

Application of a Cluster of Physical and Virtual Machines for the reduction of segmentation time in the Classification of Satellite Images using Parallel Computing and Neural Networks

Yoni D. Huaynacho, Abel S. Huaynacho, Pablo R. Yanyachi
IAAPP – Universidad Nacional de San Agustín de Arequipa - Perú,
yhuaynacho@unsa.edu.pe, ahuaynachop@unsa.edu.pe, raulpab@unsa.edu.pe

Abstract— *There are different projects, theses, articles and others that need a cluster of computers for testing with a large number of machines. The students, investigators and others don't have many machines to study in the big data field. In addition, if you get machines, you usually have different operating systems, processor, hard disk and others. A very important issue is to reduce the time to classify satellite images that have very high data sizes and when they are processed, the computer becomes slow. For all this, the article uses virtual machines as a solution; these are used to test the reduction of response times in the classification of satellite images using also a classification tool with neural networks. As a result of this article, we have tested qualifying 1000000 lines of data in sequential form and as distributed form using virtual machines to compare response times.*

Keywords- *Virtual Machine, VMware, Neural Network Cluster, Images Satellites, Message Passing Interface (PMI)*

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2018.1.1.514>
ISBN: 978-0-9993443-1-6
ISSN: 2414-6390

Aplicación de un Clúster de Máquinas Físicas y Virtuales para la Reducción de Tiempo de Segmentación en la Clasificación de Imágenes Satelitales Utilizando Computación Paralela y Redes Neuronales

Yoni D. Huaynacho, Abel S. Huaynacho, Pablo R. Yanyachi
IAAPP – Universidad Nacional de San Agustín de Arequipa - Perú,
yhuaynacho@unsa.edu.pe, ahuaynachop@unsa.edu.pe, raulpab@unsa.edu.pe.com

Resumen – Existen diferentes proyectos, tesis, artículos en el campo de Big-Data que requieren un Clúster de computadoras de un número considerable de máquinas para realizar pruebas. Debido a que no hay muchas máquinas físicas, hay gastos para obtenerlas, una solución lógica sería usar máquinas en la nube, sin embargo, las solicitudes de permisos para usarlas ocasionan pérdidas de tiempo y dinero. Además, al adquirir el servicio, su configuración es complicada y puede dificultar la realización de pruebas en estas computadoras de alto rendimiento. Por esta razón, en el presente artículo, las máquinas físicas y virtuales se utilizan en realizar pruebas para reducir los tiempos de respuesta en la clasificación de imágenes de satélite utilizando computación paralela y redes neuronales. Como resultado de este artículo se ha reducido en un 64.6% el tiempo de clasificación secuencial de imágenes multispectrales de 4 bandas para 1'000,000 de datos a nivel digital (pixel), utilizando el clúster diseñado con 4 máquinas virtuales y 2 físicas.

Palabras Clave-- Big Data, Máquina Virtual, VMware, Red Neuronal, Clúster, computación paralela, Sistema de Información Geográfica (SIG).

I. INTRODUCCIÓN

En la actualidad un investigador no cuenta siempre con muchas máquinas para trabajar con grandes cantidades de datos, y en el campo de Big Data, Teledetección, Modelado en 3D y otros, se necesita trabajar con equipos que cumplan los requisitos requeridos. Sin ellas, no se podría realizar las investigaciones propuestas y se necesita utilizar Clústeres por la demanda de procesamiento, los cuales incluyen una gran cantidad de máquinas.

Sin embargo, para manejar estos equipos se aconseja haber entrenado con algún clúster creado por uno mismo y haber realizado pruebas a menor escala, se debe considerar también el tamaño de espacio requerido y el tiempo de procesamiento para obtener resultados. Ya que, dejar una máquina procesando varios días y luego darse cuenta de que se quedó a mitad del proceso, son hechos que debemos evitar. Como podemos ver, existen varios problemas cuando se trabaja con altas cantidades de datos o se utilizan clústeres.

Actualmente podemos aprovechar el uso de las máquinas que existen, como las computadoras Core i7, i5, i3 que tienen varios núcleos. Podemos utilizar la diversidad de estos equipos que están a nuestro alcance, incluso aquellas con procesadores antiguos, para lo cual debemos tener en cuenta el sistema operativo que tengan instalados como Windows XP, Windows 7, Windows 10, Centos 7, Ubuntu 17, etc.

