

Design of a Technology Management Infrastructure for Large Volumes of Data in an Intelligent Power Network

Guillermo Escobedo Briones, Norma E. Jácome Grajales, Gustavo Arroyo-Figueroa

National Institute of Electricity and Clean Energies, Cuernavaca, Morelos, Mexico

{gescobedo, nejacome, garroyo}@iee.org.mx

Abstract. In Mexico, high - priority energy generation target, associated with clean energy sources, have been established through the promotion of other abundant natural resources such as solar or wind energy. In addition to this, there is a global trend supported by the development of renewable energies that is leading the electrical systems to integrate distributed generation capacities. This new operation perspective of the electrical network is called: "Smart Grid". The objective of this research is to present a technological infrastructure for the management of large volumes of information through Big Data tools to support the integration of renewable energy. The infrastructure includes a methodological architecture for the acquisition, processing, storage, management, analysis, monitoring and forecast of large amounts of data. The development of a Big Data application for the analysis and monitoring of the information generated by photovoltaic systems is included as a case study. The goal is to have timely information to make better decisions to improve the integration of renewable energy in the Smart Grid.

Keywords: Big Data, Smart Grid, Renewable Energy, Distributed Generation, Photovoltaic Systems, Electric Power Utility, Information Systems, Data Analytics.

1 Introduction

As part of efforts to promote energy efficiency and sustainability and reduce dependence on fossil fuels as a primary energy source, Mexico has increased its efforts to promote the use of renewable energy sources and clean technologies to generate electricity. However, this effort is just beginning. It has a legal goal to increase the percentage of non-fossil energy in the portfolio of primary energy sources for electricity generation by at least 35% by 2024. There is also a global trend, supported by the development of renewable energies, cogeneration in situ and the development of communication infrastructure and information technology, which is leading electrical systems to integrate capabilities of distributed generation in small capacities, but in a lot of points. This is a phenomenon that is changing the way in which electrical networks are designed and operated. This forces to consider a

wide and competitive portfolio of projects, as well as bold steps to achieve this legal goal, which includes removing barriers and promoting renewable energy sources [1].

The growing demand for clean and sustainable electricity and the strong dependence of modern life to have a high quality power supply, reliable and continuous, has forced conventional electric systems to incorporate distributed generation with renewable energy sources. Here are some reasons that gave rise to the concept of Smart Grid (SG). In addition, the high level of incident solar radiation in Mexico is somehow favoring the installation of Photovoltaic Systems interconnected with the power grid (PG) for the production of electricity at the point of consumption. Likewise, given the benefits that the user and the power grid potentially have by the use of this technology, there will be more users of electrical service motivated to invest in it [2].

Moreover, considering that the information technology and communications play an important role in the use, transfer of knowledge and clean energy supply, Big Data is a trend in Mexican companies, in order to obtain value of data and maintain a competitive edge in the business. Big Data will allow to process and analyze large amounts of information in order to generate projects that benefit organizations to better understanding of their environment, innovate and improve their products or processes and define new business models.

Therefore, this work integrates, processes and analyzes information generated by photovoltaic cells sensors with Big Data, concentrating the necessary concepts for the understanding and implementation of projects related to the subject. In the same way, it seeks to create an infrastructure or platform that can be used in SG future projects.

2 Challenges of the Traditional Electrical System

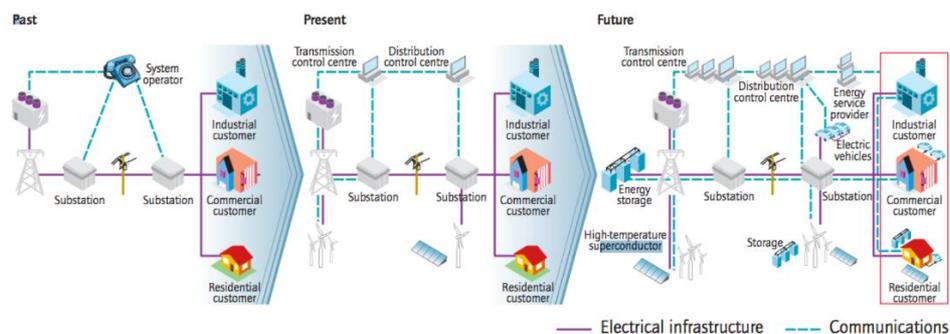


Fig. 1. IEA Smart Grid [4].

