APPLICATIONS OF BIG DATA IN RENEWABLE ENERGY SYSTEMS BASED ON CLOUD COMPUTING

Tarun Shakthi Sreedhar, Saiful Islam, Meron Atmosa, Elaheh Yazdandoust, Mohamed Suliman Elnaim, Shomesh Mishra, Venkata Naresh, Vemparala Rupali Bajpai*

SRH University of Applied Sciences
Germany

* Corresponding Author, e-mail: tarunshakthi.sreedhar@stud.srh-campus-berlin.de

Abstract: This study examines the potential of microgrids (MG), which utilize renewable energy sources to provide sustainable power solutions. To conduct the analysis, we examined load and photovoltaic (PV) data, calculated minimum and maximum averages, and visualized the correlation using big data tools. We cleaned the data by removing unnecessary rows, merged the tables, converted them into CSV format, and uploaded them to the Databricks file distribution system (DBFS). Subsequently, we processed the data by creating a pipeline and using ETL (extract, transform, load) processes. We analyzed and visualized the data using tools such as Power BI and Tableau. The analysis identified the maximum and minimum PV production, assessed the impact of weather patterns on production, and measured the energy shortage between load demand and PV generation. Our research demonstrates the steps involved in handling and analyzing data, uploading it to the Hadoop ecosystem, transforming it into different file formats, connecting it to a relational database management system (RDBMS), and visualizing it using BI tools. In this study, we utilized cloud infrastructure to perform analytical tasks, including the use of business intelligence (BI) tools.

Key words: microgrid, cloud computing, Spark & Azure services, data visualization, renewable energy systems (RES).

1. INTRODUCTION

The demand for sustainable and reliable energy solutions has led to the development of microgrids, which integrate renewable energy sources and storage systems. In this paper, we use data from the TwInSolar Consortium at the University of La Reunion to scrutinize load and photovoltaic (PV) data from several university buildings. Our analysis reveals underlying patterns in energy demand (load) and PV production. To prepare the data, we have applied a multi-step process that involves cleaning the data and transforming it into CSV files. Next, we have transferred these files to the DBFS for further processing. To optimize storage productivity, we have converted the data into Parquet files. We have analysed and visualize the data using Power BI and Tableau, a high-powered data processing engine inside cloud platform. Our study aims to recognize periods of high and minimal PV production, analysed the impact of weather patterns, and measure the energy gap between load demand and PV generation.

2. LITERATURE REVIEW

The technology architecture using Big Data tools for managing massive volumes of information to enable renewable energy integration is presented in the study. Big Data technologies can significantly improve the integration of renewable energies, contributing to grid efficiency and sustainability [3]. Big Data analytics used in smart grids and renewable energy networks were thoroughly reviewed in [4]. The PV Energy Management (EM) as an Internet of Things (IoT) Application study investigates how solar energy management can use defect detection, power optimization, and remote monitoring. Integrating IoT and artificial intelligence (AI) technology can greatly open the door to smarter, more resilient PV energy infrastructure by providing strong tools for fault detection, real-time monitoring, and system optimization [10]. A cloud computing-based approach to evaluate solar energy potential for improved energy planning is presented in [5].

An Observation Report on Cloud Computing in EM of Smart Grid demonstrates how cloud computing may improve smart grid data processing, energy management, and operational effectiveness. They concluded that cloud computing provides scalable, dependable, and reasonably priced computational resources and smart grid energy management greatly benefits from it [1]. The cloud computing framework for smart grid operations is examined also in [6]. Dynamic Load-Shifting Program (LSP), a study, provides a dynamic LSP on a cloud computing framework for effectively managing renewable energy. It presents a multi-agent system that improves real-time monitoring and control of energy use by adopting demand-side management (DSM) at the household level. The program showed an 18% increase in PV energy consumption at the consumer level, demonstrating the value of dynamic, real-time load-shifting [7]. The study EM of Smart Grid using cloud computing focuses on load scheduling and real-time electricity usage monitoring in smart grids through cloud computing. To achieve optimal energy usage, the study highlights the use of cloud-based real-time data monitoring and analysis to reduce peak energy demand. It offers a process for implementing EMS in educational facilities and shows possible ways to increase efficiency and save energy [8].

