

Reconstrucción del patrón de consumo eléctrico a partir de Big Data mediante técnica de MapReduce

(Reconstruction of the Electric Consumption Pattern from Big Data using MapReduce Technique)

Esteban Inga¹, Juan Inga², Estuardo Correa³, Roberto Hincapié⁴

Resumen:

El trabajo presenta el rendimiento de la técnica de MapReduce para reconstruir la curva de carga a partir de una cantidad de información previamente almacenada proveniente de medición inteligente de energía eléctrica y considerada en la actualidad como Big Data. La gestión de información en la etapa de una red eléctrica inteligente considerada como Sistema de Gestión de Datos Medidos o MDMS necesita reducir los tiempos respecto de los reportes que se requieran en un determinado instante para toma de decisiones en relación a la respuesta de la demanda eléctrica. Por lo tanto, este trabajo propone el uso de MapReduce como técnica para conseguir información de la curva de carga en un tiempo adecuado para obtener tendencias y estadísticas relacionadas con el patrón de consumo eléctrico residencial.

Palabras clave: big data, mapreduce, sistema de gestión de datos medidos, red eléctrica inteligente, medición inteligente.

Abstract:

The work presents the performance of the MapReduce technique to reconstruct the load curve from a previously stored amount of information coming from smart metering of electrical energy and currently considered as Big Data. The management of information in the stage of an intelligent electrical network considered as a System of Management of Measured Data or MDMS needs reducing the times with respect to the reports that are required in a certain moment for decision making in relation to the electrical demand response. Therefore, this paper proposes the use of MapReduce as a technique to obtain information of the load curve in a suitable time to obtain trends and statistics related to the residential electric pattern.

Keywords: big data, mapreduce, meter data management system, smart grid, smart metering.

1. Introducción

A partir del concepto de red eléctrica inteligente en la actualidad para llegar a una red eficiente y confiable se han generado varias etapas y subetapas con una misión definida y concreta. Así tenemos que la etapa de medición inteligente conformada por los medidores inteligentes obtiene la información de consumo eléctrico de los usuarios o consumidores (residenciales, comerciales e industriales), para esto se ha generado una infraestructura de medición inteligente que a partir de telecomunicaciones inalámbricas y de fibra óptica permite garantizar la conectividad de los medidores inteligentes y la oficina central de las empresas eléctricas según lo demuestran (Campaña, Inga, & Hincapié, 2017; Esteban Inga, Céspedes, Hincapié, & Cárdenas, 2017; Juan Inga, Inga, Hincapié, &

¹ Universidad Politécnica Salesiana, Quito, Ecuador (einga@ups.edu.ec).

² Universidad Politécnica Salesiana, Cuenca, Ecuador (jinga@ups.edu.ec).

³ Universidad Politécnica Salesiana, Quito, Ecuador (ecorrea@ups.edu.ec).

⁴ Universidad Pontificia Bolivariana, Medellín, Colombia (roberto.hincapie@upb.edu.co).

Cristina, 2017; Arturo Peralta, Inga, & Hincapié, 2017). La información recolectada de millones de medidores inteligentes es almacenada en silos de información para su respectivo análisis y minería de datos para conocer el patrón de comportamiento eléctrico de cada usuario y con esta información tener una respuesta a la demanda más confiable con la cual se podrán *a posteriori* reducir los costos por generación, ya que al conocer el patrón de consumo se genera electricidad en estrecha relación con lo que se va a utilizar y se minimizan las pérdidas de energía (Inga-ortega, Inga-ortega, & Gómez, 2017).

Entonces, se prevé que la infraestructura de medición inteligente está estrechamente relacionada al sistema de gestión de datos medidos; es donde la cantidad de información va tomando una real importancia y un problema no trivial que requiere de un tratamiento de información con el menor costo por los recursos empleados en el proceso de tratamiento de información (E. Inga & Hincapié, 2016).

Para transportar la información desde los medidores inteligentes, se han generado modelos de planeación de redes inalámbricas heterogéneas representadas en un multigráfico-multicapa para llevar información con la mejor cobertura de los medidores inteligentes y considerar tecnologías de bajo costo en relación con otras opciones como la red celular, la misma que es limitada por la falta de cobertura en zonas suburbanas y rurales. En este sentido, el despliegue óptimo de una red inalámbrica a través de multisalto (Lichtensteiger, Bjelajac, Müller, & Wietfeld, 2010) ha sido considerado como la mejor opción, incluso en relación con trabajos iniciales que proponen el uso híbrido de fibra y red inalámbrica donde el costo por despliegue de una red soterrada de fibra óptica encarece la inversión para ser tomada en cuenta para una medición inteligente de energía eléctrica, esto se aprecia en el trabajo de (Inga-Ortega, Peralta-Sevilla, Hincapié, Amaya, & Tafur Monroy, 2015; J. Inga, Inga, Hincapié, & Gómez, 2016; A Peralta, Inga, & Hincapié, 2015).