The traditional power grid requires changes in terms of restructuring, with the establishment of an electricity market that encourages competition and efforts to ensure energy efficiency in a context of environmental sustainability. These challenges have led

governments, electricity companies and scientific community to find solutions that enable efficient, reliable and responsible energy use, appealing to an optimal and flexible design of the electrical network. This modernization of the electricity network is known as the Smart Grid [3].

Figure 1 depicts a smart grid taken from the Smart Grid Technology Roadmap of the International Energy Agency (IEA) published in 2011. In this figure the evolution experienced by the electricity sector is shown. The vision corresponds to an electrical network in constant development, in real-time, with a two-way flow of energy and information, between the power generator, the network operator and the end consumers.

3 Integration of Distributed Generation Through Renewable Resources in the Smart Grid

The general objectives of SG are reducing the negative impact on the environment, reduce dependence on non-renewable natural resources and increase energy security through the diversification of distributed generation sources with renewable energies. SG technology can significantly reduce the barriers to large-scale integration of renewable resources in the electrical network [5].

The IEA estimates, in their study "New Policies Scenario", that globally, the deployment of renewable resources can save about 4.1 Gigatons of CO₂ emissions in 2035 in contrast with the emitted in 2010 [6].

The main challenge lies in the amount, type and location where renewable resources can be interconnected with the transmission and distribution systems and end consumers. This integration task of distributed generation from renewable resources in the SG requires careful and systematic approach.

Distributed generation technologies can be divided into two big groups: those that use fossil fuels and those that use renewable energy resources. In the first group there are internal combustion engines, micro gas turbines, Stirling engines and fuel cells. In the second, it can be found the generation with biomass, wind turbines, tidal power, geothermal and photovoltaic cells [7].

In this work, it is consider the analysis and monitoring of the information generated by photovoltaic cells.

4 Operation of a Photovoltaic System

There are two types of photovoltaic systems, autonomous systems and interconnected with (PG) power supply systems. It is important to note that for purposes of this work only PGs will be considered.

Figure 2 illustrates the operation of a PG. Sunlight is converted into direct current electric power by solar cells. This direct current passes through the different components until the

investor turns it into alternative, which can provide domestic consumption or be sending directly to the mains.

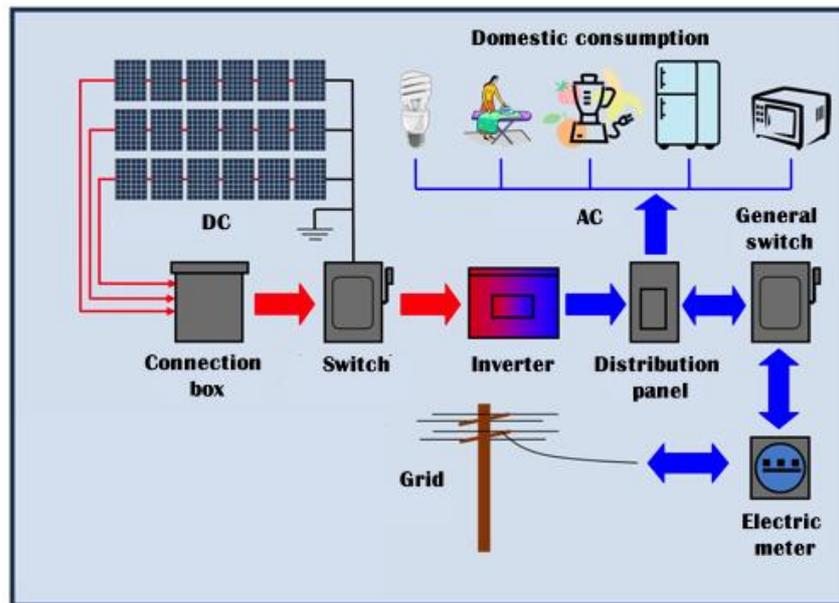


Fig. 2. Operation of a PG [8].

