

Solar Energy Portal Service using IoT

Author, Vishwajeet Yadav

A Data Science Foundation White Paper

May 2018

www.datascience.foundation

Copyright 2016 - 2017 Data Science Foundation

Solar Energy Portal Service

using the Internet of Things

I. Introduction

Over the last few years, India has been suffering from huge power cuts. The Government of India is working to help resolve this problem by developing the country's renewable energy capability. The World Bank Group is helping India to scale up solar energy from installing rooftops systems to massive solar farms. Government of India's mission is to generate 40% of energy through renewable energy sources by 2030. So one can say that solar energy is the one of the ultimate future of energy sources[1][3].

Currently, most of the solar energy is generated at the solar farms or solar power plants. Efficiency is the most important aspect of a system and this needs to be constantly maintained. To maintain efficiency, each power plant must check for fault in their system and monitors them manually. In the near future when many solar rooftops system have been installed, individuals will need to do the same. With such an increase in the solar energy generation project, there must be an overall monitoring system which will help to analyse the complete system to boost efficiency. Around 14% of the total solar system experiences major faults every year and stops working. Half of resident solar panels experience the same problems[2].

Also inverter, batteries, transformers are the factors which are also needed to analyse side by side. These components are not 100% efficient, they cause around 30-40% power loss. Hence this analysis will increase not only the efficiency, but also the life of these components.

The data storing, processing and analysing is done with advanced technology: Big Data. Data visualization will help the user to understand the numbers. The connection between the solar grids and Big data can be done using IoT.

II. data collection

The collection of data requires internet technology. Since solar grids are located at remote locations, we need to access the data using the Internet of Things. IoT processes the data to system from things which are embedded with sensors through a network. In solar power grids, there are many sensors installed in the system like irradiance meter, wind sensor, temperature sensor etc. To supervise and pull data from these sensors, SCADA(Supervision Control and Data Acquisition) is used. SCADA is for remote access to various local modules. It uses the network data, computers and GUI(Graphical User Interface) for supervisory management. For process plant and data acquisition, it uses peripheral devices like PLC(Programmable Logic Control) and RTU(Remote Terminal Units).[4]

Remote terminal units gather information from the remote site from various intelligent electronic devices. They convert the signal from devices into a language used to transmit data[5]. This data is collected at plant level monitoring. In our case, we are collecting data from the different plant level monitoring system. This data is transferred from plant to database using 3G network. This data is transferred using lightweight MQTT(MQ Telemetry Transport) protocol. All data is collected at a single place and generate enough data-set to perform the analysis. At database, data is converted into the Comma-separated Values(CSV) format(Semi-Structured data). After that, it is processed to Hadoop Distributed File System(HDFS).

Hadoop is the open source system used for storing and processing Big data.

III. big data and hadoop

A. Big data

The solar system generates a tremendous amount of data. In our situation, we are receiving data from plants with capacity more than 100MW. And each solar panel generates 200W energy. Hence approximately we have 5,00,000 solar panels in a power plant. While transferring data, the minimum size of the MQTT packet is 200 bytes with payload data. If system latency is 1 second then 1.6Gb/s data will be generated[2].

Every solar panel, inverter and battery has sensors which provides status about the system in particular interval of time. In such case, the data generated is huge and incremental. Handling such an amount, the velocity of generation, variety and uncertainty of data in convention way is impossible. The solution for managing such data is Hadoop.

B. Hadoop

Hadoop is open-source, java based technology which works on commodity hardware. Hadoop provides a distributed file system and framework for analysis and transformation of the large data-sets to different user applications. It uses a Map-reduce algorithm. Hadoop Distributed File System(HDFS) is designed to store a very large data-set and it partitions data into its many hosts. HDFS do execution of the application computations close to data. Hadoop always try to reduce the network and disk I/O(input/output). Due to this strategy, Hadoop is the one of the fast Big data processing tools.

