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 – 33
Key Enablers of Industrial IoT: Processing – Part 1

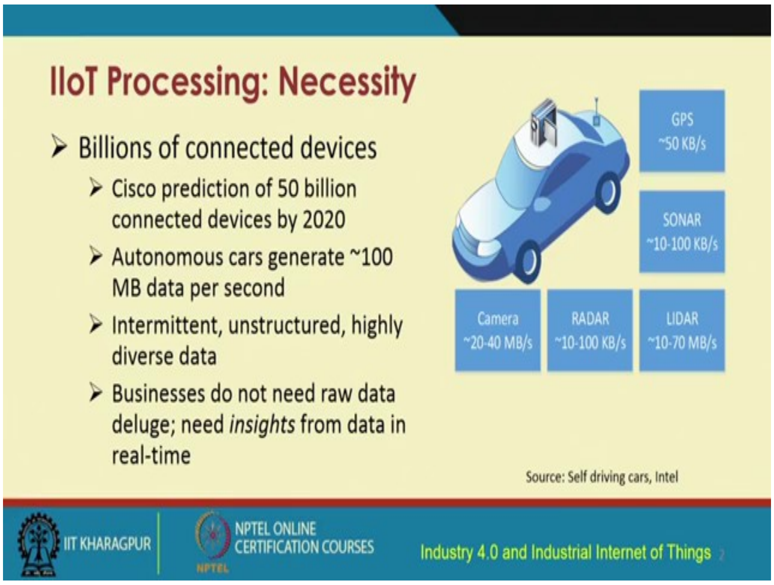
In this lecture and the next, we are going to talk about the processing of the data, that are received through the IIoT systems, that are deployed in the companies. Let us go back and think about what we have covered so far, in this particular module. We have talked about the deployment of different types of sensors, in industrial scale. In different parts of the manufacturing plants, different industries, the different types of sensors that could be used.

We started with that then we talked about in length, we talked about the different types of connectivity mechanisms, that could be used. We talked about the different protocols, the communication protocols, the network infrastructure, and so on the different types of topologies that could be used, and so on.

We have talked about all of these connectivity issues. So, what happens is these different sensors and the actuators are all connected. Through this particular network that we have gone through, in the last few lectures. So, the data basically are received at somewhere, it is received let us say in the cloud or in some server or somewhere it is received, and this data will have to be processed.

So, it has to be processed as quickly as possible in most of the cases. So, how this data has to be processed and what are the different mechanisms in place, what are the different research works that are going on in terms of processing of IoT data and so on is what we are going to look at in this particular lecture and the next.

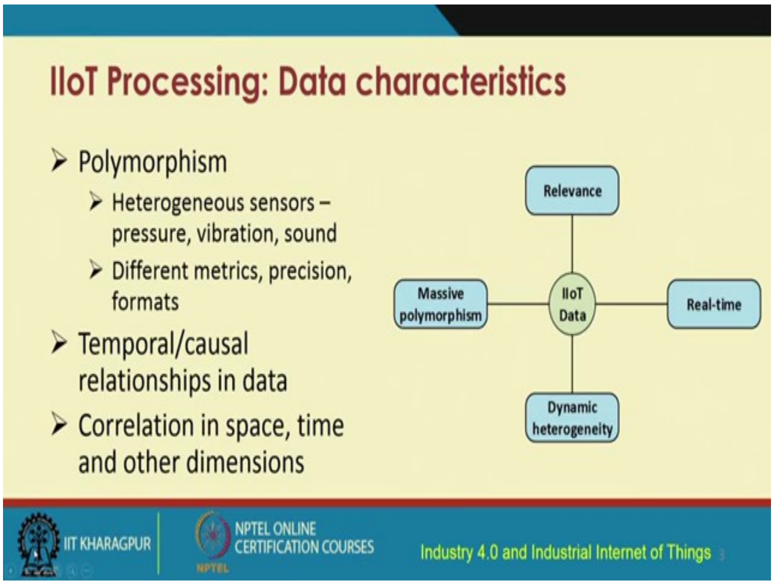
(Refer Slide Time: 02:00)



Let us take a look at the self-driving cars, which has very become very popular in the recent times. So, these self-driving cars or forget about the self-driving cars, if we are talking about the smart cars, in general. These smart cars are equipped with different sensors and different communication devices. And these different cars it is expected that we are going to have these connected cars in huge numbers in the next few years. These cars they will generate large volumes of data and it is estimated as per one of the estimates, it is estimated that these cars and particularly the autonomous cars; they are going to generate data at 100 MB per second. At 100 mbps, it is estimated that this much of data is going to be generated per second. So, this data will have to be dealt with. If, we take the case of the self-driving cars, the self-driving cars are equipped with GPS. The GPS transmits data in these cars at the rate of roughly 50 kbps, sonar these cars could be equipped with sonar modules and which could be sending data at the rate of 10 to 100 kbps. LiDAR 10 to 70 mbps, radar 10 to 100 kbps, cameras 20 to 40 mbps, and so on.

So, as we can see that individual components of these autonomous cars, they are going to throw in lot of data at huge rates. So, you have to deal with this kind of data and in order to get proper meaning out of this particular data that is collected, there has to be some kind of processing engine, that is going to be executed, and that will derive the meaning out of the data, that are collected and received.

(Refer Slide Time: 04:16)



How we can deal with this data, we need to first understand the type of data, that we are receiving. The type of data, that we receive can be characterized in different ways. First of all this data has massive polymorphism in place. Massive polymorphism means, we are talking about the use of large number of different types of sensors, pressure sensors, vibration sensors, sound sensors, and large number of different types of sensors could be used.