5 Handling Large Volumes of Data from Distributed Generation Sources in the Smart Grid

To face the enormous challenge of handling large volumes of data from distributed generation sources in the SG, there are different technologies that fall under the term Big Data. These technologies allow obtaining of value through the intelligent interpretation of data. However, these strategies to obtain that potential for smart grid in Mexico are still at an early stage of development.

Big Data will enable to process and analyze large amounts of information in order to generate projects that benefit the institutions and organizations to achieve a better understanding of their environment, innovate and improve their products or processes and define new business models.

In our modern and digital society, create digital data at anytime and anywhere has become an unnoticed everyday task. The digitalization era has led to an explosion of data. As an example: it is estimated that since the beginning of mankind until 2003, it has been generated about 5 million gigabytes of data. In 2011, the same amount of data was created

almost every two days. By 2013, the same amount of data was created every 10 minutes [9]. Hence the need to convert data into information to generate knowledge as a platform to make better decisions.

This work integrates, processes and analyzes information generated by photovoltaic systems with Big Data, concentrating the necessary concepts for the understanding and implementation of projects related to the subject, in the same way it seeks to create a technologic platform that can be used in future SG projects.

6 Technologic Platform

6.1 Defining Big Data Architecture

According to Mysore, Khupat, & Jain [10] Big Data architecture consists on different dimensions and logical layers. Thus, the layers provide an approach of components of the organization with specific functions. In Figure. 3 Big Data architecture developed in this paper is presented.

6.2 Analysis of Hardware and Software Infrastructure for Big Data

With the emergence of the Big Data concept, it has been created a variety of tools and products to support the management and exploitation of information. Therefore, the tool selection process becomes a key factor for the viability and success of a Big Data solution. Technologies and tools of Big Data can be classified according to their functions and features:

- Processing. Technologies dedicated to mass data processing.
- Real time data acquisition technologies (streaming).
- Technologies for information storage.
- Data visualization and analytics.
- Tools for statistical development.
- Extraction, transformation and loading (ETL) tools.
- Frameworks, which include an integration of different components for acquisition, processing, analysis and visualization of data.

In Figure 4 an overview map of Big Data tools classified according to their functionality is shown. Open source tools are included [11].

