E-Business.
Professor Mamata Jenamani.
Department of Industrial and Systems Engineering.
Indian Institute of Technology, Kharagpur.
Lecture-43.
Basic Analytics and Big Data.

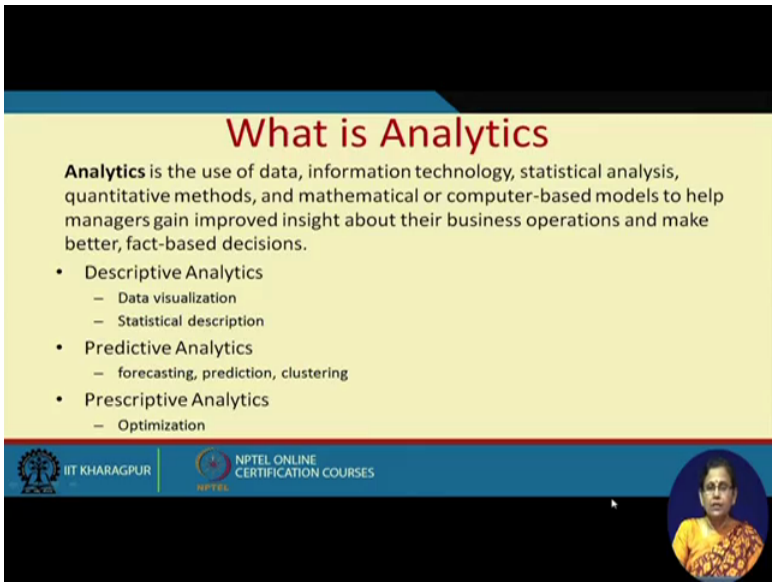
(Refer Slide Time: 0:31)

Welcome back, this is the last lecture of this section, in fact from the beginning I told you 1st few lectures, 1st few weeks, 1st 4 weeks in fact were devoted for knowing the functional e-business functional areas. Next 4 weeks we devoted, including this class we devoted for knowing the technology. Now over, we did not go to the details of this technology but knowing this overall technology is necessary for taking business decisions. And next 4

weeks, we are going to talk about various decision support systems which are resulted, mostly because of these ICT technologies.

So right now we are going to focus on the business analytics and big data. Here we are going to learn what is business analytics and role of big data. In fact, so far we have discussed about the technologies and various subsystems. Now, data is generated from every source, from the transactions, when you carry out transactions, starting from a transaction level, you collect the data. That data we know is used for management decision support, your management information system and decision support system uses data to take tactical level decisions.

(Refer Slide Time: 2:32)



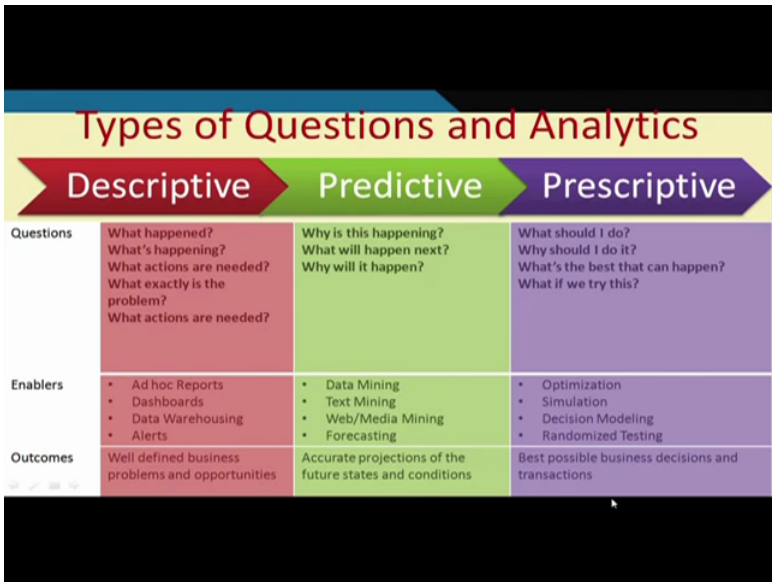
Then this data is further refined and used at the strategic level as well. And in last few lectures we saw not only these transition data, we also, we are also getting location data. Not only location data, you are also getting the data about the quality of the product. So this huge amount of data that we are getting about the transactions, about the products, about the locations, etc. what are we going to do with this data. Because now we have become informed with the available of the data, while discussing about the difference between data and information we know that data is something which just exists.

