



# Solución con Pentaho para análisis de datos de la planta fotovoltaica de estudio del CIES

## Solution with Pentaho for data analysis of the CIES study photovoltaic plant

Yasser Quintana Casulo<sup>1</sup>  
Roger Anner Proenza Yero<sup>2</sup>  
Luis Vázquez Seisdedos<sup>3</sup>

<sup>1</sup>Departamento de Informática. Universidad de Oriente. yasser.quintana@uo.edu.cu

<sup>2</sup>Centro de Investigaciones de Energía Solar. rproenza@cies.cu

<sup>3</sup>Departamento de Automática. Universidad de Oriente. lvazquez@uo.edu.cu

## Resumen

Un problema que se enfrenta hoy en día en los parques fotovoltaicos es la incapacidad de poder realizar pruebas para diagnósticos y análisis ante fallos de cualquier índole de la energía solar fotovoltaica por este motivo con la colaboración del Centro de Investigaciones de Energía Solar de la provincia de Santiago de Cuba se obtienen datos en cuanto a mediciones de variables eléctricas como la energía activa entregada, la energía inductiva y capacitiva reactiva, la tensión alterna y corriente efectiva, la tensión y la corriente en su punto de máxima potencia, también la potencia entregada y en ambientales se consiguen la humedad, la irradiación solar, la velocidad del viento, la temperatura ambiental y del módulo fotovoltaico que permitieron extraer información útil para observar comportamientos de los paneles e inversores. La información fue extraída de diferentes archivos Excel y cargadas al almacén de datos para así mantener un almacenamiento histórico para la toma de decisiones. La metodología empleada fue HEFESTOS y se utilizaron las herramientas de la Suite Pentaho.

**Palabras clave:** energía solar fotovoltaica, Pentaho, almacén de datos, inteligencia de negocios.

## Abstract

*A problem that is faced today in the photovoltaic parks is the inability to be able to carry out tests for diagnoses and analysis in case of failures of any kind of solar photovoltaic energy for this reason with the collaboration of the Solar Energy Research Center of the province of Santiago de Cuba data are obtained in terms of measurements of electrical variables such as active energy delivered, inductive and reactive capacitive energy,*



*alternating voltage and effective current, voltage and current at its point of maximum power, also the power delivered and In environmental studies, humidity, solar irradiation, wind speed, environmental temperature and the photovoltaic module were obtained, which allowed extracting useful information to observe the behavior of the panels and inverters. The information was extracted from different Excel files and loaded into the data warehouse in order to maintain a historical storage for decision making. The methodology used was HEFESTOS and the tools of the Pentaho Suite were used.*

**Keywords:** *photovoltaic solar energy, Pentaho, data warehouse, business intelligence.*

## Introducción

El peligro global y la crisis energética se ha incrementado en pocas décadas por el uso de fuentes no renovables de energía las cuales provocan un gran impacto en las políticas económicas, condiciones climáticas y seguridad energética; debido a esto se ha promovido el uso y desarrollo de alternativas de energías limpias para el reemplazo de la producción de energía actual (Kumar, Majumder y Nayak, 2018) (Sahu, 2015). La estructura energética mundial no es sustentable y se requiere dar un cambio de paradigma energético basado en la eficiencia energética y con el uso de las fuentes renovables de energía (FRE) para así poder controlar y mitigar los daños ocasionados por el uso de fuentes fósiles por varias décadas (Estrada, 2013).

En realidad, el incremento en el consumo de electricidad procedente de fuentes de energía renovables no convencionales, es una tendencia que únicamente se manifiesta allá donde han sido emprendidas políticas energéticas encaminadas a eliminar barreras de entrada a los mercados eléctricos a las nuevas tecnologías. Las medidas emprendidas han hecho posible activar inversiones de manera sostenida y recurrente en esta actividad económica, y en la generalidad de los casos estas medidas no han incluido apoyos a los precios de la generación eléctrica de estas fuentes (FuturEnergy, 2017).

