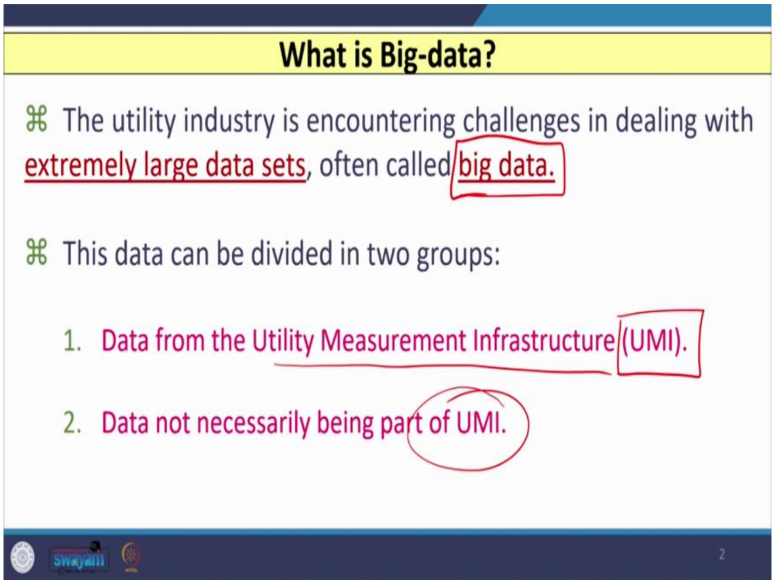
Digital Protection of Power System
Professor Bhaveshkumar Bhalja
Department of Electrical Engineering
Indian Institute of Technology, Roorkee
Lecture 37
Application of Big-Data Analytics in Power System Protection

Hello friends. So, in this lecture, we will discuss about the Application of Big-Data Analytics in Power System Protection. So, we will discuss what is the big data, how it can be used in the application of protection area and maybe in some other areas also. So, that we will discuss in this lecture. So, let us see first what is big data.

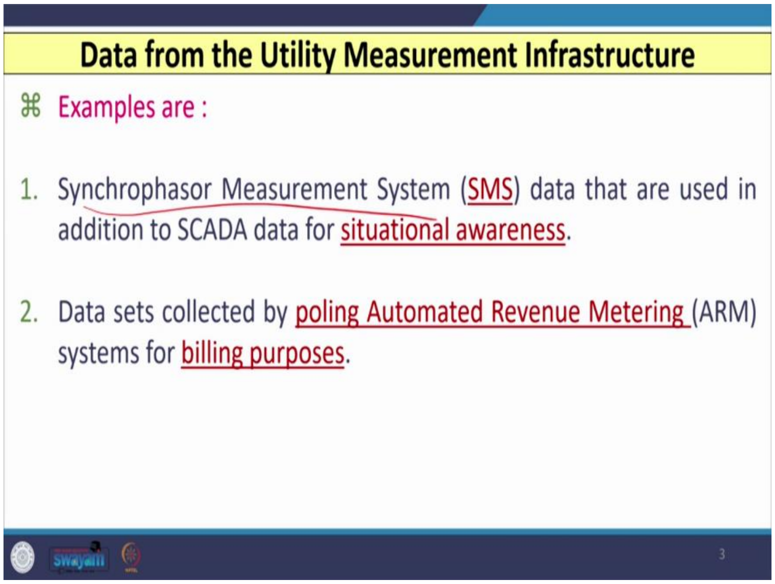
(Refer Slide Time: 00:48)



So, we know that the utility industry is encountering several challenges in dealing with extremely large data sets. And this extremely large data sets are often called big data. And this large data sets which we call as a big data that is normally divided into groups, the first group of data that is from the utility measurement infrastructure, we can also call it as UMI. So, these data are available from utility measurements, several measurements are carried out by utility using several devices.

And whatever data obtained by these devices, those data fall under this category that is the data available from the utility measurement infrastructure. The second group of data that is available other than utility measurement infrastructure. So, these data are not necessarily the part of utility measurement infrastructure. Of course, those data are not required by utility, but if those data is integrated with the first group of data, that is utility measurement infrastructure data, then whatever combined algorithm we have that will be useful for several purposes.

(Refer Slide Time: 02:14)

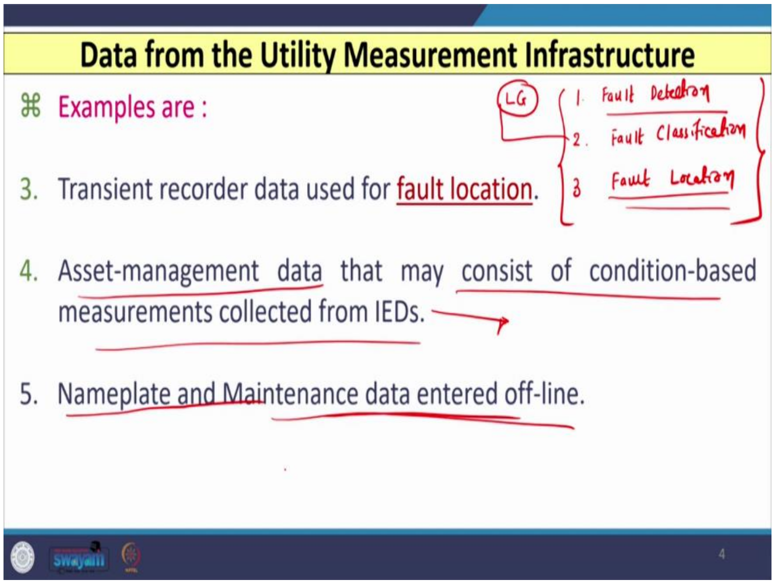


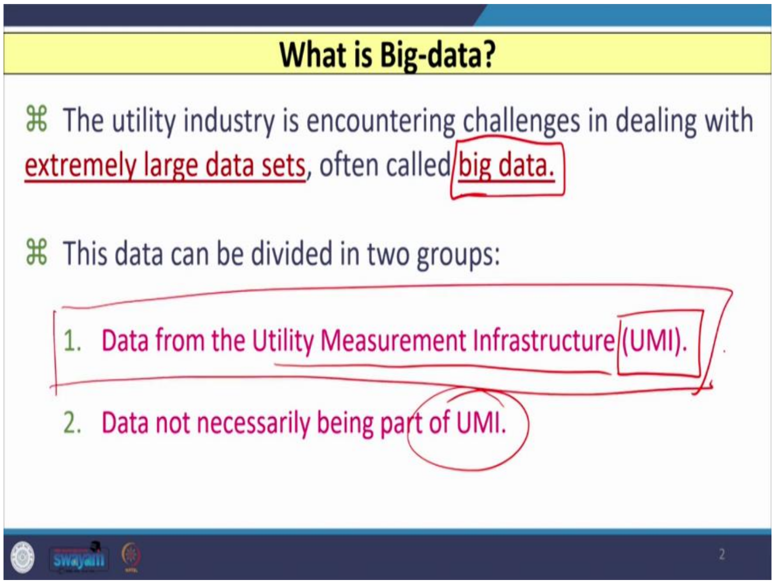
So, the data available from utility measurement infrastructure, there are several examples. The first example is the data that is captured or available from Synchrophasor measurement systems. So, we have phasor measurement units, we have phasor data concentrators. So, based on this whatever data are used, and this data are used, in addition with the data available from the SCADA system, that is supervisory control and data acquisition system.