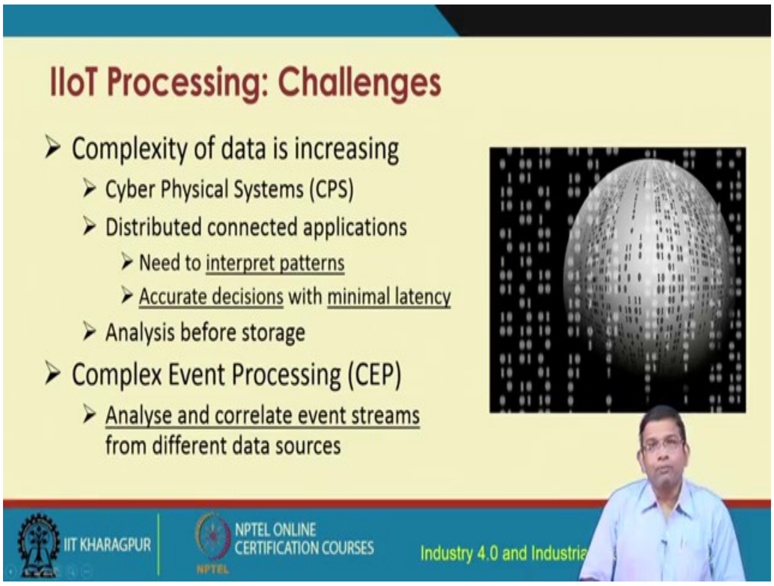
This heterogeneous say data from these different types of sensors will have to be dealt with, because they come in different rates, they are of different types their data types are different and so on. So, they have different varieties, different types of data coming through a single pipe. Additionally, this data will come in different formats their precision levels will be different; the metrics that characterize this data are going to be all different.

So, this kind of variety of data having different forms, of having different types of varieties, will have to be dealt with. The next thing is that you have to deal with data in real time, to be able to make most value out of the received data. So, this real time dealing of the data will have to be performed. So, the data will have to be dealt with in real time and we have already seen that there is heterogeneity in the type of data that is received through this common pipe, conceptually, if we can think of that way. And this heterogeneity itself is going to be dynamic with respect to time.

The type of heterogeneity is also going to change with over time plus this relevance of this particular data. The relationships--the causal relationships, the temporal relationships between these different data components, these are also very important and will have to be understood in order to make this data useful. There has to be some kind of correlation, which will have to be established in space in time and other possible dimensions of the data that is received.

So, these are some of these different attributes or the characteristics of the data and we have seen that these this type of data will have to be dealt with appropriately.

(Refer Slide Time: 07:02)

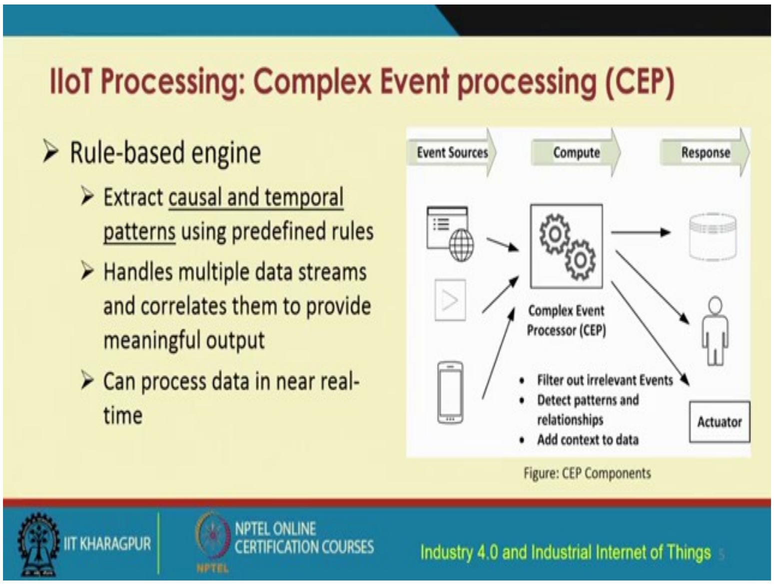


Now, the so, we have to process the data, but there are different challenges with respect to this processing. The complexity of data in industrial systems is increasing day by day. These industrial systems exhibit the behavior of cyber physical systems, where there is an intricate relationship, that is there between the cyber component and the physical component of the different machines and there and the interconnected ones.

So, what is required is to interpret through different applications, it is required to interpret the patterns that are there, out of the data that are coming from these different cyber physical systems. Accurate decisions will have to be taken and these decisions will have to be taken with minimal latency; that means, in least possible time, this decision will have to be taken about decision means like what is the the data that are coming, what it is going to do what if there is any kind of thing that is going to happen in the future. So, those kind of decisions will have to be taken.

So, even before the storage so, first of all, the data will have to be stored. And one can analyze, the data after storage, but before storage also the in many cases the data may need to be analyzed. And this is particularly true for real-time applications, where there are real-time decisions will have to be taken in real-time. So, there is some kind of complex event processing, that will have to be executed, the data will have to be analyzed correlations between different event streams will have to be found out from these different data sources.

(Refer Slide Time: 09:07)



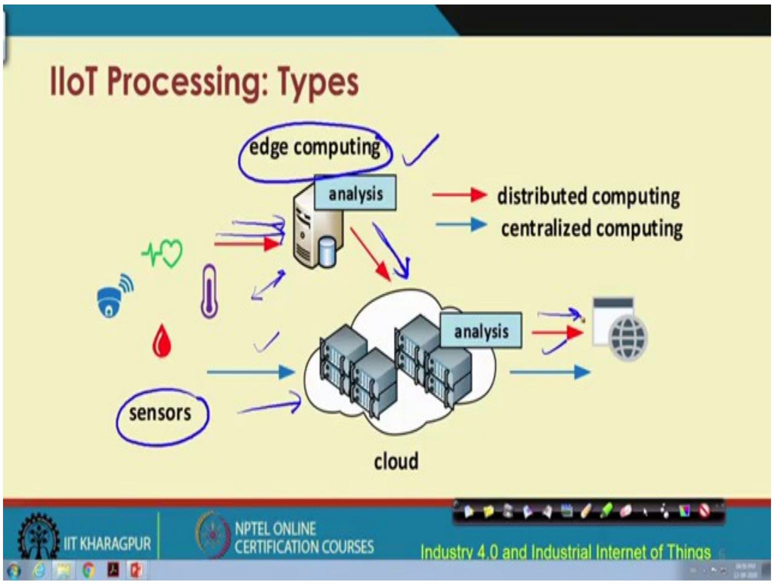
So, let us look at this complex event processing. So, one of the ways to do it is with the help of a rule based engine and this is shown over here in this particular figure. So, let us look at the complex event processing, one of the ways to implement it is using rule-based engine.