3. PROPOSED METHOD

We used two approaches for Data Processing and Visualization for MG Data Analysis: one uses Apache Spark on Databricks and visualizes it using Tableau, and the other way is analysing the data using Azure services and visualization using Power BI. The steps are stated below as:

1. Creating table to add the collected date from the source.
2. Transforming Data and Merging: Use *PySpark* to read the Hive tables.
   a) For each PV production dataset: We Created a temporary table by adding a new column named "Campus" with the corresponding campus name (e.g., "Dept 1 & 2").
   b) We created a temporary table for the weather data.
   c) We have used Spark SQL to execute JOIN operations for merging the PV production tables with the weather data table according to a common date/time column (e.g., "Datetime").
   d) This creates three different merged datasets: Dept_1_2_PV_weather, ESIROI_PV_weather, ENERPOS_PV_weather.
3. Analysis and Visualization using Spark & Azure:
   a) Maximum and Minimum PV Production: We Used Spark SQL functions like `max` and `min` to calculate each campus's maximum and minimum PV production over the entire dataset.

b) Month of Lowest Production: We Utilized Spark SQL functions to identify the month(s) with the lowest PV production for each campus.

c) GHI Visualization: We used Tableau and Power BI for Data visualization.

In Figure 1, we summarize and illustrate the approach using Apache Spark on Databricks and visualize it using Tableau. In Figure 2, we summarize and illustrate the approach with Apache Spark on Databricks and visualize it using Tableau.

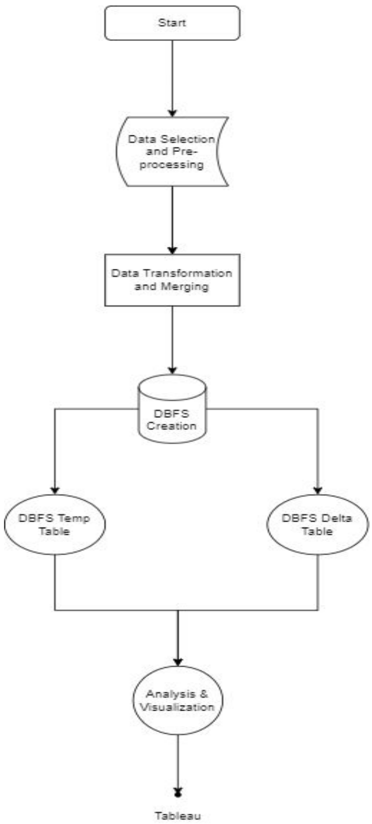

Figure 1. Flowchart for Approach 1

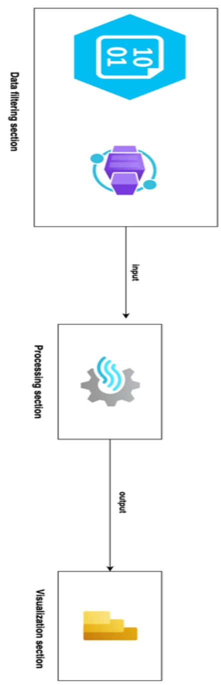

Figure 2. Flowchart for Approach 2

4. RESULTS AND ANALYSIS

We have analysed metadata extracted from the file Meteo_Terre_Sainte.txt. The data was collected from the University of La Reunion's weather station situated at coordinates 21°20'S, 55°29'E and an elevation of 75 meters, located at the Terre Sainte Campus. The station operates under GMT/UTC+4h time zone. It's important to note that the data was subjected to consolidation, filtering, and gap-filling methods, although it did not undergo any formal quality check.

4.1. Visualization and Analysis using Tableau

Figure 3 shows GHI (Global Horizontal Irradiance, GHI: Global Horizontal Irradiance in W/m^2) values related to solar energy production over a year. GHI is a measurement of solar power per unit area that hits a horizontal surface. Here's a more detailed breakdown of the image: (the x-axis is labelled month, and the y-axis is labelled as a value which shows a range of GHI values from 400 to

1400). This graph suggests that GHI values tend to be higher in the middle of the year (around June to September) and lower at the beginning and end of the year.

