Introduction to Industry 4.0 and Industrial Internet of Things

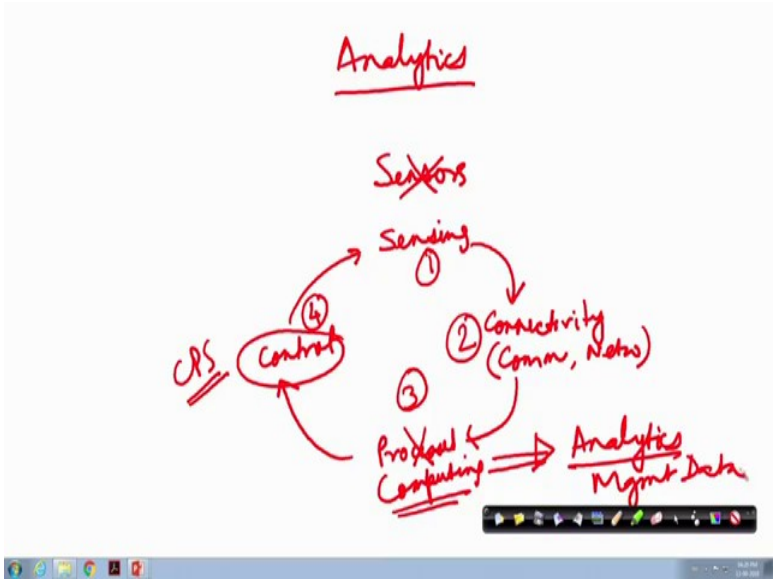
Prof. Sudip Misra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 36
IIoT Analytics and Data Management: Introduction

This particular module of this course on IIoT and industry 4.0 focuses on the analytics and data management. Particularly, in this particular lecture we are going to focus on knowing some of the introductory concepts of how to analyze the data that is received. In the previous lectures we have seen that IIoT based systems generate lot of data. And some of this data will have to be processed as soon as or as close as possible to the point of generation and the rest of the data can be sent for further processing at a later point in time. But, the analysis of the data is very important because otherwise there is no point in installing different sensors and actuators to make these different industrial machinery smarter.

So, what I would like to emphasize is we need to understand; how things go on with respect to the analytics.

(Refer Slide Time: 01:47)



So, first let us start with a recap of what goes on with the end instruments, the end devices and so on. These end devices are fitted with different smart sensors. These smart sensors basically sense and they are going to generate the data. So, instead of sensors if

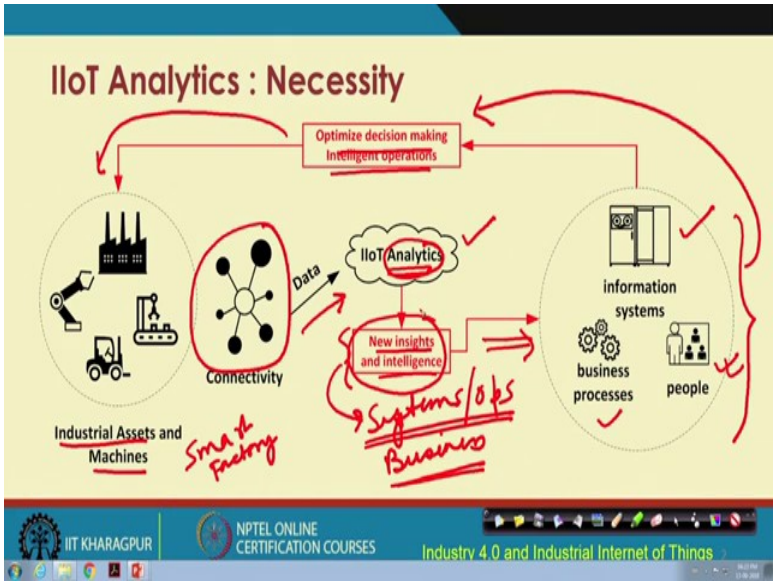
we call it sensing; I think that will be more appropriate, so, we start with the phase of sensing. From sensing we have to send the data over a network. So, let me call it connectivity. So, connectivity and in the previous module we have seen a lot about the different possibilities of connectivity; that means, talking about communication protocols network topologies and so on and so forth.

Thereafter, the data that is sent over the network will have to be processed; that means, computing. And, based on this processing in most of the industrial applications a feedback control will have to be created for feeding the signals to the source of the origination of the data; that means, these machinery fitted with the sensors.

So, this is more or less kind of a loop that continuously gets executed in a real industrial setting. So, basically it goes on like this that it starts with sensing; sensing, connectivity, processing or computing and here we have the control in most of the cases particularly with Cyber Physical Systems this control feedback is very important and most of the industrial you know industrial equipments for IIoT are CPS equipments, so, basically where there is a lot of interaction between the cyber component and the physical component and so, this control is very important.

Now, let us talk about this computing part. This is where things are very interesting. So, we have to talk about the analytics and the management of the data.

(Refer Slide Time: 04:49)



So, let us now go back and have a look at what we have in terms of analytics with respect to IIoT scenarios.

So, as I was telling you earlier we have in an industrial setting we have industrial assets and different machinery. For example, in a smart factory where there is smart manufacturing; that means, the equipments themselves are fitted with different smart sensors and so on, which continuously access the data from data of different types from this machinery. So, these industrial assets and machines these are fitted with different sensors.