De esta manera, una infraestructura de medición inteligente al articularse con el sistema de gestión de datos medidos en sus diferentes topologías como: centralizada, semi distribuida y totalmente distribuida que permiten optimizar el uso de la capacidad de la red de comunicaciones ya que en esta etapa de MDMS se puede reducir información poco relevante para la toma de decisiones (Moscoso-zea & Lujan-mora, 2017) y que no requiere de su transporte. En el MDMS también se articulan los procesos de encriptación y seguridad de información debido al uso de la información importante y que pudiera ser usado para fines delincuenciales por la importancia de conocer los detalles del patrón de consumo de energía de un usuario (Wan, Wang, Yang, & Shi, 2014), (Arian, Ameli, Soleimani, & Ghazalizadeh, 2011; Zhou, Hu, & Qian, 2012).

Es así que la cantidad de información almacenada con el consumo de energía de los usuarios superan los gigabytes (Correa, Inga, Inga, & Hincapié, 2017), lo cual incide en la necesidad de buscar un proceso computacional en paralelo que permita reducir el costo en término de recursos empleados por el tiempo empleado en este tipo de tratamiento de información; en otras palabras, hacer uso de una técnica de minería de datos nos facilita reducir los tiempos en procesamiento de información y por consiguiente obtener un rápido reporte por análisis para una subsiguiente toma de decisiones para mejorar la respuesta de la demanda de acuerdo con lo propuesto por (Ahmed, 2014; Grolinger, Capretz, & Seewald, 2016; X. Liu, Golab, Golab, & Ilyas, 2015; Munshi & Mohamed, 2017).

La *Figura 1* muestra el esquema conceptual de la técnica de MapReduce para reconstruir el patrón de consumo de energía eléctrica de la información que fuera previamente tratada y almacenada por el MDMS.

En adelante este artículo se organiza de la siguiente manera. En la *Sección 2* se introducen los fundamentos de MapReduce en medición inteligente de energía eléctrica. En la *Sección 3* se plantea la formulación del problema. En la *Sección 4* se realiza el análisis de los resultados a partir de métricas y tendencias. Finalmente, en la *Sección 5* se concluye este artículo.

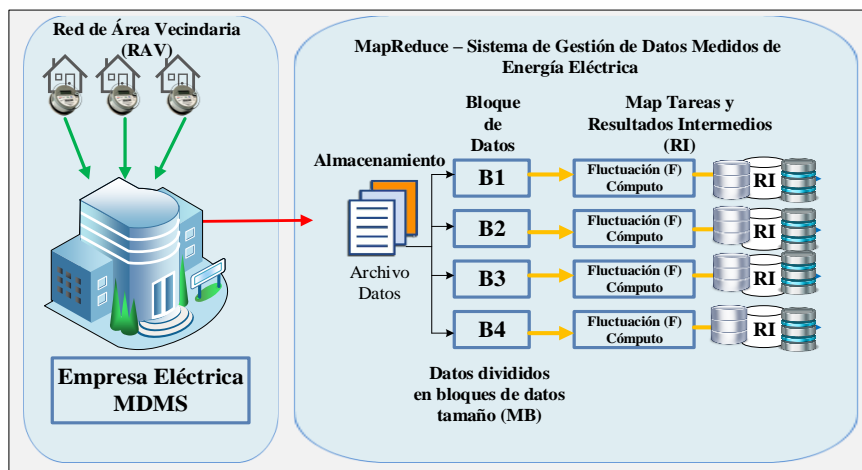


Figura 1. Esquema de MapReduce para SGDM

2. Procesamiento de datos paralelo en el Sistema de Gestión de datos medidos

Un sistema de gestión de datos medidos en cualquiera de sus topologías alberga una gran cantidad de información, lo cual genera información histórica con el comportamiento del consumo de energía de todos los usuarios del sistema eléctrico. Por esta razón, se advierte una posibilidad para la minería de datos que permita seleccionar, filtrar, leer y analizar la información recolectada desde los medidores inteligentes aporte sustancial de este trabajo.

Para contemplar la necesidad de realizar un proceso en paralelo es importante justificar que el crecimiento de la información en un MDMS realmente cumple con las características de *big data*; por lo tanto, si se consideran medidores inteligentes de energía eléctrica con intervalos de lectura cada 15 minutos, en 24 horas se tendría 96 lecturas por día y en un mes 2880 lecturas. Además, si se considera que cada lectura representa 5 bytes; en un escenario de 4'000.000 de abonados, al multiplicar por 5 bytes genera una cantidad aproximada a 20'000.000 bytes por lectura. La *Tabla 1* muestra el crecimiento de la información para un intervalo entre cada lectura.

