

Identificación de los aspectos a tener en cuenta para la integración de IoT, Cloud Computing y Big Data

María A. Murazzo, Nelson R. Rodriguez, Diego Medel, Gustavo Fernandez

Departamento e Instituto de Informática, Facultad de Ciencias Exactas, Físicas y Naturales – UNSJ

Complejo Islas Malvinas (CUIM), San Juan, Argentina

marite@unsj-cuim.edu.ar, nelson@iinfo.unsj.edu.ar, mdiego88@gmail.com,
gustavofernandez.gdf@gmail.com

Resumen

El éxito del IoT (Internet of Things) radica en contar con una infraestructura que provea ubicuidad, fiabilidad, eficiencia, escalabilidad y capacidad de almacenamiento y cómputo. Con el objetivo de lograr esto es que se considera que el cloud puede ofrecer una infraestructura que permita soportar las grandes cantidades de datos generados por los dispositivos de IoT y la capacidad de cómputo necesaria para procesar esos datos de manera dinámica y global. De esta manera los usuarios no deberán preocuparse por la escasez de recursos locales, sino centrarse en aprovechar la potencialidades de la aplicación. El presente trabajo tiene por objetivo identificar algunos de los aspectos a tener en cuenta en la integración del IoT, el cloud y el big data, así como las características de esta convergencia.

Keywords: IoT, Cloud Computing, Big Data, Hadoop.

1. Introducción

Debido a que la cantidad de dispositivos conectados a Internet crece exponencialmente, el volumen de los datos generado por las aplicaciones de usuario en esos dispositivos aumenta cada día. En definitiva se tienen cada vez más dispositivos conectados a Internet, tales como computadoras, teléfonos móviles, sensores, cámaras de vigilancia, GPS, consolas de videojuegos; cada uno de ellos corriendo aplicaciones tales como redes sociales, georeferenciación, sistemas transaccionales, alertas, etc. y cada una de ellas generando datos en una amplia gama de formatos (texto, gráficos, imágenes, videos, sonidos, etc).

El primer inconveniente que se presenta en estos contextos es el almacenamiento local de los datos, debido a las restricciones físicas que poseen tales escaso almacenamiento, capacidad de procesamiento y memoria, unido a las restricciones de consumo de energía.

Esta problemática plantea la necesidad de la integración del Big Data, el Iot y el Cloud Computing con el fin de definir un nuevo paradigma denominado CoT (Cloud of Things). Esta confluencia tecnológica permitirá que el IoT aporte los medios para que el usuario interactúe mediante redes y dispositivos ubicuos. En este nuevo paradigma, cloud permitirá el acceso a un pool de recursos virtualizados capaces de almacenar y procesar bajo demanda la gran cantidad de datos generada por las aplicaciones alojadas en los dispositivos, esto es el big data.

2. Conceptos Previos

Los avances en computación móvil, redes inalámbricas, dispositivos móviles y sistemas embebidos han dado lugar al paradigma de IoT el cual consiste en una infraestructura de red global y dinámica de nodos (things) interconectados, inteligentes y auto configurables. IoT es un habilitador del ecosistema IoE (Internet of Everything) que permite no solo que las personas se interconecten y se comuniquen, sino que ahora también lo hacen los procesos, los datos y los objetos, convirtiéndose en sus cuatro pilares. Esta interrelación permite transformar la información en acciones que crean nuevas capacidades y experiencias.

La interacción entre los pilares de IoT establece tres tipos de conexiones: personas que se comunican con personas (P2P), máquinas que se comunican con personas (M2P) y máquinas que se comunican con máquinas (M2M). Sin importar cual de estos tipos de conexiones se usan, en todos los casos se trata de una convergencia de múltiples tecnologías para permitir intercambiar datos entre dispositivo conectado. Todo esto se realiza con base en tecnologías de sensores, RFID (Radio Frequency Identification), NFC (Near Field Communications) y sistemas de geolocalización, que permitirán que los objetos cotidianos se conecten y se vuelvan inteligentes, cambiando radicalmente la forma realizan las tareas cotidianas [1].