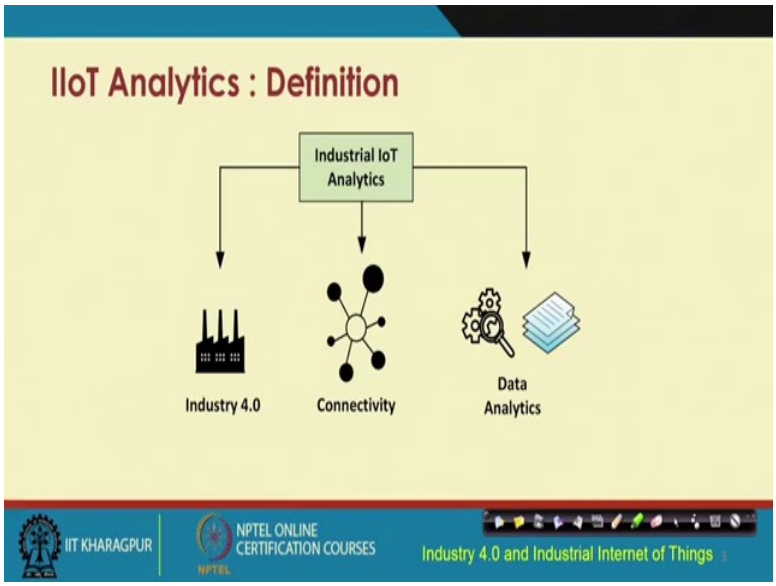
Data over a particular network are going to be sent for further processing as I was telling you earlier; the data are going to be analyzed. How it is going to be analyzed, what are the different techniques, tools and so on is what we are going to talk about in this particular module and in this particular lecture, I am going to introduce to you about the different broadly the different methodologies that are there for the analysis.

So, the purpose of this analytics is to generate new insights and intelligence, and this is this intelligence which helps in terms of the systems or the operations and business. So, overall business intelligence and operational intelligence can be derived through different types of analytics and running different analytical processes in these analytics engines is very important.

So, finally, these results are going to be all sent to either different other information systems like ERP-based systems or business processes other business processes it could be sent or to different people. So, these are the different possibilities where the new insights and intelligence that are derived through these different analytics could be made useful. So, any of them could help further to feedback and provide optimal decision making and intelligent operations and those decisions are again fed back through these machines for improved operations improved decision making and further intelligence and so on.

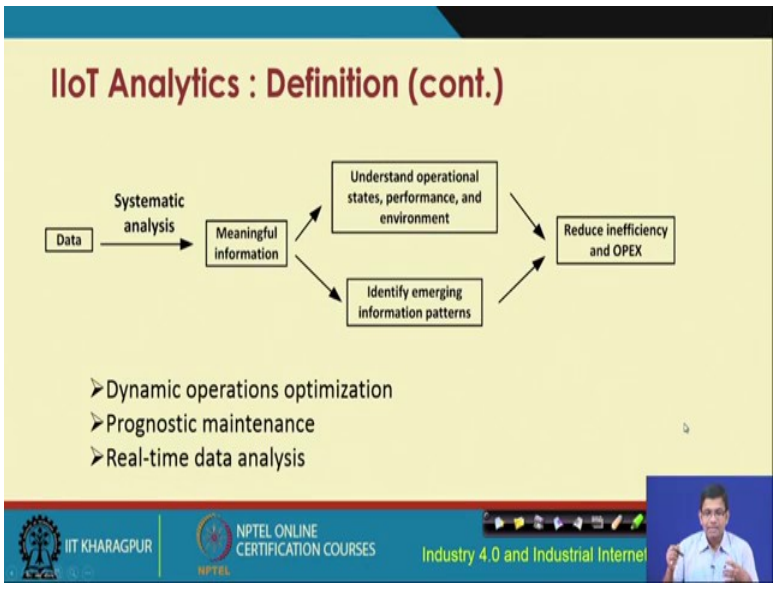
So, this is more or less this cycle where analytics finds lot of use and the core thing over here is to derive new insights and intelligence for which these different methods of analytics are going to be used.

(Refer Slide Time: 07:53)



So, talking about IIoT analytics; analytics basically is a core component for industry 4.0. In industry 4.0 we are talking about connectivity. So, basically what we see is there is an interplay between the connectivity, the machinery, the Cyber Physical Systems and this analytics.

(Refer Slide Time: 08:21)



Now, this analytics can be performed in different places for getting different types of insights. The analytics could be performed close to the point of generation; that means, at

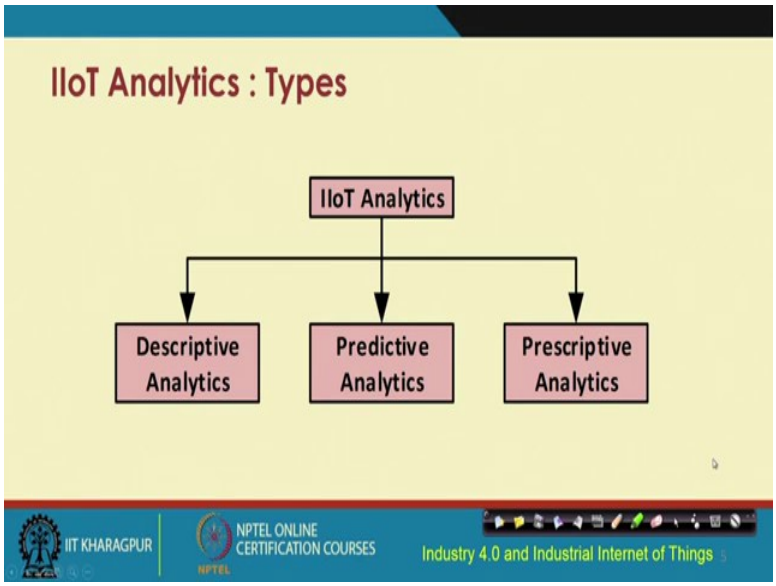
the gateway node or at the edge or the analytics could be performed completely at the back end, in the cloud, the server, the server firm and so on or it could be a mix of both.