Tabla 1. Cantidad de información para un intervalo de lectura de 15 minutos

Número de medidores	Datos	Lecturas	Intervalo de tiempo	Cantidad de datos
4M	5 Bytes	96	1 Día	1.92GB
4M	5 Bytes	96	1 Mes	57.6GB
4M	5 Bytes	96	1 Año	691.2GB

De los datos preliminares se puede apreciar que el crecimiento de la información es considerable, más aún si se requieren obtener reportes de esa cantidad de datos de manera eficiente para una posterior toma de decisiones.

La medición inteligente puede incluir ciertas variaciones respecto de la decisión del intervalo de lectura del suministro eléctrico, en este caso podría ser inferior o superior a los 15 minutos antes mencionados. Esta decisión la estipula la empresa de distribución eléctrica en virtud de requerir una información en un intervalo reducido según las necesidades vinculadas al sistema de gestión de datos medidos, en caso de no requerir un intervalo reducido puede optar por 30 minutos. Sin embargo, empresas eléctricas advierten que la medición inteligente podría a futuro requerir una lectura de tiempo real, este particular encarece el sistema tanto a nivel de red de comunicaciones como en almacenamiento y procesamiento de información.

En la *Figura 2* se muestra un análisis de datos con lecturas cada 7 minutos, se obtiene un total de 206 lecturas diarias. Si se toman como referencia lecturas cada 15 minutos, se tiene un total de 96 lecturas diarias. Finalmente, con lecturas cada 30 minutos, se tiene 48 lecturas diarias. Si advertimos un número de 4 millones de clientes, caso ecuatoriano, provocaría el verdadero efecto de Big Data; el almacenamiento; entonces, el tratamiento de la información tiene mucha validez para las empresas eléctricas especialmente en la toma de decisiones, patrón de consumo eléctrico y la respuesta de la demanda, para esto se debe recordar que una generadora no cuenta con almacenamiento de energía y su generación se da a partir de la demanda por lo cual conocer el patrón de comportamiento de los usuarios reduce costos por generación eléctrica. Los valores en color azul se encuentran en MB; mientras que en color anaranjado y rojo en GB.

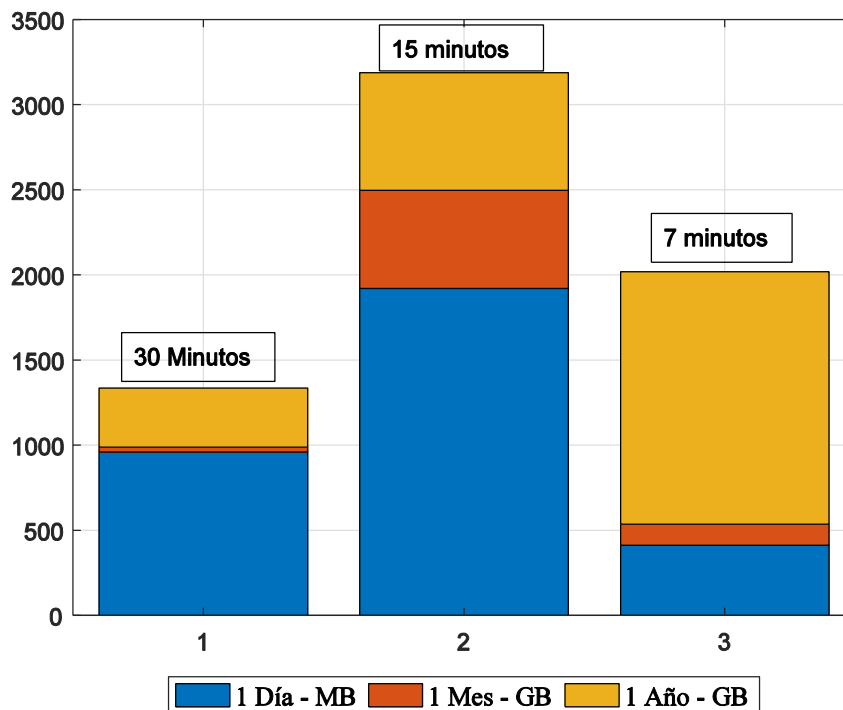


Figura 2. Intervalo de Lectura y crecimiento de información. Fuente: Autor

Un MDMS integra y almacena información de una medición inteligente; por lo tanto, lo que se aprecia es una necesidad de escalabilidad en cuanto al almacenamiento y procesamiento de la información y al crecimiento de los recursos físicos; pero el problema de este trabajo se centra en la gestión misma de la información previamente almacenada en silos. Esta información requiere de un tiempo oportuno con la finalidad de obtener los reportes y/o estadísticas para determinar el comportamiento del consumo de energía eléctrica, y a partir del empleo de MapReduce como técnica para manejo de *big data* que a través de programación paralela o paralelismo incrementa el rendimiento computacional y reduce los costos por procesamiento como expresan (Chihoub & Collet, 2016; Dai, Chai, Qiu, Zhang, & Jiang, 2016; Han & Yang, 2016; Shi et al., 2016; Zhang, 2017).