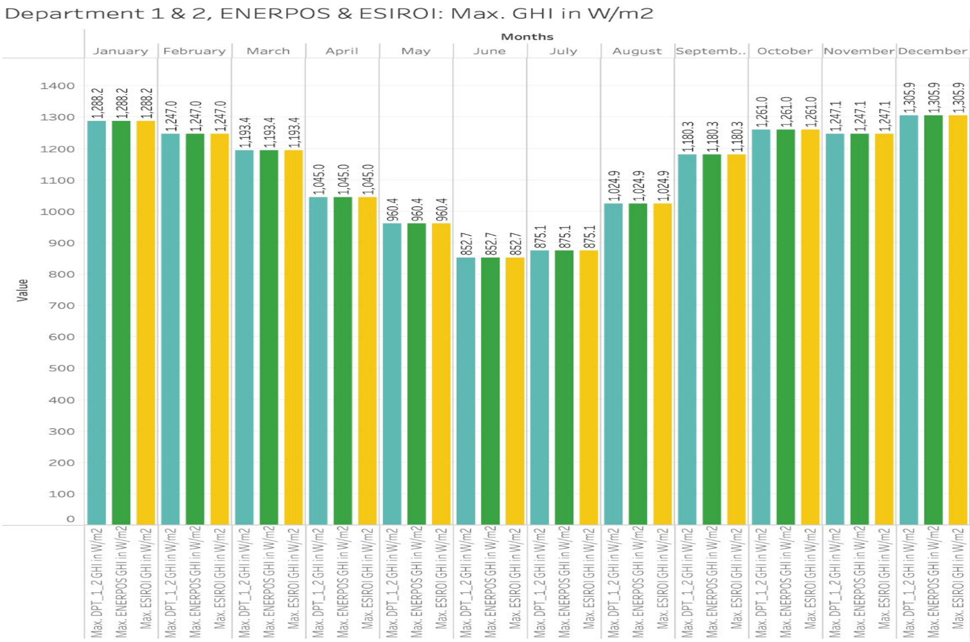

Figure 3. GHI: Global Horizontal Irradiance in W/m²) values related to solar energy production over a year

Figure 4 depicts the maximum monthly production in kW for Department building 1 & 2 (DPT_1_2), ENERPOS, and ESIROI over a year. January has the highest maximum production in all the department buildings. ENERPOS has the highest overall optimum output of 100.98 kW in January and December. Department building 1 and 2 (DPT_1_2) had the minimum peak production of 8.18 kW in December.

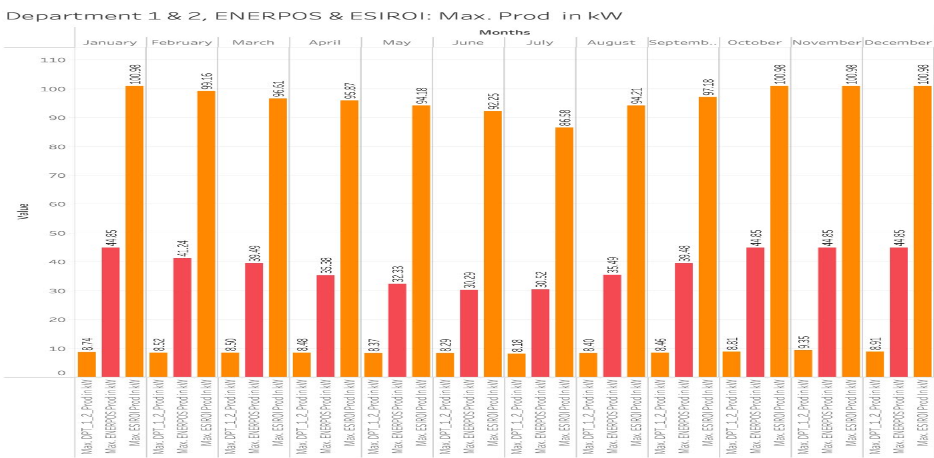

Figure 4. Maximum monthly production in kW for Department building 1 & 2 (DPT_1_2), ENERPOS, and ESIROI over a year

The stacked bar graph from Figure 5 illustrates the maximum expected load versus maximum production in kW for ENERPOS during a year.