So, let us look at this particular figure, we have this complex event processing mechanism, which does the computation. Data come from different event sources and this data will have to be processed, with respect to filtering out irrelevant events, with respect to detecting patterns and relationships, and adding context to the data.

So, all of these things will have to be done appropriately in this compute engine. And, based on that either there would be some kind of an actuation or, this data will have to be sent appropriately to certain actor or, it has to be stored. So, a complex event processing will have to be performed. So, it is very important to extract the causal and temporal patterns causal and temporal patterns, will have to be found out, through complex event processing in this particular component.

So, it is very important to process the data in real near real-time; that means as much real-time as possible in order to be able to have value out of the decisions, that are made out of the processing of the data.

(Refer Slide Time: 11:05)



So, if we are talking about this processing, where do we do this processing? So, let us say that we have different sensors like the ones that are shown in this figure.

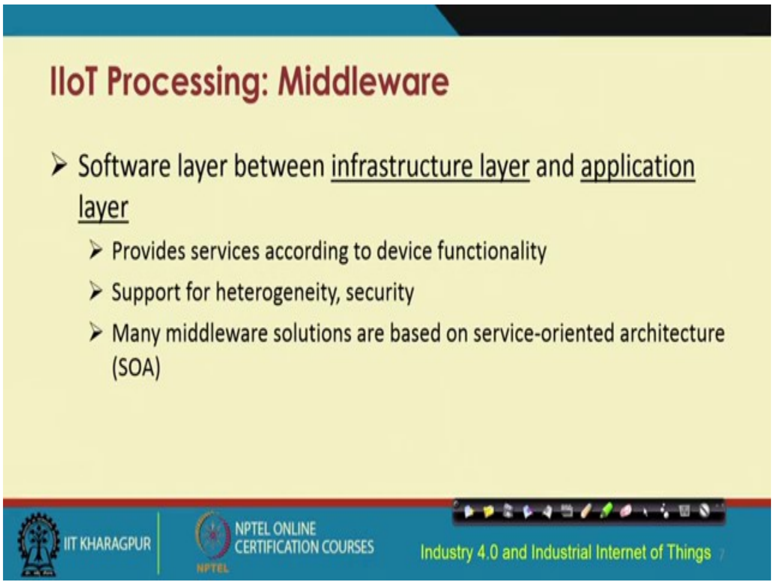
So, this these sensors will be sending lot of data. This data will have to be processed. This processing can be either done through some server's small servers or, something of that type, close to the sensors, as much close as possible to the sensors, in order to reduce the latency in processing or, the data could be sent to the cloud. It is also possible that partially some part of the data will be processed at the edge over here, and the rest of the processing would be done at the cloud.

So, this is the flow of distributed processing or distributed computing the red colored arrow shows it and the centralized processing pathway is shown, through the blue colored arrow. So, this is basically the centralized processing, and this one is the pathway for the distributed processing of the data.

So, what we see that the data that are produced from the sensors, will either have to be analyzed will have to be sent completely to the cloud, and will have to be processed at the cloud, or it could be sent to the edge partial processing, will take place at the edge. And the rest of the processing will be done at the cloud or it is also possible for small jobs, the processing will be done completely at the edge.

So, all the 3 different possibilities are possible to be implemented. Now, what is important is to identify which type of processing, what computation will be done at the edge, which processing, what computation will be done at the cloud? And more importantly, if you have to distribute the processing between the edge and the cloud, then how do you make that differentiation and correspondingly the distribution.

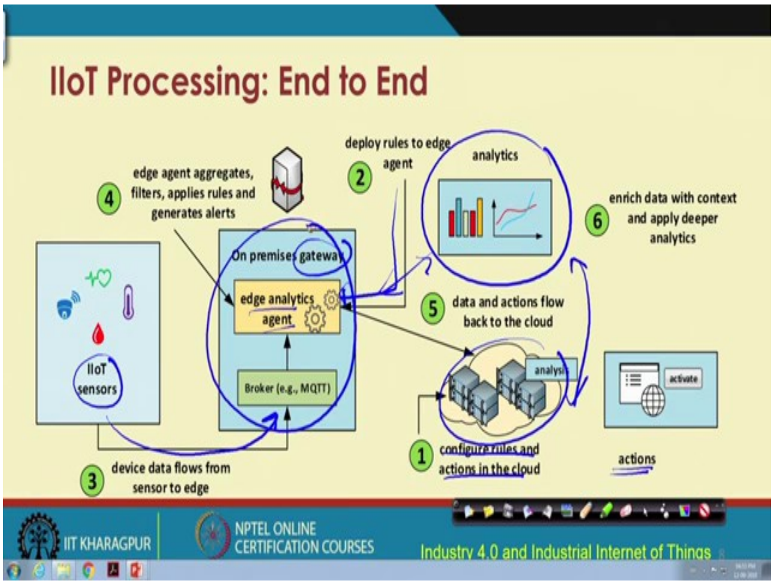
(Refer Slide Time: 13:35)



So, it is proposed that we could use some sort of a middleware. Middleware means it is a software layer, which will be standing between the infrastructure layer and the application layer. So, this particular middleware software will provide different services according to the device functionalities. This middleware could also implement security mechanisms; it could also handle device heterogeneity and correspondingly data heterogeneity.

So, there are different middleware solutions, that are proposed and many of them are based on the service orientation, service oriented architecture based middleware.