El modelo de programación MapReduce propone una función map que se encarga de procesar en partes la información, pero identificadas con una llave/valor que genera un conjunto de llave/valor intermedio y a través de la función reduce se encarga de unir o fusionar todos los valores intermedios asociados con la misma llave intermedia, este trabajo es paralelizado de manera automática. Es decir, la información es fragmentada con un identificador con el fin de reducir el tiempo de procesamiento ya presentado por (Lang & Patel, 2010; Selamat, 2016; Sieverts & Systems, 2015; Vaccaro, Troiano, Vaccaro, & Vitelli, 2016; Wang, 2014; Yildiz, Ibrahim, Phuong, & Antoniu, 2015).

MapReduce presenta una característica importante para resolver este tipo de problemas debido a su programación escalable, capaz de procesar un volumen masivo de información proveniente de la medición inteligente de energía eléctrica bajo una ejecución en un gran número de nodos de computo de productos (Ahmed, 2014; Fanibhare & Dahake, 2016; Grolinger et al., 2014, 2016; R. Liu, Kuo, Yang, Chen, & Liu, 2016; Mashayekhy, Member, Nejad, & Member, 2015; Nghiem & Figueira, 2016).

3. Formulación del problema

El problema de gestión de *big data* en MDMS requiere de un proceso recursivo en paralelo que permita adquirir el total o parte de la información según sean los requerimientos por el análisis o reportes deseados para la posterior toma de decisión de la respuesta de la demanda eléctrica; por lo tanto, el análisis de big data proveniente de medición inteligente de energía eléctrica no es un problema trivial, requiere una gestión de la información que responda de manera oportuna y escalable y que permita realizar una asignación de recursos con el menor costo, considerando las restricciones de capacidad de cada MDMS, a través de un adecuado gestor de reportes que permita obtener las tendencias del comportamiento del consumo eléctrico.

El algoritmo GDM-Mapreduce toma como entrada un archivo de extensión *csv* que tiene aproximadamente 11Gb de tamaño y lo almacena en la variable denominada *ds*. Esta variable contiene el criterio incorporado de búsqueda; a partir de ese momento todas las referencias de programación se orientan hacia la variable *ds*. Posteriormente, se debe preparar estos datos para ser mapeados; es decir, se debe dejar lista la data (sin espacios en blanco dentro de los campos) para que forme paquetes de datos más pequeños y pueda trabajar de forma oportuna. Se hace uso de las funciones propias de MapReduce para crear los paquetes mediante el método de clave-valor. La *Tabla 2* muestra las variables presentadas por el algoritmo.

Tabla 2. Variables Algoritmo – GDM-Mapreduce

Nombre	Interpretación
<i>Entradas</i>	
Bigdata.csv	Archivo de datos
ds	Variable de datos
Campo1	Campo del archivo de datos
Campo2	Campo del archivo de datos
sumLen	Sumador interno de mapreduce
<i>Salidas</i>	
Outds	Variable de datos
KeyValueStore	Variable interna para formar los paquetes

Una vez formado el paquete (mapeado) se procede con la etapa reduce que se encarga de realizar las operaciones y a su vez guardar los resultados parciales para posteriores cálculos. Este proceso se vuelve a iterar hasta que se termine con todos los datos almacenados en el *ds*. El pseudocódigo del algoritmo GDM-Mapreduce se presenta en la *Figura 3*. Es importante tener en cuenta que las restricciones o criterios de selección que formarán un subconjunto de registros que se requieren analizar, incurrirán en un mayor tiempo de gestión de mapreduce, es decir a más restricciones de búsqueda el algoritmo empleará mayor tiempo de procesamiento y por tanto, mayor consumo de energía empleado en el proceso.