En los últimos años una de estas fuentes que ha captado la atención es la energía solar fotovoltaica (ESF) capaz de brindar ventajas ambientales y un desarrollo energético sostenible. Sin embargo, los parques fotovoltaicos tienen un régimen de generación intermitente, una imposibilidad del control y limitado conocimiento de la producción eléctrica, debido a su alta dependencia de las condiciones climáticas.

La amplia gama de datos que influyen y son adquiridos a través de los sensores e inversores pertenecientes a la planta de estudio día tras día nos lleva a la paradoja de que “cada vez se tiene más información y menos tiempo para analizarla”. Esta problemática conlleva al término tan usado en la actualidad de inteligencia de negocios (IN), el que se puede considerar todo un proceso continuo que intenta explorar y analizar la información estructurada, persiguiendo un objetivo concreto para mejorar la toma de decisiones, lo que se traduce en una ventaja competitiva. El primer paso para poder elegir un tipo de solución de IN y una herramienta que nos proporcione las respuestas a todas estas preguntas, consiste en entender, que lo que buscamos, es usar la información de nuestro negocio con el objetivo de mejorarlo (Goyzueta, 2015). En el desarrollo de estas soluciones una de las herramientas más usadas es Pentaho por las funcionalidades que brinda y por ser de código abierto líder del mercado nacional e internacional.

El Grupo experimental del Centro de Investigaciones de la Energía Solar (CIES) tiene definido un proyecto de instalación e investigación de la generación de ESF, en el cual se pretende conocer, el comportamiento de las variables eléctricas y meteorológicas involucradas en la cuantificación de energía eléctrica



y detección de fallas de sistemas fotovoltaicos. Además de cómo influyen su relación en la demanda de energía para la toma de decisiones por parte del equipo de Energía Solar de la empresa.

## Materiales y métodos

### Componentes de la planta de estudio

Desde el punto de vista de hardware la planta como se muestra en la Figura 1 esta provista de 30 módulos fotovoltaicos monocristalinos de fabricación española los cuales están basados en secciones de una barra de silicio perfectamente cristalizado en una sola pieza. Donde la máxima potencia por módulo es de 250 W para una entrega total de 7500 W.

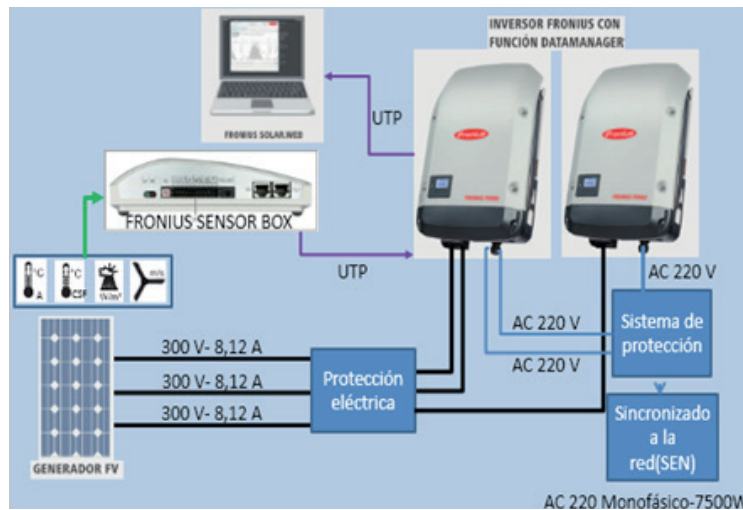


Figura 1. Esquema general de funcionamiento

Además, cuenta con 2 inversores Fronius Primo 5.0-1 que manejan hasta 5 kW de potencia de salida. Este es un inversor compacto monofásico sin transformador ideal para aplicaciones residenciales y su diseño está basado en el sistema de instalación SnapINverter, el cual permite instalaciones y reparaciones sencillas y seguras.