Como demostración, se ha realizado el procesamiento de datos satelitales para obtener la reducción de tiempos de segmentación en su clasificación, donde se muestra la aplicación de un clúster, que utiliza un conjunto de máquinas físicas disponibles y aprovecha la creación de máquinas virtuales para incrementar el número de máquinas, consiguiendo usar redes neuronales y programación paralela implementadas en el software CIS [3].

En la sección 2 se muestra el estado del arte, en la sección 3 se encuentran los conceptos importantes utilizados, en la sección 4 la propuesta del clúster, en la sección 5 se detallan las pruebas realizadas, en la sección 6 se presentan las conclusiones y por último en la sección 7 se presenta la propuesta de trabajos futuros.

II. ESTADO DEL ARTE

Para operaciones de grandes cantidades de datos se utiliza un Clúster de máquinas, que soporta datos de altos tamaños como los Petabytes (equivalente a 1024 Terabytes) [2]. En el campo de la teledetección se realiza el procesamiento de Imágenes Multiespectrales e Hiperespectrales con el fin de obtener información de la imagen satelital. Se cuenta con múltiples Sistemas de Información Geográfica (SIG) comerciales como el ENVI, ARGIS, ERDAS, etc., y SIG libres como el Spring, QGIS, gvSIG, etc., los cuales usan bases de datos relacionales como repositorios de datos, por otro lado, existen otros que usan la computación paralela como el Análisis Multimedia y Sistema de Recuperación (IMARS) [36], que se creó para realizar la clasificación de imágenes satelitales con Máquinas de Vectores de Soporte (SVM). La diferencia de estos SIG está en el tiempo de demora para

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2018.1.1.514>
ISBN: 978-0-9993443-1-6
ISSN: 2414-6390

realizar las operaciones, siendo el sistema SIG en paralelo, altamente escalable y más rápido.

Para la implementación de SIG paralelos se necesitan utilizar múltiples máquinas para comprobar o hacer las pruebas de funcionamiento, se cuenta con máquinas de alto desempeño por ejemplo el del IMARS que está configurado con 224 CPUs, 448 GB de RAM, y 14 TB de almacenamiento HDFS (Hadoop Distributed File System) [6] y [9].

Se ha implementado un algoritmo Backpropagation (BP) de redes neuronales para el procesamiento de imágenes satelitales, donde se demuestra que su aplicación en un grupo de máquinas reales reduce los tiempos de respuesta [3]. Es posible optimizar las clasificaciones de datos, para lo cual se han implementado diferentes técnicas, una de ellas, es la fusión de estas técnicas, como la generación automática donde se usa la técnica de Redes Neuronales que ajusta sus parámetros con la técnica de Algoritmos Genéticos [4].

Existen diferentes administradores de bases de datos, con SQL (Structured Query Language) y No SQL donde se utilizan bases de datos no SQL para manejar grandes cantidades de datos [5].

Los algoritmos de fusión de imágenes requieren mucho tiempo para su procesamiento, para su uso, utilizan la computación paralela [7].

Para programar en paralelo, está disponible MPI (Message Pass Interface), esta programación permite que el procesamiento sea más rápido pero las máquinas deben tener las mismas características [8]; por otro lado, tenemos el modelo de programación MapReduce que se usa en plataforma Hadoop y se puede utilizar con máquinas de diferentes características [6].

En el año 2002, se realizó la instalación, configuración y pruebas de una supercomputadora basado en un Clúster en el laboratorio de Cibernética de la Universidad Nacional de Entre Ríos de Argentina (FI-UNER), donde se utiliza el clúster para realizar experimentos numéricos [10].

III. CONCEPTOS BÁSICOS

A. Máquina Virtual

Es un Software que simula a una computadora real y se puede instalar y usar sobre otro Sistema Operativo de forma simultánea [11].

B. Clúster

Conjunto de computadoras conectadas a través de una red, que trabaja un gran problema de cómputo que ha sido dividido en varios sub-problemas pequeños [10].

C. Algoritmo Back Propagation (BP)

Algoritmo de redes neuronales, que ha sido implementado en bases de datos para afrontar problemas como la clasificación en el procesamiento de imágenes [12], [13]. El aprendizaje se logra con el ajuste de los nodos entre el nodo de salida y el resultado, ajustando los pesos recursivamente con la propagación del error a través de la red.

D. Imágenes de Satélites

Para obtener una imagen satelital, se utilizan cámaras ubicadas en los satélites, estas imágenes se diferencian por la cantidad de bandas espectrales, tamaño del píxel o resolución espacial, resolución espectral, resolución temporal y resolución radiométrica [14]. Por ejemplo, las imágenes del satélite Perusat1 llegan a tener un tamaño de pixel de 0.7 metros o una resolución espacial de 0.7m.

