Chapter

Solar-Wind Energy Assessment by Big Data Analysis

Vikas Khare and Aaquil Bunglowala

Abstract

Big data refer to the massive datasets that are collected from a variety of data sources for business needs to reveal new insights for optimized decision-making. The solar and wind energy system is the modernization of electrical energy generation systems due to the pollution free nature and the continuous advancement of photo-voltaic and wind turbine system technologies. In the solar and wind energy surroundings, the application of big data analysis based decision-making and control are mainly in the following three aspects: data stream side management, storage side management and load side management. The objective of this research is to present a technological framework for the management of large volumes, variety, and velocity of solar system related information through big data tools such as Hadoop to support the assessment of solar and wind energy system. The framework includes a modeling of system, storage, management, monitoring and forecast based on large amounts of global and diffuse solar radiation and wind energy system. This chapter also includes market basket model, the concept of solar and wind depository and application of the Map Reduce algorithm.

Keywords: solar energy system, wind energy system, big data, Hadoop, Map Reduce

1. Introduction

Big data refer to the massive data set that are collected from a variety of data sources for implementing solar energy and wind energy system at a particular place and to reveal new insights for better decision-making. Based on different data analysis of the any study area, it is observed worldwide a lot of places are available where solar radiation and wind velocity available in abundance. The accessibility of non-conventional energy resources at any place is a significant feature to develop the solar energy and wind energy system for different purpose. Big data capable to generate values related to solar energy system and wind energy system for the storage and processing of very large quantities of information that cannot be analyzed with traditional computing techniques. Big data are categories into three part volume, velocity, and the variety and assess the pre-feasibility assessment with the help of these three features which is shown in **Figure 1**.

With phenomenal development in the field of electricity generation through renewable energy system, solar and wind power data sources have risen sharply. Exhaustive use of wind power big data can provide an effective way for safe operation of high quality power supply of a wind energy system. The effective goal

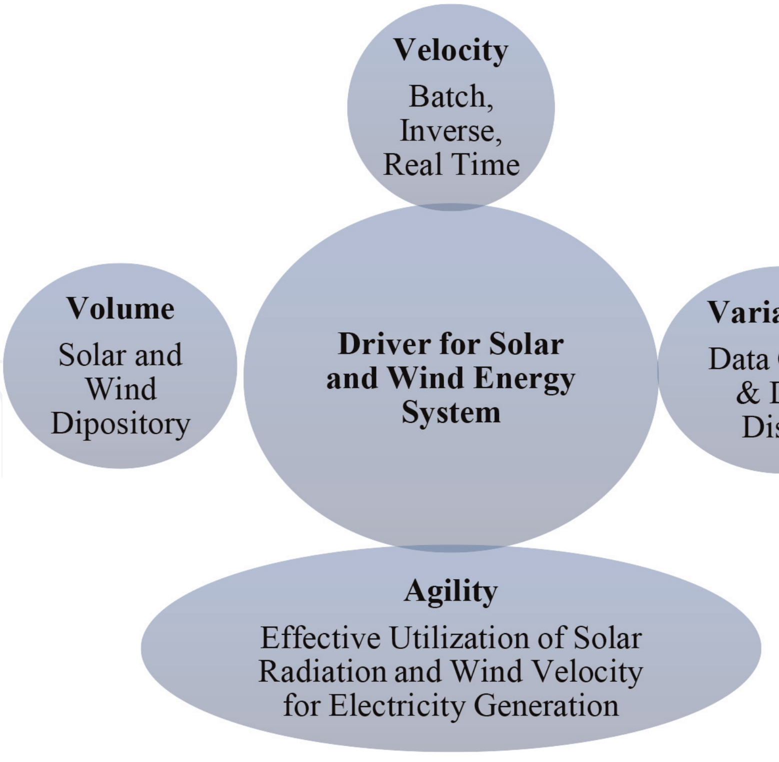

Figure 1.
Big data driver for solar and wind energy.

of big data research in the field of wind energy system is to "become aware of unawareness" and objectively there are well known facts and unknown facts which affect the working of wind energy system. Big data helps us in learning what we do not know and is done in two ways: from small to big and from big to small. It is necessary to develop a proper insight view of big data in the field of wind energy system and model out the wind system based on big data. Most big data applications in wind system consists two parts data corral and data dispel, in a data corral lot of information about the wind system project is collected and stored in wind cargo space, which is used to store the data and is utilized for further suitable application [1, 2].

Big data creates data from the processing of a large number of information which is related to digital information and such type of digital information cannot assess by normal computing processes. Wal-Mart handles near about 1.5 million consumer transaction in half an hour. Facebook handles 75 billion photos from its user base. Twitter develops 10 TB of data per day. By the big data analysis all of these processes are done in 1 week, which collect and process infinite number of data in a week, maybe without big data it was done in approximate 1 year. Hejazi et al. [3] described challenges and solution of electrical power system by big data analytics. In this paper different assessment method of power system is assessed by big data analytics. The paper also provides a holistic outline, classification and concise discussion on the technical approaches, research opportunity and application areas for energy big data analytics. Yao et al. [4] explained application of big data in the field of smart meter. In this article running path of smart meter is assessed by collection of lot of data, which is related to the operation and maintenance losses and their power information collection data. Shyam et al. [5] described assessment of smart grid through Apache spark based big data analytics. Apache Park is used to store the data related to the smart grid and respond according to the consumer requirements. Yang et al. [6] described different aspects and parameter of power to the gas energy system through big data analytics. In this paper big data assessment is also used for the operation and assessment of wind energy system.

Bersa et al. [7] predicted application of big data in power system. This paper also represented different technical parameter such as operational efficiency, losses and other parameter which is related to the power system is assessed by big data analytics. Percuku et al. [8] presented consumer load forecasting of electrical transmission system through big data analysis. The purpose of this article is to analyze a framework which is designed by Neo4j graph technology, which is a part of big data NoSQL data storage system. The Big data method is also used for prediction of load forecasting. Junaidi et al. [9] assessed electrical energy system through big data analytics. In this chapter, phasor measurement unit, modern days automated electric meter, and atmospheric measuring unit are assessed by big data analysis. Hangxun et al. [10] analyzed measuring power system quality by big data analytics. In this article a huge number of data is used to predict some power quality parameter such as sag and tension of power system, voltage, and current imbalance, which create lots of problems during the transmission and distribution of electrical energy. Huang et al. [11] analyzed different technical aspects of electrical energy system through big data. In this assessment author analyzed transient and steady state analysis of energy apparatus such as transformer, induction motor and synchronous condenser, etc. Wanxing et al. [12] described reactive power analysis of power system by big data topology. In this paper 8760 hours data of voltage and current by sending and receiving end is utilized for assessment of the reactive power of the electrical energy system. Zhan et al. [13] explained different prospects of smart grid or micro grid with the help of big data analytics. Guan et al. [14] described security and stability analysis of bulk power system. In this paper author assessed equal area criteria and other stability parameter with the help of huge number of data of load angle and consumer demand. Qing et al. [15] described impact of big data on electric power industry. The information in a specific sense has turned into another financial resource class. Step by step instructions to utilize big data to make more esteem will be another undertaking looked by all businesses, particularly the power business. In this paper, big data stage model of the power business and atomic power review and structure industry is planned.

This chapter is different from all above mentioned research in the application of big data in electrical power system. In this chapter, solar-wind energy system is analyzed through the big data analysis. Pre-feasibility analysis, modeling and financial analysis of solar-wind energy system is assessed by big data analysis. Market basket model and data retrieval is also used in this paper for assessment of solar-wind energy system.