So, let us look at this particular diagram. So, as we can see here; we have the raw data that are generated by the different sensors. These raw data are have to be analyzed because these raw data are going to be all useless, if meaningful information cannot be generated out of the data. So, basically this transformation of the data through the different analysis, a systematic analysis, transforming the data raw data into information has to be done, it is crucial otherwise it is that the data becomes useless. This meaningful information will help us to understand the operational states, the performance and the environment in the industrial setting; that means, the operational states of the machine, the performance of the machine and the environment in which the machine operates.

It also helps, this information also helps in identifying different information patterns that emerge out of this execution or the running of the system. Overall the idea is to reduce the inefficiency and also to reduce operational expenses that might otherwise get incurred if you have a bare bone kind of system deployed for the firm for manufacturing and process execution.

So, the overall idea is we need analytics for reducing inefficiencies of all sorts, overall inefficiency of the product, the process; the process that is being used in order to manufacture the product and the operational efficiency and the expense reduction of the expenses. These are the different considerations that are taken into these are the ones that are taken into consideration in order to motivate the use of analytics in that particular loop which I stated at the very beginning which basically collects the data from these different machinery fitted with different sensors, these IoT devices and send the data over the network.

(Refer Slide Time: 11:25)



So, I think we have so far a fair bit of idea about why analytics is important. Let us now try to understand and get an over overview of the different types of analytics that could be executed. We could have analytics of different types which can be classified into three; first is the descriptive analytics, second is the predictive analytics and the third is the prescriptive analytics.

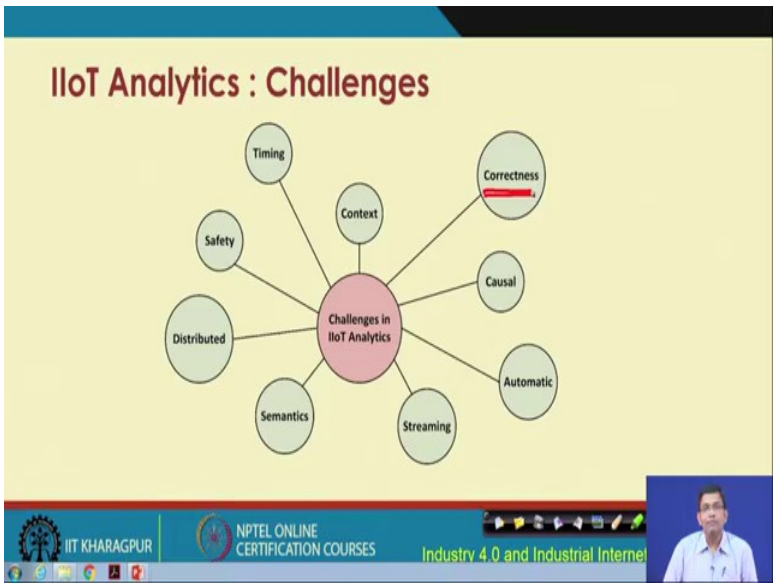
Descriptive analytics is about analysis in the current time scale. So, basically the data that are received will have to be analyzed in order to basically improve whatever is being executed at present to get more insight, to get more meaning of the data that is being received from this different machinery to get a descriptive idea of what is going on at present that is descriptive analytics.

The predictive analytics on the other hand talks about use of different tools and techniques in order to predict in the future what is likely to happen because if we can predict what is likely to happen, for example, a part of a machine component might go down in the future, maybe after 1 hour and if that information is obtained beforehand then that will be very much useful because that way you could prevent that particular component from completely going down and consequently the process or the machine that is being operated will also be autonomously taking care of the complete down time and will improve upon the overall efficiency. So, the predicting what is going to happen in the future is what predictive analytics talks about.

Prescriptive analytics basically is related to the predictive analytics. So, here we are going little bit further. In prescriptive analytics we are talking about that we predict that what is going to happen, but then based on that prediction of what is going to happen in the future, you prescribe the different actions that needs to be taken. So, that is this prescriptive different actions that need to be taken in order to maybe you know control a particular scenario a disaster from happening or something of that sort.

So, there is a link between the predictive and the prescriptive; prescriptive goes beyond what predictive does and so, there are different tools and techniques to be used for each of these descriptive, predictive, and prescriptive and so on. Descriptive analytics pair basic we could use different statistical methods for the descriptive analytics. Predictive one could use different machine learning techniques and so on and prescriptive could be a combination of machine learning statistics and different other AI techniques. So, these are the different types of analytics that could be used in an IIoT scenario.

(Refer Slide Time: 15:15)



So, there are different challenges in implementing analytics in an industrial scenario. Here are some of these challenges that are mentioned. I would start with the first one which is correctness. Correctness is very important. We are talking about machines which typically operate in high speeds where precision is very important; precision of a job that is being done is important. If precision, correctness, accuracy is not taken into account then that might lead to machine failures, process failures and including hazards.

So, safety; safety issues come into picture if correctness, position and accuracy are not taken into account.

So, your analytics that you implement will have to do the things correctly based on what is actually required and is suitable for improving the operations. Causal relationships are very important; understanding the causal relationships, the cause-effect. So, what causes something that is very important if you do not have that causal understanding of the cause-effect and the causal relationships between different events then that will not help you to implement analytics properly. Automatic with minimum human intervention everything should continue, the analytics and the different processes that get executed without any human intervention ideally should be able to control and communicate with the different parts of the system.

Streaming is a very important thing in industrial machinery, IIoT scenarios, Cyber Physical Systems we are talking about machines which run continuously; which because of the continuous operations collect lot of data, stream those data, that will send those data continuously over the channel and such kind of data coming in huge streams, large streams continuously in real time will have to be analyzed and that is not a very trivial task. Semantics and getting more insights meaning out of the data and trying to relate these meanings of one component with another and their interrelationships is also very important and should be taken into consideration while proposing an algorithm to be used for the analytics.