Para que se genere una imagen se debe convertir la energía electromagnética, en datos digitales mediante la conversión analógica a digital. Una imagen es la representación digital de un objeto, y las representaciones digitales son datos numéricos [15].

La imagen de satélite resultante es un conjunto de matrices, la unidad básica es una celda llamada píxel, que contiene las propiedades físicas y químicas de la superficie de observación reflejada, dependiente de la resolución espacial, espectral y radiométrica del sensor de observación.

Entre los tipos de imágenes satelitales por el número de bandas [16] y [17], tenemos:

- Imágenes Pancromáticas: Son las imágenes de blanco y negro o escala de grises estas tienen mejor resolución espacial.
- Imágenes Multiespectrales: Estas tienen varias bandas, como las imágenes Landsat-8, que tiene 11 bandas.
- Imágenes Hiperespectrales: Tiene cientos de bandas.

En la Figura 1, se muestra una imagen satelital pancromática del satélite LandSat 8.

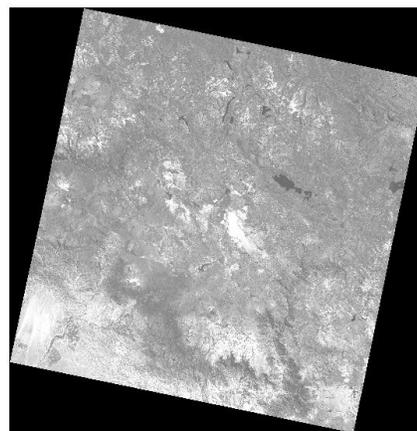


Figura 2. Imagen de Satélite LandSat 8 de Arequipa-Perú.

1) *Clasificación de Imágenes Satelitales*: Entre los métodos para clasificar una imagen satelital se encuentra el convencional que es de píxel por píxel y el de segmentación, píxeles homogéneos. El primero depende principalmente por la respuesta espectral de píxeles individuales y el segundo por la resolución espacial y espectral [18]. Según el conocimiento del área de estudio [19], se pueden diferenciar como:

- *Clasificadores no supervisados*: No requieren entrada de información para clasificar.
- *Clasificadores supervisados*: Requieren entradas de información antes de clasificar, llamado conjunto de entrenamiento, para identificar y diferenciar las clases.

2) *Formatos de Imágenes Satelitales*: La capacidad de los sensores en el desarrollo de los satélites, representa generalmente la información adquirida de las imágenes en dos formatos: vectoriales y ráster. Se diferencian en la dependencia de la resolución, entre ellos tenemos: Arc Digitized Raster Graphics (ADRG), Band Interleaved by Line (BIL), Band Interleaved by Pixel (BIP), Band Sequential (BSQ), Digital Elevation Model (DEM), PC Paintbrush Exchange (PCX), Spatial Data Transfer Standard (SDTS) y Tagged Image File Format (TIFF) [17].

E. Macrodatos

Se llama también Big data, se refiere a grandes cantidades de datos o cantidades masivas de datos [17].

Las características son: Volumen, Velocidad, Variedad, Valor, Complejidad [17], [20], [21], [22].

1) *Tipo de Macrodatos*: Los tipos de Macrodatos se clasifican según las fuentes de Datos, formato de contenido, almacenamiento de datos, etapas de los datos y procesamiento de datos [17] [23].

2) *Gestor de Base de Datos*: Un gestor de base de datos es un software creado para administrar los datos, tanto para la escritura, lectura, actualización, o eliminación de datos. Para administrar grandes cantidades de datos se utilizan bases de datos NoSQL que son sistemas de gestión de bases de datos y no usan SQL como lenguaje principal [17], [24]. La clasificación de Gestores de bases de datos NoSQL son: las Bases de datos Clave-Valor, las Orientadas a Columnas [25], [26], [27], [28], las de tipo Orientado a Documentos [29], y las Orientadas a Grafos [30].

F. Segmentación Satelital

Es la clasificación orientada a objetos, grupo de píxel, a diferencia del procedimiento tradicional de píxel por píxel para extraer información supervisada o no supervisada de una imagen satelital [31]. Esta metodología muestra resultados más exactos para la clasificación de imágenes de alta resolución espacial que es el área cubierta por un píxel [32], [33], se basa en diferentes parámetros de homogeneidad para la obtención de información significativa, representada en la calidad del objeto [34], que son criterios de índices espectrales, forma,

textura y relación contextual, que constituyen la eficacia de la clasificación primaria.