Algoritmo: Gestión de datos Medidos con Mapreduce – GDM-Mapreduce

Entrada: Bigdata.csv, ds, campo1, campo2, sumLen

Salida: outds, KeyValueStore

Paso 1: Carga de datos

ds= Almacenar(bigdata.csv)

Paso 2: Proceso de datos

leer_por_grupos(ds)

Visualizar cabeceras (ds)

ds=Convertir datos en blanco a NaN (bigdata.csv,NA)

ds=Convertir datos faltantes/perdidos a

NaN(bigdata.csv,missingvalue)

Paso3: Clasificación de datos (por clave_valor)

add(intermKV, campo ,ds.campo)

addmulti(intermKV,{campo1 campo2},{ds.campo1 ds.campo2})

Paso 4: Mapeo

Utilizar las funciones hasNext y getNext para recuperar los valores

asociados por una clave única denominada KeyValueStore

While hasNext(sumLenIter)

sumLen = sumLen + getNext(sumLenIter)

End

Paso 5: Reducción

Determinación de variables por cada grupo

outds = mapreduce(ds)

volver a iterar para otro grupo

Go paso 2

Paso 6: Resultados

Agrupación de resultados por grupos

Determinación de resultados globales

Figura 3. pseudocódigo del algoritmo GDM-Mapreduce

4. Análisis de resultados

La construcción de un proceso recursivo en paralelo ha permitido que el tiempo computacional sea reducido en función del tratamiento de la información deseada, de esta manera se estima que existen procesos que pueden tomar un tiempo mayor o menor según el análisis a realizar. Este trabajo ha encontrado como aporte sustancial que un proceso que retarda o minimiza el tiempo de procesamiento se encuentra en la etapa de filtro de la información; por lo tanto, a más restricciones del espacio de búsqueda el tiempo de procesamiento se incrementa. El trabajo ha sido elaborado usando datos del orden de 11GB y procesado bajo entorno de Matlab del cual se obtienen resultados claros y pertinentes para el análisis de la respuesta de la demanda de energía eléctrica como se detallan a continuación en la *Figura 4*.

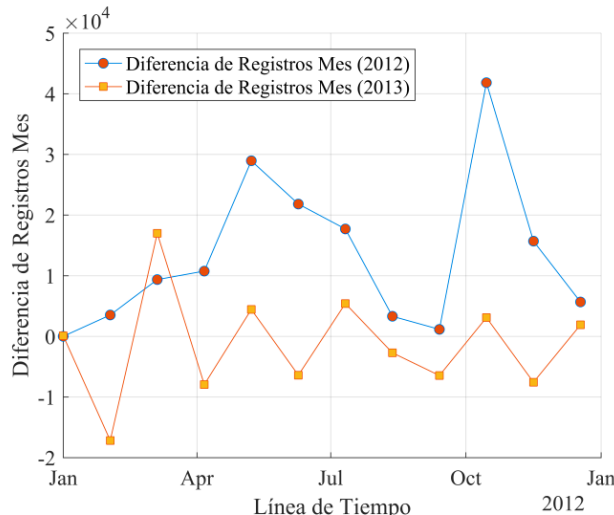


Figura 4. Análisis de variabilidad del consumo eléctrico. Fuente: Autor

En la *Figura 4* se visualiza el comportamiento del consumo promedio en KWh respecto de la variación en la línea del tiempo, obtenidos a partir de la técnica de mapreduce que permite a partir de datos históricos conocer la variación del consumo de energía; esto se puede aplicar a los tipos de clientes de cada empresa de distribución eléctrica y generar un análisis de demanda de energía en función de los datos históricos, de esta manera se consigue una línea de tendencia determinística para la respuesta de la demanda de energía eléctrica, que permitiría reducir los índices de generación eléctrica, al trabajar únicamente sobre la demanda debido a que el costo por energía no consumida es significativo para una empresa generadora.

En la *Figura 5*, la variación de registros que fueron efectuados a los medidores inteligentes para el usuario tipo1, se puede apreciar el patrón de consumo de energía obtenidos de 11GB de información en un proceso de *big data* mediante mapreduce.

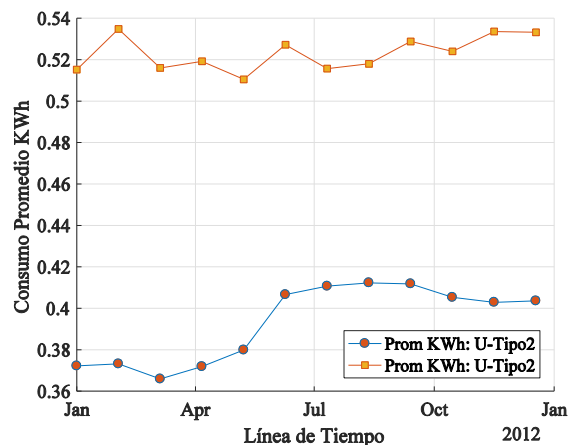


Figura 5. Patrón de consumo de energía. Fuente: Autor

En la *Figura 6* se presenta la variación de registros que fueron efectuados a los medidores inteligentes para el usuario tipo 1 en cada mes. Se puede apreciar que no tiene un comportamiento singular o uniforme en el número de muestras recolectadas, entonces se tiene meses en donde el número de registros fue mayor que otros meses en el mismo año.