Unless and otherwise we have appropriate tools and techniques and models to use this data, this is basically useless. So here is the role of analytics. What is analytics, analytics is the use of data information technology, statistical analysis, quantitative methods and mathematical or computer-based models to build, models to help managers gain improved insights about their

business operations and make better, fact-based decisions. So analytics is about making fact-based decisions using the data the technology and responding mathematical models.

So broadly we can classify these analytics into 3 categories descriptive, which is about data visualisation and statistical descriptions, predictive analysis which like you are forecasting, predicting, descriptive analysis prescriptive analysis like telling you what is the right way of doing something is through optimisation.

(Refer Slide Time: 4:55)



This in fact in one of the earlier classes we have already told little bit, what are various types of questions that are answered by these analytics. Now coming to this disruptive analysis, the question such as what happened in the past, what is right now happening, what actions are needed, what exactly is the problem, what actions are needed and getting answers to this are your descriptive analytics. So here the tools are use of ad hoc reports, dashboards, data warehouses, genetic alerts are some of the enablers.

What is the outcome, outcome is a well-defined business problems and opportunities. Now analytics can be predictive, which answers questions such as why something is happening, what will happen next and why is it happening, why will it happen. The answers to these questions can be found by using data mining, text mining, web data mining and forecasting tools. Again, accurate projections of the future states and conditions which helps in forcing certain business situations and take some decisions are the outcome of this kind of analytics.

The prescriptive analysis, as the name indicates, prescribes something, so the question such as what should I do, why should I do it, what is the best that can be, that can happen, what if

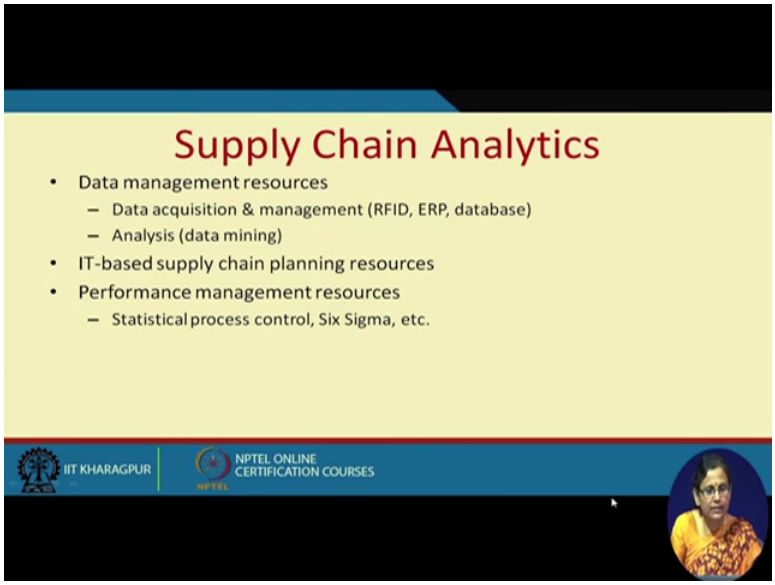
if we try this and so on are answered here. So after this, after knowing about these 3 descriptive 3 types of analytics that this descriptive, predictive and prescriptive, let us try to see one example.

(Refer Slide Time: 7:08)

Let us say this example is about some retail markdown decisions, in fact most department stores clear their seasonal inventory by reducing prices. Now the question is when to reduce the price and by how much. Through descriptive analysis we can examine, we can view descriptive analysis is basically viewing the data. To examine the historical data for similar products, prices, units sold, advertising that has happened, we can visually see it and think of taking some decisions. Then next step is we can use some kind of predictive analysis for predicting sales based on price and taking some decisions accordingly.