La aplicación de segmentación en múltiples resoluciones a permitido desarrollar niveles de clasificación a diferentes escalas, generando y relacionando objetos con diferentes parámetros de homogeneidad, estableciendo propiedades de herencia entre las clases a través de una red jerárquica de objetos, lo que ha permitido extraer diferentes patrones y comportamientos de acuerdo con la escala [35].

IV. PROPUESTA

Consta de los siguientes recursos que corresponden a máquinas virtuales creadas en máquinas reales diferentes.

A. Máquinas Físicas

La Tabla I, muestra las características de los equipos que conforman el clúster.

TABLA I
ESPECIFICACIÓN DE EQUIPOS

Nro. Equip	Sist. Operativo	Memoria RAM (GB)	Disco Duro (GB)	Procesador (GHZ)	Máquina Virtual (Cant.)
1	XP Profesional	3.25	320	2.21 ^a	2
2	7 Ultimate	2.00	296	2.30 ^b	1
3	XP Profesional	2.00	80	1.53 ^c	-
4	10	4.00	195	3.50 ^d	1

^aAMD Sempron LE, ^bPentium(R) Dual Core, ^cAMD Athlon XP, ^dCore i3.

La red del diseño del Clúster se muestra en la Figura 2 donde se puede identificar a la máquina que será el máster, las que son virtuales y las físicas.

B. Máquinas Virtuales

La Máquina Virtual fue creada con el software VMware y está formado por las siguientes características:

TABLA II
CONFIGURACIÓN DE MÁQUINA VIRTUAL

Cant	Sist. Operativo	Memoria RAM (GB)	Disco Duro (GB)	Conexion de Red
.	Windows	(GB)	(GB)	(Ethernet)
1	XP Profesional	3.25	20	Bridge Automatico

Esta máquina virtual es la que se replica en las diferentes máquinas físicas, que deben de tener instalado el Hipervisor VMware para ser reconocidas y ocupa un espacio de 2.06GB.

C. Diseño del Clúster

Como se puede ver en la Figura 2, se muestra la arquitectura del clúster, que está formado por un Máster que controla 4 máquinas virtuales y una máquina real, el modelo de máquina virtual se puede replicar para distintos sistemas operativos. Las máquinas deben estar en un grupo de red y

cada uno debe identificarse con su nombre de equipo. Los puertos asignados para cada máquina van desde el puerto 8081 al 8085 para cada máquina controlada.

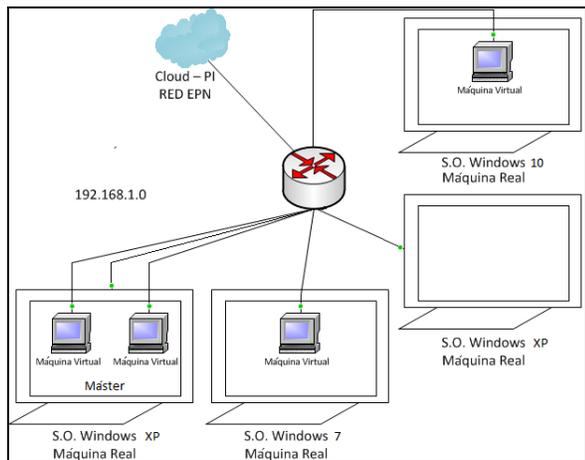


Figura 2. Arquitectura del Clúster.

D. Funcionamiento del Clúster

Para el funcionamiento se codificó en MPI que considera el nombre de equipo y sus puertos de enlace para identificar las máquinas que serán seleccionadas por el software de clasificación. Cuando se corre el software CIS (Clasificación de Imágenes Satelitales) [3], primero se entrena la red neuronal con los datos de imágenes ya conocidos, luego se puede realizar la clasificación en forma secuencial y paralela.

1) *Entrenamiento de la Red Neuronal:* Se utiliza el algoritmo BP que relaciona las entradas de los píxeles con las salidas de 6 tipos reconocidos.

TABLA III
PARÁMETROS DE LA RED NEURONAL [3].

Parámetros	Valor	Interpretación
Capa sensorial	4	Pixel (ND ^c)
Capa de procesamiento	15	Aprendizaje
Capa de salida	6	Clasificación

^cNivel Digital

2) *Clasificación Secuencial:* Esta clasificación se realiza usando redes neuronales, pero solo con la máquina real Máster, Figura 4.

