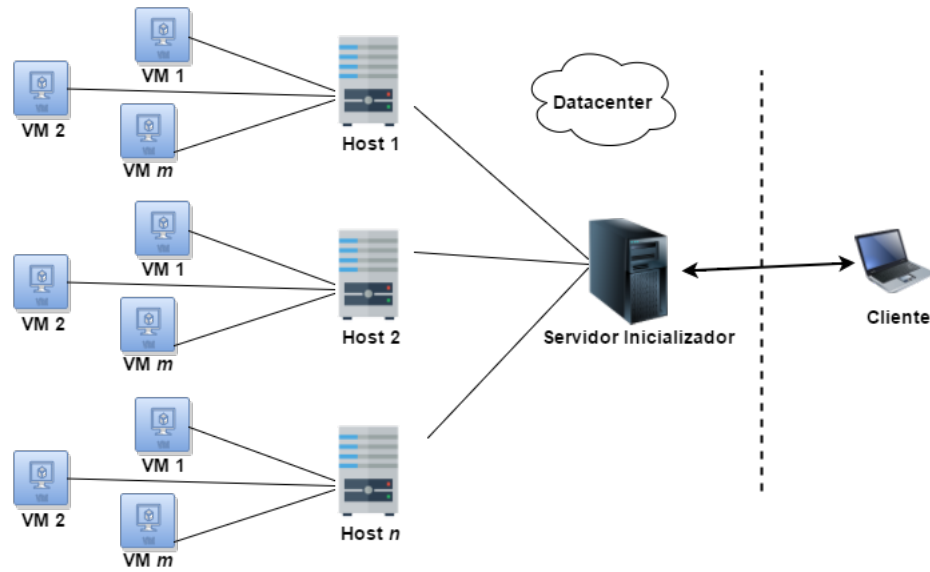


Figura 1: Propuesta de arquitectura en la nube, Fuente: Elaboración propia.

- **Servidor administrador:** Es el servidor que tendrá el rol de *front end* en el centro de datos, teniendo una interacción directa con las peticiones de los usuarios, que son especificados en éste documento como trabajos. El principal objetivo de éste elemento es delegar los trabajos a las máquinas virtuales en los distintos *hosts* del centro de datos.
- **Host:** es el recurso de *hardware* que se tiene en el centro de datos. Una de las características de éste elemento es que es finito.
- **Máquinas Virtuales:** son instancias dentro de los host, tienen como objetivo resolver los trabajos que se les sea asignado.

Objetivos

Objetivo general

Determinar y proponer un esquema que permita resolver de manera eficiente el problema de la calendarización de un centro de datos y satisfacer las necesidades de un sistema *ERP* en la nube.

Objetivos específicos

- Definir una arquitectura para un centro de datos en la nube con las necesidades de un sistema *ERP*.
- Realizar la simulación en base a la arquitectura del centro de datos.
- Realizar e implementar diferentes esquemas de distribución de trabajos, rescatando sus características principales.
- Reunir las características principales y proponer el nuevo esquema de distribución.
- Aplicar el nuevo esquema de distribución a las necesidades de un *ERP*.
- Comparar los resultados obtenidos del nuevo esquema propuesto y de los anteriores para determinar si hubo una mejora.

Hipótesis

La modificación de un esquema de distribución mejorará el rendimiento de la calendarización para un sistema *ERP* en la nube.

Para realizar la verificación de validez de la hipótesis se usarán:

- **Pruebas de *hardware y software*:**
 - **Prueba de desempeño:** Validar el tiempo de respuesta para las tareas bajo condiciones normales y máximas.
 - **Prueba de carga:** Verificar el tiempo de respuesta del sistema, bajo diferentes condiciones de carga. Se mide el tiempo de respuesta y otros requisitos sensibles al tiempo.
- **Estadísticas:** en base a los resultados que se obtendrán en las pruebas de desempeño y de carga, se va realizar comparaciones y posteriormente visualizarlas por medio de gráficas y tablas.

Capítulo 1

Marco contextual

Introducción

En éste capítulo se describen algunos marcos conceptuales necesarios para fundamentar y analizar los requisitos de la investigación, las secciones son las siguientes: Situación actual, Estado del Arte, Marco Teórico y Marco Metodológico. En la primera sección se muestra cómo se encuentra la investigación en la actualidad, cómo se encuentra nacional e internacionalmente y cuál es el punto de partida del proyecto; en la siguiente sección se presenta la información sobre los diferentes esquemas de distribución que serán implementados durante la elaboración del proyecto. El Marco teórico nos describe los conceptos necesarios para fundamentar y comprender mejor el proyecto y su propósito, finalmente en la última sección se describe los pasos que se van a seguir para realizar y alcanzar los objetivos del proyecto.

1.1. Situación Actual

Actualmente, son muchas las organizaciones que utilizan el cómputo en la nube, ya que trae beneficios en cuanto a gastos y recursos. Ya no hay necesidad de invertir tanto en la infraestructura de *TI (Information Technology)* y existe una mejor capacidad de almacenamiento para los recursos que el ámbito empresarial necesita. El cómputo en la nube utiliza los servicios *SaaS (Software as a Service)* y *HaaS (Hardware as a Service)* para hacer más eficiente y flexible el uso de la infraestructura de la nube (Mariscal and Gil-Garcia, 2013).

México utiliza, en cuanto a la infraestructura de almacenamiento virtual, los servicios ofrecidos por *Amazon, Google y Triada (TELMEX)*. Los cuales proporcionan almacenaje de información virtual, plataformas de hospedaje de aplicaciones y servicios en línea (Mariscal and Gil-Garcia, 2013).

Al proveer el servicio se tiene dos escenarios diferentes: cuando hay una mayor demanda de recursos y cuando no la hay, eso va dependiendo de las necesidades de las empresas. Para que se tenga una mejor eficacia durante los servicios, se tiene que mejorar el uso de los recursos en el centro de datos. Dentro del centro de datos hay muchos servidores virtuales que están recibiendo trabajos, mientras la nube las mantiene en los lotes de solicitud de trabajos. Es aquí, donde resaltamos la importancia de la distribución de trabajos dentro del centro de datos (Shimpy and Sidhu, 2014).

En este momento, la distribución de trabajos es un problema NP- difícil, pues que no se ha encontrado una solución óptima. Sin embargo, existen diferentes propuestas para encontrar una mejor solución, utilizando diferentes esquemas de distribución o dicho de otra forma algoritmos (Shimpy and Sidhu, 2014).

Existen una gran variedad de algoritmos utilizados en el problema de distribución de trabajos. Los algoritmos más utilizados en las investigaciones son: *FCFS, Round Robin, Min-Min, Max-Min y Metaheurística*. Cabe mencionar que en las investigaciones solo hacen una selección de una serie de algoritmos para que, por medio de pruebas, determinen quién es el mejor de ellos.