So, this data are normally used for situational awareness. So, if we wish to go for online monitoring, real-time monitoring of the entire network, then we can use the data of Synchrophasor along with the data of SCADA system. The second group of data which we can collect that is from the automated revenue metering systems and this type of systems are widely used for billing purpose and such type of meters are installed by several utilities at various locations.

And so, there is no need to send the linemen to capture the data and then billing will be done instead of that, automated revenue metering system type of arrangement is carried out by the utility, where utility installed several meters we can call it as smart meters, at several locations and these smart meters will send the data at a control center and based on the data available at the control center billing will be done. So, this type of data are also fall under the first category that is the data available from the utility measurement infrastructure.

(Refer Slide Time: 04:03)





The third example, we can say that is the data available from the transient recorder and this data are related to the fault location. So, we know that when we consider any power system network, three tasks are very required the first task that is known as the fault detection. So, whenever anything goes wrong, any abnormal condition is there in the network or in the system, then the first task is to detect the fault.

Sometimes this type of condition is also known as abnormal detection condition or algorithm. So, this algorithm will tell you whether it is a fault or it is some other things. So, first task is the fault detection task. The second task that is the, we can call it as fault classification. And the third task is known as the fault location. So, in fault classification, we are able to classify the different types of faults let whether it is line to ground or double line to ground or triple line faults.

So, why fault classification is important because based on fault classification, let us say we have identified the fault is single line to ground fault. So, if fault is single line to ground fault then that this type of information is useful for the auto reclosers. And based on that auto reclosers can perform the reclosing attempt. So, this way the information is very useful for such a system. The third task is the fault location task.

So, in fault location algorithm the estimation of fault is carried out and this algorithm will tell you that from this substation or from the sending end or from this local end at a distance of let us say x kilometer, there is a probability that the fault has occurred and then you can carry out the required maintenance by sending the lineman or the crew. So, these three tasks are very important and these three tasks need to be performed by any protection algorithm especially for the transmission network or EHV, long EHV and UHB lines.

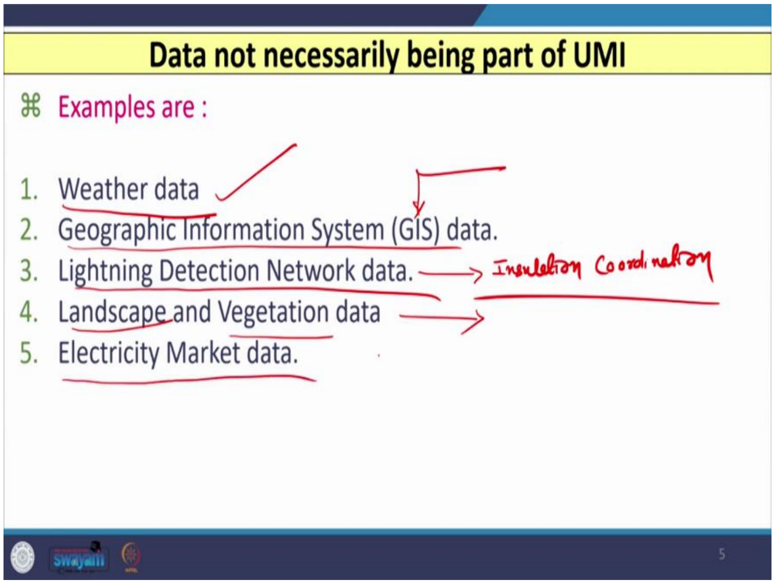
So, in that case, transient recorded data are also important for fault location. So, those data are also captured by utility measurement infrastructure. The fourth type or example of the utility measurement infrastructure type of data that is the data related to asset management. And this type of data consists of condition-based measurement collected from the intelligent electronic devices. So, we have installed several intelligent electronic devices.

And we discussed earlier that the function of IED is that it can perform measurement, control, monitoring, protection, all these functions. So, whatever IEDs we have installed, and those data of those IEDs are available or grouped at particular control center. And by assessing those data analysis related to asset management that can be carried out. And this asset management task includes the management related issues.

Maybe it can also include financial related issues like electrical pricing, other issue related to these technical things. So, all these are available or comes under the asset management task. The fifth important example is the nameplate and maintenance data entered offline. This is normally offline because when you installed any devices, the nameplate data you need to capture normally it is provided by the manufacturer.

So, those data are also important and some data which you have captured or which you observed at the time of schedule maintenance. So, those type of data are also very important. So, these five datas which are available from the let us say, Synchrophasor, SCADA systems, maybe automated revenue metering systems or smart meters available maybe from transient recorders or maybe from IEDs or maybe some offline data may be collected from nameplate or during the scheduled maintenance, all those data are available or fall under the category of this first that is the data from the utility measurement infrastructure.

(Refer Slide Time: 08:55)

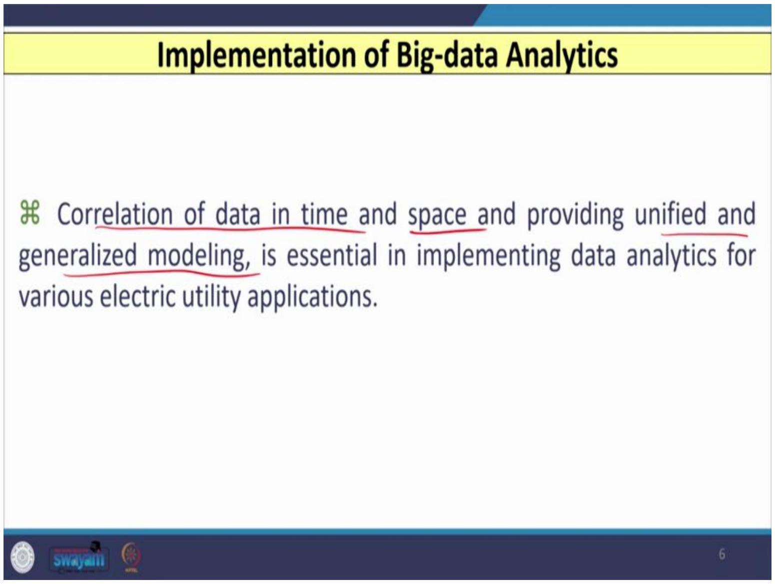


Now, there are some other datas which are not the part of utility measurement infrastructure. And the examples of those data are the first example is the weather data. So, we need weather data, let us say to forecast what is the actual production of, let us say we have solar based power generation. So, for forecasting of this we can go forward if we have wind-based power generation. So, weather data plays an important role.

Second type of example, those data which are not part of utility measurement infrastructure, those are available from geographical information system and normally known as GIS data, Geographic Information System data. The third type of example is the lightning detection network data. So, lightning detection network data are very useful for studies related to the insulation coordination.