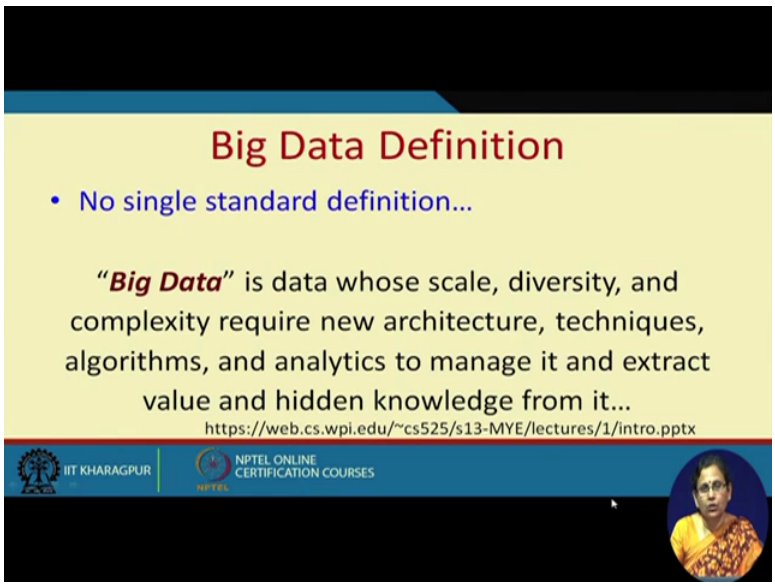
Next we can use some kind of prescriptive analysis to decide what is the best pricing and what is the advertising strategy that can maximise my sales revenue. So for the same situation, now we can use descriptive analytics, predictive analytics and prescriptive analytics.

(Refer Slide Time: 8:27)



Okay, now we, coming to our supply chain situations, if we would like to have such kind of analytics in the supply chain, we need some data management resources for acquisition and management of the data from your RFID, ERP and other databases. We can use some kind of analysis which includes data mining, other analysis of course are there. We need to, we need to have IT based supply chain planning resources, support of this and support of. And we can also help in performance management for statistical process control, for managing your 6 Sigma process and all.

(Refer Slide Time: 9:15)