2. Pre-feasibility assessment of solar-wind energy system by big data analysis

Electricity generation through solar and wind energy system is mainly depends on the amount of solar radiation and wind velocity at a particular site area. When the millions of solar radiation (kWh/m^2/day) and wind velocity (m/s) data are split into batches, sparse, interval and real time data, then this is the terms related to velocity of big data. When we consider a variety of big data on solar radiation and wind velocity data includes different types of data in the form of 3D data, audio, video, and unstructured text. When we store data in solar or wind depository, it is stored in the form of kilobytes, megabytes, gigabytes, terabytes, etc., and this represents the data volume of solar and wind energy system. For the feasibility assessment of solar and wind energy system in the study area data assessment of solar radiation, wind velocity, and hourly load consumption is categories in following two ways:

2.1 Modus operandi

Data that represents the real time status of solar data, wind data performance assessment, and loading of solar and wind energy equipment. This is the very basic information related to pre-feasibility assessment used by the system engineer to assess and manage the solar and wind power plant.

2.2 Non-modus operandi

A data file consists of data elements such as longitude, latitude and other data, and it also consists of a vendor specific property. Ancillary climatic data which is required in the database are in the form of air temperature, wind speed data, and water vapor content. Data of air temperature is used for calculating the PV module temperature, which is subsequently used to calculate PV power, after that wind cools the PV modules and modifies the temperature and hence the PV power. On the other hand, wind power is also depends on the specification of wind turbine, hub height, etc. **Figure 2** shows the number of applications of big data in solar and wind energy system.

The field of electricity generation through solar and wind energy source is the highly competitive world today and the downtime equates to real dollars lost and is deadly to company reputation. The NoSQL database environment is able to provide solar and wind system related data continue with operations without data loss and it is working like a solar and wind depository. Systems updates can be made dynamically without having to take the database offline. **Figure 3** shows the content and capacity of solar and wind depository. Solar and wind depository is divided into four parts for proper pre-feasibility assessment of solar and wind energy system and these are ERP, CRM, WEB and big data and data range consider from MB to PB. ERP module content data related to capital, replacement and operation and maintenance cost of photo-voltaic panel, generator, inverter, and battery system. CRM is the collection of nontechnical data which is related to the vendor information, client data, tender information, different types of solar and wind energy company data and data related to government policy, government subsidy, government renewable energy framework which is essential to develop a solar and wind energy system at study area. After the collection of all the three module information last one module is called big data module of solar and wind depository. It is a brain and heart of solar

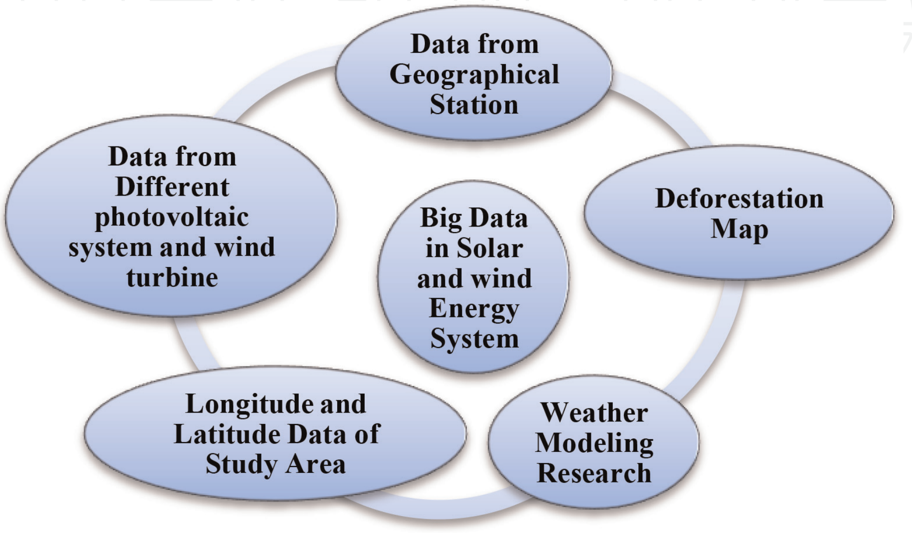

Figure 2.
Application of big data in solar and wind energy system.

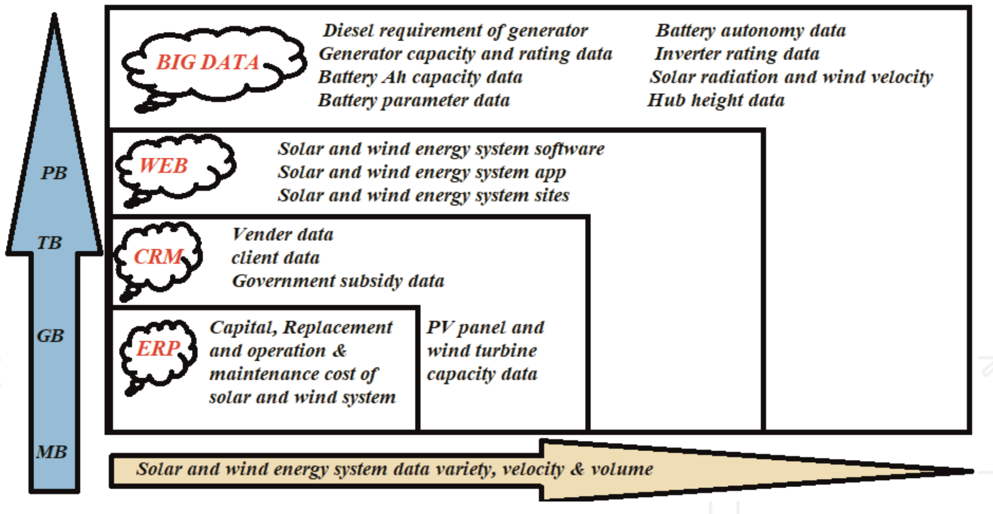

Figure 3.
Layout of solar and wind depository.

and wind depository because it accumulates all the data which is necessary to develop a solar and wind energy system in an efficient manner [16, 17].

The capability to read and write solar and wind system related information and in that position solar radiation and wind velocity input and electrical energy output physically occurs is termed as location transparency which is affected by longitude and latitude of the particular location. In a solar and wind depository document database key access pattern access data through a single key and navigates to another document through a related key.

3. Solar-wind energy system by Hadoop environment

Hadoop ecosystem is a framework of various types of complex and evolving tools and components. Some of these elements may be very different from each other in terms of their architecture. If we apply the concept of Hadoop ecosystem in the field of solar and wind energy system, then it is categorized into four types, data management, data access, data processing, and data storage. **Figure 4** shows Hadoop solar and wind energy ecosystem is also categorized into three types of pre-feasibility assessment in terms of longitude and latitude data assessment, solar radiation and wind velocity data assessment and load demand data assessment. In Hadoop system data management is done by Oozie, Chukwa, Flume, and Zoo-keeper, where Oozie is an open source Apache Hadoop service used to manage and process submitted tasks. Hadoop works by the divide and conquer approach. Once a problem is divided, it is approached and processed by using distributed and parallel processing technique across Hadoop cluster. Big data problems are approached with distributed applications and Zookeeper helps in coordinating all the elements of the distributed applications. Flume aids in transferring large amounts of data from distributed resources to a single centralized repository. It is robust and fault tolerant and efficiently collects, assembles and transfer data. Apache Chukwa is an open source information gathering framework for checking the enormous conveyed frameworks. Apache Chukwa is based over the Hadoop Distributed File System (HDFS) and Map/Reduce structure and acquires Hadoop versatility and heartiness. Apache Chukwa additionally incorporates a flexible and an incredible toolbox for showing; observing and dissecting results utilize the gathered information [18, 19].