(Refer Slide Time: 14:26)



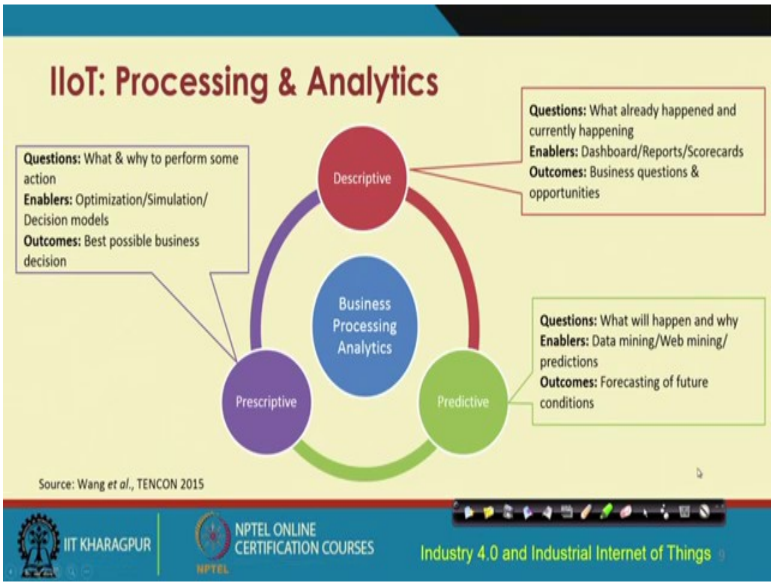
So, let us look at this end to end mechanism and let us try to highlight stepwise, how this processing is coming into picture and its importance. So, let us look at this particular figure and as shown over here. We are talking about; we are talking about different steps; step one is configuring the rules and actions in the cloud.

So, after this configuration is done at the cloud, it is required to deploy these rules at the edge. This deployment will have to be done. Next, the device data will flow from these sensors these IIoT sensors, these device data are going to flow from the sensors or, the end devices to the edge device. Let us assume, that this is the edge device or the gateway, which is doing partial processing.

There after this edge agent analytics agent will do things like aggregation filtering applying the rules generating alerts. There after the data and the action flow are sent back to the cloud, for further processing, where initially we had started with the configuration of the rules and the corresponding actions, that are going to be taken. And thereafter, the actual actions are going to be taken. And also it is important to analyze the data based on this analysis, basically the data, the actions will be taken.

So, basically this analysis of the data will have to be performed and this will have to be done mostly at the cloud, but partly this analytics could also be done at the edge; that means the gateway device.

(Refer Slide Time: 16:34)

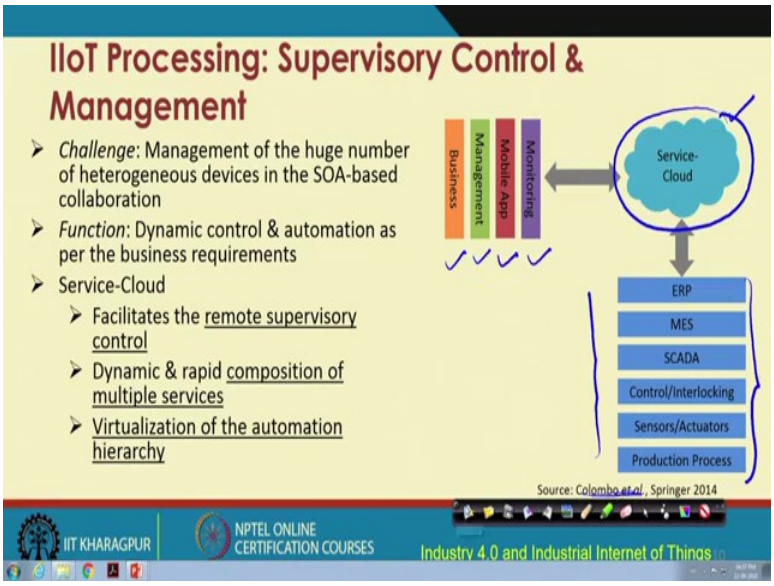


So, we talked about processing, we talked about analytics. So, some business processing analytics will have to be performed. So, this business processing analytics could be of 3 different types, it could be prescriptive, descriptive or predictive.

Prescriptive basically questions things like, what and why to perform some of the actions. And the enablers for doing it, are some optimization techniques simulations, use of different decision models, and so on. And finally, the outcome is the best possible business decision and this is prescriptive. As this name suggests prescriptive means that it will prescribe, that what is the best possible business decision.

Descriptive talks about what has already happened and is currently happening. And based on that coming up with outcome such as the business questions and the opportunities, the enablers for descriptive analytics are dashboards, reports, scorecards, and so on. Prescriptive analytics questions, what will happen and why it will happen? The enablers for it are data mining different data mining techniques, web mining techniques, and predictive techniques. The outcomes from predictive analytics are forecasting for future conditions

(Refer Slide Time: 18:16)



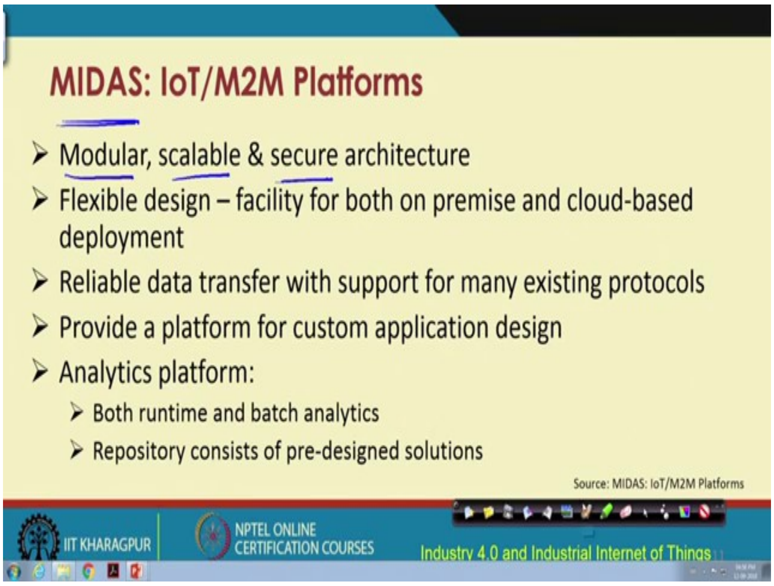
So, there are different solutions, that are placed that are in place, with respect to analytics and processing. So, one of the solutions the source is given over here. So, this particular solution was proposed by Colombo et al, talks about supervisory control and management in the context of processing. So, as we can see over here its 3 different components in this particular solution mechanism.