So, that is why this data plays an important role and they are useful for insulation coordination studies. Then the fourth example is the data related to the landscape and vegetation. So, normally this data are very useful for the utilities. And that is why this type of data are also there. And the last but not the least, the electricity market data, which will be used for pricing and some other applications. So, this data are also important. This data are normally not captured by the most of the utilities, but this type of data are available from some other organization. Now, if we wish to implement big data analytics, then what is required?

(Refer Slide Time: 10:53)

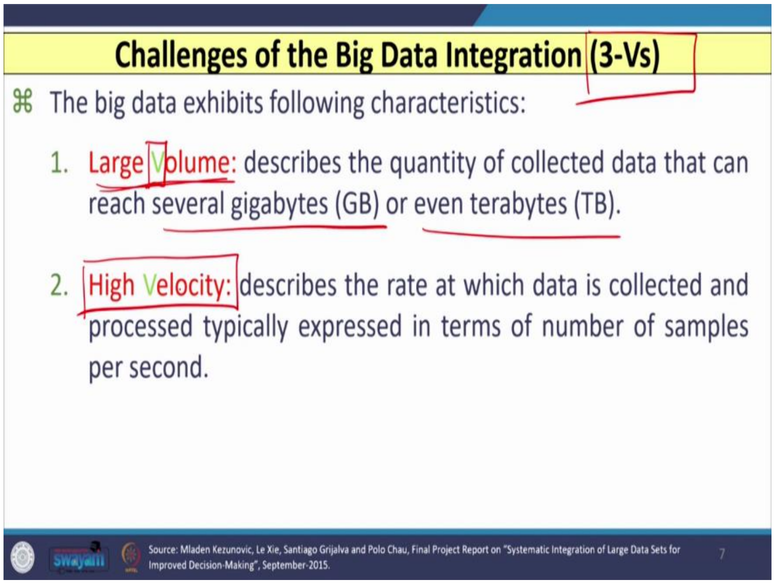


So for the successful implementation of big data analytics, correlation of data in time and space, and providing unified and generalized modeling is very important. So, we need to coordinate the data which are available in time domain or which are available in space domain, let us say whatever data captured by the smart meters, let us say data captured by the PMU synchrophasor's data, SCADA data, all those data are in time space, whereas the data available in space like the weather data, maybe GIS data, they are available in images or lightning detection data.

So, all these data are available in some different domain or form, we need to correlate this data that means both the group of datas, those data which are part of a utility measurement infrastructure, and those are not part of utility measurement infrastructure, we need to correlate this data and then we have to carry out a generalized modeling.

And using that, we can definitely use this generalized model which we have developed for implementing data analytics for various electric utility applications. So, we will discuss maybe one or two applications also in this case, if we correlate this two data in time and space, and if we achieve a generalized model, then what we can obtain. Now, before we discuss this thing, let us see what are the challenges of the big data integration.

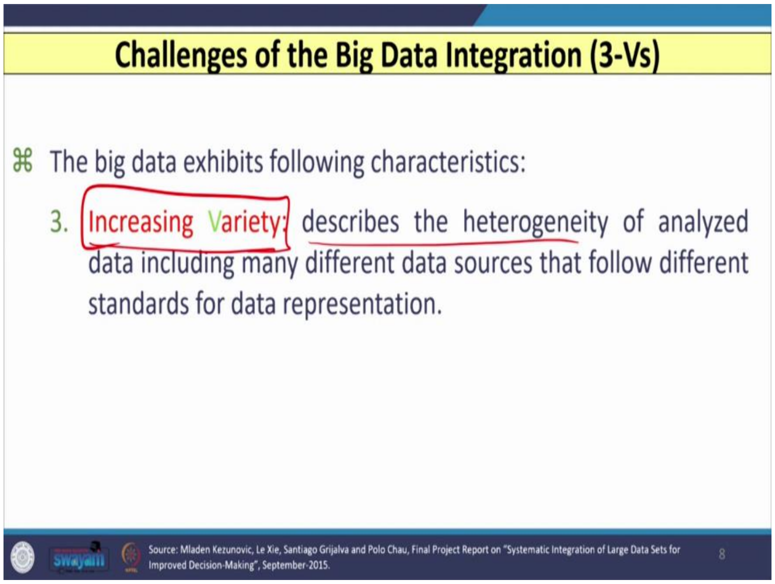
(Refer Slide Time: 12:28)



And these challenges are in the form of 3-Vs. So, the first challenge is related to the large volume, because normally, whenever the data are captured by the utility, large volume is related to the quantity of data collected by several devices, and size of this data that can be in several gigabytes, or even sometimes in terabytes. So, this large volume, so the first V comes from the volume of the data that is related to the size of the data that is captured by the utility.

The second V comes from the velocity of the data that is high velocity. And this is basically related to the data collection. And processing of those data expressed in terms of samples per second. So, whatever data we capture, if we capture the data, let us say at every 1 minute, and if we are capturing the data, let us say at every half an hour, in both the cases when we capture the data in less number of samples per second means if we have more samples per second if we have more data in one cycle, if we are capturing then the velocity of those data, that is really very high even volume of those data, storage requirements, so that also increases. So, the second V comes from the velocity of the data whereas first V comes from the volume of the data.

(Refer Slide Time: 13:58)

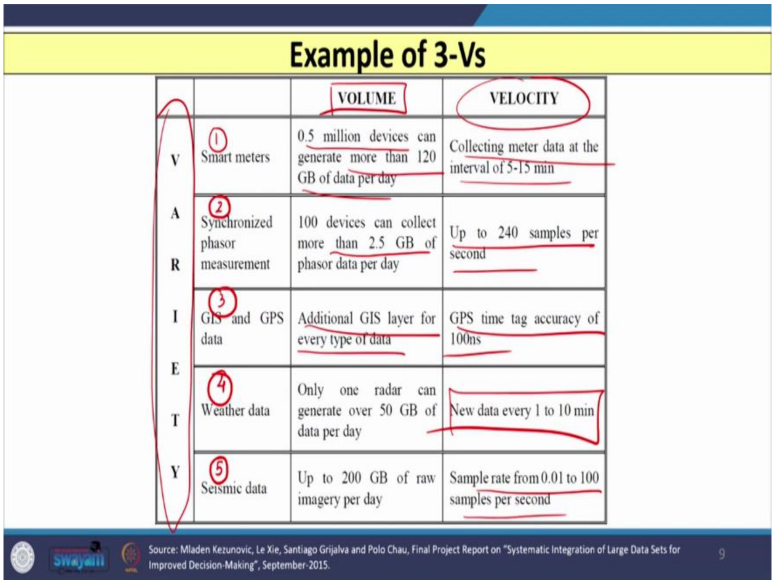


The third V comes from the increasing variety of the data. Because we know that we can capture or utility can capture data in time domain, let us say the data of Synchrophasors or data of SCADA system or maybe the data available from the IEDs or smart meters. Whereas some other datas are also there in the form of let us say, latitude or longitude or maybe in the form of images, like GIS data, weather data, so many other things.

So, this type of increasing variety that is third V which is because of the variety of the data that describes the heterogeneity of analyzed data, including many different sources, as I told you, let us say we have whether, we have GIS, let us say we have the information related to the lightning. So, all those different sources are there.