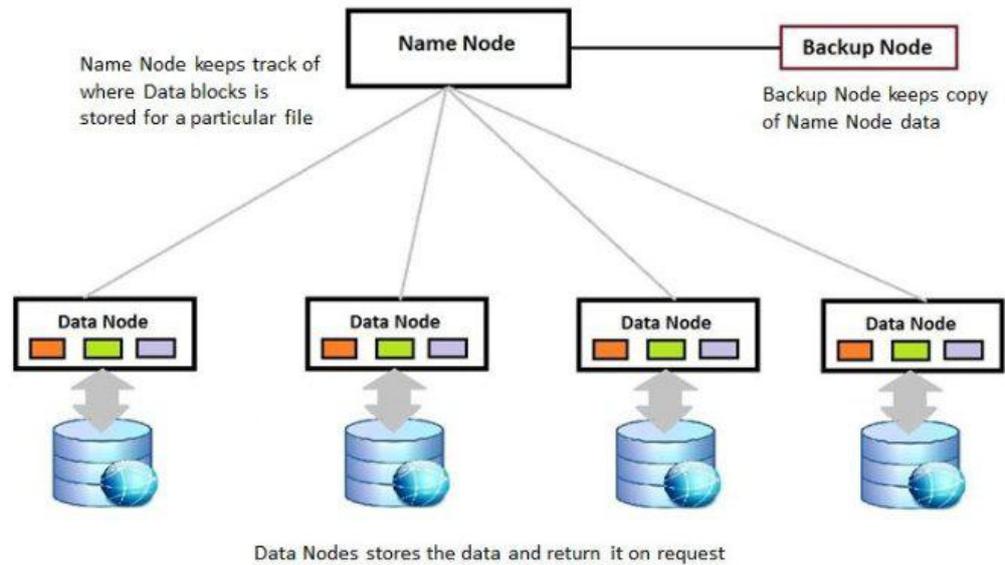


Figure 1.0 Hadoop Data storage architecture

There are some components in the Hadoop ecosystem which play an important role in keeping Hadoop robust. HDFS namespace is a hierarchy of files and directories. These files and directories are represented by NameNode. NameNode records attributes like permissions, modification, disk space, etc. It works as the manager node. In HDFS, files are chopped into blocks. The size of the blocks may vary according to system to system. But by default, it is 128 megabytes. These blocks are replicated independently in the DataNodes. NameNode maintains the namespace and mapping of the blocks in the DataNodes.[6]

When a DataNode starts, it scans the local file system and generates the list of all HDFS blocks that correspond to the local system. This is called the block report. Every DataNode generates this report and sends it to NameNode using the heartbeat service. The Hadoop file system can hold data up to petabytes of data.

The data received at HDFS is stored and replicated using the above pattern. HDFS is for storage. For analysis, the data set needs to be clean. The data cleansing process can be done using the Extract-Transform-Load (ETL) process. The Hadoop system provides the MapReduce algorithm for analysis of the data.

Using Hadoop, the system is able to do:

1. Store the solar data in Hadoop
2. Analyse the energy produced
3. Fault detection in the solar system
4. Recognize patterns in the data

5. Visualize the real-time data

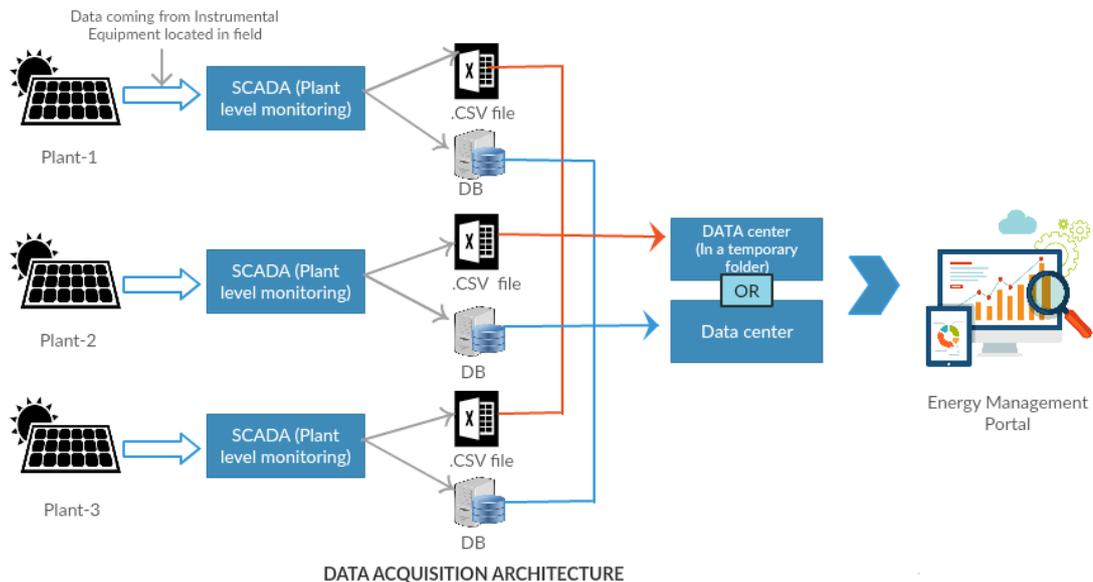


Figure 2.0 Data acquisition architecture

IV. Data model

The understanding data is the most important part of the project. Analysis of power generation, fault detection depends on the relation of the data with power, panels/grids, inverters etc. All terminologies in the data should be understood before we analyse data.

At the time of data acquisition, SCADA pulls data in the form of string from different blocks of solar panels, inverters in solar files, inverter files respectively. These files contain various fields as shown in fig 3.0. There are various variables under the 'parameter name' field such as AJB1_STR2 in the data-set. These variables play an important role while analysing.