3) *Clasificación Paralela:* Se realiza en las máquinas físicas o virtuales controladas por la máquina Master, primero se divide el proceso según la cantidad de máquinas a utilizar, luego se envía los datos particionados y por último se ejecuta la clasificación en cada máquina. El resultado del tiempo de demora de la clasificación se muestra en porcentajes en la máquina Master, Figuras 5, 6 y 7.

La clasificación consiste en encontrar el tipo de Clase para los datos de clasificación (Para una imagen Multiespectral de 4 bandas, el dato de clasificación está formado por 4 píxeles).

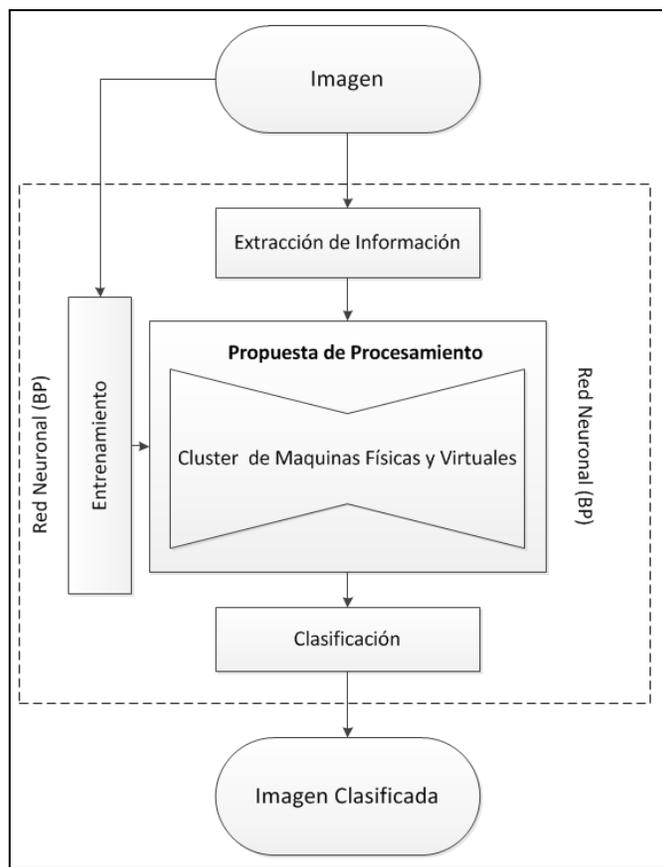


Figura 3. Diagrama de Bloques de funcionamiento del Clúster.

V. PRUEBAS

Las pruebas fueron realizadas usando el software CIS y datos de imágenes satelitales en formato “.csv”. Las pruebas de entrenamiento con 4435 casos, 2000 entradas para la validación en formato “.csv” de fuente del software Weka, y 1'000,000 de líneas de datos para clasificar [3]. Obteniendo los siguientes resultados mostrados en la Tabla 4, donde se aprecia la reducción de tiempos de procesamiento en porcentajes, considerando como referente la clasificación secuencial.

TABLA IV
RESULTADOS

Clasificación	Cantidad de Maquinas		Valor (min)	Reducción de tiempo (%)
	Físicas	Virtuales		
Secuencial	1 ^d	-	3.56	-
Paralela	-	1	4.20	-18.0
Paralela	-	3	2.18	38.8
Paralela	1	4	1.26	64.6

^dMaster

En la Figura 4 se muestra la interfaz para realizar la clasificación secuencial y paralela, donde se señala el tiempo de demora de la clasificación secuencial.

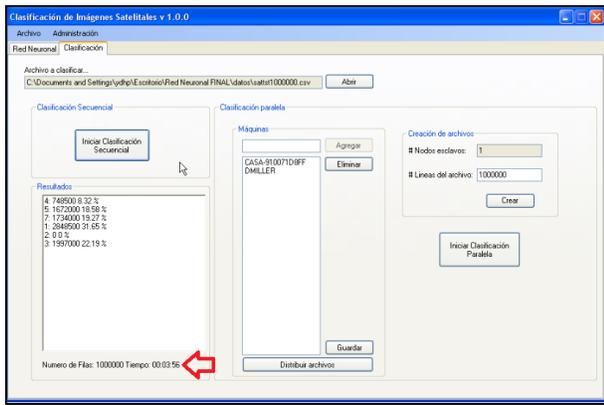


Figura 4. Clasificación Secuencial.

En la Figura 5, 6, y 7 se detalla los resultados de la clasificación paralela que corresponden a 1, 3 y 5 máquinas del clúster y se señala el tiempo de demora del procesamiento.