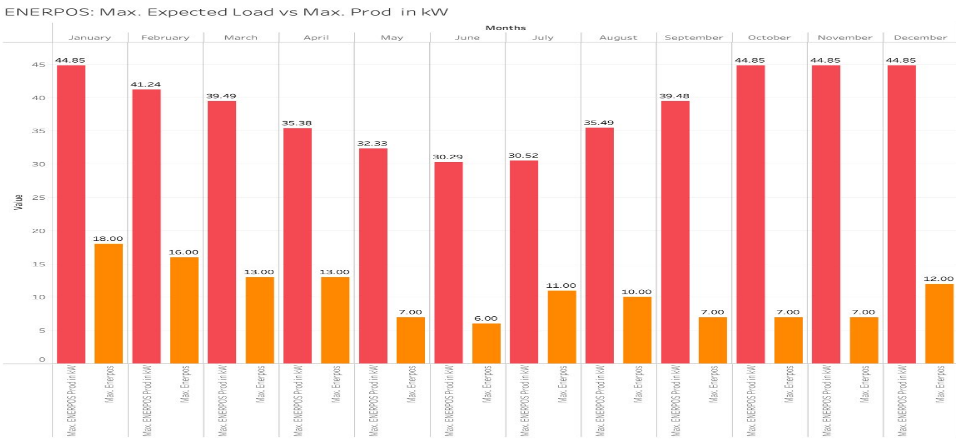

Figure 5. Maximum expected load versus maximum production in kW for ENERPOS during a year

4.2. Visualization and Analysis with Azure and PowerBI (Using Azure Blob Storage and Power BI for Cloud-Based Data Visualization): This method provides a three-step process for using cloud services to achieve data visualization. The process involves the use of stream job analytics services. **Step 1:** Azure Blob Storage Setup involves creating a container within your Azure Blob Storage account. All the necessary files for analysis are uploaded into this container, which will serve as the source data for the visualizations. **Step 2:** Power BI Account Creation involves creating a connection between services inside azure environment. This account provides access to Power BI's cloud-based functionalities. **Step 3:** Data Analysis and Visualization is achieved through an analytic job that retrieves data files from the Azure Blob Storage container and directs its output to Power BI for analysis and visualization. It can be done by stream analytics job services too inside azure.

The stacked bar graph presented in Figure 6 compares the average of production per datetime 4 to average of production per datetime 5 to average of production per datetime 6 in kW over the month of January 2021. average of production per datetime 6 is highest and constant throughout the month where average of production per datetime 5 can be seen fluctuating with lowest on 12th January.

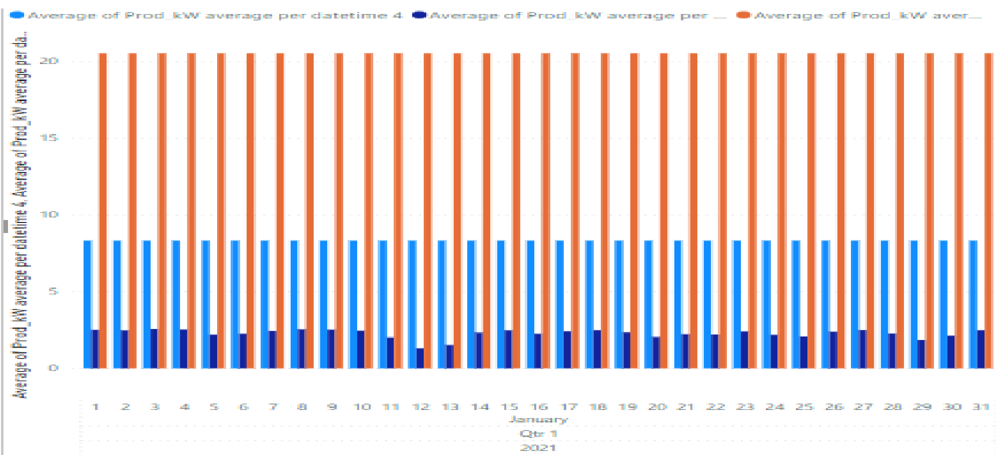

Figure 6. Comparison of the average of production in kW over the month of January 2021

The stacked bar graph presented in Figure 7 compares the average GHI per datetime to average GHI per datetime 3 and average GHI per datetime over the month of January 2021. The average Global Horizontal Irradiance (GHI) per datetime remains constant at a consistent level throughout the month. Notably, the average GHI per datetime peaks on specific dates, namely the 1st, 3rd, 10th, 18th, and 27th of January 2021. Conversely, the lowest average GHI per datetime is observed on the 12th of January 2021.