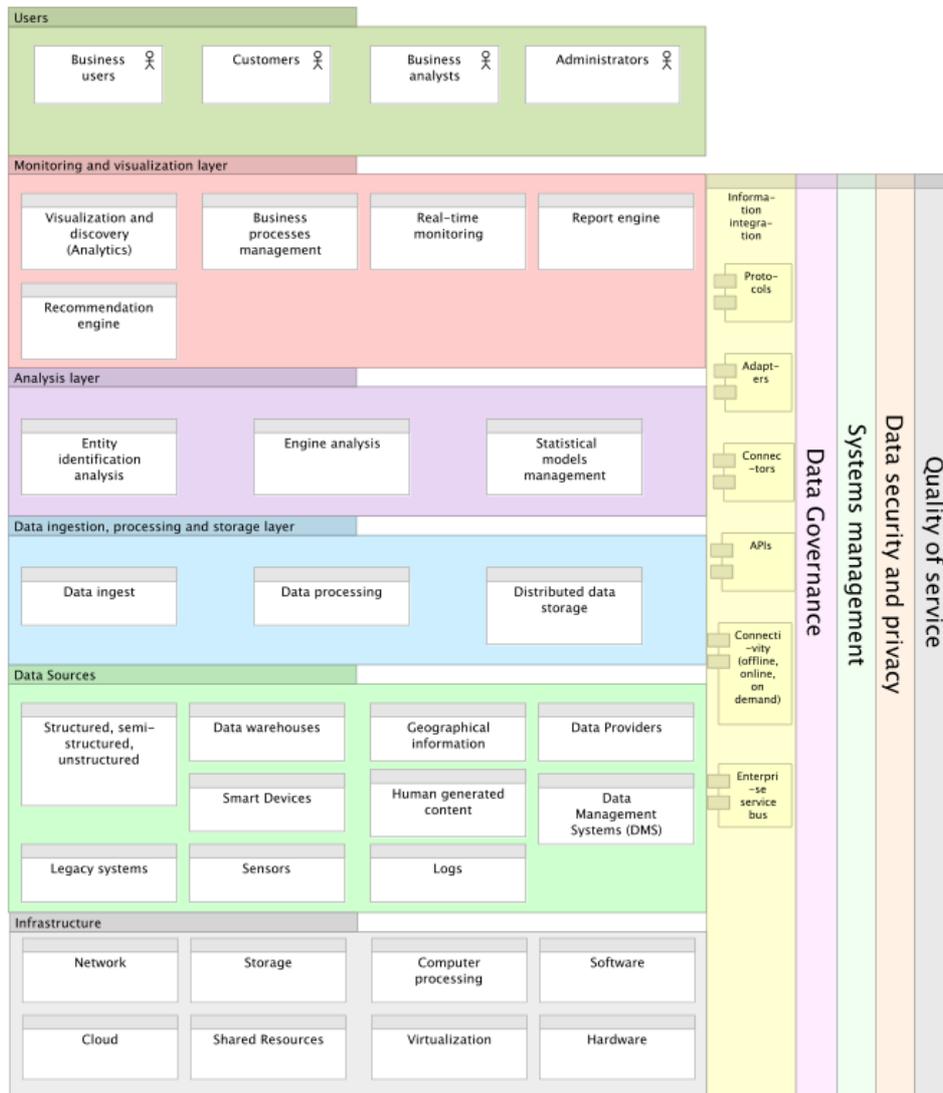


Fig. 3. Components of the logical and vertical layers.

For the selection and evaluation of the required platform. For this work the following criteria was considered:

- Open source platforms.
- Free Software Licenses.

- Real-time processing.
- Distributed processing.
- Distributed storage.
- Scalability.
- Simplicity in writing programs.
- Easy integration with other tools.
- Hardware Resources.

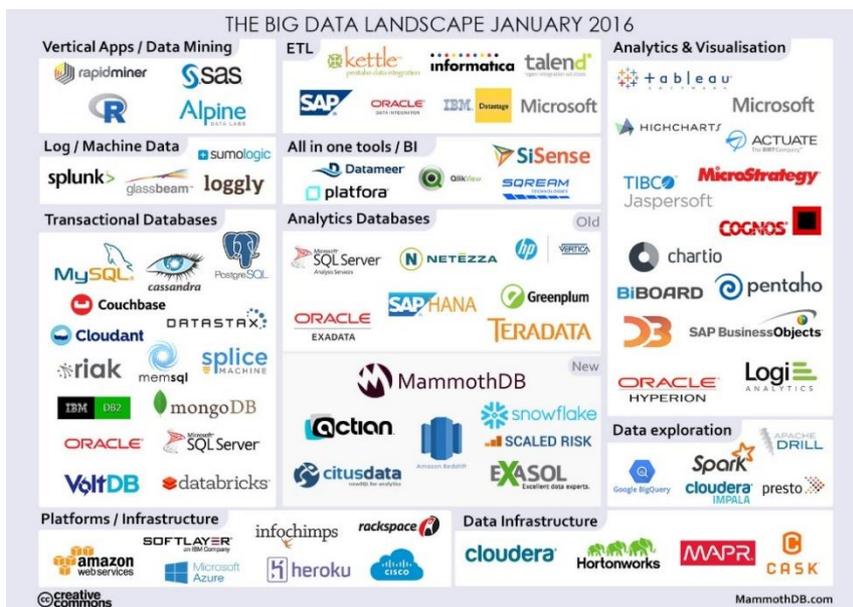


Fig. 4. Big Data map [12].

As a result of the evaluation of Big Data platforms and according to the need of real time PG information monitoring, the Spark platform was chosen as the ideal option according to its free licensing, real-time processing, speed, scalability, simplicity in creating programs and ease of integration with other components.

7 Methodology and Case Study

The objective of Big Data is to provide new knowledge to the company from the processing and analysis of information. This new knowledge will help to support the business decision-making. This section presents the results obtained from applying Big Data in the integration of renewable energies into the Grid.