Distributed nature is very important. Nowadays we are talking about not centralized systems, but distributed systems which have certain parts or components in one location and other parts are geo distributed located in another geographic location. So, you need to have distributed analytics; analytics which will take into account the differences in the geo-locations and their corresponding executions of the different parts of the same system or maybe of a different system, but are synchronized together that has to be taken into consideration.

So, distributed analytics is also very important because certain parts of the analytics might have to be done close to the component located in one location, certain other parts close to the component located in another location and certain analytics will have to be done in synchrony putting together the bigger picture all the components together and

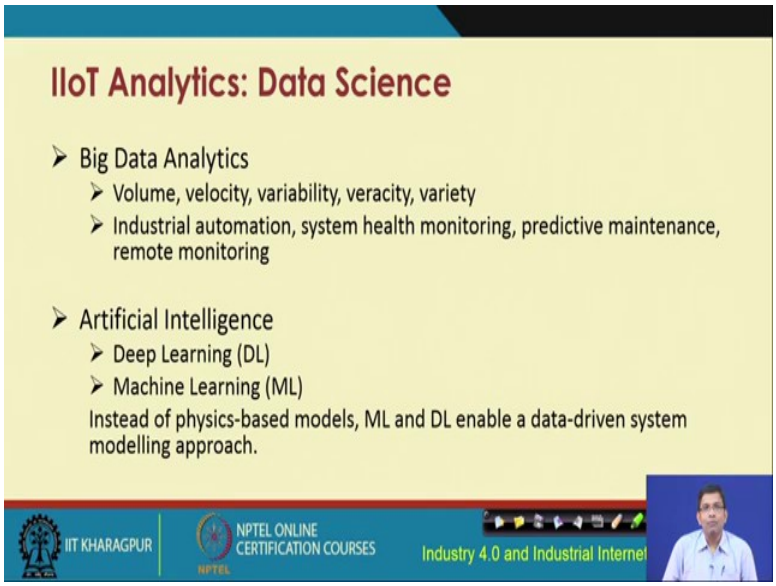
making an overall idea about what is happening or what is going to happen and making certain prescriptions if it is required.

Safety is very important in industry scenarios and I cannot emphasize this more. Safety analytics is very important because you know talking about plain bare basic ergonomics safety to safety from different disasters machine failures or consider something like if the operator of a crane makes certain mistake what might so happen is basically there might be some disasters underneath. So, people might die, right. So, there might be deaths that might be associated with certain mistakes.

And, all of these can happen in autonomous systems where things are happening due to certain algorithms that are implemented behind the scene and are getting executed and if some something goes wrong or is imprecise then basically that might lead to certain failures which can lead to accidents in the industrial settings. So, starting from ergonomics safety to accidents etc. safety is very important and paramount in most of the industries particularly in industry 4.0. So, safety is a very important consideration. So, safety reliability and so on are very very important and so, analytics considering safety issues are very important.

Timing is very important because if a particular process gets delayed in execution and that particular signal coming from that particular routine or that algorithm under execution gets delayed by even a millisecond that might also lead to disasters. So, timing is very important and the context. Context is very important taking into account taking into account the different contexts in which the operations are going on and holistically trying to have the analytics that is very important. So, these are the different challenges with respect to analytics for IIoT scenarios.

(Refer Slide Time: 21:33)

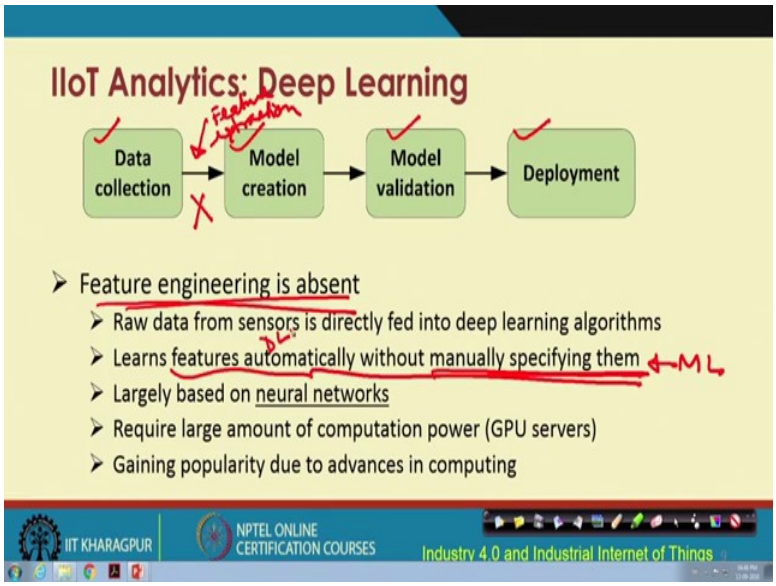


So, we have IIoT analytics; the concept nowadays people are talking about data science, right. So, basically data science is nothing, but trying to derive meaning or insights from data, insights that may not be very apparent and so on. So, this is a new trend in terms of trying to gain meanings out of the data. So, big data analytics is very popular. Big data we are talking about data coming in huge volumes, coming in high velocities in real time streams of data; data of which have high variability and veracity, variety of data and so on. So, all of these are different attributes of big data.

So, big data analytics is a huge challenge, but is a reality and analytics of these types of data, data which are unstructured, data which cannot be stored in relational tables. So, that you could use the relational methodologies for extracting meaning out of the data. So, those things are not available. So, in the absence of that for a data which is unstructured and bearing these kind of characteristics which is a reality in the industrial settings, how do you analyze the data in order to have insights or meaning out of the data as soon as possible; that means, as close as possible to the time when this data gets generated.

AI techniques such as machine learning, deep learning etc. are becoming very popular. So, these techniques together can help in getting more insights about the data.