1.2. Estado del Arte

Los siguientes algoritmos de calendarización actualmente están prevaleciendo en la nube:

- ***Resource-Aware-Scheduling Algorithm (RASA)***: Parsa, Entezari-Maleki (2009) proponen el algoritmo *RASA*, el cual utiliza las ventajas de dos algoritmos tradicionales (*Max-min* y *Min-min*) y cubre sus desventajas. Aunque el tiempo límite, la tasa de llegada, costo de ejecución y costo de comunicación no están considerados (Parsa and Entezari-Maleki, 2009).
- ***RSDC (Reliable Scheduling Distributed In Cloud Computing)***: Delevar, Javanmard, Shabestari y Talebi (2012) proponen un algoritmo confiable en un entorno en la nube, en este algoritmo los trabajos importantes son divididos en sub trabajos, de tal manera que se puedan balancear las peticiones (Delavar et al., 2012).
- ***An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment***: Dakshayini, Gurupasad (2011), proponen un nuevo algoritmo de calendarización que se basa en la prioridad y un esquema de control de admisión. En este algoritmo, la prioridad se asigna a cada proceso admitido en la cola (Dakshayini and Guruprasad, 2011) .
- ***Pre-emptable Shortest Job Next Scheduling algorithm (PSJN)***: Este algoritmo se propone en una nube privada. Utiliza la técnica de suscripción preferente del algoritmo de *Round Robin* junto con el siguiente proceso más corto (*PSN*). Brinda beneficios de costos y mejora tiempo de respuesta y tiempo de ejecución (Nishant and R., 2015).
- ***User-priority Guided Min-min scheduling algorithm***: Se realiza una mejora para el algoritmo de balanceo de cargas, a través del algoritmo *Min-min* para la calendarización de trabajos con el fin de minimizar el tiempo de terminación del último trabajo (*makespan*) y maximizar la utilización de los recursos (Chen et al., 2013).

1.3. Marco teórico

El cómputo en la nube es una tecnología emergente, la cual está compuesta por un grupo de recursos heterogéneos que proveen servicios a través de internet (Agarwal et al., 2014). Esta tecnología permite a los consumidores y negocios utilizar aplicaciones sin necesidad de instalación y con acceso a sus archivos personales en cualquier computadora con acceso a internet (Chowdhury et al., 2012). El cómputo en la nube se puede clasificar de dos maneras:

- Por la ubicación:
 - **Nube Pública:** La infraestructura de cómputo se puede compartir entre cualquier organización (Chowdhury et al., 2012).
 - **Nube Privada:** La infraestructura de cómputo es dedicada a una organización en particular y no se comparte con otras organizaciones (Chowdhury et al., 2012).
 - **Nube Híbrida:** Las organizaciones pueden albergar sus aplicaciones críticas en nubes privadas y las aplicaciones con menos problemas de seguridad las puede albergar en nubes públicas (Chowdhury et al., 2012).
- Por el tipo de servicios ofrecidos:
 - **Infrastructure as a Service (IaaS):** En éste nivel, la infraestructura se ofrece como servicio hacia los solicitantes en forma de Máquinas Virtuales (VM) (Agarwal et al., 2014).
 - **Platform as a Service (PaaS):** Es una plataforma de desarrollo de aplicaciones que se provee como servicio hacia los desarrolladores para crear aplicaciones basadas en web (Agarwal et al., 2014).
 - **Software as a Service (SaaS):** En éste nivel el proveedor en la nube provee las aplicaciones de *software* (Agarwal et al., 2014).

Centro de Datos

Un centro de datos (*datacenter*) es un espacio dedicado donde las organizaciones almacenan y operan la mayoría de su información y la infraestruc-

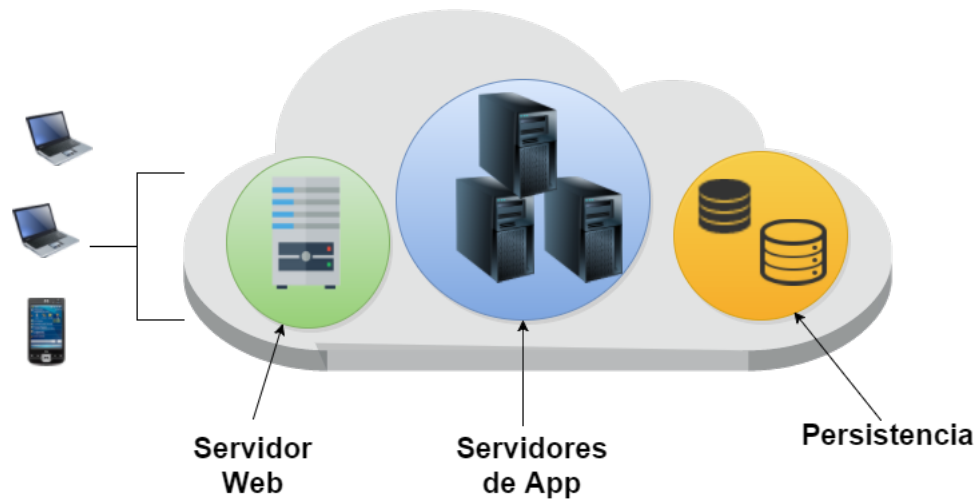


Figura 2: Esquema general del Cómputo en la Nube, Fuente: Elaboración propia.

tura de las tecnologías de comunicaciones que respaldan sus negocios (EPI, 8 31).

Esquemas de Distribución de Trabajos

La distribución de trabajo (calendarización) es uno de los puntos principales además de los más retadores problemas en el cómputo en la nube (Li et al., 2014). Esta consiste en un conjunto de políticas para controlar el orden en el cual los procesos serán ejecutados por el sistema (Agarwal et al., 2014).

Existen varios tipos de algoritmos de calendarización, su principal ventaja es obtener un alto rendimiento al cumplir con las tareas. Los principales ejemplos de algoritmos de calendarización son (Salot, 2013):

- **FCFS (*First Come First Serve*) Algorithm:** significa que el trabajo que llegue primero será el primero en ser ejecutado.
- **Round Robin Algorithm:** consiste en agregar un tiempo de ejecución para cada trabajo, si dicho tiempo se agota y el trabajo actual no ha finalizado, pasa a un estado de espera y continúa con el siguiente trabajo, al finalizar el último trabajo regresa a revisar si existen trabajos en espera para proceder a ejecutarlos.

- ***Min-Min Algorithm:*** se va ejecutando una por una iniciando por los trabajos de menor tamaño.
- ***Max-Min Algorithm:*** se va ejecutando una por una iniciando por los trabajos de mayor tamaño.
- ***Most fit task scheduling algorithm:*** este algoritmo busca los elementos más aptos en la cola de trabajos para ejecutarlo primero. Tiene un alto radio de error.
- ***Priority scheduling algorithm:*** la idea básica es que a cada proceso se le asigna una prioridad, y dependiendo a las prioridades de los demás procesos es cómo se ejecutarán cada uno de ellos.

Balanceo de Cargas

Dentro de nuestro entorno en la nube cada *host* tiene recursos de *hardware* finitos y son susceptibles a fallos. Para mitigar contra las fallas y el exhaustivo uso de los recursos, los *host* son agrupados en *clusters*, los cuales son esencialmente un agrupamiento de recursos compartidos. El administrador (*manager*) es capaz de asegurarse que ninguno de los *host* dentro del *cluster* es responsable de todas las máquinas virtuales dentro de ese *cluster* (Dover et al., 2013).

Un entorno de virtualización responde a los cambios en la demanda para los recursos en cada *host* haciendo uso del balanceo de cargas, calendarización de trabajos y migración.

Una política de balanceo de cargas es configurada para un *cluster*, que a su vez contiene multiples *hosts*, cada uno de ellos con distintos recursos de *hardware* y disponibilidad de memoria (Dover et al., 2013).

Existen tres políticas para el balanceo de cargas:

- Sin Balanceo.
- Distribución Uniforme.
- Ahorro de Energía.

Migración

La migración se utiliza para hacer cumplir las políticas dentro del balanceo de cargas. La migración de una máquina virtual toma lugar de acuerdo a las políticas de balanceo de cargas para un *cluster* y la demanda actual sobre los *hosts* dentro de dicho *cluster*. La migración de igual manera puede ser configurada para ocurrir automáticamente cuando un *host* es bloqueado o movido a modo de mantenimiento (Dover et al., 2013).