At the center is basically the service cloud and at the bottom are a stack of different layers. So, it starts with the production process; production process, then the sensors and the actuators control and interlocking SCADA, MES, and ERP. So, these are the different things, that are implemented as per their solution. And at the same time different functionalities such as monitoring having mobile apps, running different mobile apps, management strategies, business logic implementations, are also performed.

So, this service cloud that we talk about over here, as per the solution by Colombo et al, facilitates some kind of supervisory remote supervisory control. It also does service composition. Service composition means like we will have multiple different services, which will be aggregated together into different clusters of services. So, this is the service composition. And the third component of the service cloud or this or its functionality is the virtualization of the automation hierarchy.

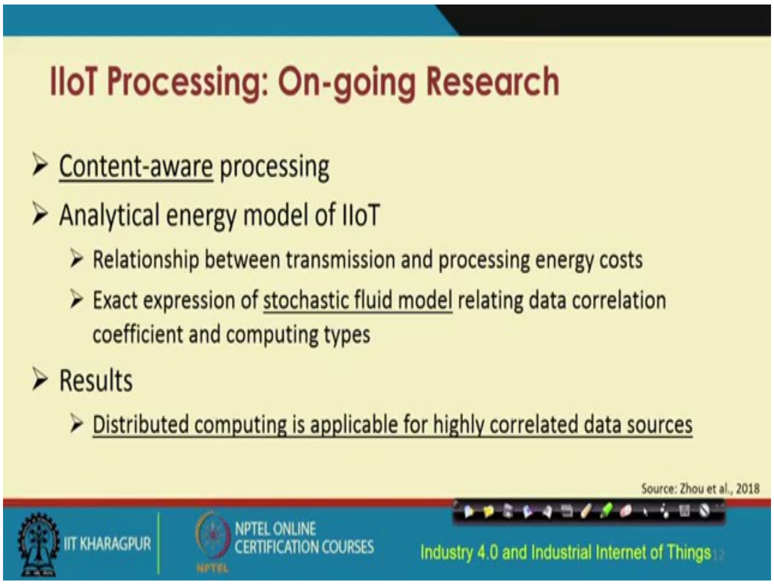
So, these are the 3 primary functionalities of the service cloud as per the proposal by Colombo et al.

(Refer Slide Time: 20:25)



There is another solution named as MIDAS, which provides some kind of IoT/M2M platform. So, this MIDAS platform is a modular, scalable, and secure architectural platform. It provides a flexible design and offers a facility, for both premise and cloud based deployment. So, there are different types of analytics that can be performed through the MIDAS platform, runtime analytics could be performed or batch analytics could be performed and the repository in MIDAS consists of different predesigned solutions, that have been implemented

(Refer Slide Time: 21:21)



Now, let us look at some ongoing research works with respect to processing, with respect to processing one of the important research works by Zhou et al. the corresponding complete reference is given at the end of the slides. So, this particular work talks about content-aware processing. So, the processing is done in a content-aware. So, what is there the content the processing is done, content aware.

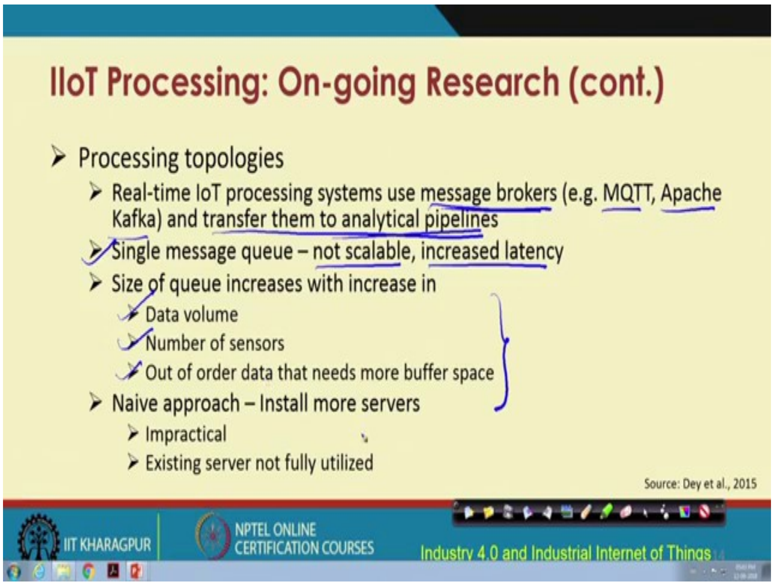
So, they talk about an analytical energy model for IIoT, where the relationship between the transmission and processing energy costs are established. And some kind of a stochastic fluid model is used to correlate the data and try to infer the correlation basically try to infer the decisions, based on that particular correlation.

So, the results of this particular system is to enable distributed computing, for highly correlated data sources. So, that was content-aware. The next one is context-aware. And this particular mechanism was proposed by Akbar et al., and they talk about the use of contexts, different context the information about these contexts are used in order to do the processing. So, here they are talking about stream processing.

So, different streams of data coming from different sources tagged with different contexts are going to be used for processing. So, the limitations of the current CEP -based systems; that means, the context-aware systems is or the traditional systems I am sorry that limitations of the traditional systems are that they use some kind of manual thresholding. So, some kind of a threshold is specified beforehand manually and based on that the decisions are made, and those are not context-aware. So, Akbar et al they proposed some kind of context aware engine which they have named as the micro-CEP engine, which uses adaptive clustering techniques to dynamically detect the boundaries, between the CEP values and find optimal rules.