Actualmente, el 99% de los objetos que se encuentran en el mundo físico no están conectados aún a Internet sin embargo se estima que para el año 2020 habrá 50 millones de objetos inteligentes

conectados a Internet, los cuales proporcionarán billones de gigabytes de datos que deben ser almacenados, procesados y presentados en forma eficiente y fácilmente interpretables. Se considera que el 90% de los datos existentes se han generado en los últimos dos años, esto está dando lugar a la Era del Exa y Zetta Byte [2]. Este aumento en la cantidad de datos demanda nuevas estrategias que permitan su almacenamiento, procesamiento y análisis de manera eficiente; esto conlleva un cambio de paradigma en las arquitecturas de cómputo, los algoritmos y los mecanismos de procesamiento.

Frente a esta problemática se ha popularizado el término Big Data [3], el cual se usa para describir grandes conjuntos de datos, los cuales exhiben las siguientes propiedades: variedad, volumen, velocidad, variabilidad, valor y complejidad; todas ellas denotan a datos multidimensionales. Estas propiedades hacen que los sistemas de cómputo convencionales sean muchas veces inapropiadas para lograr un procesamiento adecuado [4] [5].

Es aquí donde la integración de la IoT con Cloud Computing permite que esta gran cantidad de datos puedan ser alojados en internet, favoreciendo que recursos, servicios y datos estén disponibles para su uso y para la prestación de servicios end-to-end para empresas y usuarios que accedan desde cualquier lugar, proporcionando la infraestructura virtual de integración para dispositivos de almacenamiento, herramientas de análisis, visualización y plataforma.

En este sentido, Cloud Computing, es un modelo para permitir el acceso en red omnipresente, conveniente y bajo demanda a un conjunto compartido de recursos informáticos configurables tales como, redes, servidores, almacenamiento, aplicaciones y servicios que pueden ser rápidamente provisionados y liberados con un mínimo de gestión esfuerzo e interacción con el proveedor o administrador de servicio [6].

Gracias a estas características una infraestructura cloud permite contar con una cantidad de recursos computacionales virtualmente infinitos, administrados por terceros y accedidos bajo demanda pagando por el uso, de esta manera es posible lograr calidad de servicio (QoS) garantizado por parte del proveedor de la infraestructura cloud.

Según [7], cloud es un modelo de prestación de servicios informáticos cuya principal orientación es la escalabilidad. Esto es, que desde el punto de vista de los usuarios, los servicios son elásticos, o sea, pueden crecer o recuperar su tamaño original de manera rápida y sencilla. Esta orientación permite que los usuarios que acceden a los servicios, perciban que todo funciona de manera simple y rápida, dando como resultado una experiencia más gratificante.

En función de esto, cloud se ha convertido en un enorme repositorio de recursos computacionales, lo

cual es una buena posibilidad para construir una plataforma para las aplicaciones que necesitan una gran cantidad de recursos. Esta capacidad del cloud se debe principalmente a la habilidad de escalado elástico de recursos en función de las necesidades de las aplicaciones y el presupuesto del usuario. Esta es una tecnología centrada en ofrecer cualquier recurso (bases de datos, red, procesador, etc.) y ofrecerlo como un servicio (AaaS, Anything as a Service) bajo demanda, inclusive el cómputo [8].

3. Características de la Integración

IoT promete un mundo donde todo este conectado, para lograr esto es necesario: accesibilidad y conectividad ubicua, orquestación dinámica de dispositivos, óptima utilización de recursos y personalización de servicios. Para lograr que estos servicios sean prestados de manera eficiente y efectiva es necesario asegurar altos niveles de disponibilidad y escalabilidad.

Debido a que los dispositivos de IoT están geográficamente distribuidos sobre plataformas heterogéneas y tienen limitadas capacidades de procesamiento, energía y almacenamiento, aspectos tales como performance, seguridad, confiabilidad e integridad no se pueden asegurar pues no existen recursos necesarios. Esto lleva a la necesidad de integración con el cloud, el cual puede ofrecer sus capacidades ilimitadas (esta es una sensación dada por la virtualización de recursos) de almacenamiento y poder de cómputo.