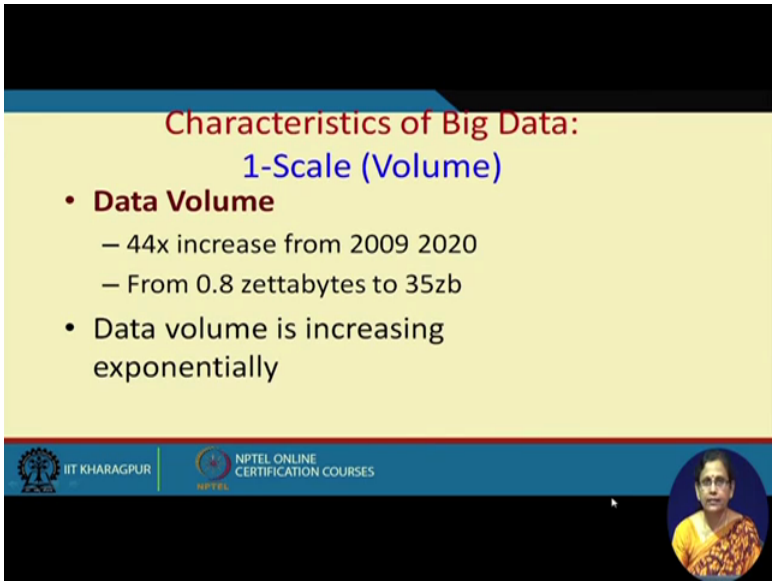


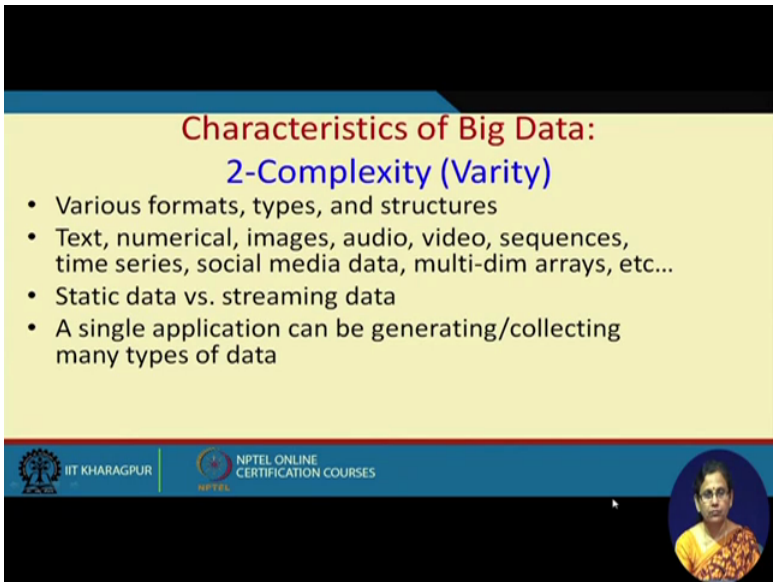
Now, as we have been talking about, this analytics is about making informed decisions, informed fact-based decisions. Now when it comes to fact-based decisions, the facts are

basically what, about the data. And as I have told, the data is no more limited to transaction data, volume of the transaction data is already increasing but the kind of source that we have discussed more recently, like that of sensors, IOT, your RFID and other IOT devices etc., GPS and so on are generating huge amount of data. Those data are different from your transaction data.

However there is no standard definition so far for big data, but this definition I got from some other source. In fact some of the facts I have got this from the source itself. This big data is a is data whose scale, diversity, complexity requires new architecture, techniques, algorithms and analytics to manage it and extract the value and hidden knowledge from it.

(Refer Slide Time: 10:39)