So, that they do with the help of different rules, which are adaptive in nature, and which are able to extract the causal and contextual relationships between these different data.

(Refer Slide Time: 24:17)

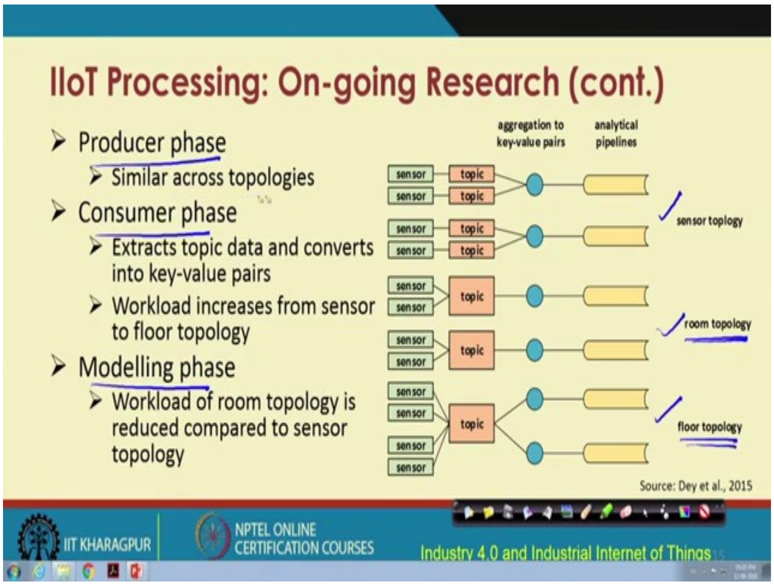


Another work by Dey et al. they talk about the use of different processing topologies. So, here they are talking about real time IOT processing systems using some kind of message broker. Message brokers such as MQTT, which we have studied before or similarly Apache Kafka could be used as message brokers in their system. And those could be used to transfer the data to different analytical pipelines.

In this work, they are talking about the use of single message queues and which may not be scalable and single queues; obviously, will increase the latency. Alternatively, if you are using single message queues, that increases the data volume, the number of sensors that are connected to them, the number of sensors could be reduced, and it is also possible that out of order data will come because the buffer space is very limited. So, all these problems could be dealt with appropriately.

So, a naive approach to deal with this kind of single message queue based system is to install more servers, but installing more servers is impractical, and it might also lead to not proper utilization of this existing server infrastructure.

(Refer Slide Time: 25:58)

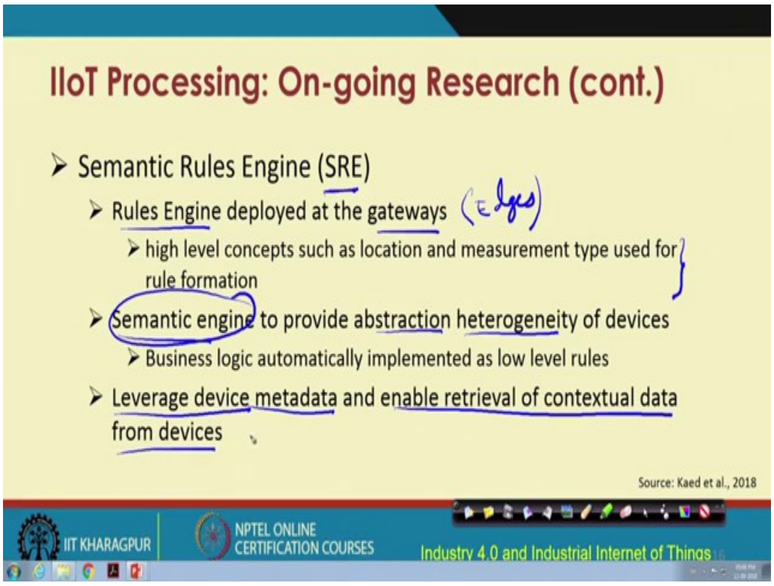


So, Dey et al. they came up with their architecture to deal with this particular problem. In their architecture they have the producer phase, they have the consumer phase, and the modelling phase. The producer phase is similar across different topologies and as we see over here. These are the different topologies that they have analyzed, they have the sensor topology, where we can see that individual sensors are connected or, mapped with individual separate topics, in this manner. And those data are basically aggregated at this junction.

So, this is the sensor topology, they have also analyzed the room topology, room topology basically, what it does is that the topic level this aggregation is performed. So, in a particular room with different sensors this data are going to come and topic wise this aggregation is going to perform in the room level. They also analyze the flow topology, in the flow topology as we can see over here different sensors are installed in the different floors of the building and so, all these sensor data are analyzed and aggregated topic wise, in the floor level.

So, these are the different types of sensor topology, room topology, and flow topology, that they have analyzed. So, in the producer phase the producer phase, basically, is similar across all these different topologies. The consumer phase extracts the topic data and converts them into different key value pairs. And the modelling phase basically does this workload modelling of the data, that are received.

