Big Data Analytics of solar system centre using DataStax and Cross Correlation methods

S.Samundeeswari
Head of the Department, Department of Computer Science and Engineering
Periyar Maniammai Univeristy, Thanjavur, TamilNadu, India

Sharanya Seetharam
Student, Department of Computer Science and Engineering
Periyar Maniammai Univeristy, Thanjavur, TamilNadu, India

K.Pavithra
Student, Department of Computer Science and Engineering
Periyar Maniammai Univeristy, Thanjavur, TamilNadu, India

Abstract- There are lot many techniques for searching and the world of Big Data Analytics has widespread all over the era and an emerging tool for handling big data's, but what we are going to do in this paper is designing a fast and efficient and quick search method for solar system for quick accessing of data to be searched in database using the Apache Cassandra database using NoSQL statements with the following software of DataStax provided by them and which is open source like Solar cloud or Elastic searching. Using cross correlation platform we can derive a formula for big data in Spatial data which makes the search easier using these tricks and using this we can use geohashing technique for where the solar system has to be setup for deriving where it will be suitable in a particular place. It would perform the process faster and will be a good idea for the upcoming world to be broad and specific on handling the data.

Keywords - Apache Cassandra, No_SQL, DataStax, Cross Correlation, Geohashing.

I. INTRODUCTION

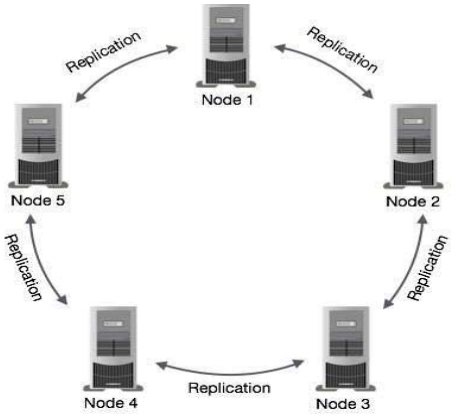
We all know what Big Data Analytics is? It's actually a sector in which the world is growing to store the databases. Here are using this Big Data Analysis for solar generated power system[2] where a high quality of data needs to be recorded based on which we will be able to generate the place where it has to setup for more utilization than a place where it cannot get much radiation from the sun as to use its power for general uses. It will be useful for prediction of data in the near future for setting up solar panels[1].

II. RELATED WORK

The unpredictable seasonal of solar and wind resources should be factored to get proper prediction of climate changes all over the world. Wind and Temperature plays major role in the generation of solar energy. Thermocouples are used for the measurement of temperature in traditional methods. New approach is flexible and eases of changing the type of sensors and records the data which is suitable for very large data and also less cost[4]. The algorithms like Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbour (Knn) and Multivarient Linear Regression (MLR) methods are used to predict the model[5] . The numerical simulations for monthly calculation of solar generated data are demonstrated for solar radiation values. Mobile data gathering approach solves the above said problem with longer data collection latency, to improve the throughput energy harvesting is introduced[6]. Large scale data mining needed for graphical visualizations to understand better the data and the results. The techniques which are performed to calculate the mining process like Big Data Analytics are used for searching higher level of data which are huge in amount and all the details are collected using this technique.

III. PROPOSED METHOD

We are using NOSQL databases like Apache Cassandra for working out in Big Data Analytics. Apache Cassandra is distributed and decentralized open source storage system, to manage large data of structured data. It provides high availability of data with no single point of failure (uses Dynamo-style replication). It is a key-value as well as column-oriented database, has elastic scalability, eventually consistent and fault-tolerant and is created by Facebook. It gives quick response time by increasing the throughput as the number of nodes increases in the network (linear scalable). The structured, semi structured and unstructured of data formats are included and it dynamically allocates the datatypes and update the changes to the data structure created. It supports the ACID property of the database and it is designed to run on cheap commodity hardware. Cassandra Query Language (CQL) is used for interacting with the database. Each node in the cluster can read and write the requests regardless of the location, if any node gets down the data will be taken from the other node. There are many advantages of Cassandra databases which provide its main characteristics in creating the useful application and in designing most of the useful application. Cluster can read and write the requests regardless of the location, if any node gets down the data will be taken from the other node.



Architecture of Apache Cassandra

DataStax is a software which has highly scalable and has capability of intelligently distributing load across huge clusters of machines spanning multiple data centres. DataStax Enterprise (DSE) motivates your ability to deliver real-time value at epic scale by providing a comprehensive and operationally simple data management layer with a unique always-on architecture built on Apache Cassandra. DataStax Enterprise software provides the distributed, responsive and intelligent foundation to build and run and function cloud applications. So using DataStax we can develop huge amount of data and it can be stored in cloud for spatial data requirements.

What is actually geohashing? Geohashing is mainly an adventure, comprising journeys to random places within a given area. It is also Spontaneous Adventure Generator. The coordinates is generated for each 1°×1° latitude/longitude zone (known as a graticule) in the world. The coordinates can lie anywhere in the graticule — in the forest, in a city, on a mountain, or even in the middle of a lake! or anywhere. Everyone in a given region gets the same set of coordinates relative to their graticule. It's actually a type of search used for identifying in geospatial data for locating the place for setting up a solar system.