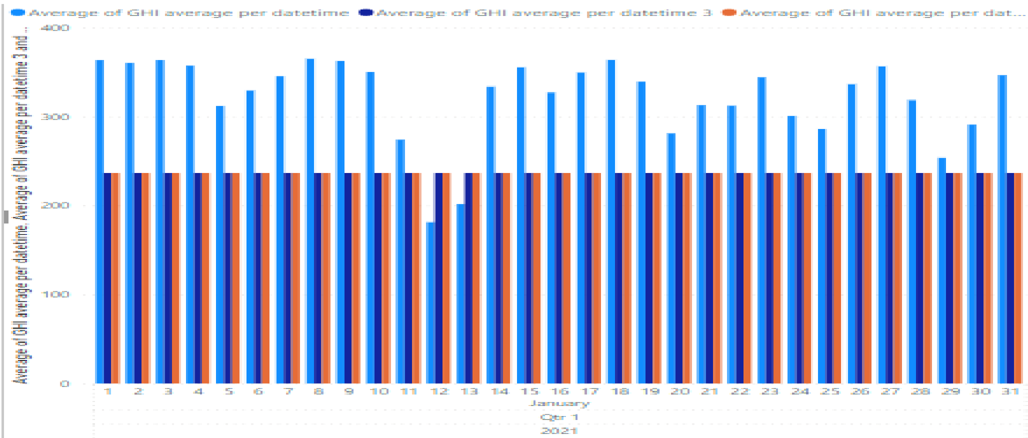

Figure 7. Comparison of the average over the month of January 2021(as GHI is correlated with PV production which has been identified during data analysis process)

5. CONCLUSION

This research investigated the effectiveness of utilizing cloud-based data processing and visualization techniques for analysing MG data. Our analysis involved several key steps: Data preprocessing, where unnecessary data records were removed, and relevant tables were merged before converting them into a CSV format suitable for upload to DBFS. Following that, the ETL process was implemented for optimized storage. Data Processing and Analysis were carried out using Apache Spark on the Azure Databricks platform. Various analyses were conducted, including identifying the maximum and minimum PV production across all campuses within the dataset. It explored the month(s) with the lowest PV production for each campus, visualizing the GHI trend to recognize its effect on energy production, and calculating and visualizing the energy gap between PV generation and preset load demand.

Cloud computing has been explored deploying the Spark application to HDFS (Hadoop environment) for increased scalability and cloud-based execution. Additionally, Azure Databricks visualizations were used for interactive investigation of the processed data. This research demonstrates the effectiveness of cloud-based data processing and visualization tools in analysing MG data. The approach provides valuable insights into PV production patterns, weather influence, and potential energy shortfalls, ultimately aiding in informed decision-making for optimized MG operations.

In the future, it would be beneficial to utilize a higher volume of data and develop a rapid energy production scheme to help mitigate load scheduling in an MG system. Additionally, it is important to investigate issues in MG such as inverter fault detection, PV string fault detection, and implement real-time scenarios using big data tools.

AKNOWLEDGEMENTS

Data has been collection from the University of La Reunion and TwinSolar. The University was the source of all information regarding the case study and data preparation. The date is available in internet as an open-source data.

REFERENCE

[1] Allahvirdizadeh, Y., Moghaddam, M. P., Shayanfar, H. A survey on cloud computing in energy management of the smart grids. *International Transactions on Electrical Energy Systems*, vol.29, no.10, 2019, e12094.

[2] Şerban, A. C., Lytras, M. D. Artificial intelligence for smart renewable energy sector in Europe—smart energy infrastructures for next generation smart cities. I*EEE access*, vol. 8, 2020, pp.77364-77377.