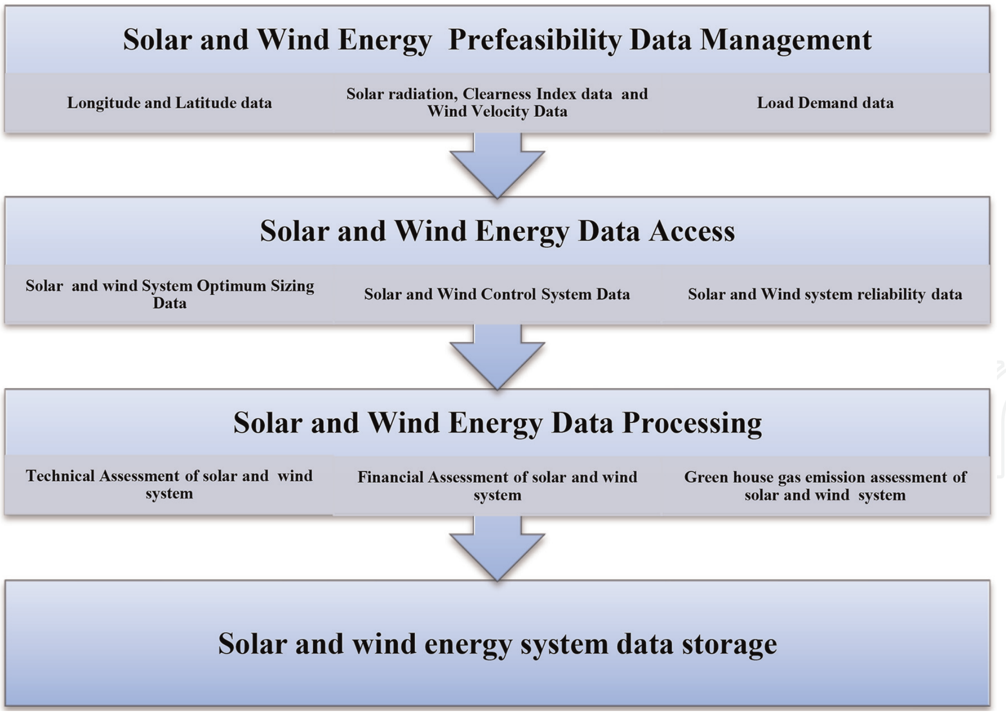

Figure 4.
Hadoop ecosystem for solar and wind energy system.

Data management of solar and wind energy system through the Hadoop system is always part of resource characterization and site assessment. In solar and wind energy system site examination is normally completed previously mentioned to utilization to set up the thorough physical encompassing for a specific sustainable power source venture, with the following objectives:

- To analyze the solar and wind energy generation life cycle of the given plan.
- To assess moderate global and diffuse solar radiation and high and low wind velocity conditions.
- To describe the clearness index and a hub height of the site area.

Objective and nature of resource assessment, solar and wind energy conversion characteristics, site condition, constraints, and physical boundary of assessment is the part of the solar and wind project description and which is the first step of resource allocation. If the project is in its initial stages and consists mainly of site screening, the resource assessment should be qualified as regional if the area of study is very large and incorporates many potential sites, whole country, or a large portion thereof.

4. Energy aware cluster node management of solar-wind energy system

Cluster node management is one of the parts of big data analysis and a cluster manager usually is a backend graphical user interface or command line software that runs on one or all cluster nodes. The cluster manager works together with a cluster management agent. A cluster is the process of making a group of abstract objects into classes of similar objects. The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.

Clustering method in solar and wind energy system:

- **Partitioning method**:

Suppose we are given a database of pre-feasibility assessment of solar and wind energy system which is represented by "n" database and the partition method construct 'k' partition of data. Then, a pre-feasibility assessment of the data is partially in the form of project description, estimation of current speed, result presentation, data analysis, and available and extractable energy data; so, "n" database is partitioned into k = 5 partitioned.

- **Hierarchical methods**:

In the hierarchical method of clustering, we identify step by step process of data gathering. In this method data collection and decomposition is done in two ways, first one is agglomerative and another one is divisive approach. In the agglomerative process merging the object and group that are close to one another and in the case of solar and wind energy system data are collected in the following manner:

1. Collection of pre-feasibility data of solar and wind energy system
2. Collection of data for modeling of solar and wind energy system
3. Collection of data for controlling of solar and wind energy system
4. Collection of data for reliability assessment of solar and wind energy system

In divisive approaches, we start with all of the objects in the same cluster and in the continuous iteration, a cluster is split into smaller clusters [20, 21]. According to the divisive approach agglomerative data is distributed and divide into following manner.

1. Collection of prefeasibility data of solar and wind energy system
    i. Location of the site
    ii. Geographical condition of the site
    iii. Data on solar radiation
    iv. Data on wind velocity
    v. Data of temperature
    vi. Data of rain fall
    vii. Data on consumer demand
2. Collection of data for modeling of solar and wind energy system
    i. Data of electricity required
    ii. Data of specification of solar panel and wind turbine
    iii. Data of specification of DC generator and AC generator
    iv. Data of specification of battery

3. Collection of data for controlling of solar and wind energy system

    i. Data of different types of errors

    ii. Data of different control strategies of solar and wind energy system

    iii. Data for stability analysis of solar and wind power plant

4. Collection of data for reliability assessment of solar and wind energy system

    i. Data of failure distribution model of solar and wind energy system

    ii. Data of time dependent failure model of a solar and wind energy system

    iii. Data of constant failure rate model

5. Basic big data measures for solar and wind data, text retrieval

In the text retrieval we need to check the accuracy of the data because it is very necessary part in the development of solar and wind power plant. Let the set of solar and wind energy system documents relevant to a query be denoted as (relevant solar and wind data) and the set of retrieved documents as (retrieved solar and wind data). The set of solar or wind documents that are relevant and retrieved can be denoted as.

$$(\text{Relevant solar or wind data}) \cap (\text{Retrieved solar or wind data})$$

This can be shown in **Figure 5** in the form of a Venn diagram as follows:

There are three fundamental measures for assessing the quality of solar and wind energy system retrieval:

- Precision
- Recall
- F-Score

Precision: precision is the percentage of solar and wind energy system retrieved documents that are in fact relevant to the consumer query. Precision of solar and wind power plant can be defined as:

$$Precision\ of\ renewable\ energy\ system\ data = \frac{(Relevant\ solar \vee wind\ data) \cap (Retrieved\ solar \vee wind\ data)}{(Retrieved\ solar \vee wind\ data)}$$

Recall: recall is the percentage of solar and wind energy system documents that are relevant to the consumer query and were in fact retrieved. Recall is defined as:

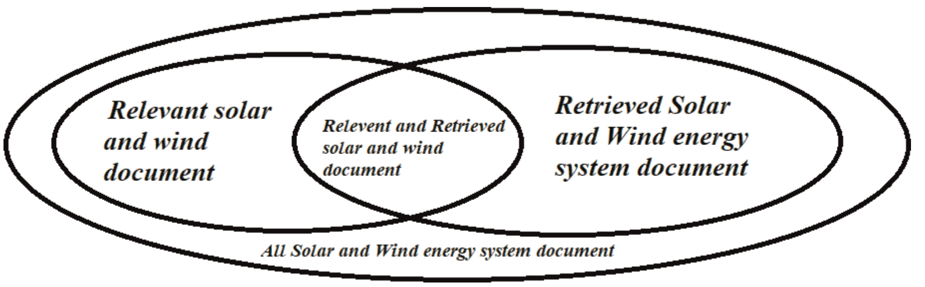

Figure 5.
Solar and wind text retrieval.