And those sources when we captured, they are in different form. So, this is also going to have a big challenge. So, volume of the data velocity of the data and variety of data, these 3-Vs are very important and because of these 3-Vs, the there are several challenges, when we are dealing with the big data analytics issue in power system protection.

(Refer Slide Time: 15:20)



Now, if we consider these 3-Vs. So, what I have done volume of the data I have considered here, the velocity of the data I have considered here and the variety of data I have considered on this side. So, if I consider these smart meters, let us say that we are utilities capturing data from smart meters, then a survey indicates that 0.5 million devices if I installed 0.5 million smart meters, then in case of 0.5 million smart meters, they can generate more than 120 GB of data per day.

So, handling of those data, volume of this data is very large and even when collecting this meter data at interval of either 5 minutes or at the most 15 minutes. So, if you collect this data at 5 minutes, then the velocity is different. And if you collect the data at 15 minutes, then the velocity is different. The second example, we can consider is the PMU data. So, if I installed 100 devices or 100 PMUs or 100 synchrophasor devices, then they can collect more than 2.5 GB of phasor data per day.

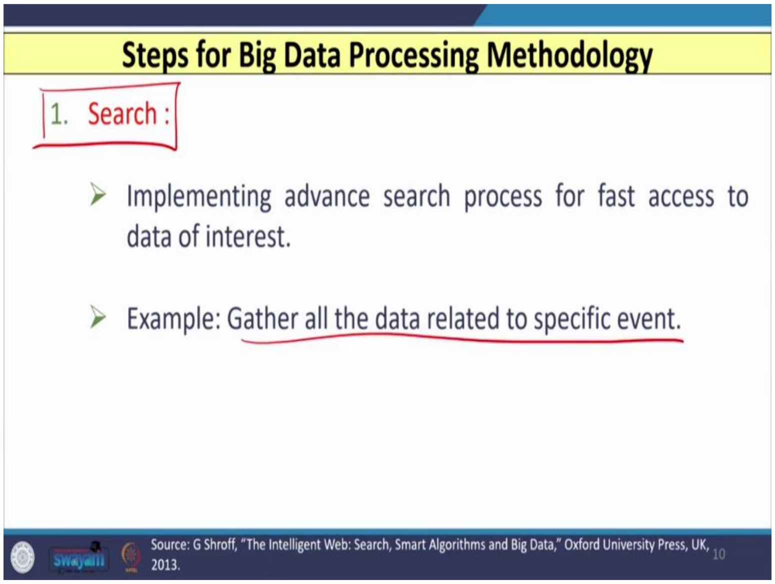
And this is up to different samples per second or frames per second. So, maybe we have highest 240 samples per second, but you can have lower values also let us say 120 or 60 or 50 samples per second. So, in that case, the volume and the velocity both are different. So, handling PMU data that is also volume wise that is size wise and the velocity wise, those are also very different.

The third example is related to the GIS and GPS data, we know that when we talk about the PMU and PDC then those devices will work based on the GPS system. So, GPS data and GIS data both are very important. And when you have GIS data, you need one additional GIS layer for every type of data. So, if you have GIS data, then you have one separate layer is required and GPS time tag accuracy is of the order of 100 nanosecond.

So, based on that, whatever data you are going to capture the size of data that is really very large, and we need to handle this issue also. The fourth data are related to the weather data. So, if I have a one radar system, then they can generate 50 GB of data per day. And this data you can capture either at every 1 minute or maybe at every 10 minutes, then the velocity and volume of this data are also different.

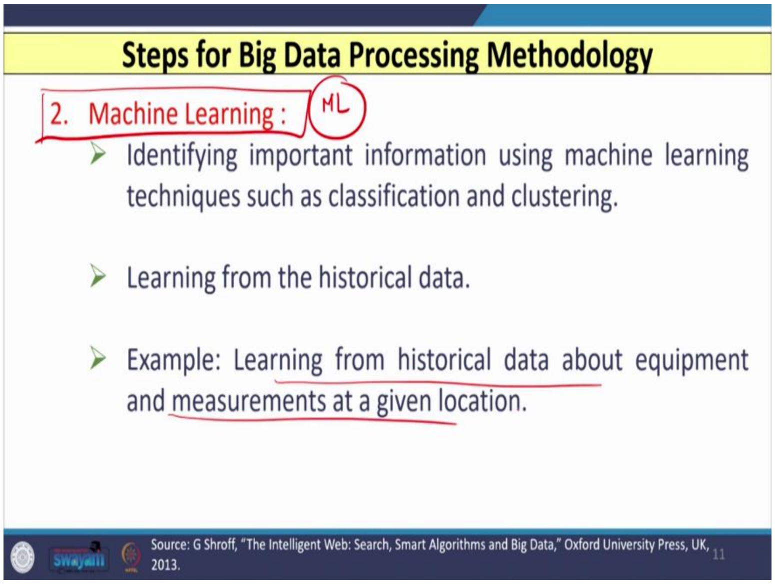
And the fifth one are the data related to the seismic data up to 200 GB of raw imaginary per day, you can have with a sampling rate of either 0.01 or 100 samples per second. So, in all the five cases, you can see the volume, velocity and variety of datas are there, which we need to handle in big data analytics issue or application related to power system protection.

(Refer Slide Time: 18:42)



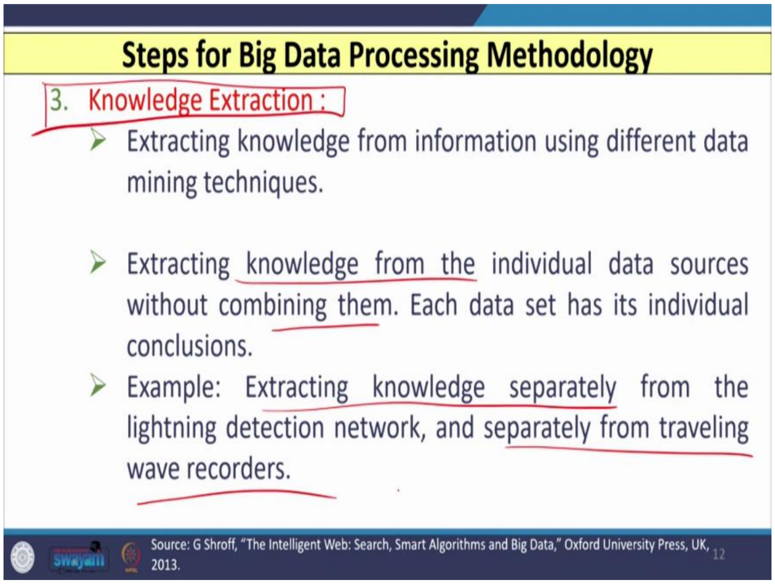
Now, let us see how we can handle this. So, let us discuss the steps for big data processing methodology. So, the first step is related to the search. So, we have to implement some advanced search process technique for fast assess of the data in which we are interested. Let us say we have number of data weather data, PMU data, smart meters data, let us say we have the some other lightning data, now out of all this data in which data we are interested for that we need to have some advanced search algorithm and the best example is we have to gather all the data related to a particular specific event. So, in that case, we need the first thing is the search part.