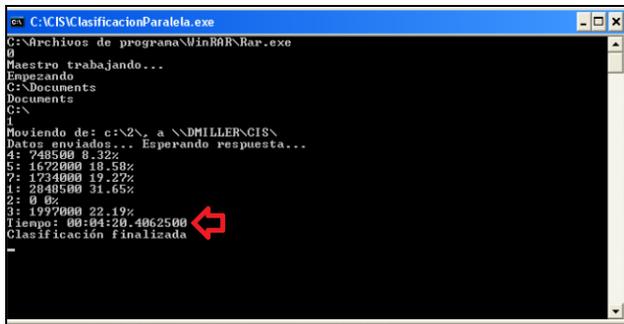


Figura 5. Clasificación paralela con 1 máquina.

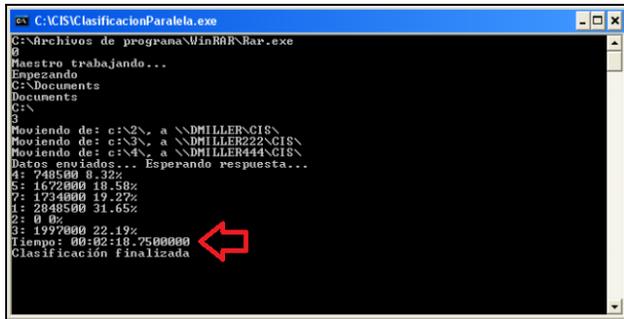


Figura 6. Clasificación paralela con 3 máquinas.

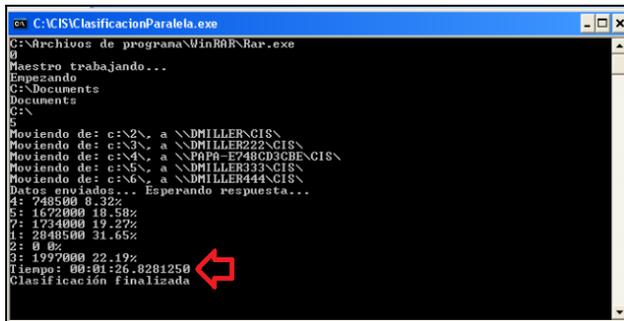


Figura 7. Clasificación paralela con 5 máquinas.

En la Figura 8, se muestra la comparación de tiempos de las pruebas realizadas. Se ha obtenido un tiempo de procesamiento de 1.26 segundos con 5 máquinas: 4 virtuales y 1 física. También se muestra los tiempos de la clasificación secuencial y paralela usando 1'000,000 de datos de clasificación.

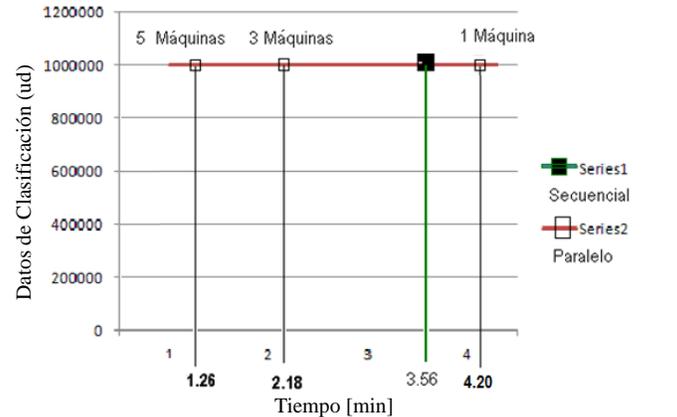


Figura 8. Detalle del procesamiento del clúster.

En la figura 9 se muestra la reducción de tiempos de clasificación en forma paralela, como resultado de las pruebas en el clúster y la cantidad de datos de clasificación, aumentando la cantidad de máquinas físicas o virtuales.

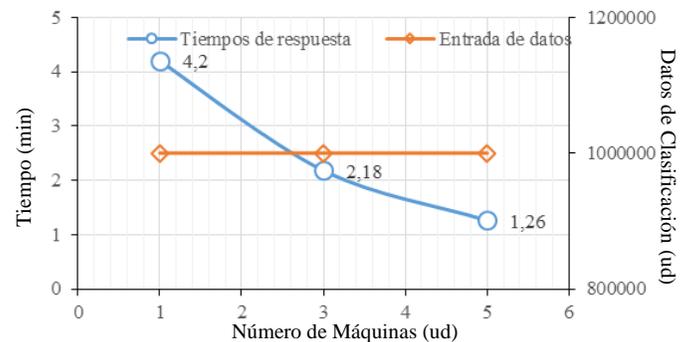


Figura 9. Tiempos de Clasificación según el Nro. de Máquinas del Clúster.