DOI: http://dx.doi.org/10.5772/intechopen.87166

$$\text{Recall of renewable energy system data} = \frac{(\text{Relevant solar} \vee \text{wind data}) \cap (\text{Retrieved solar} \vee \text{wind data})}{(\text{Relevant solar} \vee \text{wind data})}$$

F-Score: F-score is the commonly used parameter of statistical analysis and the given observation retrieved system often needs to exchange information for precision or accuracy purpose. F score is defined as the harmonic mean of recall or precision as follows:

$$F\text{-}Score = \frac{Recall \times Precision}{0.5 \times (Recall + Precision)}$$

6. Application of Map Reduces in solar and wind energy system

The Map Reduce algorithm contains two important tasks shown in **Figure 6**, namely Map and reduce, which is used to provide essential framework for any task and also reduce the time interval of completion of any task.

- The essential modeling and framework is done by Mapper Class
- To reduce the number of steps for completion of any events is done by Reducer Class.

If we assess solar-wind energy system by Mapper Class, then first is carrying the input parameter of solar or wind energy system, then it analyze further provide framework according to the certain parameter and sorted according to the requirement. The output of Mapper class is used by the Reducer class as an input parameter of solar or wind energy system, which in revolves searches identical pairs and reduces them.

Sorting: sorting is the key step of Map Reduce algorithms, which is used to analyze the parameter according to the given constraints and manipulate the data according to the requirements. Map Reduce trappings sorting algorithm to without human intervention sort the final key-value pairs from the mapper by their keys.

- Sorting methods are the first step of the mapping class.
- In the second step tokenizing the parameter which collects from the first step.
- To accumulate transitional keys, the Mapper class is designed the framework by another comparator class.
- The position of transitional parameters for a given Reducer is routinely sorted by the Hadoop system to form parameters (K2, {V2, V2, ...}) before they are presented to the Reducer.

Searching: looking assumes a significant job in the Map Reduce calculation. It helps in the combiner stage (discretionary) and in the Reduce stage. Give us a

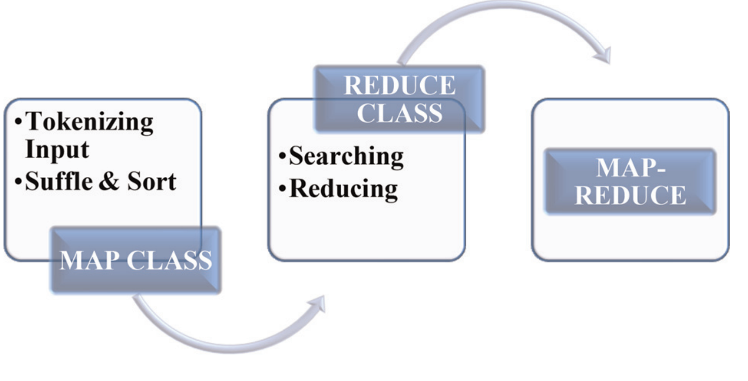

Figure 6.
Task of map-reduce algorithm.

chance to attempt to see how Searching functions with the assistance of a model. The following example shows how Map Reduce employs a searching algorithm to find out the details of the solar radiation and wind velocity and which country draws the highest solar radiation and wind velocity in a given atmospheric dataset.

- Let us assume we have solar radiation or wind velocity' data in four different files—W, X, Y and Z. Let us also assume that there are duplicate solar radiation or wind velocity' records in all four files because of importing the solar radiation or wind velocity data from all database tables repeatedly. See the following illustration and keep in mind value of solar radiation (kWh/m^2/day) and wind velocity (m/s) both are lies in the 0–1.

Country wise site area, average solar radiation or wind velocity	Country wise site area, average solar radiation or wind velocity	Country wise site area, average solar radiation or wind velocity	Country wise site area, average solar radiation or wind velocity
India, 5.2	China, 6	India, 5.2	India, 5.2
Japan, 5	Japan, 5	USA, 9	Japan, 5
Germany, 7	Germany, 7	Germany, 7	New Zealand, 8
Austria, 5	Austria, 5	Austria, 5	Austria, 5

The Map phase processes, each input file and provides the solar and wind energy system data in key-value pairs (<k, v>:<Site area, solar radiation or wind velocity>). See the following illustration.

Country wise site area, solar radiation or wind velocity	Country wise site area, solar radiation or wind velocity	Country wise site area, solar radiation or wind velocity	Country wise site area, solar radiation or wind velocity
<India, 5.2>	<China, 6>	<India, 5.2>	<India, 5.2>
<Japan, 5>	<Japan, 5>	<USA, 9>	<Japan, 5>
<Germany, 7>	<Germany, 7>	<Germany, 7>	<New Zealand, 8>
<Austria, 5>	<Austria, 5>	<Austria, 5>	<Austria, 5>

The combiner phase (searching technique) will accept the input from the Map phase as a key-value pair with site area and solar radiation or wind velocity. Using searching technique, the combiner will check all the country wise site area to find the highest solar radiation or wind velocity availability in each file. See the following snippet.

```
<k: country wise site area, v: solar current or wind velocity>
Max = the value of solar current or wind velocity of a first country. Treated as max solar current or
wind velocity
if(v(second country).solar_current or wind_velocity > Max)
{
   Max = v(solar_current or wind velocity);
}
else
{
   Continue checking;
}
```

The expected result is as follows:

<USA, 9>	<New Zealand, 8>	<Germany,7>	<China, 5>

Reducer phase: form each file, you will find the highest solar radiation or wind velocity. To keep away from excess, check all the <k, v> matches and dispose of copy passages, assuming any. A similar calculation is utilized in the middle of the four <k, v> sets, which are originating from four info documents. The last yield ought to be as per the following:

<USA, 7>

Indexing: regularly indexing is utilized to point to a specific information and its location. It performs cluster ordering on the information records for a specific Mapper. The ordering strategy that is ordinarily utilized in Map-Reduce is known as reversed file. Web crawlers like Google and Bing utilize reversed ordering method. Give us a chance to attempt to see how Indexing functions with the assistance of a straightforward model. The following text is the input for inverted indexing. Here X [0], X[1], and X[2] are the file names and their solar radiation data are in double quotes [28].

X[0] = "5.2, 7.2, 6.3, 9.4"
X[1] = "9.4, 7.2, 8.1"
X[2] = "7.2, 6.3, 8.8, 4.5"

After applying the Indexing algorithm, we get the following output of solar radiation:

"7.2":{0,1,2}
"5.2":{0}
"6.3":{0,2}
"4.5":{2}

Here "4.5":{2} implies the term "a" appears in the X[2] file. Similarly, "7.2": {0, 1, 2} implies the term "is" appears in the files X[0], X[1], and X[2].

TF-IDF: TF-IDF is a substance getting ready estimation which is short for Term Frequency-Inverse Document Frequency. It is one of the fundamental web examination estimations. Here, the term 'repeat' suggests the events a term appears in a file.

Term frequency (TF): it gauges how much of the time a specific term in a record. It is determined by the occasions a parameter shows up in a report partitioned by the absolute number of parameters in that record.

TF(He) = (number of times term the '7.2' appears in a document)/(total number of terms in the document)

Inverse document frequency (IDF): it measures the noteworthiness of a term. It is controlled by the amount of reports in the substance database separated by the amount of chronicles where a specific term appears. While figuring TF, all of the terms are considered likewise huge. In this manner we have to know the regular terms while scaling up the uncommon ones, by finishing the accompanying:

IDF(He) = log_e(total number of documents/number of documents with term '7.2' in it)

The algorithm is explained below with the help of a small example.