(Refer Slide Time: 19:31)



The second is the use of machine learning. So, using this we can identify important information such as classification or clustering. So, we know that suppose we have one data available from PMU and another data available from weather. So, whether we need to classify or we need to form a cluster for a specific type of data. So, that can be done with the help of machine learning algorithms. So, this type of algorithm has a capability to learn from the historic data or the past data, and the examples are learning from historical data about equipment and measurements at a particular given location. So, for that purpose, second step that is related to machine learning, that play an important role.

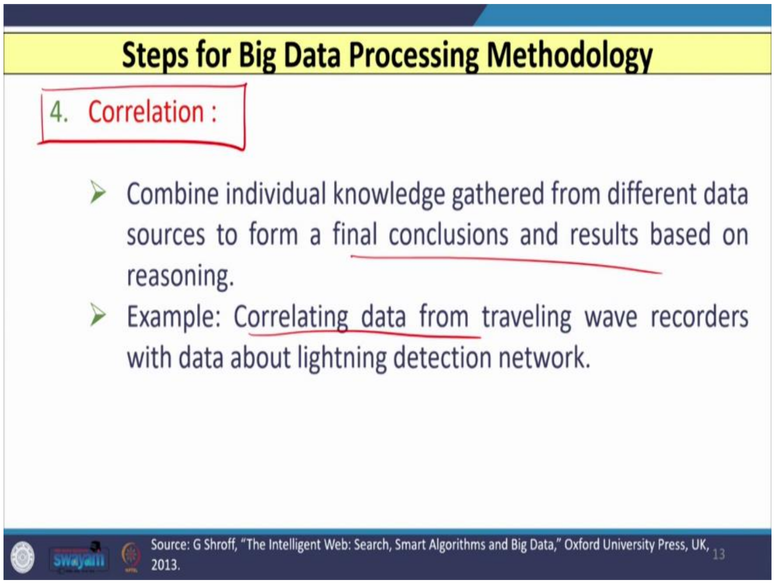
(Refer Slide Time: 20:22)



The third step is the knowledge extraction step. So, we have to extract the knowledge from information using different data mining techniques? So, we need to have data mining algorithms, where we can have several techniques and you can go for a specific technique depending upon your application. And here you need to extract the knowledge from individual data sources without combining them.

So, let us say I want to capture the specific individual data sources from PMUs. And I want to capture on the other side individual data from the lightning. In this case in the knowledge extraction part, we are going to create several data sets of with each data set has several conclusions and examples are extracting knowledge separately from the lightning detection network and separately from the traveling wave recorders. Those are the examples of these data.

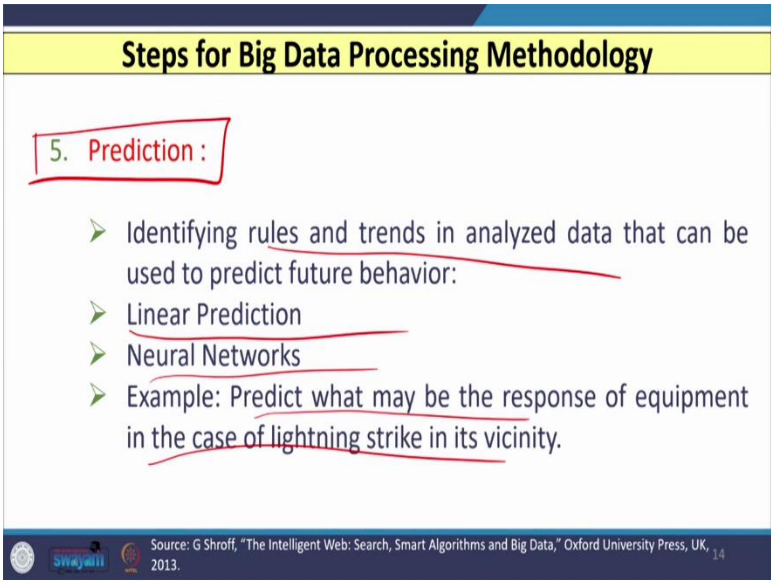
(Refer Slide Time: 21:22)



The fourth step is related to the correlation. So, here in this step, we need to combine individual knowledge gathered from different data sources from the previous part that is knowledge extraction to form a final conclusion and the results based on some specific reasons. And the examples are we have some correlation algorithms or correlating data of traveling wave recorders with the data of lightning detection network.

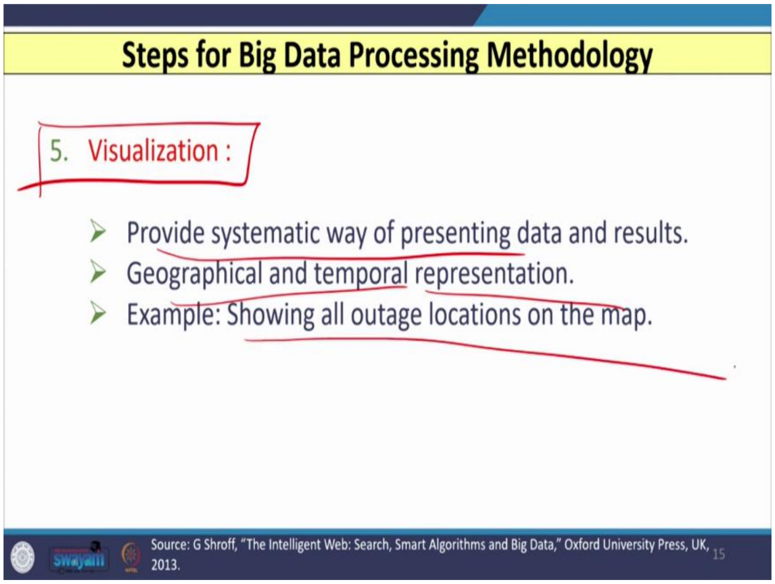
If you combine lightning detection network data with traveling wave data, then that can be really helpful for taking the decision whether the fault because of that traveling wave has generated or maybe because of some lightning or transient fault traveling waves have been generated. So, based on that you can take appropriate decision.

(Refer Slide Time: 22:13)



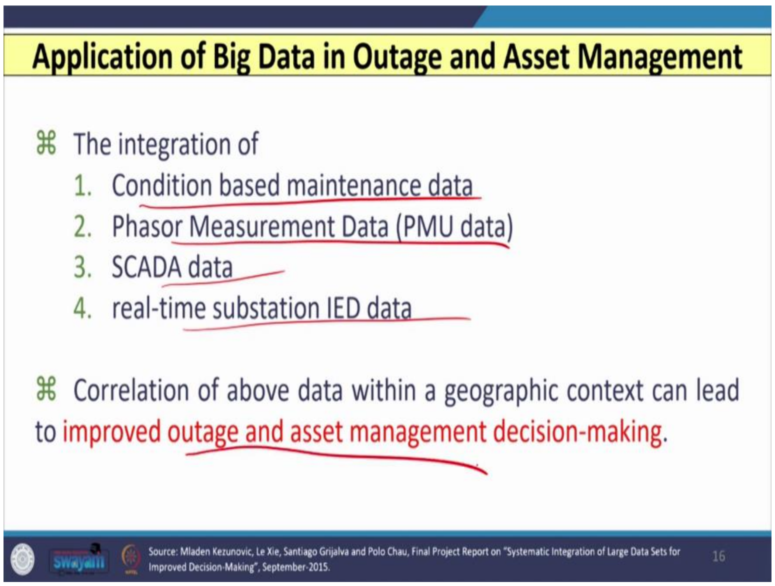
The fifth step is related to the prediction part. So, here in prediction algorithm, we need to identify rules and trends in analyze data that can be used to predict the future behavior. So, let us say for example, we have linear prediction algorithms are available, some neural networks are also available, which can predict. So, it is basically related to the interpolation. And the examples are to predict what may be the response of equipment in case of lightning strikes in its vicinity. So, if we wish to predict this thing, then we have to go for prediction algorithm.