El Fronius Primo tiene características únicas como dos seguidores de máxima potencia, alta tensión de sistema, un amplio rango de voltaje de entrada y puede instalarse en interior y exterior. Como funciones estándar incluye interfaces Wi-Fi y SunSpec Modbus para monitoreo y datalogging a través de la plataforma de Solar.web. Está diseñado para adaptarse a requerimientos futuros, por lo cual ofrece una solución completa a los cambios de normativas e innovaciones técnicas del mañana (Fronius, 2016).

También posee un Fronius Sensor Box o caja de sensores Fronius, este ofrece entradas para un total de seis señales de mediciones, la cual tiene entradas analógicas y digital (Fronius, 2014)

### Variables de la planta de estudio

La planta de estudio arroja variables eléctricas del inversor y ambientales de la caja de sensores, entre las eléctricas están la energía activa entregada, la energía inductiva, la energía capacitiva reactiva, la

tensión alterna y corriente efectiva, la tensión y la corriente en su punto de máxima potencia (mpp) y la potencia entregada. Entre las variables ambientales están la irradiación solar, la humedad, la velocidad del viento, la temperatura ambiental y la temperatura del módulo fotovoltaico.

### **Softwares empleados en la adquisición de datos**

Las instalaciones fotovoltaicas pueden ser monitorizadas, analizadas y comparadas de manera rápida y sencilla a través de diferentes softwares realizados por la Fronius. En los cuales el usuario puede acceder a la información actualizada de la instalación en cualquier momento y presentarla de forma muy clara.

El Fronius Datalogger Web es la pieza clave para la completa supervisión de los equipos Fronius. Puede integrarse de forma sencilla en redes Ethernet, a través de su interfaz LAN o mediante el stick WLAN opcional. Junto con el software Fronius Solar.access, ofrece todas las posibilidades de configuración y gestión de los datos del equipo fotovoltaico en el PC. Además, a través de su interfaz web, varios usuarios tienen acceso a la información más importante del equipo al mismo tiempo, mediante un navegador independiente del sistema operativo. Con el servicio online de Fronius Solar.web se puede acceder a los datos en tiempo real de una instalación fotovoltaica sin tener que efectuar complejas actividades de configuración (Fronius, 2015).

El Fronius Solar.web presenta un portal muy fácil de usar y tiene una amplia gama de opciones para el análisis de datos. Cuenta con una variedad de herramientas y funciones. Realiza comparaciones automáticas de rendimiento entre varios inversores o periodos de tiempo, comparación de los datos del sistema con los datos del sensor, detecta automáticamente incidencias en el campo fotovoltaico, tiene una sencilla monitorización del autoconsumo y es gratuito (Fronius, 2017).

El Fronius Solar.access es también un software gratuito y es utilizado para evaluar y archivar datos en el PC. Como herramienta de administrador ofrece todas las posibilidades de configuración de todos los componentes del equipo en Solar Net (Fronius, 2014).

### **Inteligencia de negocios**

La inteligencia de negocios (o Business Intelligence) satisface dicha necesidad. Hans Peter Luhn (1958), investigador de IBM, dió origen al concepto de IN como “la habilidad de aprehender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada”. Fue Howard Dresden (1989), analista de Gartner, quien propone la definición de IN como “conceptos y métodos para mejorar las decisiones de negocios mediante el uso de sistemas de soporte basados en hechos”. En la actualidad, la evolución hacia las sociedades más tecnológicas implicó reformula la definición. Se entiende por IN al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización (Pascal, Servetto, Mirasson y Luna, 2015).

### **Metodología HEFESTO**

El principal objetivo de la metodología HEFESTO es facilitar el arduo trabajo que significa construir un AD desde cero, aportando información que permitirá mejorar su performance. La metodología está orientada a amortiguar el tedio que provoca seguir pasos sin comprender el porqué de su ejecución. La



metodología tiene como punto de partida la recolección de requerimientos y necesidades de información de los usuarios y concluye con la confección de un esquema lógico y sus respectivos procesos de extracción, transformación y carga de datos como se observa en la Figura 2.