Por su parte, cloud se puede beneficiar del IoT al poder extender sus límites con las cosas del mundo real, de forma más dinámica y distribuida, y entregar un número masivo de servicios en tiempo real. El cloud actuará como una capa intermedia entre las aplicaciones y los dispositivos, ocultando las funcionalidades y complejidades que poseen.

Una infraestructura que integre cloud e IoT permitirá [9]:

- *Almacenamiento:* es posible alojar datos estructurados y no estructurados en múltiples formatos y con las características del big data.
- *Cómputo:* uno de los retos en ambientes IoT es la escalabilidad y el cloud lo puede proveer mediante el uso de infraestructura de altas prestaciones capaces de realizar cómputo intensivo.
- *Comunicación:* por lo general los dispositivos de IoT usan una IP para identificarse, lo cual no es un esquema global. Para solucionar esto cloud usa la georeferenciación, que permite identificar cualquier objeto en cualquier lugar que este.

- *Nuevas habilidades*: IoT se caracteriza por la heterogeneidad de dispositivos, protocolos y tecnologías, esto genera problemas de confiabilidad, interoperabilidad, escalabilidad, seguridad, disponibilidad y eficiencia que el cloud resuelve, además provee facilidad de acceso, facilidad de uso y costos de implementación reducidos.

Estas características serán capaces de proveer procesamiento en tiempo real y aprovisionamiento de recursos dinámico para el big data, mediante la orquestación y administración de instancias virtualizadas de recursos físicos.

4. Migración del cómputo al cloud

Como ya se analizó anteriormente, el principal problema de IoT es el almacenamiento de los datos generados por los dispositivos, es aquí donde cloud aparece como la infraestructura adecuada para alojar los datos y brindar la posibilidad de realizar un procesamiento de ellos con el objeto de obtener información que le sirva al usuario para realizar una toma de decisiones. Este procesamiento de datos requiere de algoritmos de cómputo intensivo que permitan realizar una extracción de conocimiento de manera eficiente y efectiva [10].

Entre los diversos modelos de programación, MapReduce es el más adecuado para el desarrollo de algoritmos de cómputo sobre grandes volúmenes de datos en el cloud debido a la tolerancia a fallos. MapReduce [11] es un modelo de programación paralela desarrollado y utilizado por Google para procesar grandes conjuntos de datos distribuidos a lo largo de clusters de servidores, los cuales pueden ser virtuales o físicos.

Unas de las características de cloud es poder definir cluster virtuales, esto permite ejecutar aplicaciones MapReduce sobre el cloud (en realidad sobre un cluster en el cloud) sin tener que lidiar con la creación y mantenimientos de cluster físicos.

La principal motivación del modelo de programación MapReduce es la delegación del cómputo intensivo en cluster físicos o virtuales (Cluster as a Service) que, mediante un sistema de archivos distribuido, reparte la carga de trabajo, optimizando tiempo y recursos. Así mismo, facilita un patrón de desarrollo paralelo para simplificar la implementación de algoritmos de cómputo intensivo en entornos distribuidos. Este modelo puede dividir un espacio grande de problema en espacios pequeños y paralelizar la ejecución de tareas más pequeñas en estos sub espacios [12].

4.1. Cómputo del Big Data en el cloud

Como se ha discutido, la cantidad de datos generado por las aplicaciones de IoT seguirá creciendo y con ello la diversidad y la complejidad. Esto hace

imposible el procesamiento con herramientas convencionales, para extraer información útil que permita realizar una toma de decisiones adecuada. Esta tarea se hace computacionalmente imposible, por parte de algoritmos de cómputo tradicional. Por ello es que se debe migrando a entornos de cómputo intensivo para mejorar los tiempos de respuesta y aumentar la escalabilidad y la eficiencia. Para lograr esto, es posible realizar una implementación que permita distribuir los datos y paralelizar el cómputo [13].