The main benefits of the implementation of Big Data technologies are:

- Make better decisions with timely and reliable information.
- Show detailed information of the process of generating electricity with renewable energy through the analysis of historical and real-time information.
- Allow the data delivery in a dynamic and flexible way through control panels.
- As a result of the above, improve the the integration of renewable energies in the Smart Grid.

The case study considers the construction of a Big Data infrastructure for the ingestion (acquisition), processing, storage, analysis (descriptive analytics) and monitoring (real time) of the information generated by photovoltaic systems.

The building of the infrastructure is based on existing technologies and available open source tools.

7.1 Architecture of the Photovoltaic System Interconnected to the Grid

In this section it is described the architecture of a Photovoltaic System interconnected to the grid (PG) that is installed at the CFE's National Training Center (CENAC) of Celaya, Guanajuato; as well as the different variables that are monitoring to measure its operating performance.

Figure 5 shows the current Celaya's PG architecture. There are described the main components of the system, the way it performs the interconnection with the CFE's distribution network and the process of obtaining data from sensors.

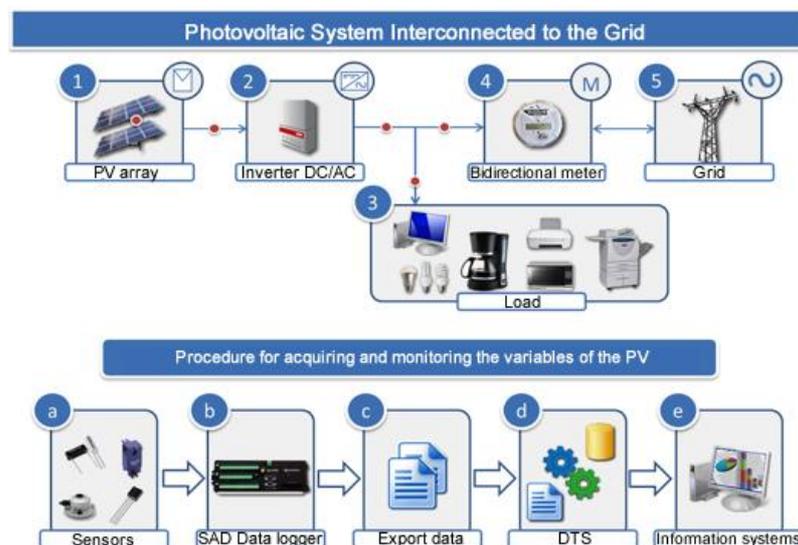


Fig. 5. Architecture of the PG [13].

Sunlight is converted into direct current electricity by the photovoltaic array. This direct current passes through the different components until the investor turns it into alternative, which can provide domestic consumption or send it directly to the mains. There is a data acquisition system (SAD) for the collection and storage of information. The SAD measures electrical signals from sensors (located in different components of the system) at a sampling rate determined and stores it in a flat file. This file is transported (exported) for processing data that will be monitored and analyzed by different users.

Importantly, the data stored in the SAD for each of the variables were scheduled according to the average value of 10 minutes of acquired values each 10 seconds [13].

According to experts of the Electrical Research Institute of Mexico (INEEL) Management of Renewable Energy, the implementation of the PG, for its operation monitoring and performance evaluation is subject to international practices and recommendations (IEC 61215, IEC 61730, UL 1703) for the execution of this type of activity in particular.

7.2 System of Photovoltaic System Measurements

For measuring the operating performance of photovoltaic systems 22 basic variables (obtained directly from the SAD), which are listed below, are monitored:

- Three climatic variables: ambient temperature, solar irradiance and irradiance on the horizontal plane of the array.
- In each PG array: current, voltage, power and temperature.
- To the inverter output current, active power and reactive power.
- In each of the grid lines (L1 and L2): active power and reactive power, both input and output, as well as the voltage between lines.

These variables are documented in the specification of requirements PG Celaya [13].

From the measured variables directly from SAD other variables that provide important additional information to measure the operational performance of PG are calculated. These derived variables are classified according to four groups: solar radiation, power, performance indicators and environmental benefits.