Figura 2. Ciclo de metodología de desarrollo de AD Hefesto v3.0

Además, cuenta con las siguientes características:

- » Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- » La piedra fundamental la constituyen los requerimientos de los usuarios, por lo cual, se adapta con facilidad y rapidez a los cambios del negocio.
- » Reduce drásticamente la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tomen decisiones respecto al comportamiento y funciones del almacén de datos, y además expone resultados inmediatos.
- » Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- » Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- » Es independiente del software/hardware que se utilicen para su implementación.
- » Cuando se culmina con una fase, los resultados obtenidos se constituyen en la entrada de la fase siguiente.
- » Se aplica en almacenes de datos y en mercados de datos. (Bernabeu y García, 2017)

### Herramientas empleadas

La suite Pentaho se presenta en dos versiones, una versión corporativa (Enterprise Edition) orientada a las empresas la que es comercializada a altos costos y otra versión comunitaria de código abierto (Community Edition) orientadas al ámbito académico que se distribuye sin costo alguno. Pentaho incluye herramientas para ETL (Extracción, transformación y carga de datos, ETL del inglés: Extract, Transform

and Load), análisis OLAP (Procesamiento analítico en línea, OLAP del inglés: On-line Analytical Processing), los metadatos, la minería de datos, generación dinámica de reportes y cuadros de mando o tableros digitales (Desten, Téllez y Vázquez, 2014).

La plataforma usada para la creación de consultas y ejecución de tareas fue Pentaho Data Integration en su versión 8. Esta herramienta posee un planificador de tareas para la ejecución automática y unión de estas, además, permite la conexión a varios gestores de base de datos, para luego transformarlos y llevarlos a un formato factible en un almacén de datos, y así sean analizados.

La aplicación web empleada fue el Pentaho BI Server en su versión 6.0 la cual cuenta con una interfaz de usuario que permite analizar todas las informaciones. Además de proveer la arquitectura y la infraestructura necesaria para crear soluciones de IN. El visor OLAP empleado fue el Saiku en sustitución del JPivot, desarrollado para código abierto la idea era mejorar la experiencia del usuario de su interfaz e interacción con el mismo mediante el aprovechamiento de las bibliotecas de interfaz de usuario libre y las tecnologías como jQuery y Ajax.

El gestor de base de datos usada para almacenar toda la información histórica fue PostgreSQL en su versión 9.5 que funciona muy bien con grandes cantidades de datos y una alta concurrencia de usuarios accediendo a la vez al sistema.

## **Resultados y discusión**

La creación del almacén viene dada por el empleo de diferentes softwares para el manejo de los datos y al emplear un almacén de datos poder centralizar en una misma base de datos toda esta información. Información que es adquirida en dos archivos uno con extensión CSV proveniente de los inversores los cuales a los 1000 días comienzan a borrar la información de las variables eléctricas del Datalogger Card y del Fronius Solar Access que me ofrece las variables ambientales y los datos de potencia en otro archivo con extensión XLS.

### **Modelo del almacén de datos**

Con la creación del AD se tienen las condiciones para la realización de cálculos estadísticos y análisis de los datos. El mismo está estructurado por 14 tablas: 10 dimensiones y 4 tablas de hechos. Para que los datos puedan ser analizados a través de la creación de consultas se necesita una estructura de datos determinada, estos son los cubos multidimensionales, los cuales permiten el acceso flexible a los datos, para explorar y analizar sus relaciones, y consiguientes resultados. A continuación, se describe el proceso ETL y la estructura e implementación del cubo multidimensional dentro del Pentaho.

### **Extracción de los datos**

Se establecen las correspondencias entre el modelo conceptual creado y las fuentes de datos, además se definen los campos que se incluirán en cada perspectiva, obteniendo así un modelo conceptual ampliado como en la Figura 3, en este caso seguimos el ejemplo de Datos Potencia (las informaciones del resto de las tablas están insertadas de forma manual en la base de datos).