Sistemas *ERP*

La sigla *ERP*, en inglés *Enterprise Resource Planning*, significa Planificación de los Recursos de la Empresa. Un sistema *ERP* constituye un marco de trabajo que incluye aplicaciones comerciales, administrativas (finanzas, contabilidad), recursos humanos, planeamiento de manufactura y gestión de proyectos (S. and Horacio, 2002).

1.4. Marco Metodológico

A continuación se propone la siguiente metodología para llevar a cabo el proyecto:

- **Simular:** Se implementará a manera de simulación un centro de datos con un entorno en la nube, las máquinas virtuales y el servidor inicializador que lo conforman.
- **Implementar:** En el centro de datos se desarrollarán los algoritmos de calendarización que se mencionan con antelación.
- **Evaluar:** Se simulará el comportamiento de las peticiones de un sistema contemplando los distintos escenarios del *ERP* y consumiendo el servicio en la nube (*SaaS*), para conocer el rendimiento en tiempo de ejecución de los algoritmos.
- **Mejorar:** Se propondrá una mejora a algún algoritmo de acuerdo a las necesidades y comportamiento de un sistema *ERP* en la nube.
- **Comparar:** Se realizará una comparativa de tiempo de ejecución entre la versión mejorada y la original para el caso de estudio de un sistema *ERP*.

Capítulo 2

Desarrollo

Introducción

En éste capítulo se muestra la arquitectura de los distintos esquemas de calendarización de trabajos en un entorno en la nube, para la implementación de dichos esquemas se hace uso de un *framework* llamado *CloudSim*, el cual será definido de igual manera.

2.1. Aplicación del marco metodológico y de actividades de experimentación

En base de los dos primeros puntos descritos anteriormente en el marco metodológico se realizaron las siguientes actividades:

- **Simular:** Se implementará a manera de simulación un centro de datos con un entorno en la nube, las máquinas virtuales y el servidor inicializador que lo conforman.
- **Implementar:** En el centro de datos se desarrollarán los algoritmos de calendarización que se mencionan con antelación.

2.1.1. Simulación del Centro de Datos en la Nube

Cloudsim es un nuevo, generalizado y extensible *framework* de simulación, que permite el modelaje, simulación y experimentación de infraestructuras emergentes de cómputo en la nube y servicios de aplicación (Calheiros, 2011, p. 2).

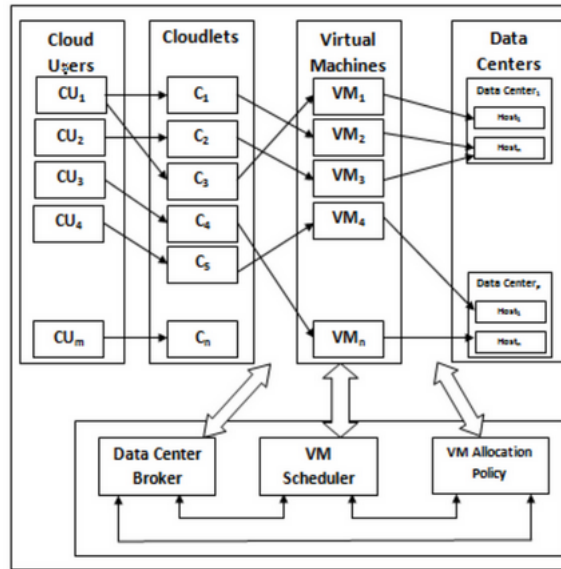


Figura 3: Estilo de Trabajo de *CloudSim*, Fuente: Chatterjee et al.

Entre los componentes que proporciona dicho *framework* se encuentran los siguientes:

- **Cloudlet:** Esta clase modela las aplicaciones de servicio basadas en la nube como pueden ser envío de contenido, redes sociales, y flujo de trabajo empresarial (Calheiros, 2011, cap. 4).
- **Datacenter:** Esta clase modela el núcleo de los servicios en un nivel de infraestructura (*hardware*) que son ofrecidos por *Cloud Providers* (*Amazon, Azure, App Engine*). Estos son encapsulados en un conjunto de *host* que pueden ser homogéneos o heterogéneos con respecto a sus configuraciones de *hardware* (memoria, núcleos, capacidad, y almacenamiento) (Calheiros, 2011, cap. 4).
- **DatacenterBroker:** Esta clase modela un *broker*, el cual es responsable de mediar las negociaciones entre el *SaaS* y los *Cloud providers*; y dichas negociaciones son manejadas por los requerimientos *QoS* (Calheiros, 2011, cap. 4).
- **Host:** Esta clase modela los recursos físicos como una computadora o un servidor de almacenamiento (Calheiros, 2011, cap. 4).
- **Vm:** Esta clase modela una Máquina Virtual (*VM*), la cual es administrada y hosteada por un componente *host* en la nube. Cada *VM* tiene acceso a un componente que almacena las siguientes características relacionadas a una *VM*: memoria accesible, procesador, tamaño de almacenamiento (Calheiros, 2011, cap. 4).

2.1.2. Implementación de los Algoritmos

Existen varios algoritmos para calendarizar los trabajos en el cómputo en la nube. La mayor ventaja de estos algoritmos es obtener el mayor rendimiento. Los principales ejemplos de algoritmos de calendarización son: *FCFS*, *Round Robin*, *Min-Min*, *Max-Min* y *algoritmos de metaheurísticas* (Shimpy and Sidhu, 2014, p. 1).

De los algoritmos mencionados anteriormente, se presentan los siguientes:

- **FCFS:** Ejecuta las tareas en orden de llegada, es decir, el primero en llegar es el primero en ser atendido.
- **Min-Min:** Selecciona las tareas más pequeñas para ser ejecutadas primero.
- **Max-Min:** Selecciona las tareas más grandes para ser ejecutadas primero.

En un entorno de trabajo normal, la forma descrita anteriormente la podemos ver de la siguiente manera, Figura (4):

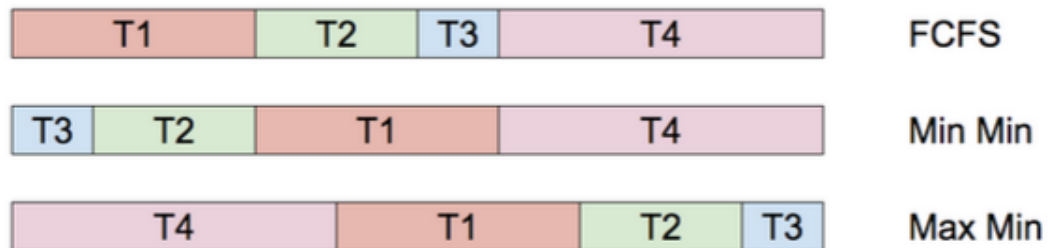


Figura 4: Esquema de trabajo de algoritmos de calendarización, Fuente: Elaboración propia.