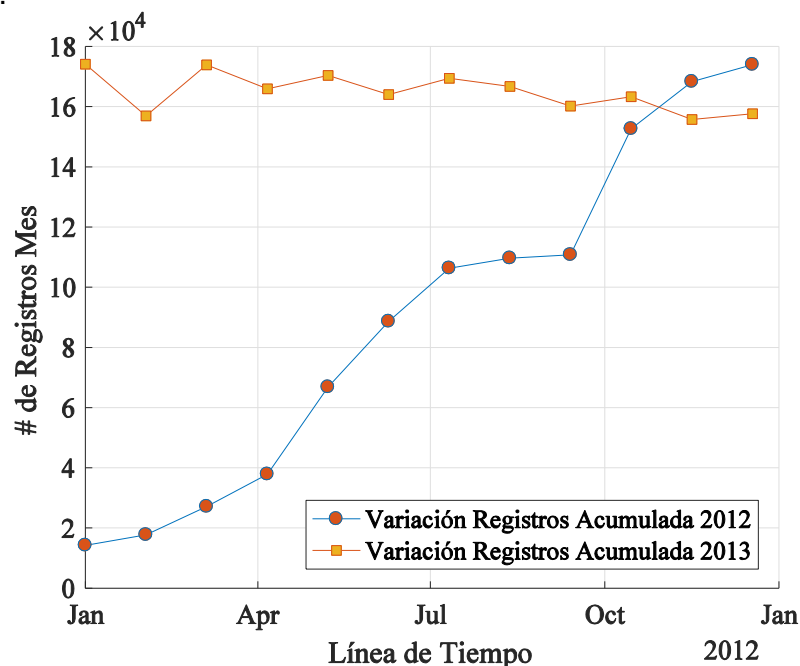


Figura 6. Diferencia de registros. Fuente: Autor

En la *Tabla 3* se presentan el tiempo de procesamiento y el consumo de energía de la etapa mapreduce para una big data de 11GB, obtenidas a partir de un caso de estudio aplicado a la big data de uso libre obtenido del sitio Web London Data Store. Ver (Networks, 2017). En la tabla se observa el tiempo requerido por la técnica mapreduce para explorar y extraer la información de tres tipos de usuarios de energía eléctrica referidos por tipo1 UT1- 2012, tipo2 (UT2) - 2013; adicionalmente, se presenta el consumo de energía en cada proceso P1-2012, P2 – 2013 en watts. El número de registro evaluados fueron 1'074.204 distribuidos de forma no uniforme entre doce meses. En el transcurso se ha identificado que la big data analizada tenía un registro cada 30 segundos, muchos de esos registros tuvieron un valor 0, debido a las fallas del sistema de comunicaciones empleado o por un desperfecto del medidor inteligente. El ordenador utilizado para ejecución del algoritmo tiene un procesador I7, 3.4GHz-Procesador, 8MB-RAM, 19.5V, 3.34-Amp y desarrollado bajo entorno Matlab R2016b.

Tabla 3. Tareas Map Reduce. Fuente: Autor.

		Tarea Map-Reduce					
		CPU Time			Consumo de Energía Watts		
		U	UT ₂		P	P ₂	
		T ₁ 2012	2013		1 2012	2013	
T	T	5.	3.03		9	6	
areas	1	54hrs	58hrs		76.95	51.3	

5. Conclusiones y recomendaciones

Un comportamiento del patrón de consumo eléctrico puede ser analizado a partir de la gestión de los datos históricos almacenados, proveniente de la medición inteligente de energía eléctrica; el patrón de comportamiento permite proyectar de forma determinística el comportamiento de años futuros del consumo de energía eléctrica.

La investigación muestra cómo la técnica de Mapreduce, a través del proceso “divide y vencerás” con una *data* de 11GB de información, logra obtener resultados de grandes cantidades de datos que no se pueden leer en un proceso simple debido a las limitaciones de memoria.

Una técnica como MapReduce aplicada al manejo de *big data* construida por datos históricos de medición inteligente de energía eléctrica presenta una gran opción para mejorar el rendimiento y así reconstruir el patrón de consumo de energía y optimizar el proceso de gestión de la información requerida en la toma de decisiones y respuesta de la demanda de energía eléctrica.

Futuros trabajos plantearán un análisis para optimizar los criterios de búsqueda asociados a la técnica de mapreduce, esto permitirá reducir aún más los tiempos de procesamiento.

Bibliografía

- Ahmed, H. (2014). Applying Big Data Analytics for Energy Efficiency. *Master's Programme in International Design Business Management Hussnain*, 85.
- Arian, M., Ameli, M., Soleimani, V., & Ghazalizadeh, S. (2011). Intelligent migration from smart metering to smart grid, 2 IEEE Power Engineering and Automation Conference. IEEE. <https://doi.org/10.1109/PEAM.2011.6134995>
- Campaña, M., Inga, E., & Hincapié, R. (2017). Optimal placement of universal data aggregation points for smart electric metering based on hybrid wireless. In *CEUR Workshop Proceedings* (Vol. 1950).