Los sistemas distribuidos, tales como cloud, proveen una infraestructura que favorece de manera eficiente y escalable al análisis de grandes volúmenes de datos, permitiendo procesar gran cantidad de datos de forma distribuida.

Las características que ofrece el cloud permite migrar los algoritmos de cómputo para big data a modelos basados en MapReduce sobre el cloud. En [14] se presenta un modelo para realizar procedimientos de clasificación *supervisada* sobre grandes cantidades de datos en una arquitectura distribuida basada en MapReduce Hadoop. Para paralelizar la clasificación se usa ICP (InterIMAGE Cloud Platform): Data Mining Package, una herramienta open source, basada en MapReduce, que implementa los cuatro clasificados de WEKA (Naïve Bayes Classifier, Decision Trees, Random Forest y Support Vector Machines) como librerías de java.

Por otro lado, clustering es otra tarea muy importante que se realiza sobre grandes conjuntos de datos. Con el incremento de los volúmenes de datos la tarea de clusterización se ha vuelto complicada debido a que los dataset se han complejizado, los algoritmos consumen mucho tiempo, el hardware usado para procesamiento se ha vuelto insuficiente, etc. Estos aspectos han llevado a proponer métodos para que se ejecuten los algoritmos de clustering en ambientes distribuidos (cloud). En [15] se propone el diseño de un framework para algoritmos de *clusterización*, tal como k-means, en MapReduce. Las pruebas realizadas demuestran que estos algoritmos basados en MapReduce poseen un buen comportamiento cuando el dataset es grande.

Otra herramienta muy usada cuando se trabaja con grandes volúmenes de datos son las reglas de asociación, las cuales permiten encontrar relaciones entre los datos. En esta categoría se encuentra el Algoritmo Apriori, el cual ejecutarlo de manera secuencial sobre dataset muy grandes es prohibitivo. En [16] se presenta una implementación del Algoritmo Apriori en MapReduce, cuyo objetivo es bajar los tiempos de ejecución y los recursos computacionales necesarios para el cómputo en ambientes de big data. Los resultados que se obtuvieron con la ejecución de la implementación del algoritmo en forma paralela, fueron

satisfactorias, sobre todo cuando el dataset era grande.

Otra opción a la migración de algoritmos de cómputo intensivo al cloud es la que se presenta en [17] donde se describe la implementación de un entorno completo de forma distribuida, tal es el caso de *DistributedWekaSpark*, que extiende las funcionalidades de Weka combinándolo con Spark de Hadoop, permitiendo a los usuarios fidelizados de Weka usar la misma interface para trabajar en forma local o distribuida.

4.2. Herramienta para el cómputo del Big Data en el cloud

Como ya se analizó, una de las potencialidades del cloud es definir cluster virtuales con el objeto de migrar aplicaciones de cómputo intensivo. Para implementar esta solución una herramienta que puede ser usada en Hadoop.

Hadoop [18] es una implementación open source de MapReduce escrito en Java y propuesto por la Apache Foundation, que maneja la distribución de datos en un conjunto de nodos de cómputo y usa mecanismos eficientes para su análisis. Hadoop posee una arquitectura *Master-Slave*, usando para almacenar los datos distribuidos *Hadoop Distributed File System (HDFS)*, creado a partir del *Google File System (GFS)*.

En la actualidad, Hadoop es un ecosistema de tecnologías que permite contar con herramientas que implementan todos los algoritmos necesarios para realizar un eficiente cómputo del big data. Entre ellas se pueden destacar *Hive* (un sistema de Data Warehouse), *Hbase* (base de datos No SQL), *Mahout* (machine learning mediante tres algoritmos recomendación, clustering y clasificación), *Pig* (data analytic), entre otros.

Con el objetivo de usar al cloud como infraestructura de cómputo y almacenamiento, se puede configurar un cluster Hadoop en el cloud de Google, denominada *Google Cloud Platform* [19]. En la figura 1 se pueden ver los componentes de la arquitectura de *Google Cloud Platform*.