Sin embargo en un entorno en la nube, al ser múltiples máquinas virtuales alojadas en distintos *hosts*, que a su vez pueden formar parte de uno o más *datacenters*, dicho esquema tiene que ser modificado para poder adoptar un estilo de trabajo similar al proporcionado por el *framework*, por lo que los resultados quedan de la distinta manera:

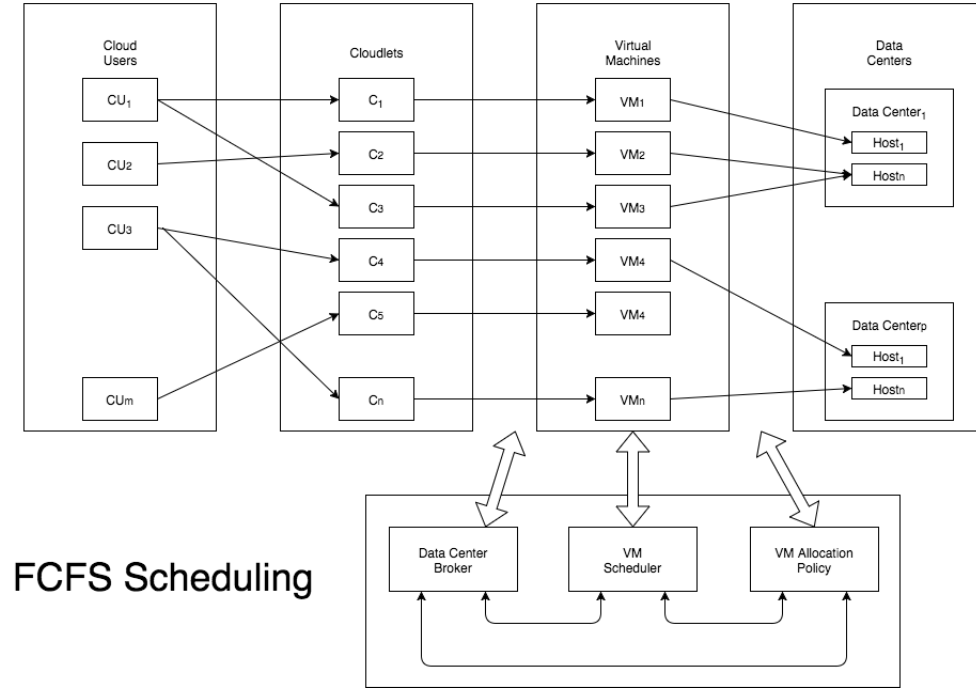


Figura 5: Arquitectura FCFS para un entorno en la nube. Fuente: Elaboración propia.

Como podemos apreciar en el diagrama (Figura 5), podemos tener m usuarios ejecutando n tareas, sin embargo la asignación de máquinas virtuales va dependiendo del orden de llegada de dichas tareas, y quien se encarga de repartir las tareas es el *datacenterBroker*.

De manera similar tenemos los siguientes algoritmos *Min-Min* (Figura 6) y *Max-Min* (Figura 7):

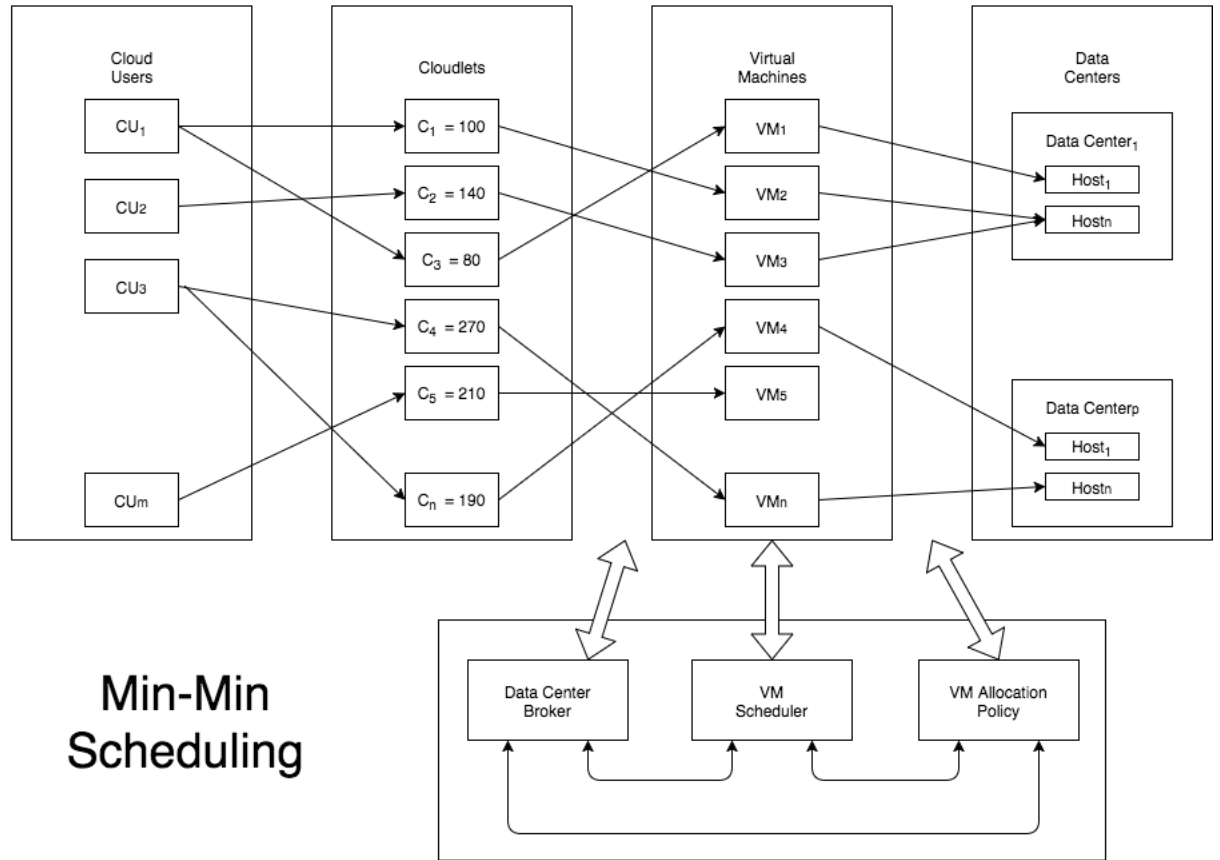


Figura 6: Arquitectura de *Min-Min* para un entorno en la nube. Fuente: Elaboración propia.

Para el algoritmo de *Min-Min* (Figura 6), obtenemos la lista de tareas a partir de la información de ellas, procedemos a ordenarlas de menor a mayor tamaño, asignando la de menor tamaño en la primera máquina virtual, la siguiente tarea, de acuerdo al tamaño, pasa a la segunda máquina y así sucesivamente.