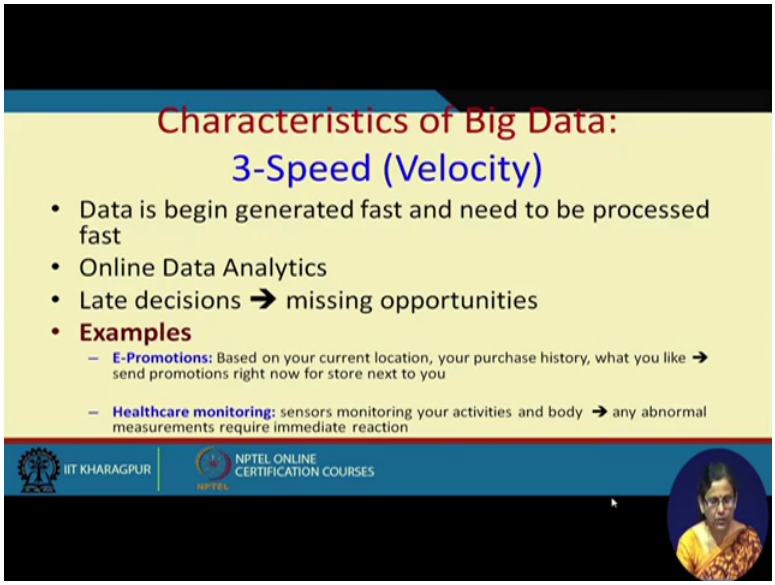


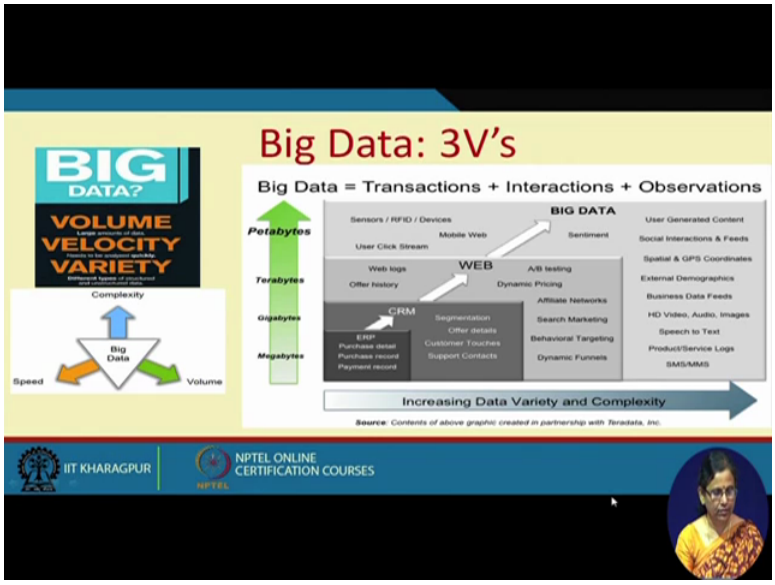
There are, as they say there are some 5 Vs associated with big data, in fact some of them countdown on 3 Vs, some of them are 4 Vs and 5th V is basically knowing about whether you have value associated with the data or not. But anyway, in technical terms people usually use 4 Vs and the way the Vs are written are also different because the literature is not yet standardised, but anyway the 1st important characteristics of this big data is this volume. In fact, there is almost 44 times the data increased between a very small period, that is from 0.8 Zettabytes to 35 Zettabytes, it is a huge change.

This volume, this data volume is, as this data generating sources keep generating the data, this volume is increasing exponentially. The 2nd characteristic of this big data is variety. It is comes in various formats, types and structures. For example, your RFID data, you GPS data, have other sensor data, data from the customers your, which comes from the social media, etc., they are in different formats, someone in text, numerics, image, audio, video, sequences, timeseries and so on.

Some of this data are static and some of them will be streaming, they will be coming continuously, their velocity is very high. Then though they are coming from various sources, finally data fusion has to happen to take certain decisions. So a single application in the scenario, big data scenario can be generating or collecting data from many many different sources.

(Refer Slide Time: 13:21)