Finally, also have varying plant (area of the settlement, azimuth array tilt array) module (technology, brand, and model module), inverter (technology, make, and model) and the PG system itself (type of system, installation type, type of arrangement). They describe technical specifications of these components.

7.3 Implementation of the Case Study

Logical layers group the various components that integrate the Big Data architecture of the prototype: infrastructure, data sources, ingestion layer, processing, storage, analysis and monitoring and visualization layer. Figure 6 shows the implementation of the Big Data

architecture proposed for this work, as well as the complete flow of information. In general, the flow of information consists of the following steps:

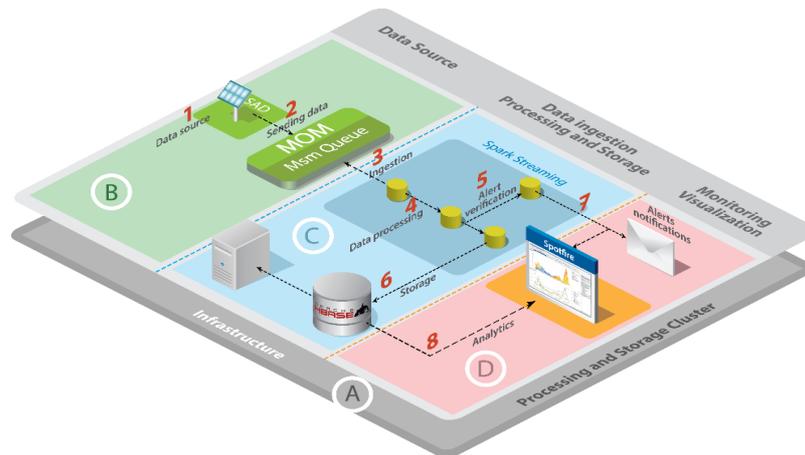
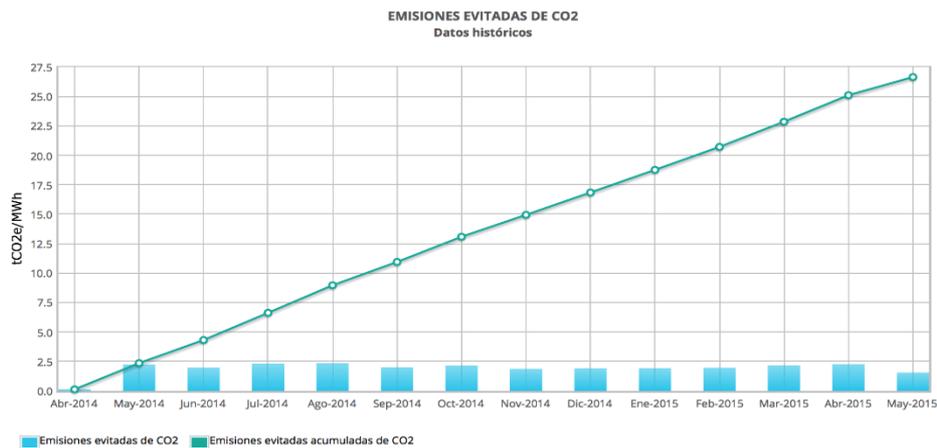


Fig. 6. Implementation of the architecture applied to the photovoltaic system.

1. **SAD-S**: simulates the data generated by the SAD in real time. As a requirement to validate and test the Big Data infrastructure, it was necessary to develop a simulator to generate real-time data and produce data for each of the 22 variables (power, current, voltage, temperature, irradiance, etc.) according to the average value of 10 minutes of measurements acquired every 10 seconds. Also, estimated failure and recovery times are simulated to generate abnormal operation alerts.
2. **Sending data**: the simulator sends the data to the MOM, Message Oriented Middleware. The messaging broker is used as an application to exchange information between the SAD simulator (SAD-S) and Spark Streaming.
3. **Ingestion**: Spark Streaming acquires the MOM data and places it in the memory of the processing cluster. The ingestion process is initiated by establishing a real-time context in order to receive continuous input data streams (DStream). In this way the data can be divided into batches to be processed by the Spark engine.
4. **Data processing**: once the data is placed in memory, Spark Streaming performs the processing of the data through the algorithms written in the “Scala” language. These algorithms calculate statistics, such as sums and averages about the measured variables of the photovoltaic system.
5. **Alert verification**: during the data processing, the measurements of the variables simulated by the SAD-S are validated to detect a failure event in a photovoltaic system (failure simulation of the electric power supply). In case that the variable measurements are out of range, an alert message, which must be distributed by the alert notification component, is created.