(Refer Slide Time: 23:25)



So, now we are talking about a world where machines are data driven, systems are data driven. Systems are data driven means like; you have lot of data and data is controlling. The data has become so important in the industry 4.0 world that it is so important that data itself is driving about the course of actions that have to be taken in terms of the manufacturing processes, the products that are being made and also the business operations.

So, let us talk about machine learning. So, I talked about machine learning and deep learning. So, machine learning basically it is a field of AI and this machine learning basically what we are talking about in machine learning is to derive some kind of intelligence to be derived, right. So, it starts with a collection of data if you look at which comes from different machines, the sensors in these different machines typically. Then there is something called feature extraction which is very important in machine learning.

So, feature extraction methods basically what they do is they convert the raw data into information that relates to the physical state of the asset. So, the feature extractions are performed and feature extraction itself is a task that requires lot of work. So, there are different algorithms that are available for feature extraction. So, once these features are extracted these are the features which will have close resemblance to the actual state of these machines and their operations.

So, following which feature extraction which is very important in machine learning. So, in machine learning feature extraction is very important. So, following feature extraction there is this model creation and model validation. So, models have to be created and then these models will have to be validated and thereafter these models will have to be deployed for further use.

So, there are two types of machine learning. Machine learning means that you are learning, you are learning from this data. The data that are collected from this you are doing certain things, certain activities, feature extraction, model creation, model validation etc. and from that you try to learn from the deployed models, you try to learn and about what is going to happen in the future. So, this is what machine learning talks about at every cross level.

So, this machine learning could be of two types supervised or unsupervised. In supervised in a nutshell I can tell you that supervised algorithms of machine learning are the ones where there is some kind of training data of different sets already available and you use that past historical training data in order to come up with your models. So, models will be created based on this training data and then these models could be also validated using other sets of data.

So, supervised learning basically uses training data. If training data is not available and if the feature extraction will have to be done in the absence of training data unsupervised machine learning algorithms are useful. So, there is no label data in unsupervised methods and these unsupervised learning methods basically extract a structure in the input data on their own. So, this is the main difference between the supervised and the unsupervised learning methods.

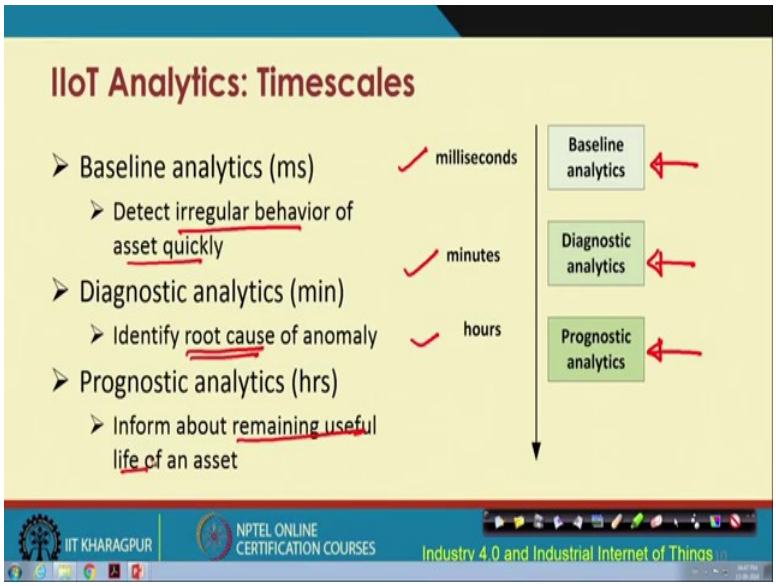
We will talk about this in little bit more detail in another lecture, but let us now talk about deep learning which is another thing that people are talking about a lot in the present day. So, people are talking about machine learning, deep learning in the industrial settings. Now, deep learning is like machine learning. It is another discipline of, we can think of that deep learning is another way of machine learning we can think that way. But, in deep learning unlike in the case of machine learning where feature extraction etc. was done; in deep learning you do not have that provision; so, there is no feature engineering aspects present.

So, if you look at this overall block diagram. So, we have the data collection, model creation, model validation and deployment. So, what is absent over here is unlike in the case of machine learning, this feature extraction which was there in machine learning is not there in deep learning. So, feature engineering is absent. So, essentially what happens is the raw data from the sensors are fed into these deep learning algorithms and from that different features will automatically get learnt without actually manually specifying those features which used to happen in the case of machine learning. So, learning different features automatically without manually specifying them this is basically the deep learning.

So, most of these deep learning methods are based on traditional neural networks; techniques such as, methodologies such as deep neural networks are quite popular and because of the structure of these deep neural networks and so on and the similar kinds of methodologies that are present what is required is to have large amounts of computation power of these different computational infrastructure which run these algorithms. So, deep learning algorithms will require GPU servers and the like.

So, these deep learning methods are getting very popular due to the fact that nowadays we have cheap computing power unlike in the previous times, cheap and much more efficient advanced computing at a very reduced price we are able to have at present and that is where deep learning is getting very popular.