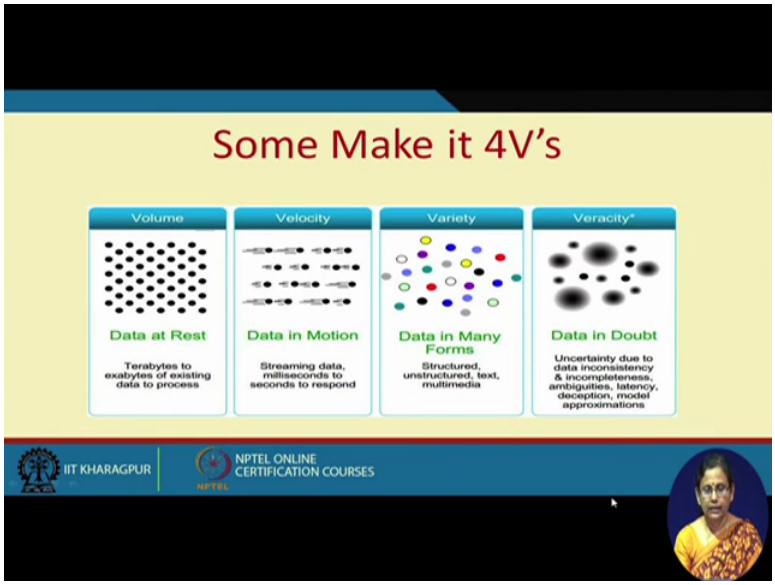


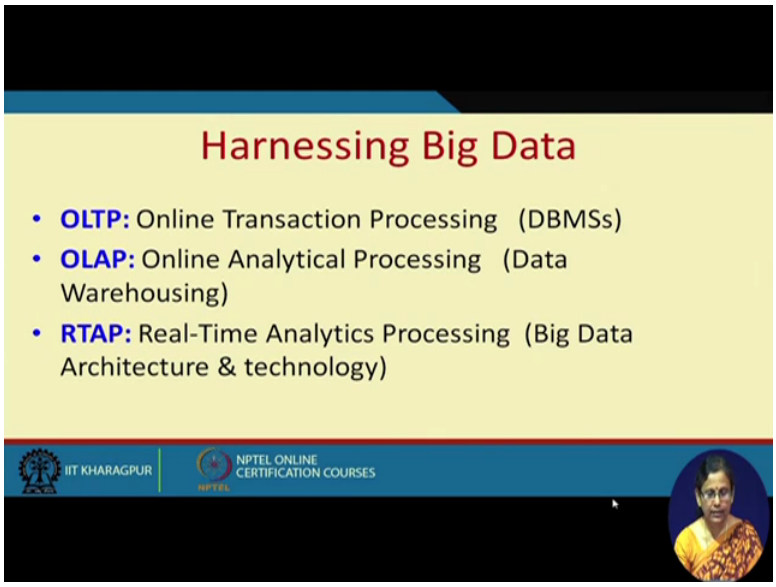
Then the 3rd characteristics is actually velocity, which is about how frequently the data gets generated. For example, think about your RFID data, GPS data, they come depending on your device's capability, they come in fact continuously, they come in, even you can adjust it to milliseconds or seconds or whatever, based on your choice. Your stock price data, it comes so frequently, varies, seconds to seconds variation you can see. So if you do not have proper online data analytics tools, as soon as this data comes, you should be able to take decisions, if you do that, in fact you lose many business opportunities.

Examples include your E Promotions: based on your current locations, your purchase history, what you like, send Promotions right now to store next to you. Healthcare monitoring: sensors monitoring your activities and body. Any abnormal measurement requires immediate reaction. These are the 3 Vs, volume, velocity and variety which are associated with how complex data, how speedily it is coming and what is the volume. If you look at this diagram, while from ERP you were getting the data in megabytes, from CRM you are getting the data in gigabytes, from where you are getting the data in terabytes and from your RFID sensors, you are getting the data in petabytes.

You can see other details from here, what are various applications of this data, what are various sources of this data and as we increase along this line, there is increase in decision-making, often very complex decision-making requires increased in not only in data volume, in data variety and complexity.

(Refer Slide Time: 15:54)