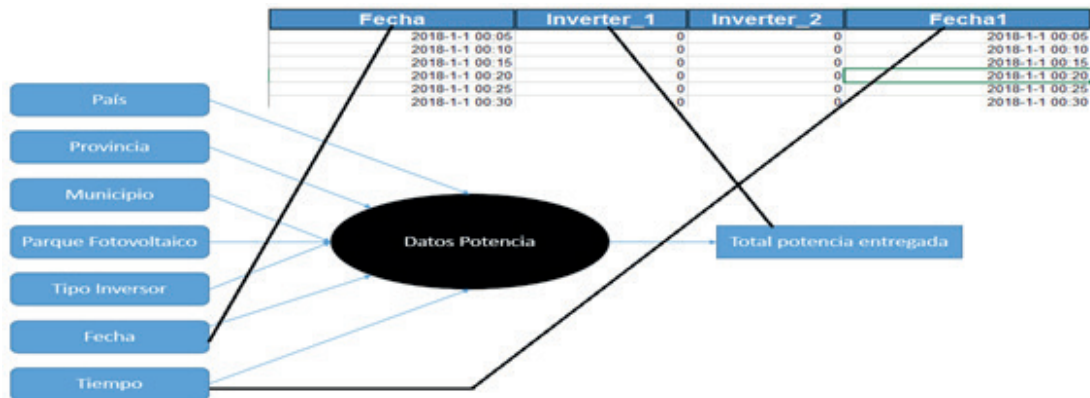


Figura 3. Modelo conceptual ampliado de Datos Potencia

En la implementación del proceso ETL se cargan los datos de las fuentes externas que se emplean para la adquisición de datos, hacia la base de datos multidimensional. Este proceso tiene tres funciones específicas: extracción, transformación y carga como se observan en la Figura 4 y Figura 5.



Figura 4. Dimensión Provincia

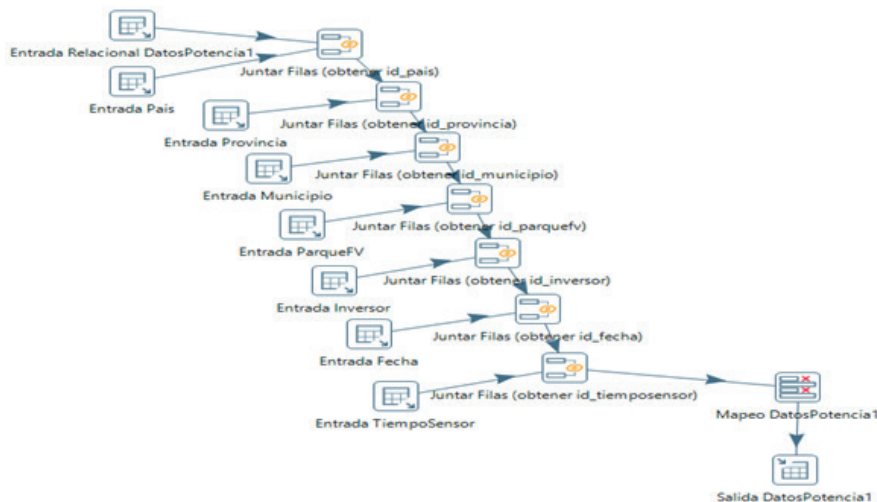


Figura 5. Hecho Datos Potencia

## Proceso OLAP

Luego de terminar el proceso ETL, se conforma el cubo de datos, este es fundamental para realizar la navegación por los datos con el fin de extraer la información deseada de forma rápida y eficiente. Los cubos se diseñan en el Pentaho BI Server en estrella por condición previa de la aplicación como se muestra en la Figura 6 tiene dimensiones globales como son país, provincia, municipio y parque fotovoltaico ya que las mismas son utilizadas en varios de los cubos. Estos cubos son los que representan cada tabla de hecho del almacén de datos, los mismos al ser creados desde el esquema base se les crea la relación con las dimensiones que le corresponden y las respectivas jerarquías para cada una. En el caso de las dimensiones globales solo necesitan ser invocadas dentro del cubo especificando el id y el nombre.