- Chihoub, H., & Collet, C. (2016). A Scalability Comparison Study of Data Management Approaches for Smart Metering Systems. *45th International Conference on Parallel Processing (ICPP)*, 474–483. <https://doi.org/10.1109/ICPP.2016.61>
- Correa, E., Inga, E., Inga, J., & Hincapié, R. (2017). Patrón de Consumo Eléctrico basado en el Sistema de Gestión de Datos Medidos usando Técnicas de Big Data. Electrical Consumption Pattern base on Meter Data Management System using Big Data Techniques. In *International Conference on Information Systems and Computer Science* (p. 6).
- Dai, J., Chai, B., Qiu, H., Zhang, B., & Jiang, W. (2016). Probabilistic graphical model based residential energy behavioral analysis on hybrid computing platform. *China International Conference on Electricity Distribution (CICED)*, (Ciced), 1–4. <https://doi.org/10.1109/CICED.2016.7576328>
- Fanibhare, V., & Dahake, V. (2016). SmartGrids: MapReduce framework using Hadoop. *3rd International Conference on Signal Processing and Integrated Networks, SPIN 2016*, 400–405. <https://doi.org/10.1109/SPIN.2016.7566727>
- Grolinger, K., Capretz, M. A. M., & Seewald, L. (2016). Energy consumption prediction with big data: Balancing prediction accuracy and computational resources. *Proceedings IEEE International Congress on Big Data, BigData Congress 2016*, 157–164. <https://doi.org/10.1109/BigDataCongress.2016.27>
- Grolinger, K., Hayes, M., Higashino, W. a., L'Heureux, A., Allison, D. S., & Capretz, M. a. M. (2014). Challenges for MapReduce in Big Data. *Proc. of the SERVICES - IEEE World Congress on Services*, (Services), 182–189. <https://doi.org/10.1109/SERVICES.2014.41>
- Han, W., & Yang, X. (2016). Security, Privacy, and Anonymity in Computation, Communication, and Storage. *Springer*, 10066(March 2017), 1–11. <https://doi.org/10.1007/978-3-319-49148-6>
- Inga-Ortega, E., Peralta-Sevilla, A., Hincapié, R. C., Amaya, F., & Tafur Monroy, I. (2015). Optimal dimensioning of FiWi networks over advanced metering infrastructure for the smart grid. *2015 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, 30–35. <https://doi.org/10.1109/ISGT-LA.2015.7381125>
- Inga-Ortega, J., Inga-Ortega, E., & Gómez, C. (2017). Electrical Load Curve Reconstruction required for Demand Response using Compressed Sensing Techniques. *IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*.
- Inga, E., Céspedes, S., Hincapié, R., & Cárdenas, A. (2017). Scalable Route Map for Advanced Metering Infrastructure Based on Optimal Routing of Wireless Heterogeneous Networks. *IEEE Wireless Communications*, (April), 1–8. <https://doi.org/10.1109/MWC.2017.1600255>
- Inga, E., & Hincapié, R. (2016). Matched Channel Allocation for Advanced Metering Infrastructure based on Cognitive Mobile Virtual Network Operator. *Latin America Transactions, IEEE (Revista IEEE America Latina)*, 14(4), 1780–1785.
- Inga, J., Inga, E., Hincapié, R., & Cristina, G. (2017). Optimal Planning for Deployment of FiWi Networks based on Hybrid Heuristic Process. *Latin America Transactions, IEEE (Revista IEEE America Latina)*, 15(9), 1684–1690. <https://doi.org/10.1109/TLA.2017.8015053>
- Inga, J., Inga, E., Hincapié, R., & Gómez, C. (2016). Evaluación de la Infraestructura de Medición y la Respuesta de la Demanda. *Revista Técnica Energía*, (12), 262–269.
- Lang, W., & Patel, J. M. (2010). Energy management for MapReduce clusters. *Proceedings of the VLDB Endowment*, 3, 129–139. <https://doi.org/10.14778/1920841.1920862>
- Lichtensteiger, B., Bjelajac, B., Müller, C., & Wietfeld, C. (2010). RF Mesh Systems for Smart Metering: System Architecture and Performance. *Smart Grid Communications (SmartGridComm), First IEEE International Conference on*, 379–384. <https://doi.org/10.1109/SMARTGRID.2010.5622071>