6. **Storage:** the processed information is sent to the HBase database, which relies on distributed HDFS storage. Internally, the data storage is managed by the Master through the RegionServers, which store the data in files and send it to the HDFS. These files are divided into one or more blocks to be stored in a set of DataNodes [14].
7. **Alerts notifications:** alerts notifications are responsible for coordinating the distribution and publication through different media of the alerts generated. For this work the alerts are notified through the dashboards and via e-mail to the operators of the grid.
8. **Analytics:** dashboards and visualization boards provide users with the ability to understand, search and navigate data. This ability to create analysis through reports and dashboards allows stakeholders to make decisions and design appropriate strategies to improve the operational efficiency of photovoltaic systems. The boards were developed using the Spotfire tool, which takes the information stored in HBase.



9. **Fig. 7.** Emissions avoided CO2.

8 Results

This section shows some of the dashboards developed in this research. These dashboards include descriptive analytics and diagnostic analytics. Basically, the dashboards display the operational performance of photovoltaic systems.

8.1 Emissions Avoided CO₂

In Figure 7 the avoided CO₂ emissions are monitored for each day, month, year and historical as well as all CO₂ emissions avoided accumulated over time. This is based on the

net energy output of the inverter of the photovoltaic system. In this case, monthly historical data is from April 2014 and a cumulative emission trends observed until May 2015.

8.2 Average System Efficiency

Figure 8 shows monthly the energy production of the photovoltaic generator and the output delivered through the inverter to the load or grid, as well as the average efficiency of the inverter.

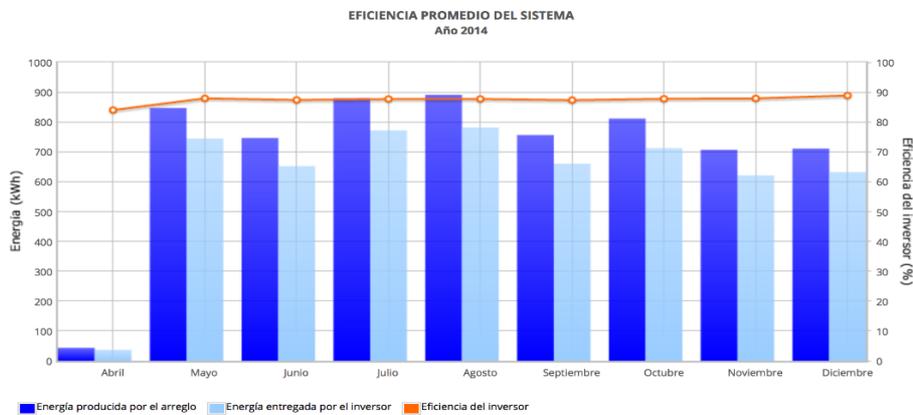


Fig. 8. Efficiency of the photovoltaic system.

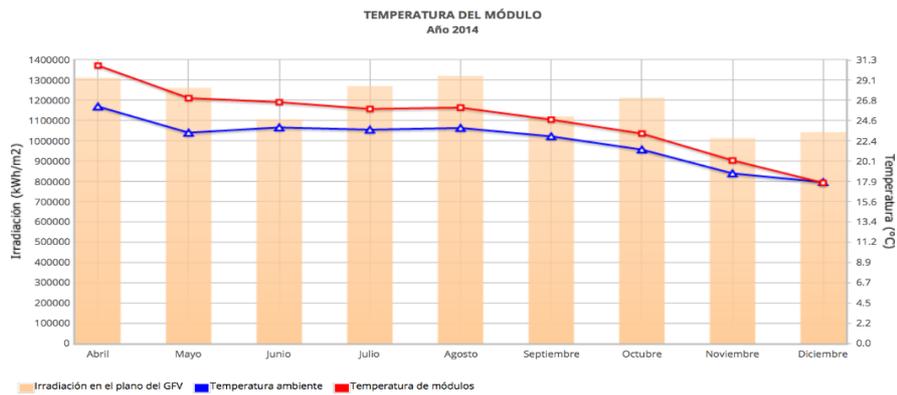


Fig. 9. Meteorological information.