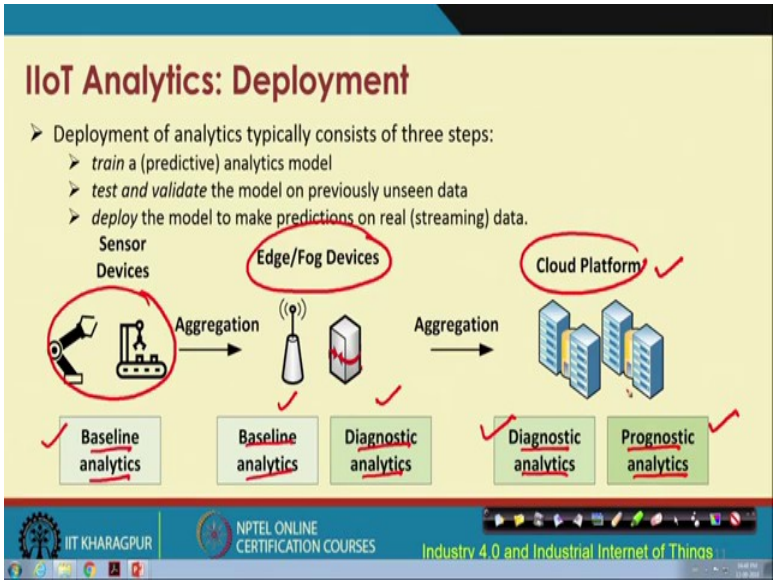
(Refer Slide Time: 30:33)



In terms of the time scales for analytics; analytics will have to be performed in milliseconds, in minutes, in hours. Milliseconds basically the baseline analytics, in minutes are diagnostic analytics and in hours our prognostic analytics. Baseline analytics detect irregular behavior of assets quickly; diagnostic analytics identify the root cause of any anomaly; and prognostic analytics basically inform about the remaining useful life of an asset.

So, these are the different types of analytics and their corresponding time skills of operation.

(Refer Slide Time: 31:19)



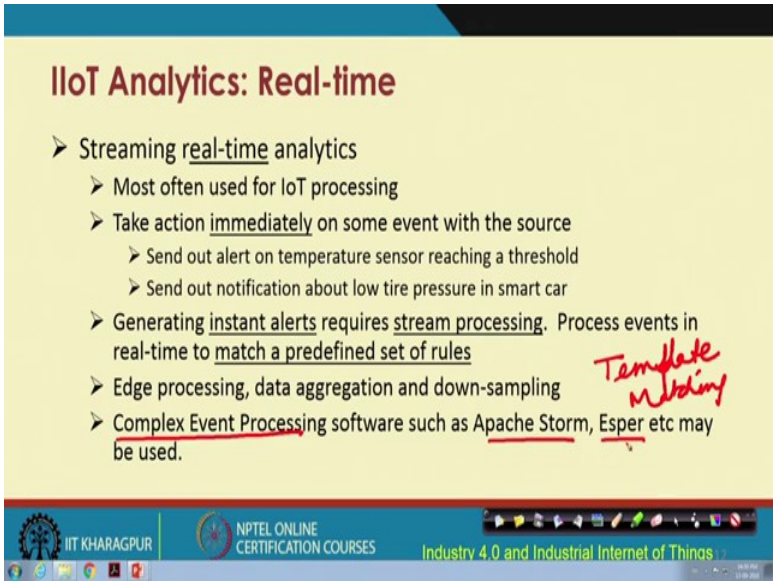
Talking about the deployment; deployment of analytics typically consists of different steps first you have to train if we are talking about some kind of predictive analytics you have to train a particular model. You have to train a model and then test and validate that particular model on some previously unseen data and thereafter deploy that model to make predictions on an actual data set which is being streamed in real time.

So, this is what is going to happen. You have this baseline analytics, from these different machines, these sensor devices this baseline analytics will be done over there that will be done very fast data will be aggregated and then at the gateway devices or edge or fog devices which are basically network equipments that are close to this machinery some baseline analytics and a combination with diagnostic analytics will be performed and

again this data are going to be further aggregated, sent to the cloud and in this cloud basically diagnostic analytics and prognostic analytics will be performed.

So, starting from baseline analytics at the source to a combination of baseline and diagnostic at the edge to the diagnostic and prognostic analytics at the cloud we have these different phases of operations of analytics.

(Refer Slide Time: 32:53)



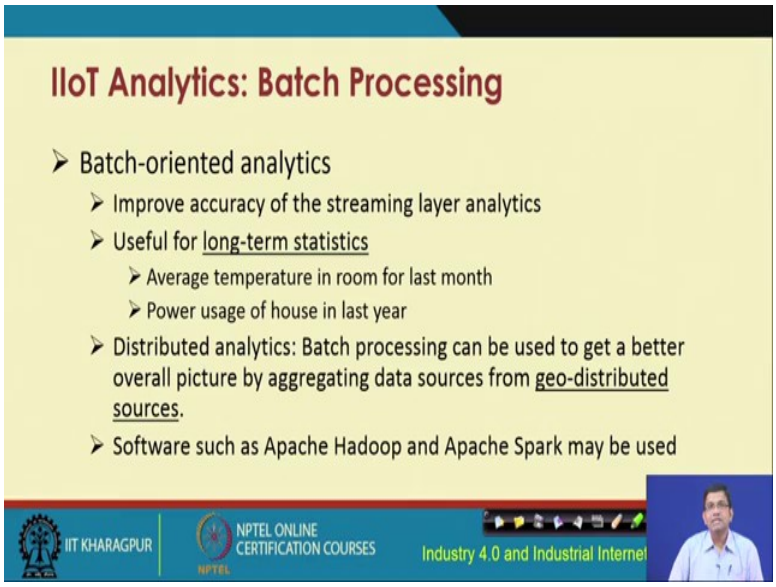
So, analytics will have to be done in real time. We have machines sending data, streaming data in real time. So, the streams of data will have to be processed and actions will have to be taken immediately as close as possible, as soon as possible and to the source of the data because based on that useful alerts can be sent. For example, if a particular machine is crossing or become coming close to a particular temperature threshold then an alert could be sent in a particular machine in operation.

Different notifications about tyre pressures for example, in a smart car could be sent immediately. So, certain baseline real time analytics could be performed and it is also very important to instantly send these alerts through different stream processing and so on. So, how that can be done? Because execution of different processes will take some time. So, that can be done with different techniques something such as matching a predefined set of rules in real time. So, you have a predefined set of rules and some kind of matching can be done. Techniques such as template matching, in machine learning could be used for doing it.