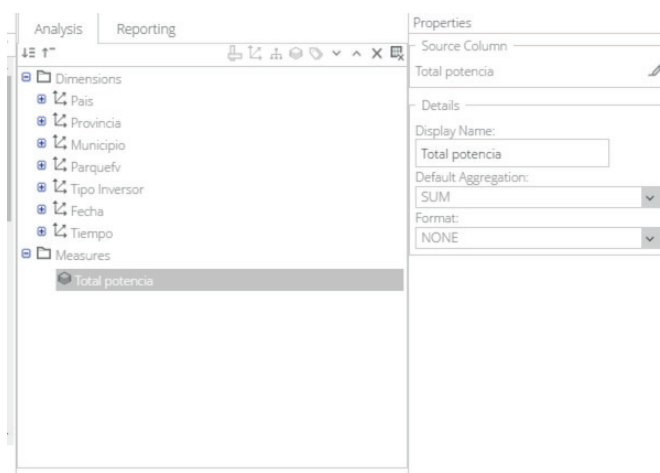


Figura 6. Cubo OLAP

## Despliegue en Pentaho BI Suite

Para realizar análisis en la Suite Pentaho se ejecuta el biserver-ce y en el navegador se inserta la url; la primera interfaz es la de autenticación como se muestra en la Figura 7 en la cual el usuario administrador accede a realizar los análisis posteriormente en el Saiku Analytics o JPivot como se muestra en la Figura 8.

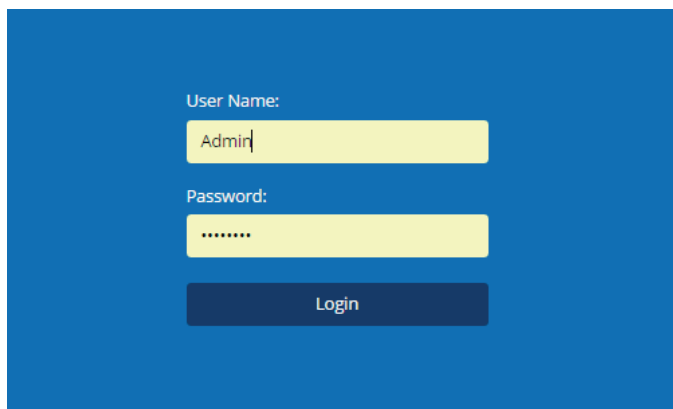


Figura 7. Autenticación de la Suite Pentaho



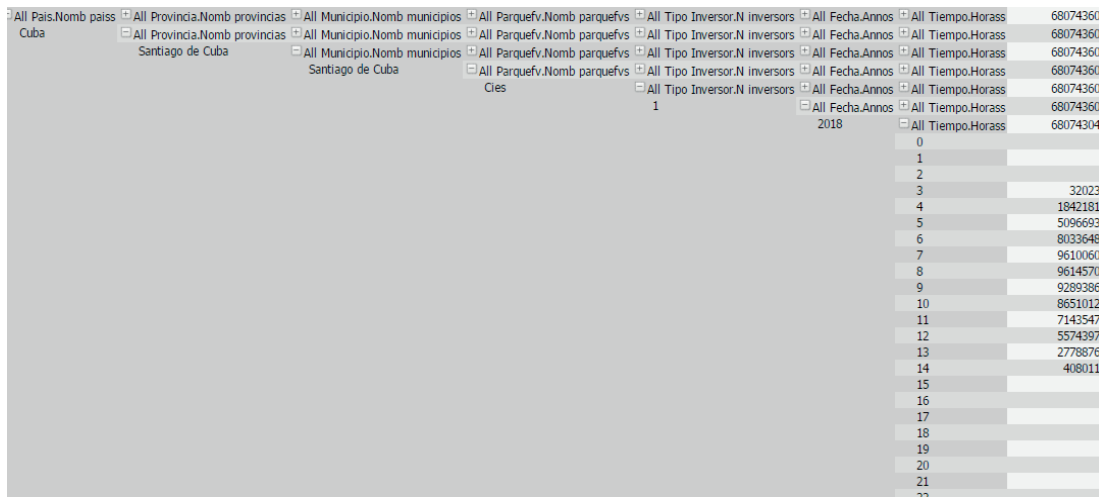


Figura 8. Análisis de datos potencia

Además de la posibilidad de obtener gráficamente información necesaria para los operadores de energía solar los comportamientos de las variables eléctricas y ambientales en la detección de fallas o de pérdidas de energía, como se muestran en las Figuras 9 y 10.