8.3 Meteorological Information

In Figure 9, the meteorological information (irradiation, ambient temperature and temperature of modules) of the site where the photovoltaic system is installed is monitored.

9 Conclusions and Future Research

As a result of this work a technological infrastructure for handling large volumes of information that can be used to implement projects related to the Smart Grid. This infrastructure includes the design of architecture of Big Data for intake, processing, storage, analysis and monitoring of large volumes of data; and the identification and evaluation of the leading technology for the development of these solutions platforms.

The case study for processing data generated by photovoltaic systems interconnected with the grid, allowed to develop boards for analysis, monitoring and display of real-time information.

Consequently, this work will serve as a basis for future research related to energy efficiency, energy for sustainable development and Big Data, which are issues that concern the State and institutions like the Department of Energy, Energy Regulatory Commission, the Institute Electrical Research and the Federal Electricity Commission.

As future work, it is currently being developed for the implementation of mathematical design models predictive and prescriptive analytics, which help define solutions for forecasting demand.

References

1. SENER.:Estrategia Nacional de Energía 2013-2027. SENER, México (2013)
2. González, R., Jiménez, H. R., Huacuz, J. M., Munguía, G.: First 3-phase Grid-connected Photovoltaic System in Mexico. In: 3rd International Conference on Electrical and Electronics Engineering and XII Conference on Electrical Engineering (2006)
3. Escobedo, G., Jácome, N., Arroyo, G.: A Framework of Business Intelligence and Data Analytics for Smart Grid Business Intelligence for Smart Grid. In: ICEI-International Conference on Enterprise Information Systems (2016)
4. IEA: Technology Roadmap Smart Grids. Available at: https://www.iea.org/publications/freepublications/publication/smartgrids_roadmap.pdf (2011)
5. CRE: Marco Regulatorio de la Red Eléctrica Inteligente (REI) en México - Para la Comisión Reguladora de Energía. ESTA International, LLC (2014)
6. IEA: World Energy Outlook 2013. Available at: http://www.worldenergyoutlook.org/media/weowebiste/2013/WEO2013_Ch06_Renewables.pdf (2013)
7. López, J., Padilha, A., Gallego, L.: Ubicación óptima de generación distribuida en sistemas de energía eléctrica. Ingeniería y ciencia, Vol. V, No. 9, pp. 9–23 (2009)

8. Gerencia de Energías No Convencionales del IIE.: Guía de usuario. Sistemas fotovoltaicos interconectados con la red, aplicaciones de pequeña escala, Cuernavaca, Morelos, México (2010)
9. Liu, Y.: Big Data and Predictive Business Analytics. *Journal of Business Forecasting*, Vol. 33, No. 4, pp. 40 (2014)
10. Mysore, D., Khupat, S., Jain, S.: Big data architecture and patterns. Part 3: Understanding the architectural layers of a big data solution. Available at: <http://www.ibm.com/developerworks/library/bd-archpatterns3/index.html> (2013)
11. Accenture: Accenture Technology Vision 2015 Digital Business Era: Stretch Your Boundaries. Available at: https://www.accenture.com/us-en/_acnmedia/Accenture/Conversion-Assets/Microsites/Documents11/Accenture-Technology-Vision-2015.pdf (2015)
12. MammothDB: Landscape Big Data. Website: <http://www.mammothdb.com/the-big-data-landscape-by-mammothdb> (2016)
13. IIE: Especificación del Sistema informático para monitoreo y análisis de los datos obtenidos del sistema fotovoltaico de 5kW con la tecnología de película delgada y silicio amorfo. IIE, Gerencia de Tecnologías de la Información (2015)
14. George, L.: HBase: The Definitive Guide. First Edition, O'reilly, USA (2011)