So, template matching is a useful technique that could be used for some kind of necessities like this template matching. Edge processing, data aggregation, and down sampling these are very important in analytics particularly for streams of data coming in real time will have to be processed in real time and so on.

So, complex event processing software such as Apache Storm, Esper etc. may be used for this particular purpose.

(Refer Slide Time: 34:53)

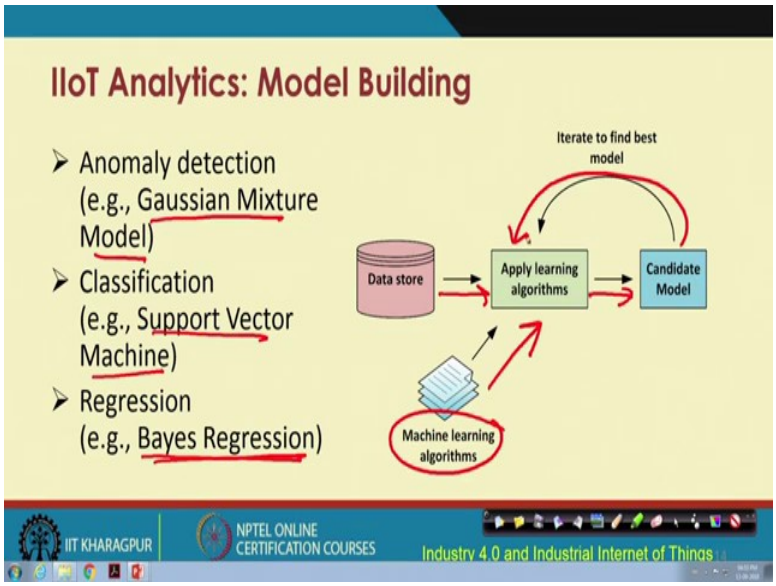


Batch processing might also be a requirement. Batch processing basically you know talks about getting lot of data, executing those data, analyzing those data for long term statistics in batches, right. For example, the average temperature of a room for the last one month that is some kind of batch processing which will give you some kind of long term statistics or the power usage of a house in the last year.

So, this is very important. Distributed analytics is very important because as I said earlier we are talking about industrial scenarios where different machines and their components are all geo distributed and some kind of data aggregation and intelligence will have to be performed as close as possible to these different components and their sources of origination.

So, software such as Apache Hadoop, Apache Spark etc. may be used for these purposes.

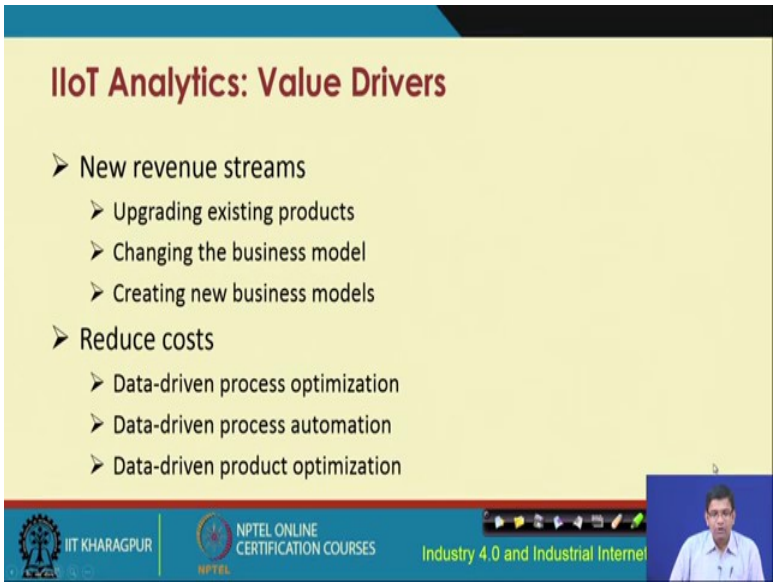
(Refer Slide Time: 35:55)



Model building: for model building, so, I told you earlier that for machine learning after feature extraction model building is very important. So, for model building different-different methodologies are there. For example, for anomaly detection GMM models are there – Gaussian Mixture Models, for classification SVM or Support Vector Machines are there, for digression Bayesian regression techniques are there.

So, basically machine learning algorithms could be implemented those algorithms could be applied on the data that comes from the data store and then these candidate models are basically built and then are fed back to these different learning algorithms to improve upon their course of action.

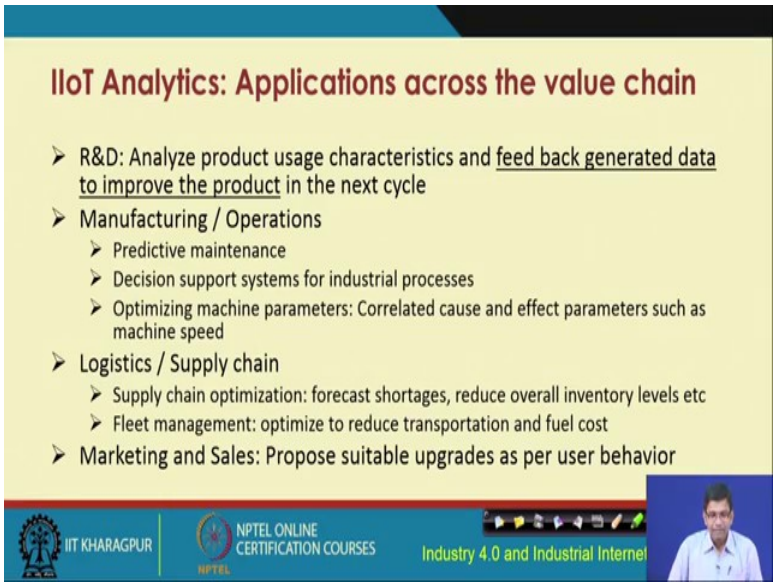
(Refer Slide Time: 36:55)



The different value drivers for IIoT analytics are like this, that it is required to derive new value streams, new revenue streams. For upgrading the existing products analytics is going to be useful, for changing the business model analytics would be useful, for creating new business models analytics would be useful. So, these are all valuable for doing different-different stuff in different business and industrial settings.