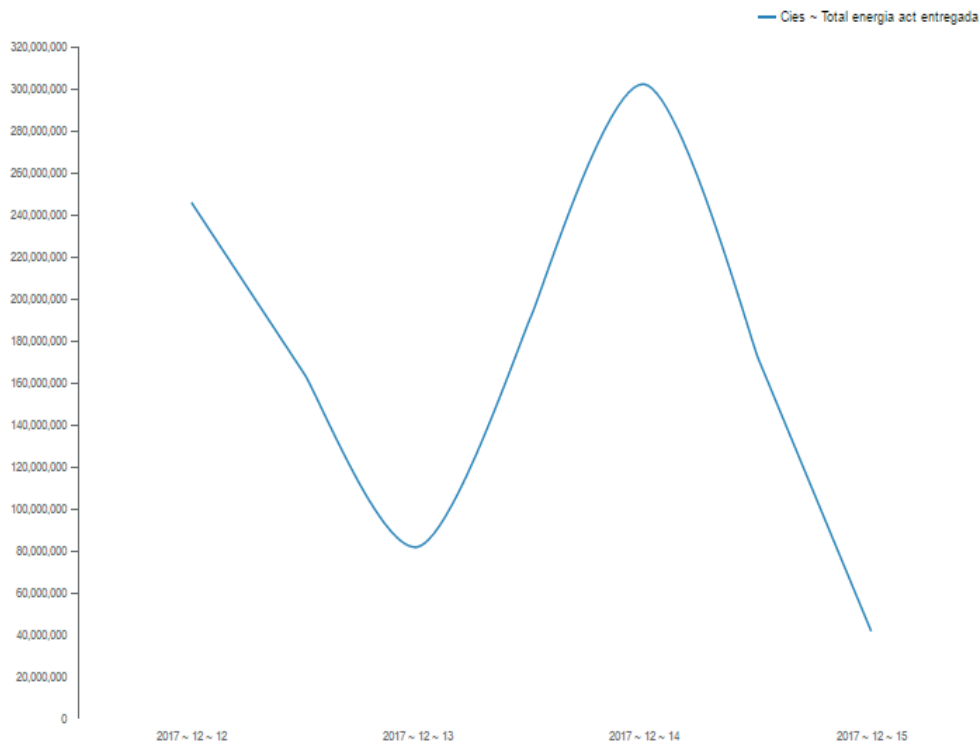


Figura 9. Total de energía entregada entre el 12 de diciembre y el 15 de diciembre del 2017