In a particular site of India, in pre-feasibility assessment data containing 8760 data on solar radiation in a year, wherein solar radiation 7.6 kWh/m^2/year appears 72 times in a year. Calculate the terms frequency and inverse document frequency of the solar data.

$$\text{Solution}: Terms\ Frequency(7.6) = \frac{Number\ of\ \times term\ the\ \text{'7.6'}\ appears \in a\ document}{Total\ number\ of\ terms \in the\ document}$$

Terms frequency (7.6) = 72/8760 = 0.008

$$Inverse\ Document\ Frequency(7.6) = \log_e \frac{8760}{72} = 2.08$$

7. Market basket model in solar and wind energy system

It shows that many relationships between two concepts "items" and "baskets" and each basket consist of a number of items. In the context of solar and wind energy system market basket model is considered in three ways, pre-feasibility assessment, modeling, and reliability assessment and it contain a lot of items which is related to the solar and wind energy system and shown in **Figures 7** and **8**. Basket of pre-feasibility assessment contains a lot of information because prior to installation and operation, the pre-feasibility study of solar and wind energy system should be done. In solar and wind energy projects an initial study undertaken to determine whether it is worthwhile to continue to the feasibility study stage. A precise feasibility study should provide a chronological background of the projects. In addition to climate condition of the application site, availability of solar and wind energy sources, the potential of solar and wind energy sources, load demand of application sites are included to find out the best location to develop a solar and wind renewable energy system. Generally feasibility precedes technical development and project implementation. It must therefore be conducted with a balanced approach to provide information upon which decisions can be based. Modeling basket contains different technique which is used to model the solar and wind energy system such as through HOMER, fuzzy logic and analytical technique. Modeling is the first step to design a system according to the different parameter and constraint. Modeling of solar and wind energy system is based on annual cost, battery autonomy function, sizing criteria and ecological statistical factor. Step by step optimization practice is used to find out the efficient result of the solar system model. Third and most important basket is reliability basket which is very important in the recent scenario because it evaluate the failure rate of individual component and overall solar energy system [22, 23].

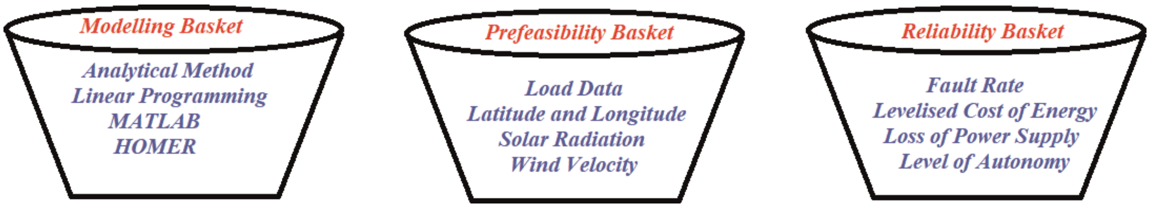

Figure 7.
Market basket of solar energy system.

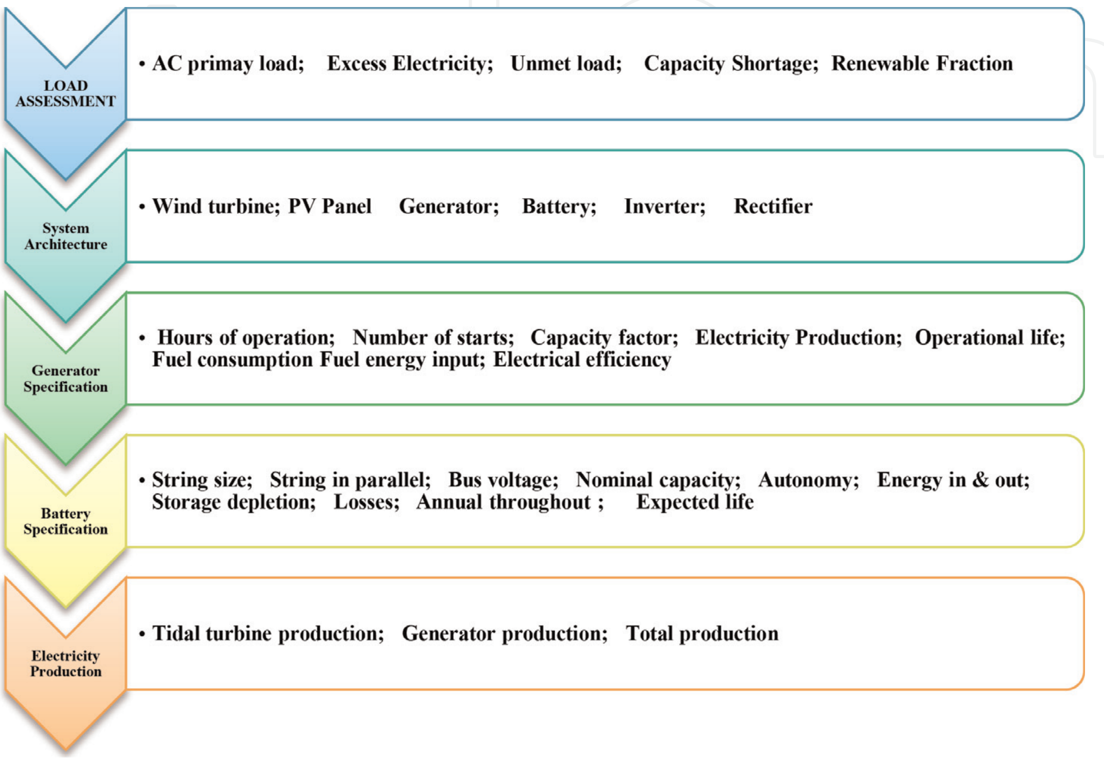

Figure 8.
Market basket model for optimum sizing and modeling of solar and wind energy system.

- **Market basket model based frequent item set mining for solar and wind energy system**:

Let K = {k_1...........k_n} be a set of parameters. Let 'E' be the task relevant data, be a set of database parameter where each parameter T is a set of parameters such that $E \subseteq K$. Each parameter is associated with an identifier, called TKE. Solar and wind energy system parameter is referred to as a parameter set. We can call a parameter K a "frequent item set" only if its support count is sufficiently large. We prescribe a minimum support 's' and any K which has support greater than or equal to 's' is a frequent parameter set.

If we consider parameters of solar and wind energy system = {solar radiation (1), Sea surface temperature (2), Wind velocity (3), Temperature (4), Rain fall (5)}; Minimum support s = 3.

Transactions

1. T1 = {1,2,4}
2. T2 = {1,3,5}
3. T3 = {1,4}

4. T4 = {2,5}
5. T5 = {1,3,4}
6. T6 = {1,2,4,5}
7. T7 = {2,4,5}
8. T8 = {4,2}

Frequent parameters: {1}, {2}, {4}, {5}, {1, 4}, {2, 4}, {5, 2}

- **Associate rule mining in solar and wind energy system**:

The main purpose to discovering frequent parameters from a large dataset is to discover a set of "If-then" rules called Association rules. The form of association rules is K-j where K is a set of parameters of solar or wind power plant. Let K = {k_1, k_2.....k_n} be a set of n distinct attributes of solar and wind energy system which is also called literals of renewable power plants. Let 'E' be a database of solar or wind energy system, where each record of supply and demand side, T has a unique identifier, and contains set of parameters from the set K. An association rules is an implication of the form $X \rightarrow Y$, where $X, Y \subseteq K$ are item sets and $X \cup Y$ is a frequent item set. Thus the frequency of occurrence in X∪Y is at least equal to the minimum support s.