We are going to create a data for solar generated system in which we are going to use Apache Cassandra database which holds the large volume of data and the spatial data for construction of solar system to be suited to be in the different places for its advantages and minimizing the power consumption. Based on the statistics of spatial data we are using Big Data Analytics in this system which improvise the manual statistics and can bring huge amount of data to us. Using DataStax we can store large amount of data in cloud which can be useful to retrieve it from anywhere based on the requirement of the spatial for a compactable solar generated system and its system are based full on the Apache Cassandra database which are mainly used for searching large amount of data. We are using the correlation method for searching to make it more interesting to find the place where humidity, rainfall and wind matters. We have derived a formula(**A**.) in which it would be easy to locate the area where the setup can be most accurately used and can give huge amount of consumption. Using this cross correlation method we can use geohashing technique for making the search far better. Depending on the formula for cross correlation method we can do a wide range of statistics of data but because of using geohashing in

terms of cross correlation can be done as searching for a particular given set of area and a specific part or location we can construct the solar system which can meet huge needs and high demands.

A. cross correlation

∑=temperature + air – humidity / wind

For Solar power wind is inversely proportional r

Temperature, humidity will increase when wind decrease it will be inversely proportional.

B. Illustration –

DATABASE CASSENDRA KEYSPACE OPERATIONS

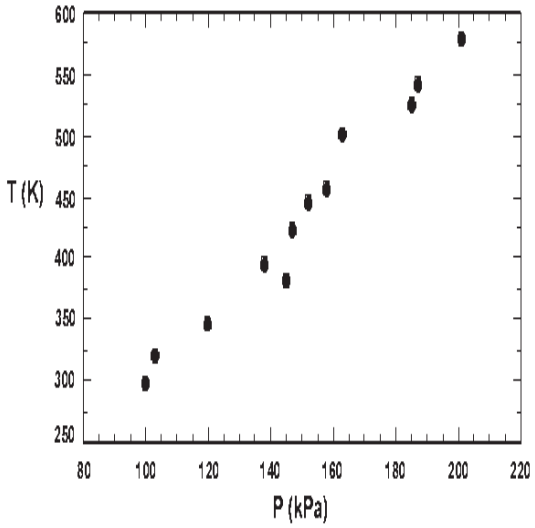
A key space is a namespace which defines the data replication on the N number of nodes, each cluster contains one key space in each node. To create key space for the solar data.
In Cassandra we write the CQL codes as

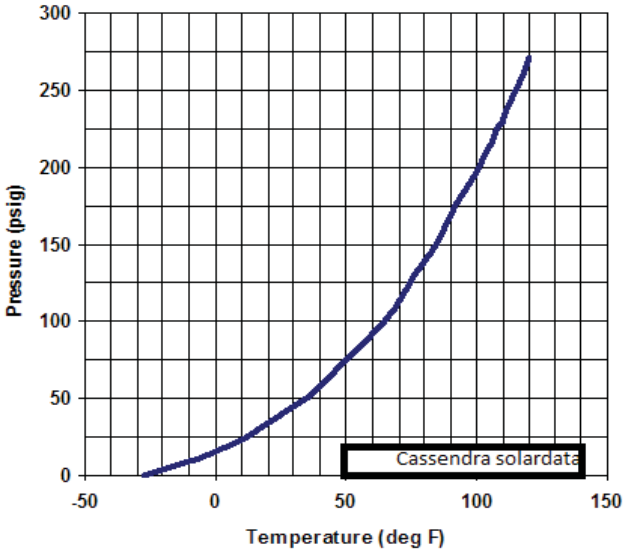
```
Cqlsh.>  CREATE KEYSPACE solar data
WITH
replication ={ 'class :'SimpleStrategy',
'replication_factor':3};
```

For the attributes Wind Speed, Air Pressure,
Wind Height, Wind period and Air Temperature, the database is created.

```
Cqlsh> USE solardata;
Cqlsh:solardata>; CRFEATE TABLE temp(
win_spint PRIMARY KEY,
air_prsint,
win_hgt variant,
win_prdint,
air_tmp.
);
```

For the above parameters graph the correlation
graphs are got.





IV. CONCLUSION

This paper address about the power generation through various part of the country and in this paper we have taken a small of our location for generation of solar generated system and it is built on the correlation of humidity, air pressure, temperature and the system is a very well applicable system which can generate data based on the various terms like Big Data Analytics, Cassandra database, DataStax, cross correlation and geohashing method. It consumes lesser power too and eliminates the cost in electricity too.

REFERENCES

[1] H.E. Gad, Hisham E. Gad, "Development of a new temperature data" It creates an acquisition system for solar energy applications, Renewable Energy, (2015).
[2] Huan Long, Zijun Zhang, Yan Su, "Analysis of daily solar power prediction with data-driven approaches". Applied Energy, 126, 29-37, (2014)
[3] TokhirGafurov, Julio Usaola, Milan Prodanovic. "Incorporating spatial correlation into stochastic generation of solar radiation data, Solar Energy, 115, 74-84, (2015).
[4] Rafael Moreno-Saez, Llanos Mora-Lopezx, "Modelling the distribution of solar spectral irradiance using data mining techniques", Environmental Modelling & Software, 53, 163-172, (2014).
[5] Ji Li,YuanyuanYang,CongWang, "Mobility assisted datagathering with solar irradiance awareness in heterogeneous energy replenish able wireless sensor networks", Computer Communications, 69, 88-97, (2015).
[6] Rafael Moreno-Saez, Llanos Mora-Lopez, "Modelling the distribution of solar spectral irradiance using data mining techniques", Environmental Modelling & Software, 53, 163-172, (2014).