Figura 1: Componentes de la arquitectura de *Google Cloud Platform*

Para lograr acceder a implementar y operar *Hadoop as a Service*, Google posee la herramienta *Cloud Dataproc* que es un administrados desatendido para Hadoop que permite crear, administrar y usar clusters, además de usar herramientas de cómputo, consulta y machine learning.

Una de las formas de administrar el cluster es mediante *Google Cloud SDK* que provee un conjunto de herramientas entre las cuales se encuentra la línea de comando *gcloud*. El proceso de instalación es simple, desde Ubuntu se debe escribir:

```
$ export CLOUD_SDK_REPO="cloud-sdk-$(lsb_release -c -s)"
$ echo "deb
http://packages.cloud.google.com/apt
$CLOUD_SDK_REPO main" | sudo tee
/etc/apt/sources.list.d/google-cloud-
sdk.list
$ curl
https://packages.cloud.google.com/apt/do
c/apt-key.gpg | sudo apt-key add -
$ sudo apt-get update && sudo apt-get
install google-cloud-sdk
$ gcloud init
```

Una vez iniciado *gcloud*, se puede interactuar con el cluster Hadoop por línea de comando. Por ejemplo para crear el cluster se debe escribir:

```
gcloud dataproc clusters create
<cluster-name>.
```

Usando *Cloud Dataproc* se accede a *Google Compute Engine* (permite configurar maquinas virtuales para cómputo intensivo mediante instancias de recursos físicos), *Google Cloud Storage* (sistema de almacenamiento de objetos) y *Google Cloud Virtual Network* (interconexión de recursos mediante Virtual Private Cloud).

5. Caso Practico

IoT tiene muchas áreas potenciales de aplicación, tales como industria, educación, agricultura, biología, agricultura, medicina, etc. Estas áreas requieren aplicaciones de monitoreo, seguimiento y control de aspectos de la vida cotidiana. De todas las aplicaciones posibles una de las mas usadas en la actualidad son las aplicaciones móviles. Este tipo de aplicaciones, generan una plataforma para desplegar contenidos rápidamente y con bajo costo. Sin embargo hay que considerar que las exigencias y requerimientos de los usuarios tanto a nivel profesional como social han cambiado y se han ampliado.

Las principales características que necesitan de los servicios móviles actualmente son tres: ubicuidad, disponibilidad y seguridad. Además hay que considerar que las necesidades de los usuarios cambian al ritmo de la tecnología por lo que ahora se esta exigiendo de las aplicaciones localización, inmediatez y personalización. Todas estas características solo se pueden lograr usando una infraestructura cloud, es por ello que este tipo de

aplicaciones se denominan *Mobile Cloud Computing* [20], ejemplo de este tipo de aplicaciones es BYOD (Bring your own device), que ofrece a los usuarios un alto grado de flexibilidad en sus tareas.

En este trabajo se presenta una aplicación colaborativa que a través del uso de dispositivos móviles, cuyo objetivo es permitir informar al usuario de eventos sucedidos en la vía pública que pueden llegar a entorpecer la circulación de los habitantes del lugar.

Para lograr esto, la aplicación muestra el mapa del sitio y permite que ante un evento el usuario pueda marcarlo en el mapa agregando comentarios y fotos. De esta forma los demás usuarios que cuenten con esta aplicación cargada, recibirán una alerta cuando se aproximen al lugar del evento.

Este sistema de alerta se logra de la siguiente manera:

1. Cada vez que un usuario informa de un evento el dispositivo se conecta con el cloud de Google donde se almacena la latitud y la longitud del evento, así como la demás información enviada por el usuario.
2. Los eventos pueden ser informados por mas de un usuario, por lo que cada uno de ellos tendrá un conteo de usuarios que lo han informado, esto permite dar grados de credibilidad a la información.
3. En el cloud además de guardar los datos se realizará un procesamiento de ellos con el objeto de clusterizarlos a fin de agruparlos por proximidad, esto permite que cada usuario reciba alerta de eventos en un radio definido en función de lo geolocalización.