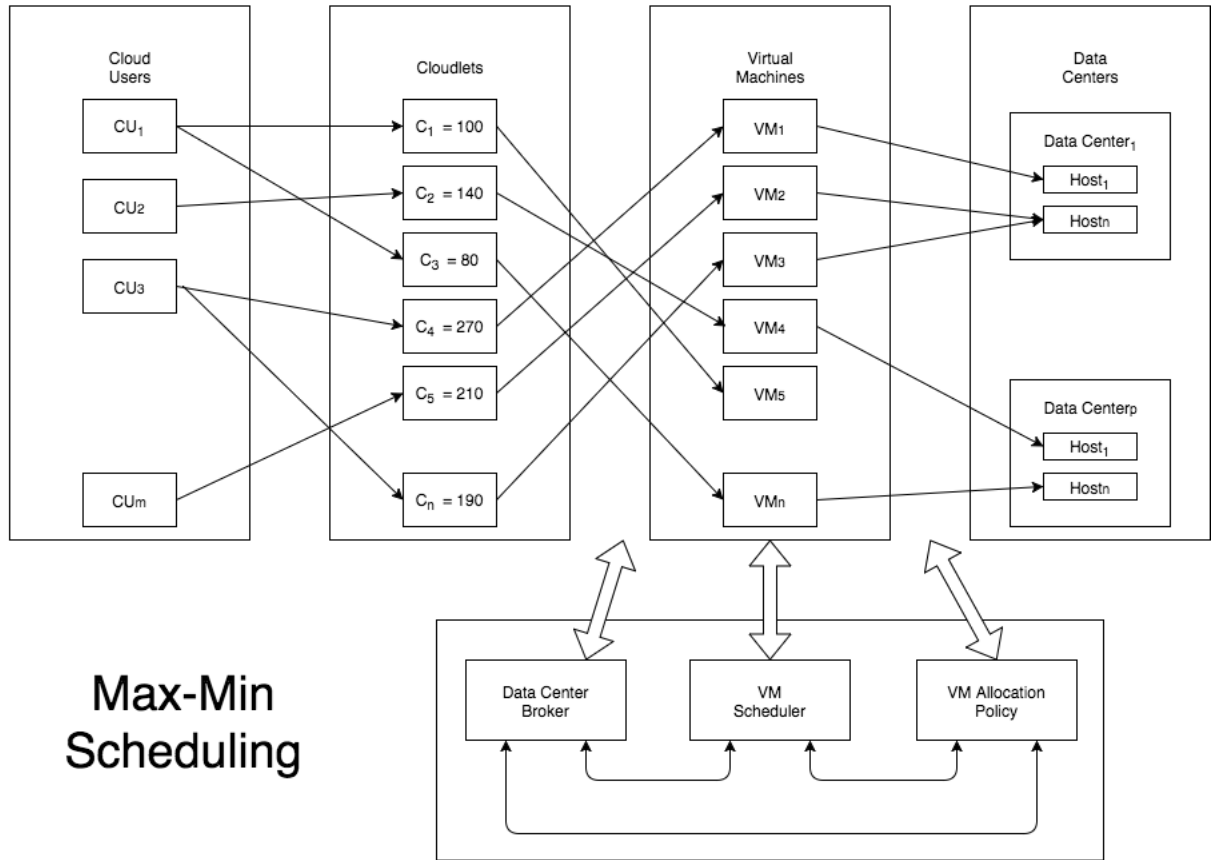


Figura 7: Arquitectura de *Max-Min* para un entorno en la nube. Fuente: Elaboración propia.

En la figura (7) tenemos el algoritmo de *Max-Min*, el cual ordena la lista de tareas de mayor a menor tamaño antes de asignarlas a las máquinas virtuales, una vez ordenada dicha lista procede a asignar la tarea más grande en la primera máquina, la siguiente en la segunda y así sucesivamente hasta terminar la asignación de las tareas.

A continuación se muestran algunas capturas del código para la implementación de los primeros algoritmos en el centro de datos:

Configuración de Elementos del *Datacenter* y *DatacenterBrokers* para Algoritmos de Calendarización

```
// 5. Create a DatacenterCharacteristics object that stores the
// properties of a data center: architecture, OS, list of
// Machines, allocation policy: time- or space-shared, time zone
// and its price (G$/Pe time unit).
String arch = "x86";           // system architecture
String os = "Linux";           // operating system
String vmm = "Xen";
double time_zone = 10.0;       // time zone this resource located
double cost = 3.0;              // the cost of using processing in this resource
double costPerMem = 0.05;       // the cost of using memory in this resource
double costPerStorage = 0.1;    // the cost of using storage in this resource
double costPerBw = 0.1;         // the cost of using bw in this resource
LinkedList<Storage> storagelist = new LinkedList<Storage>(); //we are not adding SAN devices by now
```

Figura 8: Características del *datacenter*. Fuente: Elaboración propia

Como primer paso, debemos de configurar las características que tendrá nuestro *datacenter*, para eso tenemos que definir la arquitectura, el tipo de sistema operativo, costos de utilización de memoria, almacenamiento y ancho de banda, entre otros (Figura 8).

```
List<Host> hostList = new ArrayList<Host>();
// 2. Una máquina contiene uno o más PEs o CPUs/Cores.
List<Pe> peList = new ArrayList<Pe>();
int mips = 1000;
// 3. Crear los PEs y agregarlos a la lista.
peList.add(new Pe(0, new PeProvisionerSimple(mips))); // need to store Pe id and MIPS Rating
//4. Create Host with its id and list of PEs and add them to the list of machines
int hostId = 0;
int ram = 2048; //host memory (MB)
long storage = 1000000; //host storage
int bw = 10000;

//Agregamos N máquinas a nuestro data center
for(hostId = 0; hostId < numHosts; hostId++){
    hostList.add(
        new Host(
            hostId,
            new RamProvisionerSimple(ram),
            new BwProvisionerSimple(bw),
            storage,
            peList,
            new VmSchedulerTimeShared(peList)
        )
    );
}
```

Figura 9: Características del *Host*. Fuente: Elaboración propia

De igual manera tenemos que asignar las características que tendrá cada *host* dentro del *datacenter* (Figura 9), tales como el número de procesadores, número de instrucciones por segundo, memoria RAM, capacidad de almacenamiento y el ancho de banda.

```
public static List<Vm> createVm(int userId, int vms, int idShift){
    LinkedList<Vm> list = new LinkedList<Vm>();

    //parámetros de las máquinas virtuales
    long size = 10000; //MB
    int ram = 512;
    int mips = 250;
    long bw = 1000;
    int pesNumber = 1; //Num de procesadores
    String vmm = "Xen";

    //Creamos las máquinas virtuales
    Vm[] vm = new Vm[vms];

    for(int i=0; i<vms; i++){
        //Modificamos el CloudletSchedulerTimeShared por FCFS
        if(i % 2 == 0){
            vm[i] = new Vm(idShift + i, userId, mips, pesNumber * 2, ram * 2, bw, size, vmm, new CloudletSchedulerSpaceShared());
        } else {
            vm[i] = new Vm(idShift + i, userId, mips, pesNumber, ram, bw, size, vmm, new CloudletSchedulerSpaceShared());
        }
        list.add(vm[i]);
    }

    Log.println("Máquinas virtuales creadas... Success");
    return list;
}
```

Figura 10: Características de las máquinas virtuales. Fuente: Elaboración propia

Continuando con las configuraciones, en la Figura (10) se muestran las características utilizadas en las máquinas virtuales, las cuales son: tamaño de la imagen, memoria RAM virtual, número de instrucciones por segundo, ancho de banda y el número de procesadores.

```

public static List<Cloudlet> createTasks(int userId, int cloudlets, int idShift){
    LinkedList<Cloudlet> list = new LinkedList<Cloudlet>();

    Random rand = new Random();
    //parámetros de las tareas
    long length = 10000;
    long fileSize = 300;
    long outputSize = 300;
    int pesNumber = 1;
    UtilizationModel utilizationModel = new UtilizationModelFull();

    Cloudlet[] cloudlet = new Cloudlet[cloudlets];

    for(int i=0; i<cloudlets; i++){
        cloudlet[i] = new Cloudlet(idShift + i, length * getRandomInteger(10, 1), pesNumber, fileSize,
            outputSize, utilizationModel, utilizationModel);
        cloudlet[i].setUserId(userId);
        list.add(cloudlet[i]);
    }

    return list;
}

```

Figura 11: Características de las Tareas (*cloudlets*). Fuente: Elaboración propia

Además de los recursos, se debe de hacer la configuración de las tareas a simular (Figura 11), teniendo como parámetros el tamaño de la tarea, tamaño de entrada y de salida, así como el número de procesadores que ocupará la tarea, esto último refleja la complejidad de la tarea.