To test the reliability of solar or wind energy system, we try to define "confidence" of the rule which presents performance parameter of solar or wind power plant. Let X→Y be an association rule. The confidence of the rule which assess performance of solar or wind power plant defined as the fraction of the solar or wind power plant parameters that supports the rule among those that support the antecedent:

$$\mathbf{Confidence}\ (\mathbf{X}\rightarrow Y) \coloneqq \mathbf{P}\ (\mathbf{Y}\setminus\mathbf{X}) = \mathbf{support}\ (\mathbf{X} \cup Y)/\mathbf{support}\ (\mathbf{X})$$

The confidence of the rule indicates the degree of correlation between the certain parameter of solar or wind energy system. Such that value of solar radiation also depends on the value of wind velocity, so it is necessary to determine correlation between solar radiation and wind velocity.

Consider a small database with four parameters of combined solar and wind energy system K = {wind turbine specification (TT), Photo-voltaic Specification (B), Generator Specification (G), Consumer demand in Kw (CD)} and four transactions of these parameters shown in **Table 1**. **Table 2** shows all the parameter for K. Suppose that the minimum support and minimum confidence of an association rule are 40 and 60%, respectively which is shown in **Table 3**.

- **Framework for frequent parameter mining**:

The market baskets are also organized in memory. Generally, market basket data are stored in a file basket by basket. Generally, market basket data of solar and wind energy system are stored in a memory basket by basket. **Figure 9** shows Basket of Parameter Mining.

7.1 Monotonicity and Apriori algorithm property of solar and wind energy parameters

Given a database of transaction '**E**' over '**K**' and two sets X, Y⊆K, Then

Transaction ID	Parameters
T1	Tidal turbine, Photo-Voltaic, Generators
T2	Photo-Voltaic, Generators, Consumer Demand
T3	Photo-Voltaic
T4	Tidal turbine, Photo-Voltaic

Table 1.
Transaction of solar and wind energy system database.

Parameters	Support 's'	High/Low
Wind turbine	50%	High
Photo-Voltaic	100%	High
Generators	50%	High
Consumer Demand	25%	Low
Wind turbine, Photo-Voltaic	50%	High
Wind turbine, Generator	25%	Low
Wind turbine, Consumer Demand	0%	Low
Photo-Voltaic, Generator	50%	High
Photo-Voltaic, Consumer Demand	25%	Low
Generator, Consumer Demand	25%	Low
Wind turbine, Photo-Voltaic, Generators	25%	Low
Wind turbine, Photo-Voltaic, Consumer Demand	0%	Low
Wind turbine, Generators, Consumer Demand	0%	Low
Photo-Voltaic, Generators, Consumer Demand	**25%**	**Low**
Wind turbine, Photo-Voltaic, Generators, Consumer Demand	**0%**	**Low**

Table 2.
Support for parameters in table and large parameters with a support of 40%.

Rule	Confidence	Rule-hold
Wind turbine⇒ Photo-Voltaic	**100%**	**Yes**
Photo-Voltaic ⇒Wind turbine	**50%**	**No**
Photo-Voltaic⇒ Generators	**50%**	**Yes**
Generators ⇒Photo-Voltaic	**100%**	**No**

Table 3.
Confidence of some association rule where confidence interval = 60%.

$X, Y \subseteq K \Longrightarrow \text{support } (Y) \leq \text{support } (X)$.

In the Monotonicity property of support also allows us to compact the information about frequent solar and wind energy parameters and shown in **Table 4**. First, some definition is given below:

1. Solar or wind energy system parameters closed if none of its immediate parameter has the same count as the parameter.

2. Solar or wind energy system parameter is closed frequent if it is frequently and closed.

3. Solar or wind parameter is maximal frequent if it is frequent and none of its immediate superset is frequent.

Financial analysis of solar or wind energy system is consist three parameters = {Capital cost, Replacement cost, Operation and Maintenance cost} and the following baskets:

1. {Capital cost, Replacement cost}

2. {Capital cost, Replacement cost}

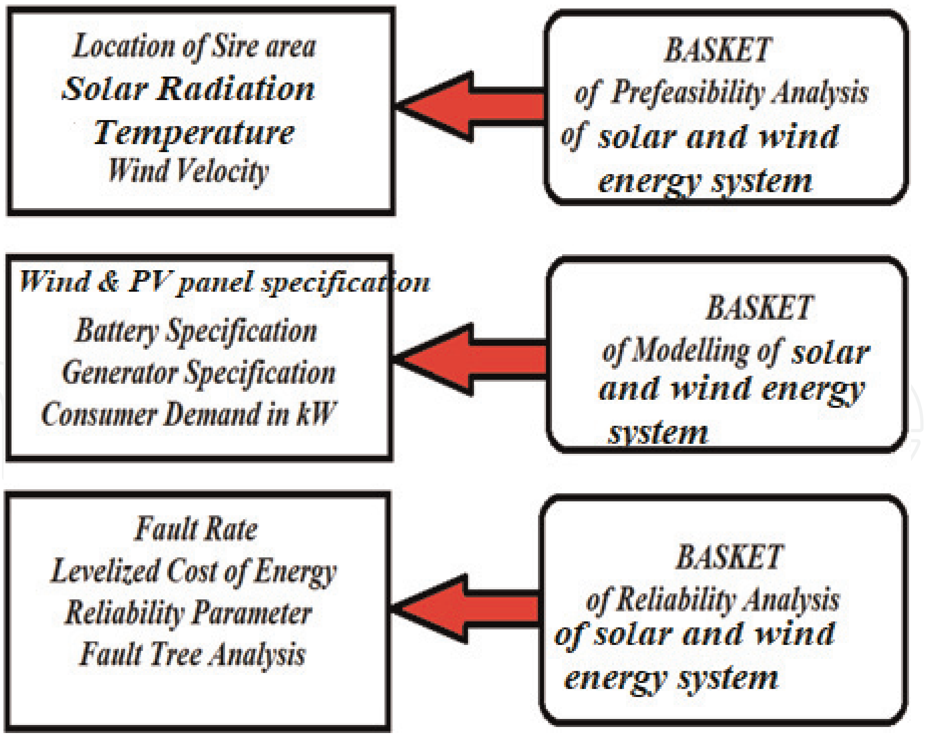

Figure 9.
Basket of parameter mining.

Parameter	Count	Frequent	Closed	Closed Frequency	Maximum Frequency
Capital cost	4	Yes	No	No	No
Replacement cost	5	Yes	Yes	Yes	No
Operation & Maintenance cost	3	Yes	No	No	No
Capital cost, Replacement cost	4	Yes	Yes	Yes	Yes
Capital cost, Operation & Maintenance cost	2	No	No	No	No
Replacement cost, Operation & Maintenance cost	3	Yes	Yes	Yes	Yes
Capital cost, Replacement cost, Operation & Maintenance cost	2	No	Yes	No	No

Table 4.
Indicating frequent, closed and maximal parameters.

3. {Replacement cost, Operation and Maintenance cost}
4. {Capital cost, Replacement cost, Operation and Maintenance cost}
5. {Capital cost, Replacement cost, Operation and Maintenance cost}

Assume the support threshold s = 3.

Monotonicity of solar and wind energy system is also done by Apriori algorithm and in this algorithm let's' be the minimum support required. Let 'n' be the number of items. In the first pass, we read the baskets and count and performance in main memory the occurrence of each parameter. In the second pass, we assess the basket again and count in main memory only those pairs where both parameters are frequent parameters. **Figure 10** shows solar-wind energy assessment through Apriori algorithm.