(Refer Slide Time: 22:55)



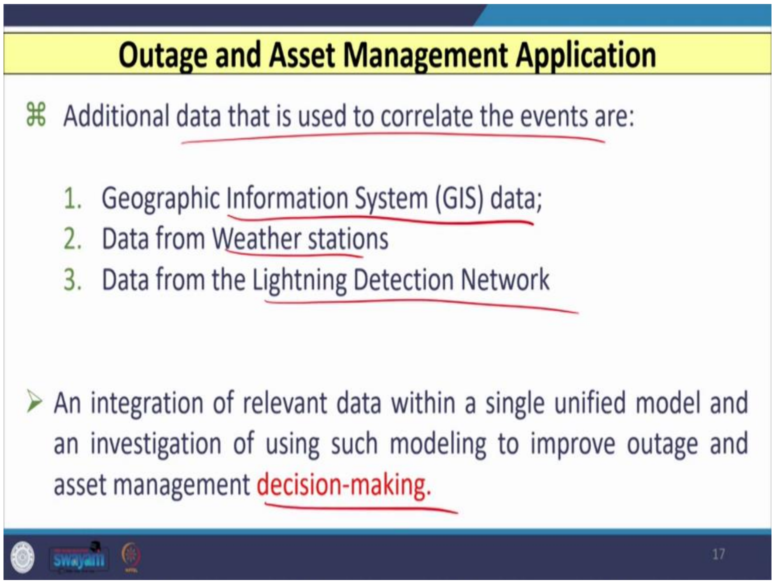
The next step that is fifth step is related to the visualization part. So, here this type of visualization algorithm provides systematic way of presenting data and the results. Geographical and temporal representation, these are the one of the way to express this thing in visualization algorithm and showing all outage locations on the map, this type of thing you can do in visualization part. Now, after these different five steps, let us see what is the application of big data in the outage and asset management.

(Refer Slide Time: 23:34)



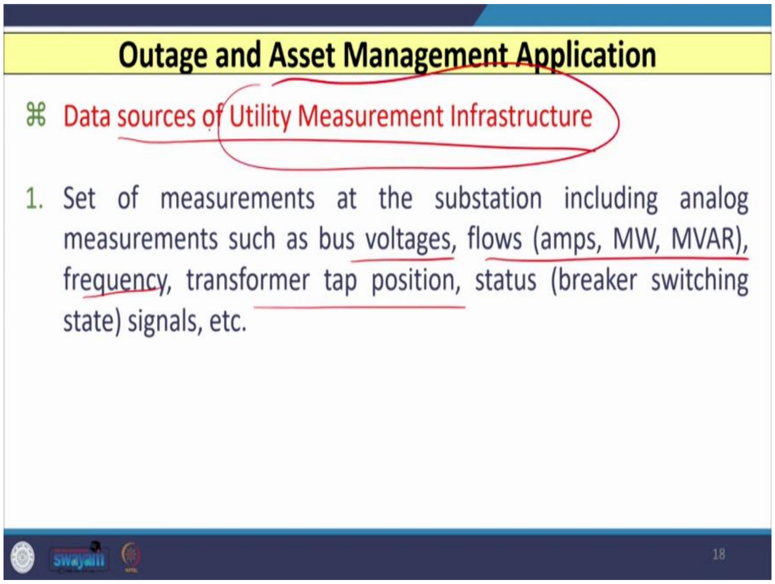
So, here, if we wish to apply a big data in some application like outage or asset management, then we need to integrate the data available from maintenance, we need to have the data related to PMU, we have SCADA data, we have IED data we need to integrate this data. And based on this correlation of the above data in geographical contexts, we will have our final decision related to the outage or asset management.

(Refer Slide Time: 24:09)



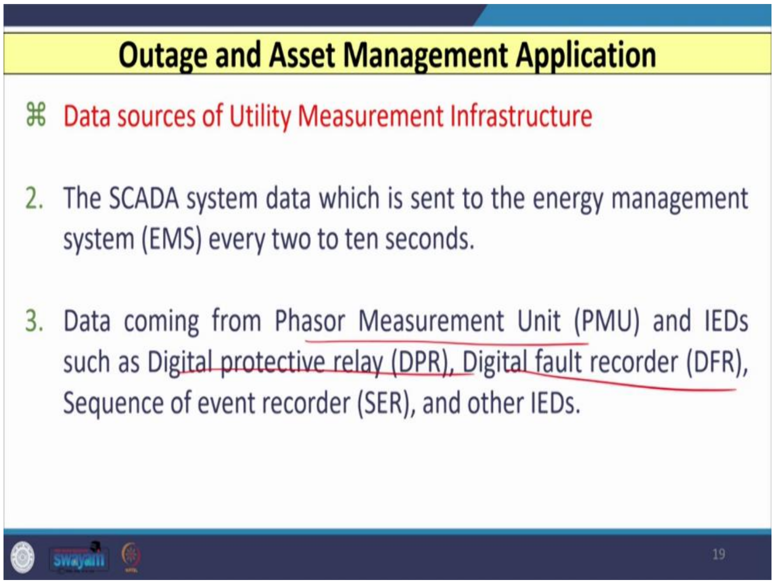
If we wish to take some decision related to outage and asset management application then we also need several other datas and those datas are available from the GIS, they are available from the weather stations or they are available from the lightning detection network. So, we need to integrate this data, say GIS data, weather data, lightning detection data with the other data captured by utilities like data available from PMUs, IEDs all that we need to integrate. So, that it can be helpful in final decision making of outage and asset management application.