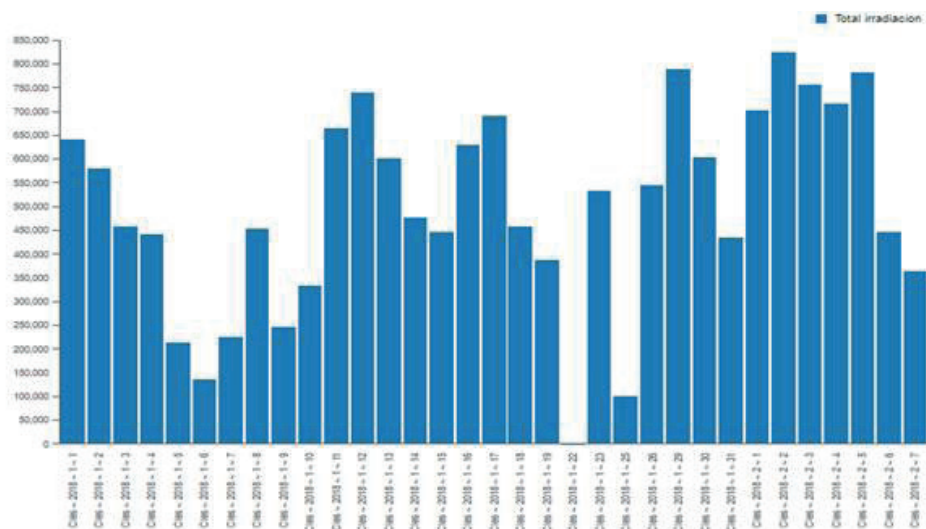


Figura 10. Total de irradiación entre el 1 de enero y el 7 de febrero del 2018

## Conclusiones

El almacén de datos creado permitirá el uso eficiente de la información al estar centralizada, con posibilidades de su explotación al haberse realizado con software libre y su posible expansión hacia otros parques con posibilidades de centralización de los datos a través del sistema gestor de base de datos de dicho almacén de datos.

## Agradecimientos (Opcional)

Se desea agradecer a la contribución de los trabajadores del CIES por el trabajo conjunto que venimos realizando. Al grupo Electricidad eficiente aprovechando las Fuentes Renovables de Energía (EREN, Energías Renovables) de la Facultad de Ingeniería Eléctrica de la Universidad de Oriente en Santiago de Cuba.

## Referencias

- Bernabeu, R. y García, M. (2017). *Hefesto Data Warehousing: guía completa de aplicación teórico – práctica; metodología Data Warehouse*. Argentina: Córdoba.
- Desten, J.C., Téllez, R. y Vázquez, J.L. (noviembre, 2014). Sistema de georreferenciación de información para Pentaho sin conexión a internet. Trabajo presentado en la 17 Convención de Ingeniería y Arquitectura, La Habana, Cuba.
- Estrada, C. A. (octubre, 2013). Transición energética, energías renovables y energía solar de potencia. *Revista Mexicana de Física*, 59(2), 75-84. Recuperado de <http://www.redalyc.org/articulo.oa?id=57030971010>
- Fronius. (2014). DATCOM. Recuperado de <http://www.fronius.com>

- Fronius. (2015). Datalogger Web: Muchas posibilidades para una supervisión completa de equipos FV. Recuperado de <http://www.fronius.com>
- Fronius. (2016). Primo 3.0-1 / 3.5-1 / 3.6-1 / 4.0-1 / 4.6-1 / 5.0-1 AUS / 5.0-1 / 6.0-1 / 8.2-1. Austria. Recuperado de <http://www.fronius.com>
- Fronius. (2017). Datamanager 2.0. Recuperado de <http://www.fronius.com>
- FuturEnergy. Desafíos y oportunidades de las energías renovables en Latinoamérica. (15 de marzo del 2018). Recuperado de <http://www.futureenergyweb.es>
- Goyzueta, S. I. (2015). Big Data Marketing: una aproximación. *Perspectivas*, 18(35), 147-158. Recuperado de <http://www.redalyc.org/articulo.oa?id=425941232007>
- Kumar Behera, M.; Majumder, I.; Nayak, N. (2018). Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Engineering Science and Technology, an International Journal*, 21, 428 - 438. doi: 10.1016/j.jestch.2018.04.013
- Pascal, G., Servetto, D., Mirasson, U. L. y Luna, Y. (2017). Aplicación de Business Intelligence para la toma de decisiones en Instituciones Universitarias. Implementación de Boletines Estadísticos en la Universidad Nacional de Lomas de Zamora (UNLZ). *Revista Electrónica sobre Tecnología, Educación y Sociedad*, 4(7). Recuperado de <http://ctes.org.mx/index.php/ctes/article/download/659/751>
- Sahu, B.K. (2015). A study on global solar PV energy developments and policies with special focus on the top ten solar PV power producing countries, *Renew. Sustain. Energy Rev.*, 43, 621–634.