El frontend de la aplicación se construyó con Android Studio y permite la manipulación de las APIs de Google Map.

Como base de datos para resguardar la información del usuario, información compartida y además servicios que utilizará, se usó Google Cloud Plataform que integra todos los servicios de cloud necesarios para construir el backend de la aplicación.

El framework y cliente Mobile Backend Starter de Google permite desplegar la infraestructura necesaria para usar desde la aplicación todas las ventajas de Google App Engine (GAE) y sus servidores.

En Google Cloud Plataform se a configurado un cluster Hadoop con Mahout con el objeto de realizar el cómputo en base a los algoritmos de machine learning que permita implementar la clusterización de los datos para dar las alertas. En la Figura 2, se pueden ver las interacciones necesarias para la implementación.

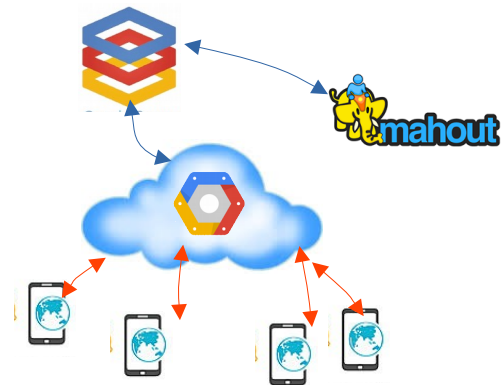


Figura 2: Interacción de componentes

En la figura 3 se puede ver la pantalla de la aplicación donde se a marcado un evento, al pulsar sobre un marcador concreto, la aplicación ofrecerá al usuario el suceso ocurrido y compartida por los demás usuarios que estén utilizando la aplicación.

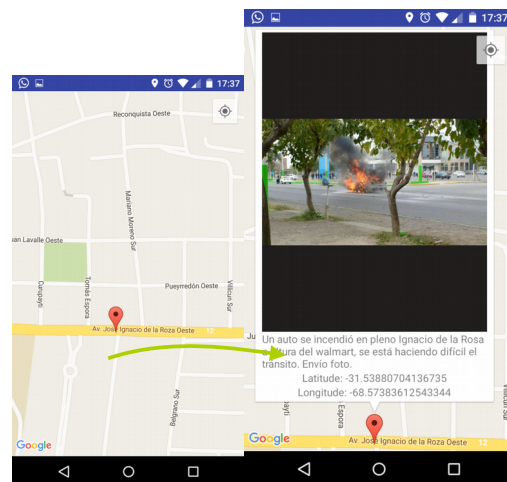


Figura 3: Aplicación mostrando un evento marcado

Esta aplicación pretende mejorar la experiencia de los usuarios ofreciéndole la posibilidad de enterarse de problemas sucedidos en la vía pública antes de encontrarse con ellos, de esta manera podrá elegir caminos alternativos, sin el inconveniente de pérdida de tiempo en su transporte

6. Conclusiones

Los ambientes de cómputo distribuido permiten hacer una distribución de los datos y una paralelización del cómputo, lo cual es muy importante cuando se trabaja con grandes volúmenes de datos. Además, este es un aspecto crítico sobre todo cuando las plataformas computacionales no proveen el hardware suficiente para realizar las tareas en tiempo y forma, como es el caso de los dispositivos de IoT.

El cloud computing es un paradigma que ha cambiado la ubicación de las infraestructuras

informáticas hacia Internet. Con ello se consigue que los usuarios no tengan que gestionar ni sus propios servidores ni tampoco su software, ahorrando de este modo energía, espacio físico y personal técnico. Además, los sistemas cloud ofrecen un gran rendimiento en términos de escalabilidad, mantenibilidad y procesamiento masivo de datos en entornos dinámicos y de necesidades cambiantes.