TimeStamp	12/05/2015 13:38:23
Parameter Name	AJB1_STR2
Average	7.05
Minimum	6.56
Maximum	7.48
TimeStamp for Minimum	12/05/2015 13:38:23
TimeStamp for Maximum	12/05/2015 13:38:23

Figure 3.0 Data model

The Different fields in the data models are as follows:

Time-stamp: Time at which string is pulled by SCADA.

Parameter Name: It contains the name convention of data which data is recorded.

Average: The average of the minimum and maximum value.

Minimum: Minimum value generated by the respective parameter.

Maximum: Maximum value generated by the respective parameter.

Time-stamp for minimum: Time at minimum value generated

Time-stamp for maximum: Time at maximum value generated

After data understanding, next process is analysis.

V. Analysis

For the analysis of the data, Hadoop uses MapReduce paradigm. MapReduce is the framework for working in a parallel manner on large data-set on the cluster. MapReduce performs three different functions to process the data: map, shuffle, reduce. Initially every worker node applies the map function on the local data and write the output in the temporary folder. Then worker nodes redistribute the data according to the output key(output of the map function).Data belongs to one key is collected at worker node with the same key. In the end, the reduce function processes data in parallel at each worker node [7].

The MapReduce framework works on the data structure of (key, value) pair. Map function takes one pair of one domain data and returns the list of pairs with a different domain.

Map(k_1, v_1) -> List(k_2, v_2)

The map function is applied to every key and produce list of the pairs. These pairs are grouped according to the respective keys. When reduce function applied, the collection of values generated with similar domain[7].

Reduce($k_2, \text{list}(v_2)$) --> list(v_3)

Now we need to perform different calculations on our data-set in order to get the desired result. We need to calculate power generated by every parameter on the daily, monthly, and yearly basis. Also, we need to analyse the other dependent factors like batteries, inverters, and transformers.

```
function Map is
  input: integer K1 between 1 and T, representing the all records in data-set
  for each record in the K1 batch do
    let Y be the parameter_names related to power
    let N be the value parameter_names has
    produce one output record (Y,N)
  repeat
end function

function Reduce is
  input: parameter_names in Y
  for each input record (Y,N) do
    Accumulate in Nnew the sum of N
  repeat
  produce one output record (Y,Nnew)
end function
```

Figure 4.0: MapReduce algorithm

To calculate aggregate power for every parameter, we need to use MapReduce algorithm to filter the parameters those are related to power from the whole data-set. Then we will apply the map function to generate the key value pair of individual parameter names and values. Then using reduce function, we will calculate the sum of power generated by individual, similar parameter name. We are using following the MapReduce algorithm to analyse the data as shown in fig 4.0.

In the function map, key k1 generates output with a list of parameter names related to power and its values and stored it in the record (Y, N) list. Reduce function add the values of the same parameter in the Nnew variable, and gives output with format parameter name with the total value. Similarly adding all these result's value will result in total power generated by the power station.

The other factors like fault detection, inverter status analysis can be done by applying proper MapReduce algorithm on the data-set.

VI. Data visualization

Data visualization is the important step of interpreting the results of the analysis. One can more easily understand the graphs, lines, and pie charts than the numbers. Visualization gives the user power to make a quick comparison between two entities. Here we are using Microsoft's power Bi tool for visualization.

Power Bi gives us various options like real-time streaming data-set or choosing data-set from file to prepare beautiful dashboards and reports. This helps the user to prepare custom dashboards. To prepare the dashboard of solar energy analysis, we are sending the results of MapReduce

algorithm to power bi using REST API. The data-set is visualized by various charts and graphs.

This prepared dashboard can be shared with others. Power bi gives us a facility to use mobile-optimized dashboards and reports. We can also publish the dashboard on the cloud.