VI. CONCLUSIONES

La computación paralela es necesaria para el manejo de grandes volúmenes de información como la clasificación de Imágenes Satelitales.

Con Máquinas Virtuales y Reales se puede crear un Clúster para realizar pruebas de programación paralela.

El clúster distribuido de máquinas virtuales se hace sencillo al solo configurar una máquina virtual.

Se puede usar el clúster de máquinas virtuales en diferentes sistemas operativos con diferente hardware.

La clasificación en paralelo con un Clúster mejora los tiempos de respuesta comparados con la clasificación secuencial.

A medida que se utiliza una mayor cantidad de máquinas en una clasificación en paralelo, mayor es la rapidez del procesamiento de datos.

Los tiempos de clasificación con 5 máquinas en paralelo usando el modelo de Clúster explicado, mejora en un 64.6% los tiempos de la clasificación en forma secuencial para 1'000,000 de datos de clasificación.

Se puede aprovechar la capacidad de los equipos de cómputo para crear máquinas virtuales cuyo número dependerá de su tipo de procesador.

Se pueden utilizar equipos con procesadores antiguos o de última generación para crear un Clúster.

Si se quiere abarcar la mayor cantidad de Equipos que estén al alcance, dependerá del sistema Operativo que sea más compatible con los equipos.

VII. TRABAJOS FUTUROS

Se propone la implementación de esta investigación en un SIG escalable para obtener la reducción de tiempos de procesamiento. Otro planteamiento, sería utilizar algoritmos genéticos para la mejora de tiempos de respuesta e implementar una herramienta híbrida de redes neuronales con algoritmos genéticos. Se puede probar la clasificación de imágenes en otros sistemas operativos.

AGRADECIMIENTOS

Agradecemos a la Universidad Nacional de San Agustín de Arequipa y al Instituto Astronómico Aeroespacial Pedro Paulet de la UNSA.

REFERENCIAS

- [1] J.-F. Mas, M. Graff, J. Flores, M. Arreola and F. Calderon, "Evolving artificial neural networks for binary and multi-class classification problems", 2011.
- [2] J.-F. Castañeda, "Identificación de factores tecnológicos para la aplicación de big data", Universidad Tecnológica de Pereira, Colombia, 2016.
- [3] R.C. Cueva, "Análisis de la implementación del algoritmo de Backpropagation aplicado al procesamiento de imágenes satelitales sobre un entorno distribuido", Universidad Técnica Particular de Loja, Ecuador, 2009.
- [4] A. Fizelew and R. Garcia-Martinez, "Generación automática de redes neuronales con ajuste de parámetros basado en algoritmos genéticos", CA: Universidad de Buenos Aires, Argentina, 2002.
- [5] J. Han, E. Haihong, G. Le, and J. Du, "Survey on NoSQL database", Proceedings - 2011 6th International Conference on Pervasive Computing and Applications, ICPCA 2011, pp. 363-366, 2011.
- [6] C. Gomez-Martinez, "Procesamiento de grandes volúmenes de datos en entornos cloud computing utilizando hadoop mapreduce", Universidad de Almería, España, mayo 2013.
- [7] Zhang Jixian and Yang Jinghui, "A parallel implementation framework for remotely sensed image fusion", in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. I-7, 2012.