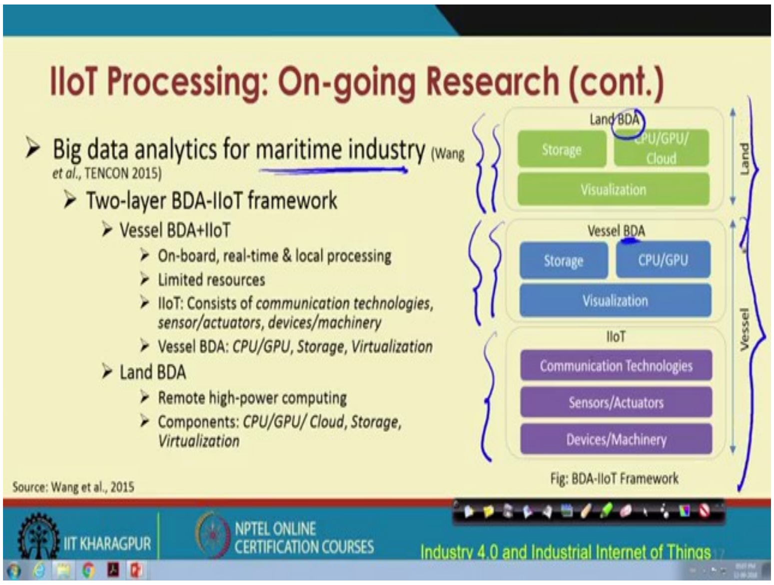
(Refer Slide Time: 27:57)



Kaed et al. talked about the use of semantic rule engines SREs, Semantic Rule Engines. These rule engines are deployed at these different gateways or the edges. So, these rule engines are going to do some kind of high level processing, with respect to the location and the measurement type, that is used for rule formation. So, this preliminary level processing is going to be done by the rule engine.

The semantic engine is going to provide abstraction heterogeneity of the different devices, business logics will automatically be implemented as low level rules, in this particular component of semantic engine. And this leveraging, the device metadata and enabling the retrieval of contextual data from these different devices, will be done by the semantic rule engine.

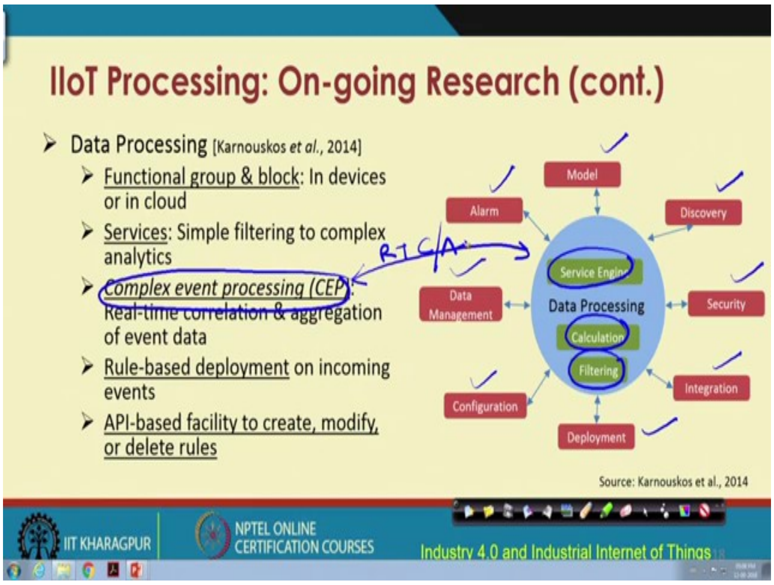
(Refer Slide Time: 29:05)



Wang et al. came up with the BDA-IIoT framework, where they are talking about the use of they are proposing a BDA; that means, the Big Data Analytics IIoT framework, for use in maritime industries.

So, in this they came up with this architecture, where they have these components the IIoT component the vessel BDA component; that means, a big data analytic component vessel; that means that the vessel level it is going to be done and the land BDA component, which does the analytics at the land level. So, part of the data will be processed at will be analyzed at the vessel and part of it will be sent to the land, and corresponding engines for vessel level processing, and land level processing are going to be done. So, this is this vessel component and this is this land component.

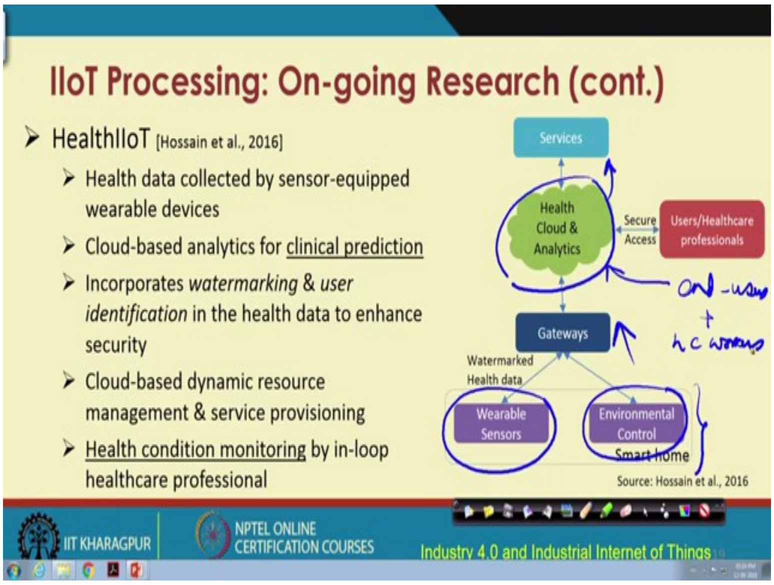
(Refer Slide Time: 30:08)



Karnouskos came up with another solution for data processing. So, for data processing they have different components, they have the service engine, a calculation engine, and filtering component. So, all of these things are going to be done at the data processing engine. Service computation, calculation of different types of computation etcetera and filtering all of these are different components of this data processing engine.

So, here basically this particular engine, they have considered to be interacting with alarms different models, discovery component, security components, integration components, deployment components, configuration components, and data management components. So, as we can see over here complex event processing will be performed by this particular service engine over here, which will take care of real time correlation and aggregation of the event data.