This algorithm is also used in solar and wind energy system for finding the most utilized parameter without counting all the necessary parameters can be extended to find larger frequent parameters without an exhaustive count of all data sets of tidal energy system. In the Apriori algorithm one step of each parameter of solar and wind energy system is taken for each set size is K [24, 25]. The pattern of moving from one size K to the next size K + 1 can be summarized as follows. For each size K, there are two sets of parameters:

1. DK is the set of parameter of size K, the solar or wind energy parameter that we must assess in order to determine whether they are in fact frequent.
2. MK is the arrangement of really visit parameters of size K.

The example of moving starting with one set, then onto the next and one size to the following is portrayed:

Assume we have assessed the reliability of solar or wind energy system and consider any parameters = {Fault rate (FR), Levelized cost of Energy (LCE), Loss of Power Supply Probability (LPSP), Level of Autonomy (LA), Minimum Time to Failure (MTF)}in the reliability basket.

1. {FR, LCE}
2. {FR, LCE, LPSP}
3. {FR, LCE, LA}
4. {LCE, LPSP, LA}
5. {FR, LCE, LPSP, LA}
6. {FR, LCE, LA, MTF}

Let the support threshold s = 3. The Apriori algorithm as follows:

1.a. Construct D1 = {{FR},{LCE},{LPSP},{LA},{MTF}}.

b. Assess the support of parameters of reliability assessment of solar or wind energy system in D1.

c. Remove infrequent parameter to get M1 = {{FR},{LCE},{LPSP},{LA}}.

2.a. Construct D2 = {{FR, LCE},{LCE, LPSP},{FR, LA},{LCE, LPSP},{LCE, LA}, {LPSP, LA}}.

b. Assess the support of parameters of reliability assessment of solar or wind energy system in D2.

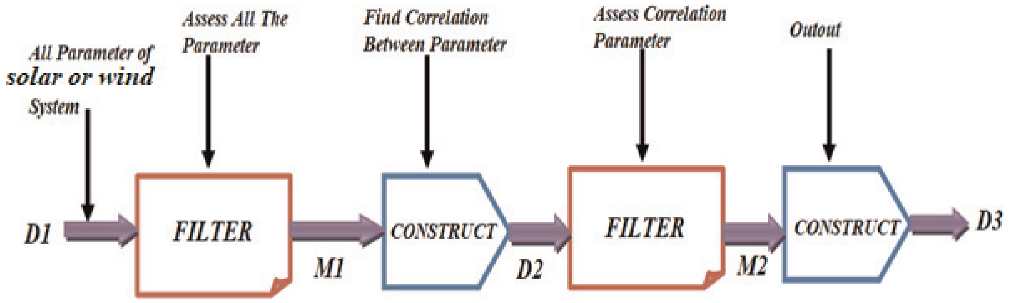

Figure 10.
Assessment through Apriori algorithm.

c. Remove infrequent parameter to get M2 = {{FR, LCE},{LCE, LPSP},{FR, LA},{LCE, LPSP}}.

3.a. Construct D3 = {{FR, LCE, LPSP},{FR, LCE LA},{LCE, LPSP, LA}}.

b. Assess the support of parameters of reliability assessment of solar or wind energy system in D3.

c. Remove infrequent parameter to get M3 = {{FR},{LCE},{LA}}.

4.a. Construct D4 = {Empty set}

Above assessment through market basket model shows that frequent item sets play an essential role in many data mining tasks of solar and wind energy system. The identification of different parameter, characteristics, and often together technical and financial assessment of solar and wind energy system is one of the basic tasks of the market basket model [26, 27].

8. Case study

In the modeling of solar-wind energy system, design a framework according to the certain parameter which is utilized for electricity generation and fulfill the consumer demand. A simulation of solar-wind energy system is an approximate imitation of the solar-wind energy system operation and developed the model with the certain boundary condition. The model of the solar-wind energy system is a well-defined description of the simulated parameter with key properties, such as technical, managerial, functional, and physical properties. Simulation through the data analysis is a key process in the recent scenario and we model out the system according to the certain parameter, where data follow the properties of big data such as volume, velocity, and variety.

In this modeling of solar-wind energy system we consider the peak load of 8KW at particular site area of India. **Table 5** shows data required for simulation of solar-wind energy system.

This information was tested each 1 hour for 365 days of a year. In an average day energy utilization is higher in the first part of the day from 5 A.M. to 9 A.M. furthermore, at night from 7 P.M. to 11:30 P.M. Numerical demonstrating is the initial phase in the structure of any solar energy sustainable power source framework and it gives an accurate perspective on any sustainable power source framework. If we consider randomly any site of particular countries and try to model out of solar-wind energy system so first necessary to define or assess certain parameter, which is essential to design the framework of solar-wind energy system. **Table 6** shows the necessary parameterization of a particular area.

According to the big data analysis following condition is satisfied for the above parameterization data:

- It should be necessary; we have all the above parameterization data in terms of terabyte or petabyte range.
- It should be necessary; all the data are well structured data.
- All the individual parameter data is structured in the form of hourly wise, day wise, month wise and year wise.
- It is necessary to find out the relationship in between two parameter.

DOI: http://dx.doi.org/10.5772/intechopen.87166

Data Required for Simulation	Velocity	Volume	Variety
Year-wise solar radiation current data	Peta bytes	Stream	Structured
Year-wise wind velocity data	Peta bytes	Stream	Structured
Energy consumption data	Peta bytes	Stream	Structured
Optimum sizing of solar wind plant data	Peta bytes	Stream	Structured
Load demand data	Peta bytes	Stream	Structured

Table 5.
Data required for simulation.

Wind velocity of the study area is definitely affecting the clearness index and the amount of solar radiation. A relation between solar radiation and wind velocity and clearness index and wind velocity is developed by regression analysis on the base of 8760 hours solar radiation, wind velocity, and clearness index data on the particular site. The relation is given by the equation:

$$SolarRadiation = 0.0039v^2 - 0.0029v + 5.9045\ (\textbf{According to the gathered data})$$

$$Cleanessindex = 0.0001v^2 - 0.0032v + 0.7643\ (\textbf{According to the gathered data})$$

The coastal vulnerability index (CVI) of the coastal area of given by:

$$CVI = 4G + 4S + 2C + 4T + 3W\ (\textbf{According to the gathered data})$$

Parameter	Range
Latitude	23’10
Longitude	79’21
Solar Radiation	4 to 8kWh/m^2/day
Wind Velocity	5 to 25m/s
Hub Height	10m
Load Demand	05 to 2.5kW

Table 6.
Parameterization of site area.