***DatacenterBrokers* para Algoritmos de Calendarización**

Partiendo del Esquema que se muestra en la figura (3), el trabajo presentado se lleva a cabo en la parte de Tareas (*cloudlets*) y Máquinas Virtuales (*Virtual Machines*) ya que se pretende mejorar un algoritmo de calendarización de tareas. Para ello según muestra el diagrama, el encargado de asignar cada tarea en una máquina virtual es el *datacenter broker*, el cual toma la lista de tareas así como la lista de máquinas virtuales, y de acuerdo a algún algoritmo realiza la asignación de cada una de las tareas en las distintas máquinas virtuales disponibles.

```
public class FcfsBroker extends DatacenterBroker {  
    public FcfsBroker(String name) throws Exception {  
        super(name);  
    }  
  
    //scheduling function  
    public void scheduleTaskstoVms(){  
        int reqTasks=cloudletList.size();  
        int reqVms=vmList.size();  
  
        System.out.println("\n\tFCFS Broker Schedules\n");  
        for(int i=0;i<reqTasks;i++){  
            bindCloudletToVm(i, (i%reqVms));  
            System.out.println("Task"+cloudletList.get(i).getCloudletId()+" is bound with VM"  
                +vmList.get(i%reqVms).getId() + "tam: " + cloudletList.get(i).getCloudletFileSize());  
        }  
  
        System.out.println("\n");  
    }  
}
```

Figura 12: *FCFS Broker*. Fuente: Elaboración propia

La Figura (12) muestra el algoritmo de *FCFS*, recorre la lista de tareas y la asigna en la máquina virtual correspondiente, calculando la tarea actual i en la máquina virtual $i \% reqVms$ donde $reqVms$ es el número de máquinas virtuales.


```
public class MinminBroker extends DatacenterBroker {  
    public MinminBroker(String name) throws Exception {  
        super(name);  
    }  
  
    //scheduling function  
    public void scheduleTaskstoVms(){  
        int numTareas = cloudletList.size();  
        int numVirtualM = vmList.size();  
  
        System.out.println("\n\tMinmin Broker Schedules\n");  
  
        CloudletList.sort(cloudletList);  
  
        //Asignamos las tareas a cada máquina virtual de manera que se atiendan primero las de menor tamaño  
        for(int i=0;i<numTareas;i++){  
            bindCloudletToVm(i, (i%numVirtualM));  
            System.out.println("Task"+cloudletList.get(i).getCloudletId()+" is bound with VM"  
                +vmList.get(i%numVirtualM).getId() + " tam: " + cloudletList.get(i).getCloudletLength());  
        }  
  
        System.out.println("\n");  
    }  
}
```

Figura 13: *Min-Min Broker*. Fuente: Elaboración propia

En la Figura (13) tenemos la implementación del algoritmo *Min-Min* en el cual podemos apreciar el ordenamiento de la lista antes de pasar a la asignación de las máquinas virtuales.

```
public class MaxminBroker extends DatacenterBroker {  
    public MaxminBroker(String name) throws Exception {  
        super(name);  
    }  
  
    //scheduling function  
    public void scheduleTaskstoVms(){  
        int reqTasks=cloudletList.size();  
        int reqVms=vmList.size();  
  
        System.out.println("\n\tMaxmin Broker Schedules\n");  
  
        CloudletList.revert(cloudletList);  
  
        for(int i=0;i<reqTasks;i++){  
            bindCloudletToVm(i, (i%reqVms));  
            System.out.println("Task"+cloudletList.get(i).getCloudletId()+" is bound with VM"  
                               +vmList.get(i%reqVms).getId() + "tam: " + cloudletList.get(i).getCloudletLength());  
        }  
  
        System.out.println("\n");  
    }  
}
```

Figura 14: *Max-min Broker*. Fuente: Elaboración propia

En la Figura (14) observamos la implementación del algoritmo de *Max-Min* el cual hace un ordenamiento de mayor a menor, véase la función *revert* utilizada, la cual se explicará más adelante.

```
public static <T extends Cloudlet> void sort(List<T> cloudletList) {
    Collections.sort(cloudletList, new Comparator<T>() {

        @Override
        public int compare(T a, T b) throws ClassCastException {
            Double cla = Double.valueOf(a.getCloudletTotalLength());
            Double clb = Double.valueOf(b.getCloudletTotalLength());
            return cla.compareTo(clb);
        }
    });
}

public static <T extends Cloudlet> void revert(List<T> cloudletList) {
    Collections.sort(cloudletList, new Comparator<T>() {

        @Override
        public int compare(T o1, T o2) {
            Double cla = Double.valueOf(o1.getCloudletTotalLength());
            Double clb = Double.valueOf(o2.getCloudletTotalLength());
            return clb.compareTo(cla);
        }
    });
}
```

Figura 15: *Algoritmos de ordenamientos*. Fuente: Elaboración propia

En la Figura (15) tenemos la implementación de los métodos de ordenamiento utilizados, el **sort** utilizado para ordenar las tareas de menor a mayor de acuerdo al tamaño y el **revert** de forma inversa.

Capítulo 3

Resultados

Introducción

En éste capítulo se describirán los resultados obtenidos en modelado y simulación del centro de datos en *CloudSim*. La experimentación fue desarrollada en una computadora con un procesador *Intel Core i5* con la siguiente configuración: 2.5 GHz, 3 MB de caché, 4 GB de RAM , Linux Ubuntu 14.04 lts 64 bits como sistema operativo y JDK 8.6.

3.1. Tiempo de ejecución y costo de procesamiento

Para evaluar los algoritmos de calendarización seleccionados se ha implementado un entorno de cómputo en la nube que consiste en un *datacenter*, un *broker*, máquinas virtuales y *host*. Con el objetivo de evaluar el tiempo de ejecución y el costo de procesamiento en el *datacenter* tras ejecutar cierto número de tareas (*cloudlets*).

Para la configuración del centro de datos en *CloudSim* se utilizaron diez *host*, cinco máquinas virtuales por *host* y las tareas fueron establecidas en un intervalo de 100 hasta 500 (Cuadro 1). Como se puede apreciar en el (Cuadro 2), cada *host* tuvo 2048 Mb de memoria RAM, dos núcleos de procesamiento que tienen un almacenamiento de 800 GB ó 1 TB, estas fueron elegidas de manera aleatoria y finalmente el ancho de banda fue de 1 GB/s.

| Datacenter | |
|------------|---------|
| Host | 10 |
| VM | 5 |
| Cloudlet | 100-500 |

Cuadro 1: Configuración *Datacenter*, Fuente: Elaboración propia.

| Host | |
|---------|-----------|
| RAM | 2048 MB |
| CPU | 2 |
| Storage | 800GB-1TB |
| BW | 1 GB/s |

Cuadro 2: Configuración de *Host*, Fuente: Elaboración propia.

| VirtualMachine | |
|----------------|-------------|
| RAM | 512 MB— 1GB |
| MIPS | 250 — 500 |
| Storage | 10 GB |
| BW | 1 GB/s |

Cuadro 3: *Virtual Machine*, Fuente: Elaboración propia.

| Cloudlet | |
|------------|------------|
| length | 1kb-10kb |
| MIPSlength | 300b-1.8kb |
| length | 300b-1.8kb |
| length | 1000 kb/s |

Cuadro 4: Configuración *Cloudlet*, Fuente: Elaboración propia.

En el Cuadro (3) se muestran los parámetros que se consideró para las máquinas virtuales, donde cada *VM* tendrá 10 GB de almacenamiento, 512 MB o 1 GB seleccionado de manera aleatoria, además para la característica del procesador se tiene la propiedad *MIPS* (*million instructions per second*) con 250 ó 500 y un ancho de banda de 1000 kb/s.

Para la configuración de las tareas, se tomó el tamaño con un intervalo de 1 kb a 10 kb, el parámetro *fileSize*, que representa el tamaño del archivo de entrada, va de 300 b a 1.8 kb así como el archivo de salida, como parámetro final se tiene los *MIPS* que irán de 1 a 3 (Cuadro 4).