(Refer Slide Time: 24:50)



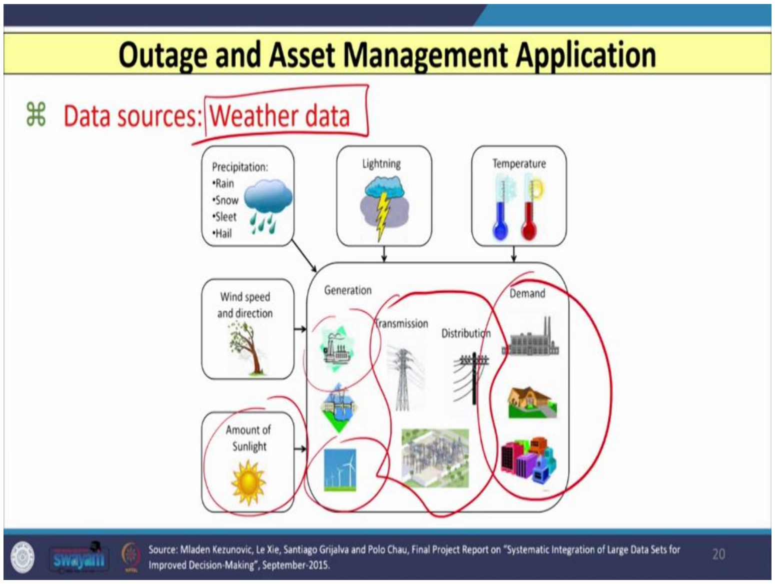
So, data sources of the utility measurement infrastructure we have divided in two groups. So, the first group of data that is captured by the utility and this data are related to the let us say bus voltages, the line and power flows, frequency, transformer tap position and some status signal like relay or breaker status is there whether relay has operated, whether breaker is operated or not. So, all those are captured by the utility.

(Refer Slide Time: 25:23)



The SCADA system data are also there and this data are useful for the energy management system algorithms. The third is the data available from PMU, IEDs or maybe you have let us say data available from digital relays or maybe digital fault recorders or maybe the sequence of event recorders and several other IEDs. So, those data are captured by the utility.

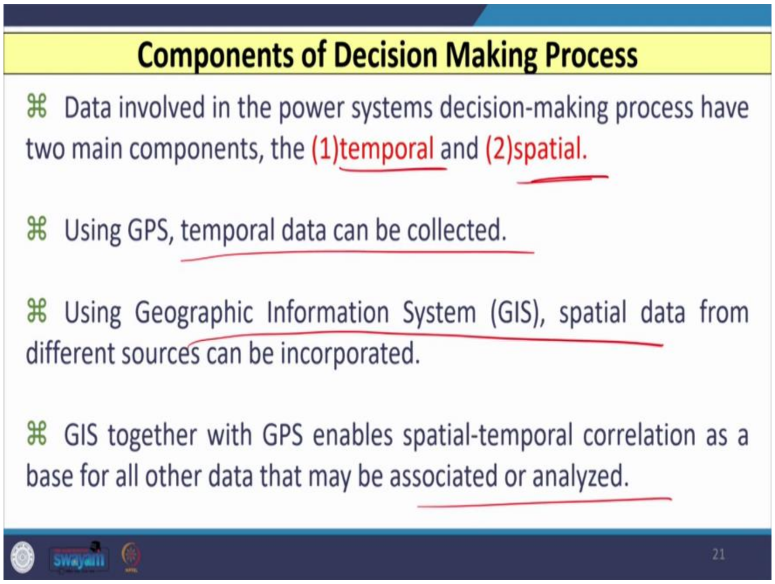
(Refer Slide Time: 25:48)



The other datas, for example, if I consider the weather data, then you can see that I have shown here one flowchart where you can have the weather data, let us say you can use the weather data then what is the wind speed you can easily predict and based on that what would be the output of this that you can have.

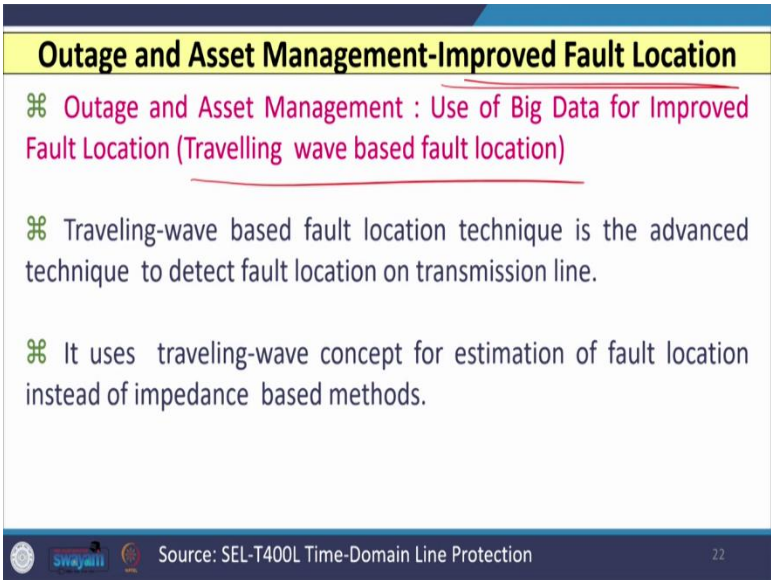
Similarly, if you have the amount of radiation of sunlight etc., available, then if you are using solar based power plants, then you can go for the forecasting of output of this power plant that can be easily done. So, here you can see that I have shown the transmission part and distribution part and in the demand part also I have shown which varies depending upon the let us say summer or winter time or rainy season. So, weather data plays an important role in several applications. Similarly, you have GIS data. Similarly, you have lightning detection data for several applications. Now, let us see what are the components of decision-making process.

(Refer Slide Time: 26:56)



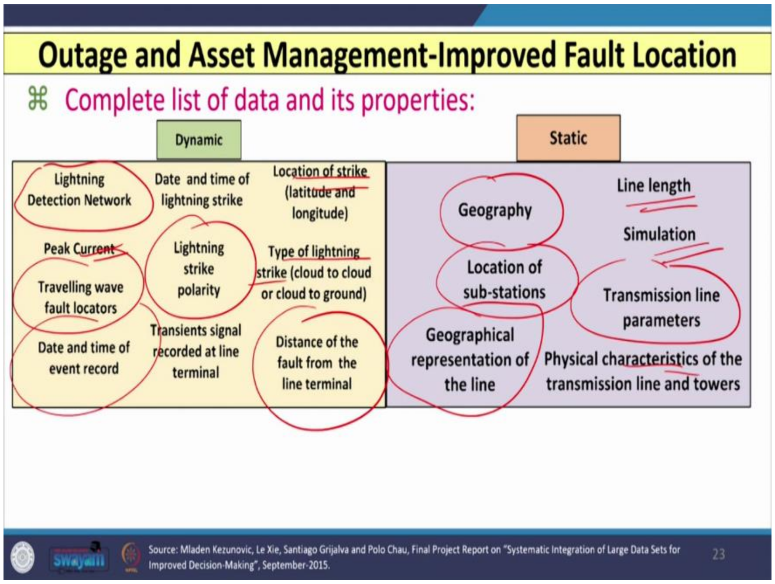
So, the data involved in the power system decision making process that has two components, the one is the temporal components and another is the spatial components. So, using GPS temporal data that can be collected like PMUs or PDC. And using GIS you can have the spatial data and once you have the data from the GIS and GPS both temporal and spatial, you can use some correlation algorithm and based on that, you can have the decision making related to a particular application.

(Refer Slide Time: 27:32)



So, if I consider one of the application in outage and asset management, then the improved fault location technique, we can have such type of algorithm and this algorithms are based on traveling wave. And normally in earlier cases, most of the utilities they use impedance-based fault location algorithms, maybe some different algorithms are available in impedance based also. So, instead of that, if we have such type of correlation of data available, let us say lightning network data, or maybe we have the data available from the traveling wave-based fault occurs, then we can go for the detection of fault location using the traveling wave-based fault location technique.