VARIABLE STATISTICS	NO. OF OBSERVATION	Mode frequency	Categories	Frequency per category	Rel. frequency per category (%)	Lower bound on frequencies (95%)	Upper bound on frequencies (95%)	Proportion per category	Lower bound on proportions (95%)	Upper bound on proportions (95%)
RADIATION	4711	2827452	4	1631941.000	15.361	15.340	15.383	0.154	0.153	0.154
			5	2177457.000	20.496	20.472	20.520	0.205	0.205	0.205
			6	2448561.000	23.048	23.023	23.073	0.230	0.230	0.231
			7	2727462.000	25.673	25.647	25.700	0.257	0.256	0.257
			8	1638324.000	15.421	15.400	15.443	0.154	0.154	0.154
WIND VELOCITY	4711	1551197	8	857143.000	8.068	8.052	8.085	0.081	0.081	0.081
			9	1149660.000	10.822	10.803	10.840	0.108	0.108	0.108
			10	1441197.000	13.566	13.545	13.586	0.136	0.135	0.136
			11	1159172.000	10.911	10.892	10.930	0.109	0.109	0.109
			12	1438055.000	13.536	13.516	13.557	0.135	0.135	0.136
			13	1144192.000	10.770	10.751	10.789	0.108	0.108	0.108
			14	579460.000	5.454	5.441	5.468	0.055	0.054	0.055
			15	1419131.000	13.358	13.338	13.379	0.134	0.133	0.134
			16	1435735.000	13.514	13.494	13.535	0.135	0.135	0.135
Clearness Index	4711	1441377	0.25	881376.000	8.296	8.280	8.313	0.083	0.083	0.083
			0.3	443280.000	4.173	4.161	4.185	0.042	0.042	0.042
			0.35	440592.000	4.147	4.135	4.159	0.041	0.041	0.042
			0.4	884256.000	8.323	8.307	8.340	0.083	0.083	0.083
			0.45	443664.000	4.176	4.164	4.188	0.042	0.042	0.042
			0.55	443856.000	4.178	4.166	4.190	0.042	0.042	0.042
			0.6	885984.000	8.340	8.323	8.356	0.083	0.083	0.084
			0.71	444240.000	4.182	4.170	4.194	0.042	0.042	0.042
			0.72	444048.000	4.180	4.168	4.192	0.042	0.042	0.042
			0.8	442704.000	4.167	4.155	4.179	0.042	0.042	0.042
			0.9	442512.000	4.165	4.153	4.177	0.042	0.042	0.042
			1	1331377.000	12.532	12.512	12.552	0.125	0.125	0.126
			1.1	1326960.000	12.491	12.471	12.510	0.125	0.125	0.125
			1.2	885216.000	8.332	8.316	8.349	0.083	0.083	0.083
			1.3	441744.000	4.158	4.146	4.170	0.042	0.041	0.042
			1.4	441936.000	4.160	4.148	4.172	0.042	0.041	0.042

Table 7.
Variable statistics of renewable energy parameter.

where G is the vulnerability ranking of geomorphology, S is the site area, slope, C is the Shore Line Change, T is the wave velocity range and W is the significant hub height. According to the tool of regression analysis **Table 7** shows the variable statistics of tidal energy system by the regression analysis tool.

Data of above table shows that by the huge number of data, we can easily analysis prefeasibility assessment of solar-wind energy system and also predict frequency, lower and higher frequency of solar radiation, wind velocity and clearness index.

9. Conclusion

In case of big data volume, variety and velocity are the three main drivers that gave a new dimension to the way analytics of solar and wind energy system. Big data helps the solar and wind power system to create new growth opportunities and entirely new categories of the solar and wind system that can combine and analyze solar and wind industry data. Following are the concluding remarks of solar and wind energy assessment through the big data analytics:

- Atmospheric prediction is always depends on the lot of data which is maybe a day-wise, month-wise and year-wise and in the case of solar and wind energy system lot of data is required for pre-feasibility assessment. So that in pre-feasibility assessment big data analytics one of the best assessment methods.

- Market basket model based big data analytics provides plentiful information about the solar and wind energy system framework, buyers, and suppliers, customer preferences that can be captured and assessed.

- Hadoop framework is used for modeling of solar-wind energy system by the huge amount of data by dividing the data into a number of parts such as overall specification of energy system into the specification of photo-voltaic panel, wind turbine, generator, battery, etc.

- Map-reduce algorithm is used to find out optimum value of technical and financial parameters of solar and wind energy system by huge number of data collection.

Author details

Vikas Khare[1,2] and Aaquil Bunglowala[3]

1 Department of Electrical Engineering, STME, NMIMS, Indore, India

2 Bureau of Energy Efficiency, India

3 School of Technology, Management and Engineering, NMIMS, Indore, India

*Address all correspondence to: vikaskharekhare@gmail.com

References

[1] Torrecilla JL, Romo J. Data learning from big data. Statistics & Probability Letters. 2018;**136**:15-19

[2] Eckroth J. A course on big data analysis. Journal of Parallel and Distributed Computing. 2018;**118**: 166-176

[3] Hejazi HA, Rad HM. Power system big data analytics: An assessment of paradigm shift barriers and prospectus. Energy Reports. 2018;**4**:91-100

[4] Yao HW, Wang XW, Wu LS. Prediction method for smart meter life based on big data. Procedia Engineering. 2018;**211**:1111-1114

[5] Shyam R, Bharati GHB, Kumar S. Apache spark based big data analytics plate form for smart grid. Procedia Technology. 2015;**21**:171-178

[6] Yang Z, Gao C, Zhao M. Utilizing big data to explore the running opportunity of power to gas in energy system. Energy Procedia. 2019;**158**:2341-2347

[7] Bersa RJ. Chapter 10: Future trends for big data application in power system. In: Big Data Application in Power System. Cambridge, USA: Elsevier; 2018. pp. 223-242

[8] Percuku A, Minkovska D, Stoyanova L. Big data and time series use in short term forecasting in power transmission system. Procedia Computer Science. 2018;**141**:167-174

[9] Junaidi N, Shaaban M. Big data application in electric energy system. In: IEEE International Conference on Computational Approaches in Smart System Design and Applications; 2018. pp. 1–4

[10] Hangxun T, Hongang W. Measuring system of power quality by big data analytics. In: IEEE International Conference on Cloud Computing and Big Data Analytics; Chengdu, China: 2018. pp. 248–252

[11] Huang J, Niu L, Zhan J. Technical aspects and case study of big data based conditioning monitoring of power apparatuses. In: IEEE PES Asia Pacific Power and Energy Engineering Conference; Hong Kong, China: 2014. pp. 1–4

[12] Wanxing S, Keyan L. The anomalous data identification study of reactive power optimization system based on big data. In: International Conference on Probabilistic Method Applied to Power System; Beijing, China: 2016. pp. 1–5

[13] Zhan J, Huang J. Study of the key technologies of electric power big data and its application prospects in smart grid. In: IEEE PES Asia Pacific Power and Energy Engineering Conference; Hong Kong, China: 2014. pp. 5–8

[14] Guan L, Zhang J. Enhancing security and resilence of bulk power systems by multisource big data learning. In: IEEE Power and Energy Society General Meeting; Chicago, USA; 2017. pp. 1–5

[15] Qing L, Boyu Z. Impact of big data on electric power industry. In: IEEE International Conference on Big Data Analysis; Beijing, China; 2017. pp. 460–463

[16] Cox DR. Big data: Some statistical issues. Statistics & Probability Letters. 2018;**136**:111-115

[17] Glushkova D, Jovanovic P. MapReduce performance model for Hadoop 2.x. Information Systems. 2017: 1-10. In Press

[18] Shankarmani R, Vijayalakshmi M. Big Data Analytics. 2nd ed. Onterio Canada: Wiley Publication; 2017

[19] Babcock B, Datar M. Maintaining variance and k-medians over data streams windows. In: Proc. ACM Symp. On Principles of Database System; 2004. pp. 234-243

[20] Aggarwal CC, Reddy C. Data Calculus: Algorithm and Applications. India: CRC Press; 2013

[21] Gracia H, Ullman J. Database System: The Complete Book. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2009

[22] Arghandeh R, Zhou Y. Big Data Application in Power Systems. 1st ed. Cambridge, USA: Elsevier; 2018

[23] Agrawal R, Imielinski T. Mining association between sets of items in massive datasets. In: Proc ACM SIGMOD International Conference on Management of Data; 1993. pp. 207–16

[24] Toivonen H. Sampling large data bases for association rules. In: International Conference on Very Large Databases; 1996. pp. 134–145

[25] Agneeswaran V. Big Data Analytics Beyond Hadoop. Pearson Education USA: 2014

[26] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters. Communications of the ACM. 2008;**51**(1):107-113

[27] Philip C, Zhang CY. Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences. 2014;**275**(Suppl. C):314-347

[28] Available from: www.tutorials_ point.com