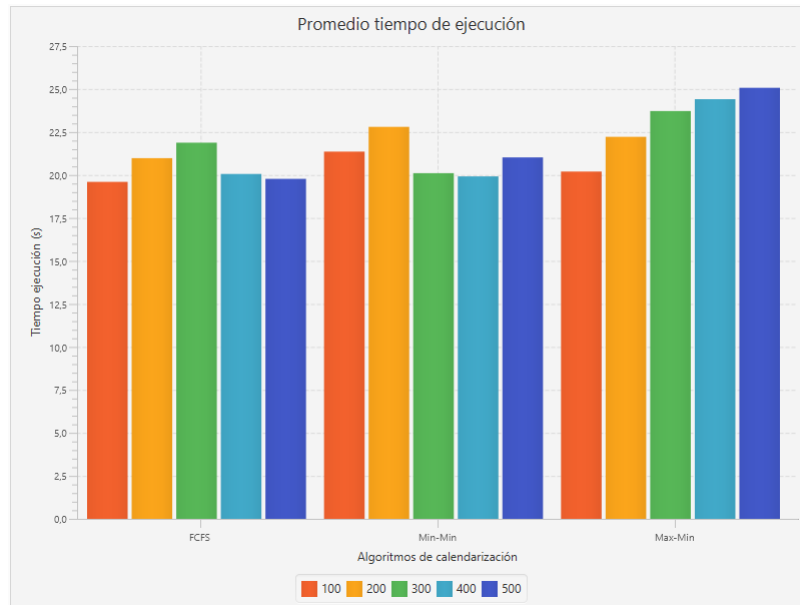


Figura 15: Promedio tiempo de ejecución con tareas 100-500, Fuente: Elaboración propia.

En la figura (15), se puede observar el promedio del tiempo de ejecución en *ms* para diferentes cantidades de tareas (de 100 a 500). A primera vista con el algoritmo *FCFS* y *Min-Min* se mantiene un tiempo de ejecución sin muchos cambios a pesar del aumento en la carga de tareas, a diferencia del algoritmo *Max-Min* que aumentó el tiempo de ejecución a medida que se incrementó el número de tareas.

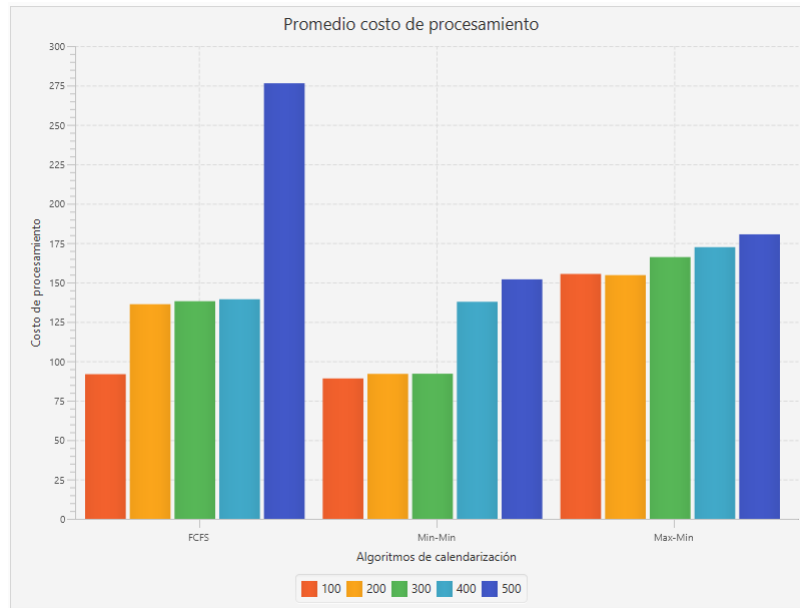


Figura 16: Promedio costo de procesamiento con tareas 100-500, Fuente: Elaboración propia.

El costo de procesamiento, de acuerdo a cada algoritmo, se puede observar en la figura (16), en donde el algoritmo *FCFS* tiene un incremento drástico al realizar la prueba con 500 tareas. El algoritmo *Max-Min* se conservó sin muchos cambios apesar de la cantidad de tareas, mientras que *Min-Min* tiene un menor costo de procesamiento cuando las tareas son inferiores a 300.

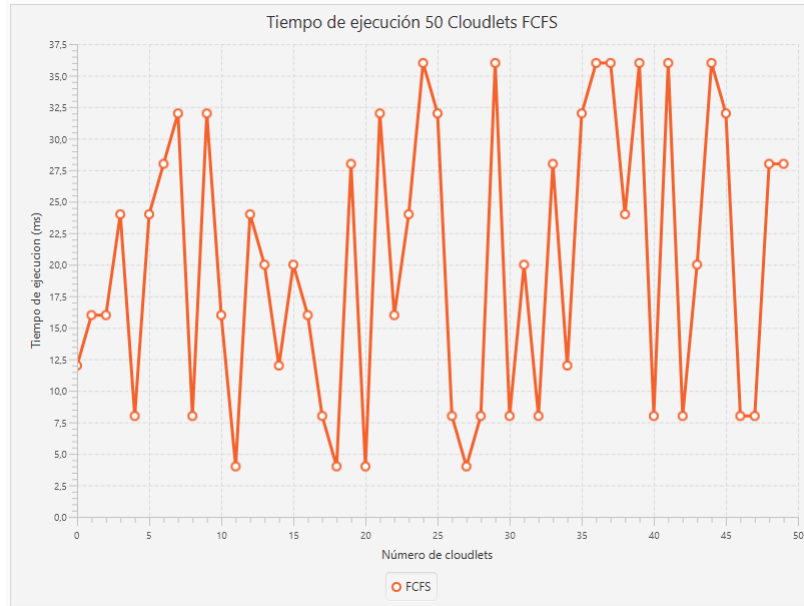


Figura 17: Tiempo ejecución 50 muestras *FCFS*, Fuente: Elaboración propia

Para mostrar el comportamiento del tiempo de ejecución por cada algoritmo, se tomaron 50 muestras de una simulación de 500 tareas. En la figura (17), se puede apreciar que el algoritmo *FCFS* tiene un comportamiento inestable ya que algunas tareas pueden tener menor complejidad o tamaño, lo que implica una respuesta rápida.

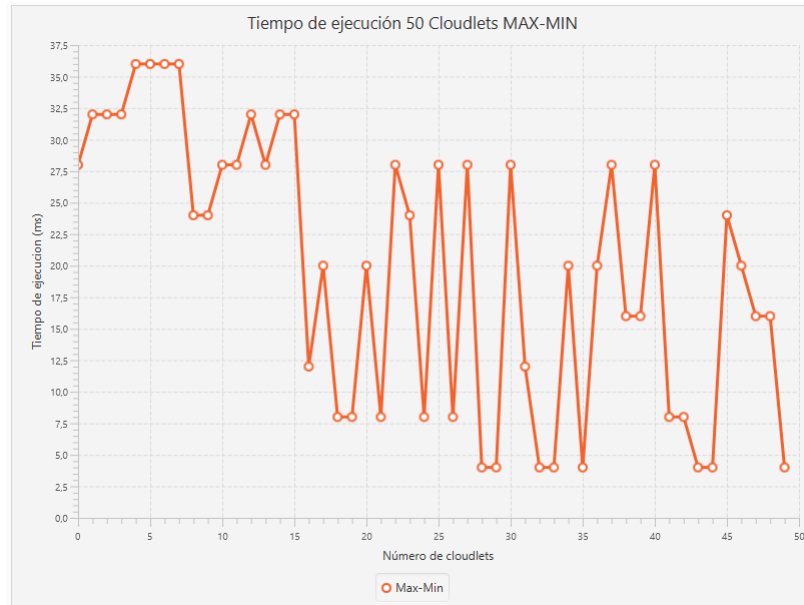


Figura 18: Tiempo ejecución 50 muestras *Max-Min*, Fuente: Elaboración propia

En la figura (18) se tiene la misma simulación pero con el algoritmo *Max-Min*, de acuerdo a las características de este calendarizador, en las primeras tareas se toma un mayor tiempo en responder y va disminuyendo de manera gradual, sin embargo aún es inestable en las últimas muestras ya que no se contempla el grado de complejidad, es decir el parámetro *MIPS* de los *cloudlets*.