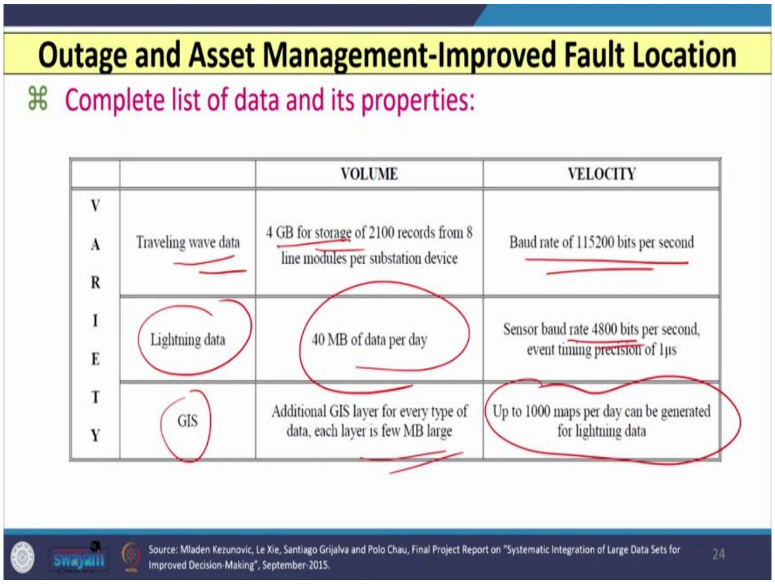
(Refer Slide Time: 28:21)



So, if I consider the complete list and its properties, then we have some static data, let us say data related to geography, line length simulation data, maybe we have parameters of transmission lines, maybe we have the location of substations, we have the geographical representation of the line, we have physical characteristic of the transmission line and towers, all those data are static they are not going to change.

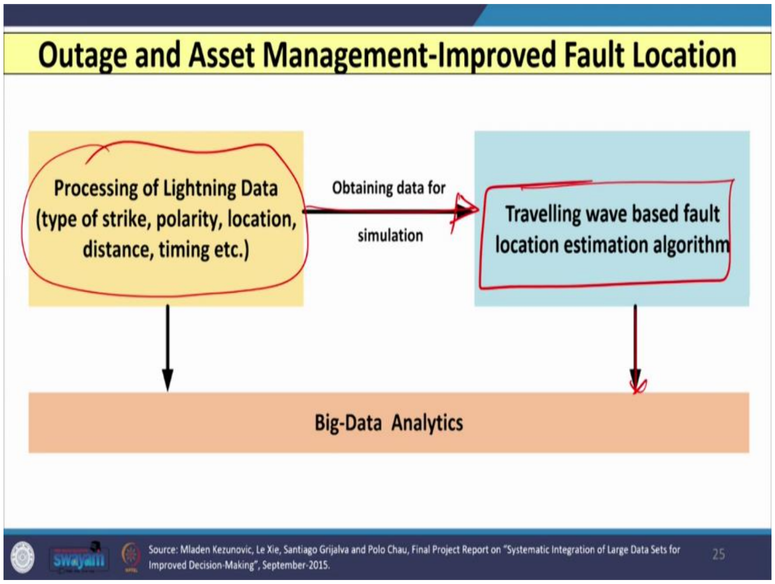
Whereas, we have some dynamic data related to lightering detection network maybe you have location of strike latitude and longitude we have type of lightning strike cloud to cloud or cloud to ground, we have distance of the fault from the line terminal, we have lightning strike polarity, date and time of lightning strike, peak current, we have the traveling wave fault locators, data and time an event of the record, such data are dynamic in nature, they may change depending upon the weather conditions and some other parameters.

(Refer Slide Time: 29:29)



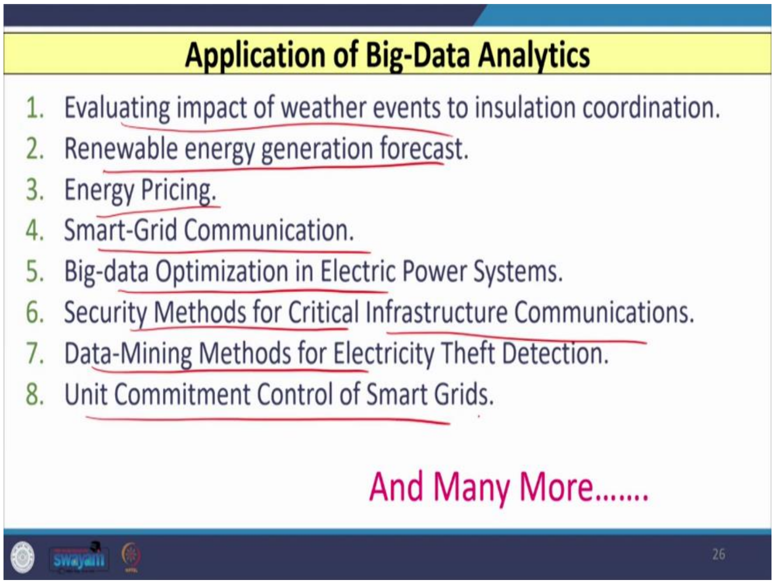
So, if we have the traveling wave data, and if we compare this with the 3-Vs that is volume, velocity and variety, then traveling wave data can generate 4 GB for storage of more than 2000 records from different line modules and with a baud rate of 115, 200 beats per second. If we have lightning data, we have 40 MB of the data per day with a baud rate of 4, 800 beats per second with a precision of 1 microsecond. And if we have GIS data, then you need additional GIS layer and that requires a few MB storage and this is up to 100 Mbps per day, this such type of data can be generated.

(Refer Slide Time: 30:18)



So, if we have fault location or improve fault location algorithm in which we need two modules, one is the traveling wave based fault location module and another is the data module which captures the data related to lightening, let us say strike polarity, location, distance, timing, etc., And this module will provide the information or required data to the traveling web-based algorithm module. And then finally, you will have the actual value of estimation of fault location. So, such type of application, you can use the big data analytics and you can perform the outage or asset management.

(Refer Slide Time: 31:02)



Big Data analytics can be also used for several other applications. Let us say you can use these big data analytics for evaluating impact of weather on insulation coordination, maybe you can use for renewable energy generation forecast, maybe you can use for pricing, you can use for the communication of smart grids devices. You can also use for optimization of the electrical power systems.

Even you can use the security methods for critical infrastructure communication systems. You can also use the data mining methods for theft detection of electricity. And maybe you can use in unit commitment of smart grids or even you can use in several other applications like when you have the electric vehicles, and then you need to capture several 1000s of data. So, in that case, you can also use big data analytics.

So, here in this lecture, we started our discussion with what is big data and big data is nothing but a large set of data captured by utility along with some other data. And we have seen that utility has to capture several data’s and some other data are provided by some other organizations. Correlating these two data, we can have a powerful decision-making algorithms and this are nothing but based on big data analytics. And then we have discussed several applications of big data analytics in different areas on which we can further explored. Thank you.