- [8] D. Buntinas P. Balaji B. Bar-rett R. Brightwell W. Gropp V. Kale T. Hoefler, J. Dinan and R. Thakur, "Mpi+ mpi: A new hybrid approach to parallel programming with mpi plus shared memory", 2013.
- [9] C. Sweeney and S. Arietta, "HIPI: A Hadoop Image Processing Interface for Image-based MapReduce Tasks. Image Processing", page 5, 2010.
- [10] A. Azar, L. Rufiner and R. Milone, "Supercomputadoras basadas en "clusters" de pcs", 2002.
- [11] A. Carrillo y K. Gonzalez, "Máquinas virtuales", 2016.
- [12] P. Sathya and L. Malathi, "Classification and segmentation in satellite imagery using Back Propagation algorithm of ANN and K-means algorithm", in International Journal of Machine Learning and Computing, vol. 1, no. 4, October 2011.
- [13] Shifali and G. Kaur, "Satellite image classification using back propagation neural network", in Indian Journal of Science and Technology, vol 9(45), December 2016.
- [14] J. Richards, "Remote sensing digital image analysis an Introduction", vol. 5, Springer, 2013.
- [15] S. Oprisescu and M. Dumitrescu, "On the regularization of segmented satellite images", In Signals, Circuits and Systems, 2005. ISSCS 2005, International Symposium on, vol. 1, pp. 83-86. IEEE, 2005.
- [16] Senores Remotos, et al, "Guía básica sobre imágenes satelitales y sus productos. Nuevas tecnologías en la gestión de espacios naturales", 2013.
- [17] J. Serrano, J. Cuba, Y. Huaynacho, A. Mamani, Y. Yari, W. Nina, R. Cruz and P. Yanyachi, "A new approach to the massive processing of satellite images", 2015.
- [18] S. Perelman, A. Cerezo y M. Conde, "Efecto de diferentes métodos de clasificación de imágenes satelitales sobre índices de paisaje", 2009.
- [19] L. Bruzzone, G. Camps-Valls, J. Benediktsson and J. Chanussot, "Introduction to the issue on advances in remote sensing image processing", in IEEE Journal of selected topics in signal processing, 5(3):365-369, 2011.
- [20] X. Zhang and F. Xu. "Survey of research on big data storage. In Distributed Computing and Applications to Business", in Engineering Science (DCABES), 12th International Symposium on, pp. 76-80. IEEE, 2013.
- [21] A. McAfee and E. Brynjolfsson, "Big data: the management revolution" in Harvard business review, (90):60-6, 2012.
- [22] J. Singh, "Big data analytic and mining with machine learning algorithm", in International Journal of Information and Computation Technology, vol. 4, no. 1, pp. 33-40, 2014.
- [23] I. Abaker, T. Hashem, I. Yaqoob, N. Badrul, S. Mokhtar, A. Gani and S.-U. Khan, "The rise of big data in cloud computing: Review and open research issues", in Information Systems, Elsevier, vol. 47, pp. 98-115, January 2015.
- [24] ABM Moniruzzaman and S.-A. Hossain, "Nosql database: New era of databases for big data analytics-classification, characteristics and comparison", in International Journal of Database Theory and Application, vol. 6, no. 4, pp. 1-14, June 2013.
- [25] P. Chaganti and R. Helms, "Amazon SimpleDB Developer Guide", Packt Publishing Ltd, 2010.
- [26] CDataStax. Apache Cassandra™ 2.0," Documentation", 2015.
- [27] Apache HBase Team HBase, "Apache hbase reference guide", 2016.
- [28] F. Chang, J. Dean, S. Ghemawat, W.-C. Hsieh, D.-A. Wallach, M. Burrows, T. Chandra, A. Fikes and R.-E Gruber, "Bigtable: A distributed storage system for structured data", in ACM Transactions on Computer Systems (TOCS), vol. 26(2), no. 4, June 2008.
- [29] The Apache Software Foundation. CouchDB, "Manual". <https://media.readthedocs.org/pdf/couchdb/latest/couchdb.pdf>, 2015.
- [30] MongoDB. for Ops Manager, "Manual", <https://docs.opsmanager.mongodb.com/current/opsmanager-manual.pdf>, 2015.
- [31] C. Delrieux, C. Pacheco, H. Del Valle and G. Bianchi, "Procesamiento inteligente de imágenes satelitales: Métodos avanzados y acelerados de segmentación y fusión de información", in XVI Workshop de Investigadores en Ciencias de la Computación, At Ushuaia, Tierra del Fuego, Volume: WICC, May 2014.

- [32] L. Reyna and B. Gossweiler, "Segmentacion y clasificación de imágenes satelitales para determinar la cobertura del suelo", Revista de Agrociencias, Universidad Técnica de Manabí, E-ISSN 2477-8982, 2014
- [33] O. Espejo and J. Ardila, "Evaluation of an object-oriented image classification methodology in satellite images", Universidad Distrital Francisco Jose de Caldas, ISSN 0121-750X E-ISSN 2344-8393, 2004.
- [34] M. Bock and R. Lessing, "Remote sensing, formation of objects and determination of quality", pp.581-594, January 2002.
- [35] M. Baatz and A. Schape, "Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation", January 2000.
- [36] A. Natsev, J. R. Smith, N. C. F. Codella, G. Hua. Towardslarge Scale Land-cover Recognition of Satellite Images. In Information, Communications and Signal Processing (ICICSP) 2011, 8th International Conference on, pages 1–5. IEEE, 2011.