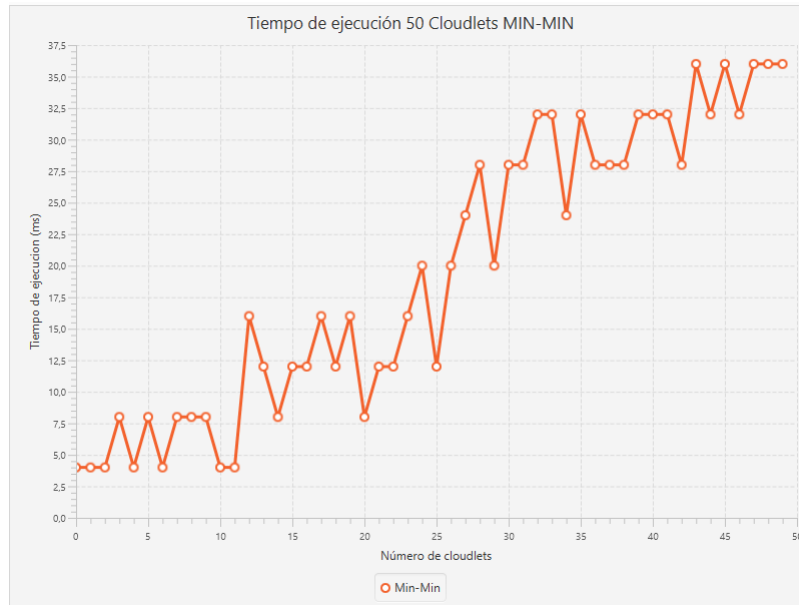


Figura 19: Tiempo ejecución 50 muestras *Min-Min*, Fuente: Elaboración propia

Por último se puede apreciar el algoritmo *Min-Min* en el que el tiempo de ejecución fue aumentando conforme se resolvían las tareas (figura 19).

| Algoritmo | Desviación estándar |
|-----------|---------------------|
| FCFS | 11.00619 |
| MAX-MIN | 8.91444 |
| MIN-MIN | 11.25613 |

Cuadro 5: Desviación estándar del tiempo de ejecución, Fuente: Elaboración propia

Observando la desviación estándar de éstas muestras anteriores, los algoritmo *Min-Min* y *FCFS* tuvieron más variaciones en las muestras con respecto a la media, mientras que el *Max-Min* tuvo las variaciones por debajo de las dos anteriores (Cuadro 5).

Conclusiones

Bibliografía

- Agarwal, S., Jain, S., et al. (2014). Efficient optimal algorithm of task scheduling in cloud computing environment. *arXiv preprint arXiv:1404.2076*. Recuperado 2015-08-21.
- Aranganathan, S. and Mehata, K. (2011). An aco algorithm for scheduling data intensive application with various qos requirements. *Int. J. Comput. Appl*, 27(10). Recuperado 2015-08-20.
- Calheiros, R. N. (2011). Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms.
- Chen, H., Wang, F., Helian, N., and Akanmu, G. (2013). User-priority guided min-min scheduling algorithm for load balancing in cloud computing. In *Parallel Computing Technologies (PARCOMPTECH), 2013 National Conference on*, pages 1–8. IEEE. Recuperado 2015-08-23.
- Chen, Q. and Deng, Q. (2009). Cloud computing and its key techniques. *Journal of Computer Applications*. Recuperado 2015-08-20.
- Chowdhury, A., Ahmed, M., Ahmed, M., and Rafee, M. M. H. (2012). An advanced survey on cloud computing and state-of-the-art research issues. *IJCSI International Journal of Computer Science Issues*, 9(1):1694–0814. Recuperado 2015-08-20.
- Dakshayini, M. and Guruprasad, H. (2011). An optimal model for priority based service scheduling policy for cloud computing environment. *International Journal of Computer Applications*, 32(9):23–29. Recuperado 2015-08-23.

- Delavar, A. G., Javanmard, M., Shabestari, M. B., and Talebi, M. K. (2012). Rsdsc (reliable scheduling distributed in cloud computing). *International Journal of Computer Science, Engineering and Applications (IJCSEA)* Vol, 2. Recuperado 2015-08-22.
- Dover, Z., Gordon, S., and Hildred, T. (2013). The technical architecture of red hat enterprise virtualization environments. Recuperado 2015-08-22.
- EPI (2015-08-31). What is a data centre. http://www.epi-ap.com/content/28/48/What_is_a_Data_Centre.
- Li, J., Feng, L., and Fang, S. (2014). An greedy-based job scheduling algorithm in cloud computing. *Journal of Software*, 9(4):921–925. Recuperado 2015-08-21.
- Mariscal, J. and Gil-Garcia, J. (2013). El cómputo en la nube en méxico: Alcances y desafíos para los sectores público y privado. cide. Recuperado 2015-08-21.
- Nishant, S. and R., K. (2015). Pre-emptable shortest job next scheduling in private cloud computing” in journal of information, knowledge and research computer engineering. In *Knowledge and research computer engineering*. Recuperado 2015-08-20.
- Parsa, S. and Entezari-Maleki, R. (2009). Rasa-a new grid task scheduling algorithm. *JDCTA*, 3(4):91–99. Recuperado 2015-08-22.
- S., S. and Horacio, R. (2002). *Sistemas de información en la era digital*. Fundación OSDE. Recuperado 2015-08-22.
- Salot, P. (2013). A survey of various scheduling algorithm in cloud computing environment. *IJRET: International Journal of Research in Engineering and Technology, ISSN*. Recuperado 2015-08-21.
- Shimpy, E. and Sidhu, M. J. (2014). Different scheduling algorithms in different cloud environment. *algorithms*, 3(9). Recuperado 2015-08-22.
- Srinivasan, S. (2014). Cloud computing evolution. In *Cloud Computing Basics*, pages 1–16. Springer. Recuperado 2015-08-31.

Glosario

A

- **Amazon:** Compañía de comercio electrónico y servicios de computación en la nube.

C

- **Cloud Computing:** cómputo en la nube, tiene como objetivo ofrecer servicios a través de Internet.
- **Cloudlet:** Es una representación de tareas en Cloudsim.
- **CloudSim:** Framework para el modelado y simulación de la infraestructura y los servicios de Cloud Computing.
- **Cluster:** Conjunto de computadoras unidas entre sí normalmente por una red de alta velocidad y que se comportan como si fuese una única computadora.

E

- **ERP:** Planificador de recursos empresariales, son sistemas informáticos destinados a la administración de recursos de una organización.

F

- **Framework:** Marco de trabajo, es un conjunto de clases que implementa todos los servicios comunes de un cierto tipo de aplicación.

P

- **Pruebas de software/hardware:** técnicas cuyo objetivo es evaluar la calidad del software/hardware.

Anexos