percentage analysis of parameters

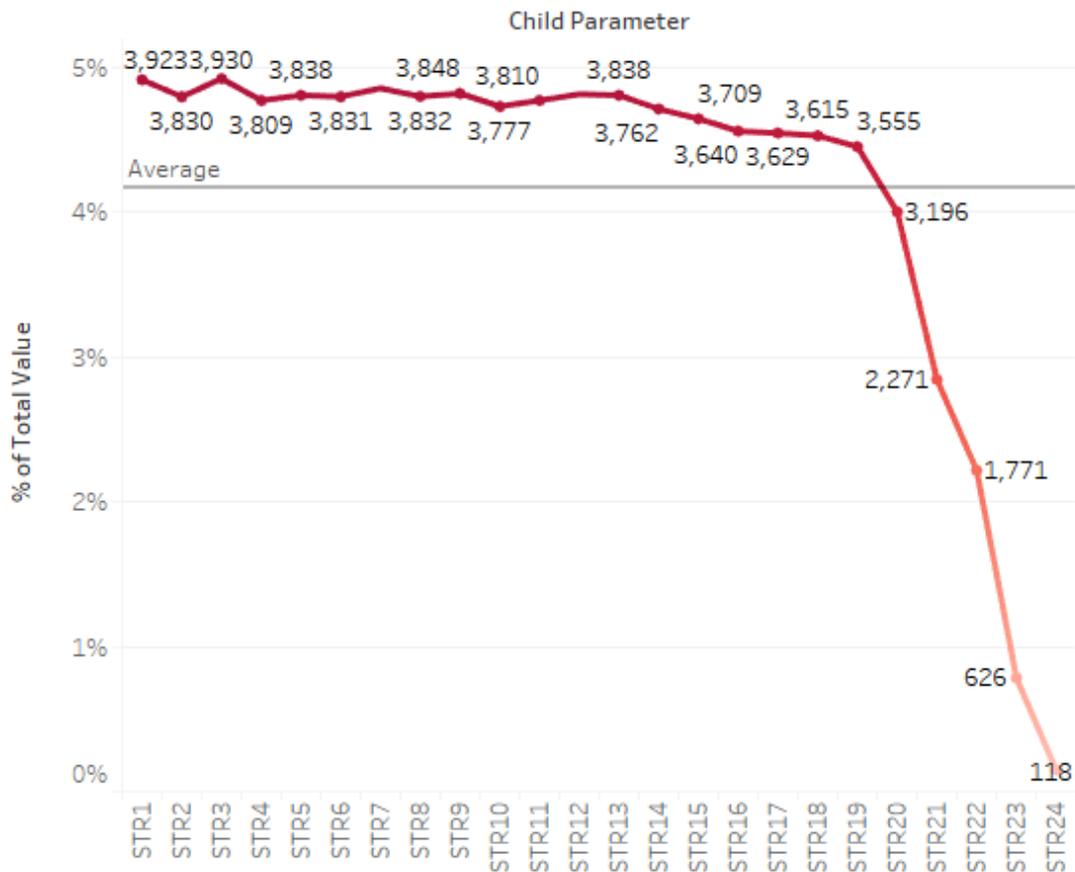


Figure 5.0: Percentage of power generation by respective parameters

VII. Conclusion

Solar energy is growing very fast and the adoption rate is increasing. As we discussed earlier, we need to analyse the data and detect the faults in the system. In this project, we tried to analyse the solar plants and grids data. But there is need to develop a system to analyse the data from rooftops solar system for small houses, hospitals, and private industries. As ARM technology is developing day by day, we can make analysis for small plants to monitor their system. Raspberry

Pi, Arduino will help to make the smart plants.

We have used Hadoop for storing and analysing the data. Hadoop is open source technology which provides their service for free of cost. To build Hadoop setup, we need high-end compatible computers in clusters. We can build Hadoop clusters using commodity hardware. There are some negative points associated with Hadoop. For small data-set, Hadoop or big data is useless. For small data-sets relational databases are the best option. But Data is continuously growing, and we need to use Big data for the optimization.

Cloud is the technology offers many benefits and reduces the need for technology investments. We can also use other options like Google Big-table. By using cloud, we can handle the custom requirement of hardware. Cost of the analysis of solar system could be reduced through the usage of cloud.

References

1. 'Solar energy to power india of the future', retrieved in March 2018, from <http://www.worldbank.org/en/news/feature/2016/06/30/solar-energy-to-power-india-of-the-future>
2. B. Vikas Reddy, Sai Prakash Sata, Satish Kumar Reddy and Bandi Jasawant Babu "Solar Energy Analytics Using Internet of Things" : International Journal Of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 7(2016) pp 4803-4806
3. Solar rooftop grid scheme, Retrieved March 2018, from <http://mnre.gov.in/solar-rooftop-grid-connected>.
4. SCADA in Wikipedia,the free encyclopedia, Retrieved March 2018, from <https://en.wikipedia.org/wiki/SCADA3>
5. Narayan Lal Purohit, Anshika "Data Acquisition of the Solar power plant using SCADA system" International Journal of Engineering Trend and Technology(IJETT) Volume 23-Number 4 - May 2015
6. HDFS architecture in Apache Hadoop, from https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
7. MapReduce in Wikipedia,the free encyclopedia : Retrieved March 2018, From <https://en.wikipedia.org/wiki/MapReduce>

About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

Contact Data Science Foundation

Email: contact@datascience.foundation
Telephone: 0161 926 3641
Atlantic Business Centre
Atlantic Street
Altrincham
WA14 5NQ
web: www.datascience.foundation

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ
Tel: 0161 926 3641 Email: contact@datascience.foundation Web: www.datascience.foundation
Registered in England and Wales 4th June 2015, Registered Number 9624670