- Liu, R., Kuo, C., Yang, C., Chen, S., & Liu, J. (2016). On Construction of an Energy Monitoring Service Using Big Data Technology for Smart Campus. *7th International Conference on Cloud Computing and Big Data*, 81–86. <https://doi.org/10.1109/CCBD.2016.66>
- Liu, X., Golab, L., Golab, W., & Ilyas, I. F. (2015). Benchmarking Smart Meter Data Analytics. *Proceedings of the 18th International Conference on Extending Database Technology*, 385–396. <https://doi.org/10.5441/002/edbt.2015.34>
- Mashayekhy, L., Member, S., Nejad, M. M., & Member, S. (2015). Energy-Aware Scheduling of MapReduce Jobs for Big Data Applications Energy-aware Scheduling of MapReduce Jobs for Big Data Applications. *IEEE Transactions on Parallel and Distributed Systems*, 1–14. <https://doi.org/10.1109/TPDS.2014.2358556>
- Moscoso-zea, O., & Lujan-mora, S. (2017). Metodologías Sugeridas de Evaluación y Selección de Software de Arquitectura Empresarial para la Digitalización del Conocimiento (Suggested Methodologies for Evaluation and Selection of Enterprise Architecture Software for Knowledge Digitization). *Revista Enfoque UTE*, 8(1), 315–328.
- Munshi, A. A., & Mohamed, Y. A.-R. I. (2017). Big data framework for analytics in smart grids. *Electric Power Systems Research*, 151, 369–380. <https://doi.org/10.1016/j.epsr.2017.06.006>
- Networks, U. P. (2017). LONDON DATASTORE. Retrieved from <https://data.london.gov.uk/dataset?format=csv>
- Nghiem, P. P., & Figueira, S. M. (2016). Towards efficient resource provisioning in MapReduce. *Journal of Parallel and Distributed Computing*, 95, 29–41. <https://doi.org/10.1016/j.jpdc.2016.04.001>
- Peralta, A., Inga, E., & Hincapié, R. (2015). FiWi Network Planning for Smart Metering Based on Multistage Stochastic Programming. *Latin America Transactions, IEEE (Revista IEEE América Latina)*, 13(12), 3838–3843. <https://doi.org/10.1109/TLA.2015.7404917>
- Peralta, A., Inga, E., & Hincapié, R. (2017). Optimal Scalability of FiWi Networks Based on Multistage Stochastic Programming and Policies. *Journal of Optical Communications and Networking*, 9(12), 1172. <https://doi.org/10.1364/JOCN.9.001172>
- Selamat, A. (2016). *Distributed Computing and Artificial Intelligence, 13th International Conference*. (Springer, Ed.) (Springer, Vol. 474). Springer. <https://doi.org/10.1007/978-3-319-40162-1>
- Shi, Y., Zhang, K., Cui, L., Liu, L., Zheng, Y., Zhang, S., & Yu, H. (2016). MapReduce short jobs optimization based on resource reuse. *Microprocessors and Microsystems*, 47, 178–187. <https://doi.org/10.1016/j.micpro.2016.05.007>
- Sieverts, P., & Systems, E. (2015). Streamlining Smart Meter Data Analytics. *Technical University Od Denmark*.
- Vaccaro, A., Troiano, L., Vaccaro, A., & Vitelli, M. C. (2016). On-line Smart Grids Optimization by Case-Based Reasoning on Big Data. *IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS)*, (September), 1–6. <https://doi.org/10.1109/EESMS.2016.7504842>
- Wan, Z., Wang, G., Yang, Y., & Shi, S. (2014). SKM: Scalable Key Management for Advanced Metering Infrastructure in Smart Grids. *IEEE Transactions on Industrial Electronics*, 61(12), 7055–7066. <https://doi.org/10.1109/TIE.2014.2331014>
- Wang, J. (2014). *Frontiers in Algorithmics*. Springer. <https://doi.org/10.1007/978-3-319-19647-3>
- Yildiz, O., Ibrahim, S., Phuong, T. A., & Antoniu, G. (2015). Chronos: Failure-aware scheduling in shared Hadoop clusters. *Proceedings of the IEEE International Conference on Big Data, IEEE Big Data 2015*, 313–318. <https://doi.org/10.1109/BigData.2015.7363770>
- Zhang, B. (2017). Analysis of residential electricity consumption behavior based on the improved Apriori algorithm. In C. P. L.: C. Press (Ed.), *Electromechanical Control*

Technology and Transportation : Proceedings of the 2nd International Conference on Electromechanical Control Technology and Transportation (1st ed., pp. 145–149). Zhuhai, China: CRC Press LLC : CRC Press.

Zhou, J., Hu, R. Q., & Qian, Y. (2012). Scalable distributed communication architectures to support advanced metering infrastructure in smart grid. *IEEE Transactions on Parallel and Distributed Systems*, 23(9), 1632–1642. <https://doi.org/10.1109/TPDS.2012.53>