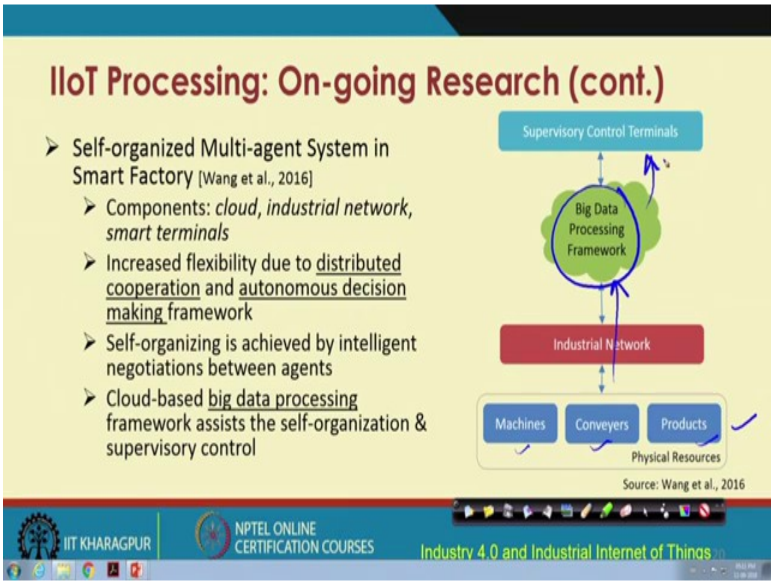
(Refer Slide Time: 31:29)



Hossain et al. came up with the health IIoT solution. And this is this solution they came up with, in their architecture they are talking about smart homes. They consider the smart homes, they consider the use of different wearable sensors, which could be owned by different humans, and also the use of different environment control sensors at homes.

So, this data from these environmental sensors and these variable sensors like different health monitoring sensors like spo2, temperature sensor, blood pressure sensor. These are all going to send the data to the cloud for analytics, but it is they are going to send through this particular gateway. And based on these analytics different services are going to be offered to the different users. And these users can get access users means it could be the end users like the people, the residents of the home, end users or, it could be the different health care workers.

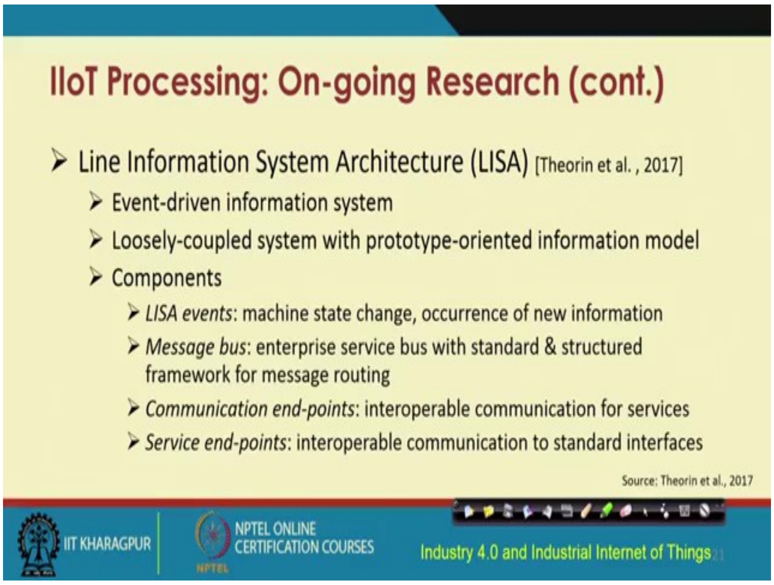
(Refer Slide Time: 32:53)



So, in another work by Wang et al. they came up with a multi agent system for smart factories, where they are talking about use of physical resources at the very bottom layer. These physical resources like machines conveyors and the products they are going to be sensor-equipped and are connected to one another. So, through the industrial network the data are going to be coming for centralized event, centralized processing.

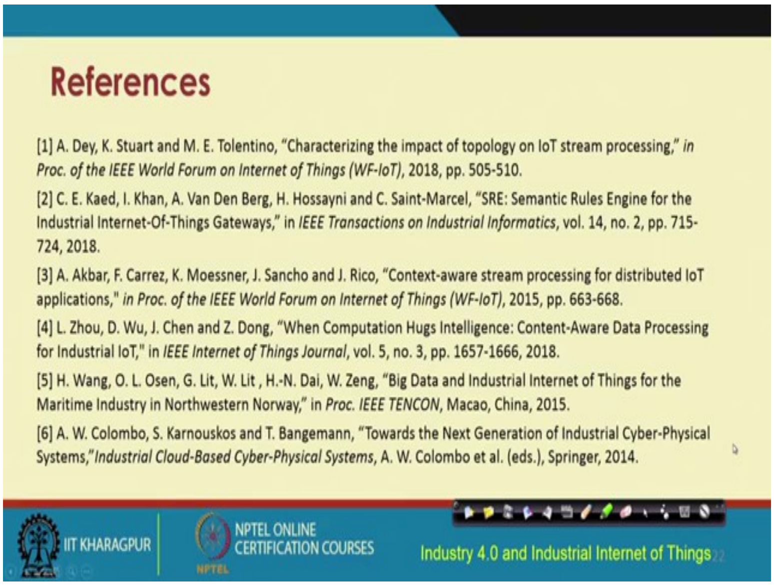
These data are typically of the nature of big data and this big data processing is going to be performed. And based on that supervisory control of the different terminals are going to be supervisory control, is going to be performed.

(Refer Slide Time: 33:50)



Another work a recent one is by Theorin et al. in 2017, where they are talking about the use of Line Information System Architecture LISA, which is an event-driven information system, which has different loosely coupled components, such as the event component, the message bus, the communication endpoint, and the service endpoint. So, this is this architecture that they have come up with these different components.

(Refer Slide Time: 34:25)



With this we come to an end of this part of the lecture. Here, what I have shown you is initially I talked about the importance of processing of the data. The characteristics of the data, that are received in these IIoT based systems. And then we worked through different solutions, that are talking about installations in different IIoT contexts installations of IIoT systems, in different contexts and their processing.

So, these different references some of which we have gone through so, far will help you to get an idea about how processing is important and how industry scale processing could be implemented, in practice. So, these are all these different references and with this we come to an end. In the next lecture, we are going to talk about a few more case studies, a few more different solutions that have been proposed in recent times focusing on IIoT processing of data.

Thank you.