[3] Escobedo, G., Jacome, N., Arroyo-Figueroa, G. Big data & analytics to support the renewable energy integration of smart grids. In *Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security (IoTBDS 2017)*, 2017, pp. 267-275.

[4] Colmenares-Quintero, R. F., Quiroga-Parra, D. J., Rojas, N., Stansfield, K. E., Colmenares-Quintero, J. C. Big Data analytics in Smart Grids for renewable energy networks: *Systematic review of information and communication technology tools. Cogent Engineering*, vol.8, no.1, 2021, 1935410.

[5] Kapoor, M., Garg, R. D. Cloud computing for energy requirement and solar potential assessment. *Spatial Information Research,* 26, 2018, pp.369-379.

[6] Bitzer, B., Gebretsadik, E. S. Cloud computing framework for smart grid applications. In *2013 48th International Universities' Power Engineering Conference (UPEC)*, Sep 2013, pp. 1-5. IEEE.

[7] Rajeev, T., Ashok, S. Dynamic load-shifting program based on a cloud computing framework to support the integration of renewable energy sources. *Applied Energy,* vol.146, 2015, pp.141-149.

[8] Hans, M., Phad, P., Jogi, V., Udayakumar, P. Energy management of smart grid using cloud computing. In *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)* Aug 2018, pp. 1-4. IEEE.

[9] Jha, S. K., Bilalovic, J., Jha, A., Patel, N., Zhang, H. Renewable energy: Present research and future scope of Artificial Intelligence. *Renewable and Sustainable Energy Reviews*, vol.77, 2017, pp.297-317.

[10] Spanias, A. S. Solar energy management as an Internet of Things (IoT) application. In *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Aug 2017, pp. 1-4. IEEE.

[11] Bin Abu Sofian, A. D. A., Lim, H. R., Siti Halimatul Munawaroh, H., Ma, Z., Chew, K. W., Show, P. L. Machine learning and the renewable energy revolution: Exploring solar and wind energy solutions for a sustainable future including innovations in energy storage. *Sustainable Development*, 2024.

Information about the author:

Tarun Shakthi Sreedhar – Currently pursuing an M.Sc. in Big Data and AI at SRH Berlin University. Working as a Data Analysis Intern at Infosys in Germany, with previous experience as a Software Engineer at CGI in India. Enthusiastic about Big Data and proficient in Python, Power BI, Azure, AWS, and SQL.

Saiful Islam – Ph.D. Student at the University of Otto von Guericke University of Magdeburg (OVGU) in the field of Computational Intelligence. Research Associate and Lecturer at SRH University of Applied Sciences.

Meron Atmosa – Enrolled as an M.Sc. student at SRH University of Applied Sciences. Enthusiastic about machine learning, with proficiency in R, Python, Power BI, and SQL.

Elaheh Yazdandoust – Enrolled as a M.Sc. student at SRH University of Applied Sciences. Obtained a B.Sc. degree in information security from Iran. Her research is centred around Artificial Intelligence, Big Data, and Data Engineering.

Mohamed Suliman Elnaim – A Master's student in Computer Science, is involved in the development of a machine learning model for recognizing Arabic handwritten text.

Shomesh Mishra – Enrolled in an M.Sc. program for Big Data and AI at SRH University of Applied Sciences, with over 4 years of experience in Data Analytics and Data Management. Currently serving as a Machine Learning Intern at AIDY Health in Munich, Germany.

Venkata Naresh – An M.Sc. in Big Data & Artificial Intelligence is currently being pursued at SRH Berlin University of Applied Sciences. Over 2 years of experience has been gained as a data analyst.

Vemparala Rupali Bajpai – Rupali Bajpai is a tech enthusiast who embraces new challenges and continuous learning. Coding, working on innovative projects.

Remark:

Manuscript has been received in May 2024 to take part in the 38th International Conference on Information Technologies (InfoTech-2024), IEEE conference, Rec. # 63258, Section C: “Networking and communication Technologies”. It has been accepted and revised based no double-blind reviewing and has not been published in full text elsewhere.