Analytics is also valuable for reducing the cost, for data-driven process optimization, for process automation and for product optimization. Data-driven process optimization, data-driven process automation and data-driven product optimization for everything we need different suitable analytical techniques to be implemented.

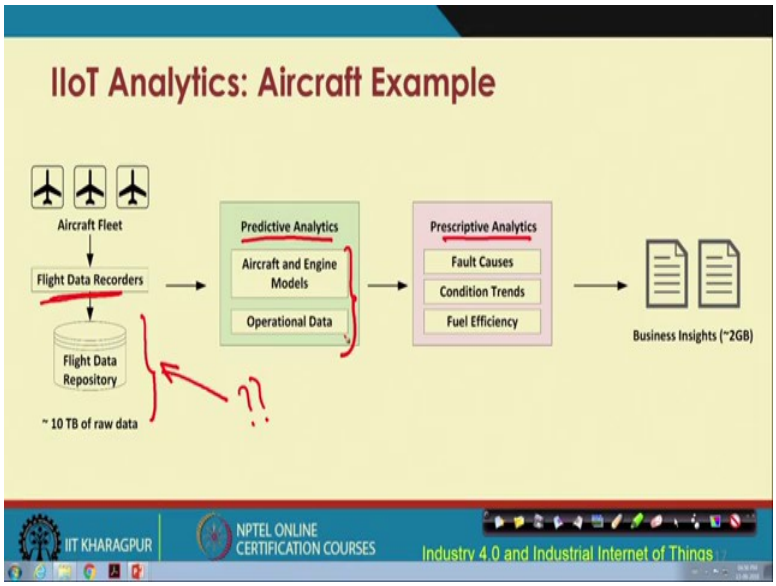
(Refer Slide Time: 37:53)



Different applications across the value chain. For R and D analytics is useful to basically understand the usage characteristics of a particular product. And, then feedback those intelligence those insights generated from this analysis to the next cycle of the product that is being manufactured. So, for R and D also analytics is useful. For manufacturing and operations, predictive maintenance, decision support systems for industrial processes and for optimizing machine parameters based on the collected data from the current operations IIoT analytics is important.

For likewise for logistics and supply chain, supply chain optimization, fleet management optimizing the reduction, optimizing the transportation and fuel costs in fleet management you need to derive intelligence out of the data. And, similarly for marketing and sales also to suggest suitable upgrades of a particular product based on the user behavior you need analytics. So, analytics is important in different fronts.

(Refer Slide Time: 39:03)

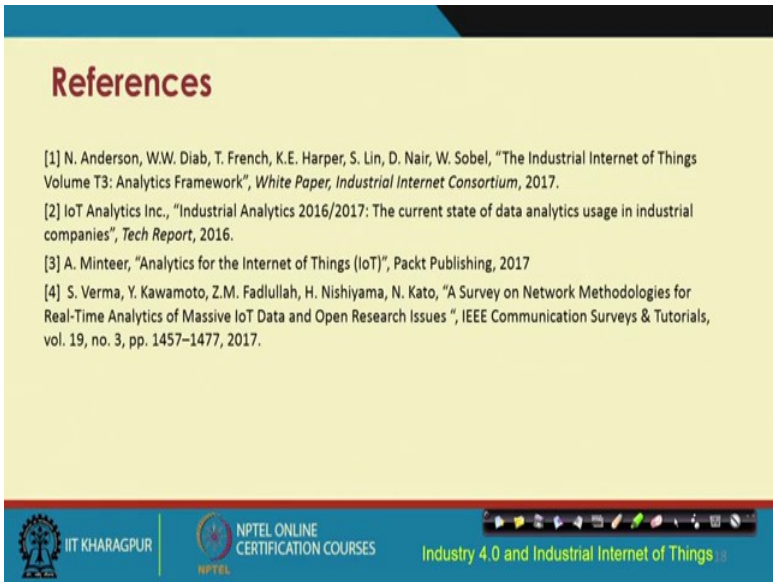


So, let me just take an example of aircrafts. So, aircraft fleet; each aircraft as you know is equipped with different flight data recorders. These flight data recorders they are going to record the data that are coming from the flight, from the different parts, different components of each of these different flights and are going to be stored in these data recorders. These data could be sensor data and different other data that are coming would be all stored in these flight data recorders. These are going to be stored; these data for every flight terabytes of raw data are produced for every flight, right terabytes.

So, huge volumes of data are stored. So, this data will be useless if you do not get meaning out of if you do not derive intelligence out of this data. So, what is done? Different types of analytics like the predictive, prescriptive, analytics, predictive analytics dealing with you know based on the aircraft and engine models and the operational data predicting about what might be happening in the future. So, that some emergency landing could be taken and a disaster would be avoided for a particular aircraft by the captain.

So, prescriptive analytics basically deals with fault causes, condition trains, fuel efficiency and so on. So, both predictive and prescriptive analytics can help in getting business insights and so on. So, this is very important and this is just an example where analytics is very important.

(Refer Slide Time: 40:57)



So, with this we come to an end of this particular lecture. These are some of these references that you could refer to. Here in this lecture we talked about the importance of analytics. We talked about the classification of different types of analytics and the different methodologies for analytics; we talked about machine learning, deep learning methods and that are available; we also saw an example where analytics would be very much useful.

So, these are the references that you could go through and with this we come to an end.

Thank you.