En función de esto y de todo lo analizado, no cabe dudas que la integración del IoT y el cloud con el objeto de realizar un adecuado almacenamiento y procesamiento de los datos, es una opción rentable, no solo desde el punto de vista económico sino también en lo que respecta a desempeño, escalabilidad, disponibilidad y calidad de los servicios ofrecidos.

7. Bibliografía

- [1] Lee, I., & Lee, K. "The Internet of Things (IoT): Applications, investments, and challenges for enterprises". *Business Horizons*, 58(4), 431-440. 2015.
- [2] Acín, Bird, Boccali, Cancio, Collier, Corney, Fuhrmann, *Architectures and methodologies for future deployment of multi-site Zettabyte-Exascale data handling platforms*. In *Journal of Physics: Conference Series* (Vol. 664, No. 4, p. 042009). IOP Publishing. 2015.
- [3] Nyikes, Rajnai, Z. *Big data, as part of the critical infrastructure*. In *Intelligent Systems and Informatics (SISY)*, 2015 IEEE 13th International Symposium on (pp. 217-222). IEEE. 2015.
- [4] Katal, A., Wazid, M., & Goudar, R. H. (2013, August). *Big data: issues, challenges, tools and good practices*. In *Contemporary Computing (IC3)*, 2013 Sixth International Conference on (pp. 404-409). IEEE.
- [5] Chen, Mao, Liu, Y. (2014). *Big data: a survey*. *Mobile Networks and Applications*, 19(2), 171-209.
- [6] Antonopoulos, Gillam. *Cloud Computing: Principles, Systems and Applications*. Editorial Springer Science & Business Media. 2010.
- [7] Mell, Grance. *The NIST definition of cloud computing*. NIST Special Publication 800 – 145. 2011.
- [8] Chee, Franklin. *Cloud computing: technologies and strategies of the ubiquitous data center*. CRC Press. 2010.
- [9] Babu, Lakshmi, Rao. *A study on cloud based Internet of Things: CloudIoT*. In *Communication Technologies (GCCT)*, 2015 Global Conference on (pp. 60-65). IEEE.
- [10] Murazzo, Rodríguez, Guevara, Tinetti. *Identificación de Algoritmos de Cómputo Intensivo para Big Data y su Implementación en Clouds*. XVIII Workshop de Investigadores en Ciencias de la Computación. 2016.
- [11] Malik, Sangwan. *Mapreduce Algorithms Optimizes the Potential of Big Data*. *International Journal of Computer Science and Mobile Computing*, Vol.4 Issue.6, June – 2015.
- [12] Carrera, Geyer. *Modeling the Performance of MapReduce Applications for the Cloud*. *Latin American Journal of Computing Faculty of Systems Engineering National Polytechnic School Quito-Ecuador*, 2(2). 2015.
- [13] Ambulkar, Borkar. *Data mining in cloud computing*. In *MPGI National Multi Conference* (pp. 7-8). 2012.
- [14] Ayma, Ferreira, Happ, Oliveira, Feitosa, Costa, Gamba. *Classification Algorithms for Big Data Analysis, a Map Reduce Approach*. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), 17. 2015
- [15] Jin, Kou, Liu, Guo. *A common framework of partition-based clustering for large scale dataset using sampling and its MapReduce implementation*. *Tehnički vjesnik*, 23(1), 25-33. 2015.
- [16] Dasgupta. *Study of various parallel implementations of association rule mining algorithm*. *American Journal Of Advanced Computing*, 2(1). 2015
- [17] Koliopoulos, Yiapanis, Tekiner, Nenadic, Keane. *A Parallel Distributed Weka Framework for Big Data Mining using Spark*. In *Big Data (BigData Congress)*, 2015 IEEE International Congress on (pp. 9-16). IEEE. 2015.
- [18] Hadoop. Welcome to Apache Hadoop. <http://hadoop.apache.org>.
- [19] Google. Google Cloud Platform. <https://cloud.google.com/>.
- [20] Murazzo, Rodríguez. *Mobile cloud computing*. XII Workshop de Investigadores en Ciencias de la Computación. 2010