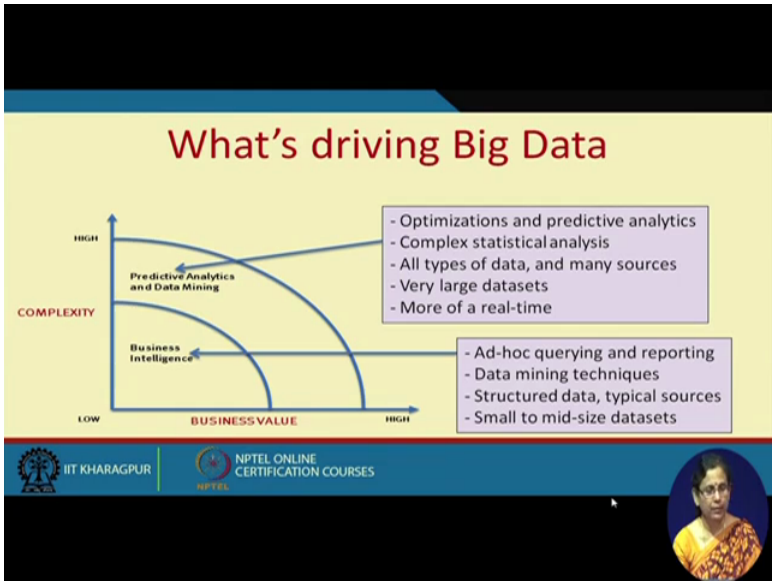


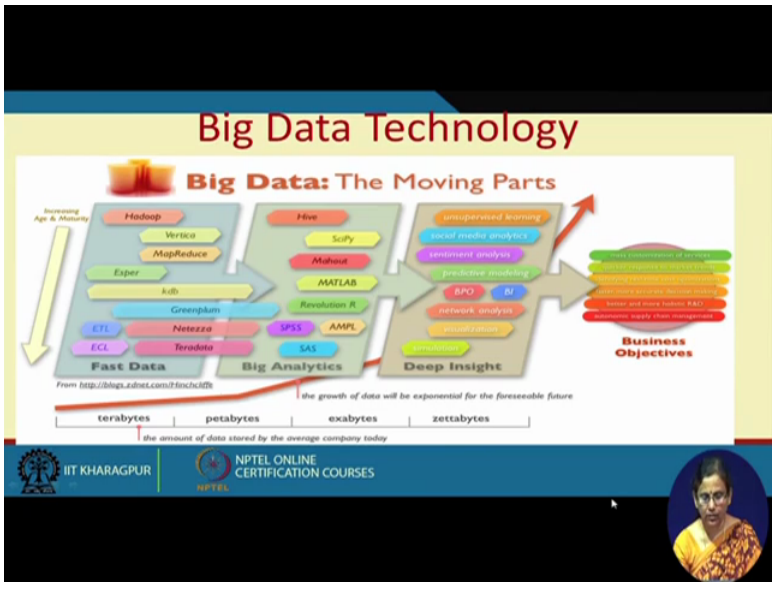
Then some even and a 4th V, that is called veracity which is about uncertainty and inconsistency associated with the data and approximately solving them. So you have these 4 dimensions, data at rest, which basically is terabytes to exabytes of existing data to process. Data in motion, this is the streaming data, this has milliseconds to seconds to respond. You have many forms in structure, unstructured text and multimedia data. You have data in doubt, which is uncertainty due to the data inconsistency, incompleteness, ambiguity, latency, deception, model and approximation and so on.

Now in order to harness the big data you have many technologies. Online transaction processing, online analytical processing, 1st one is related to database, in fact you remember the lectures that we did while managing out data resources, in fact many of these things we

have covered a little bit more detail. And about real-time analytics which has basically evolved because of this high velocity big data.

(Refer Slide Time: 17:31)

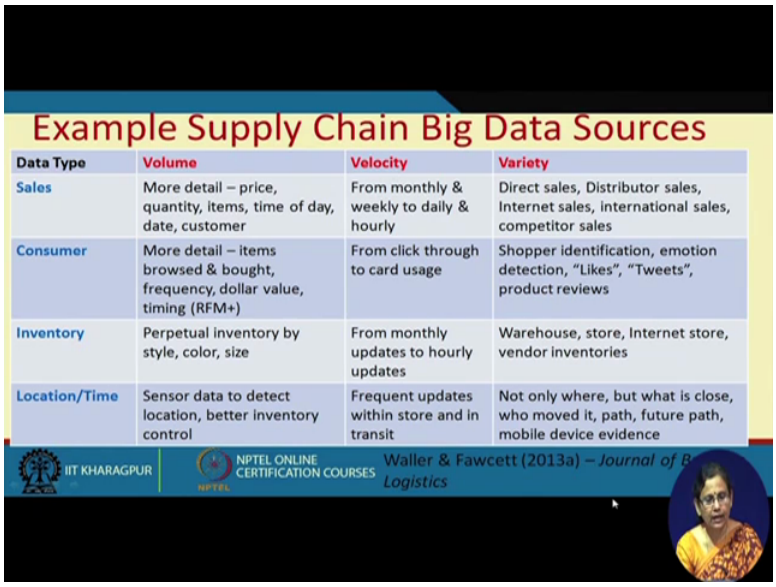




Now who all are generating, social media networks, scientific instruments, mobile devices, sensor technologies and network. Now what are the driving, what is driving this big data, is ad hoc querying and reporting, which is for business intelligence, data mining, use of data mining technique, which uses structured data, typical sources, small to medium-sized data sets, then relative analysis in data mining, which requires optimisation, predictive analytics, complex statistical analysis, data from various sources, large datasets and more of real-time data.

What are technologies? From very deep, fast data to large volume of the data, you have various technologies, you can just name this because knowing about these technologies simply does not, I mean naming this technology does not make sense. But it is basically what this diagram indicates is how what are the, what are the technologies for various types of data and what kind of insights you get and what kind of analysis you do. In fact some of these analysis and a few more data sources and little bit more on this analytics part we will be knowing while discussing about the decision supporting in e-business, specifically we will be talking about decision-making situations which are resulted because of because of new phenomena that is happening in the business world.

(Refer Slide Time: 19:51)



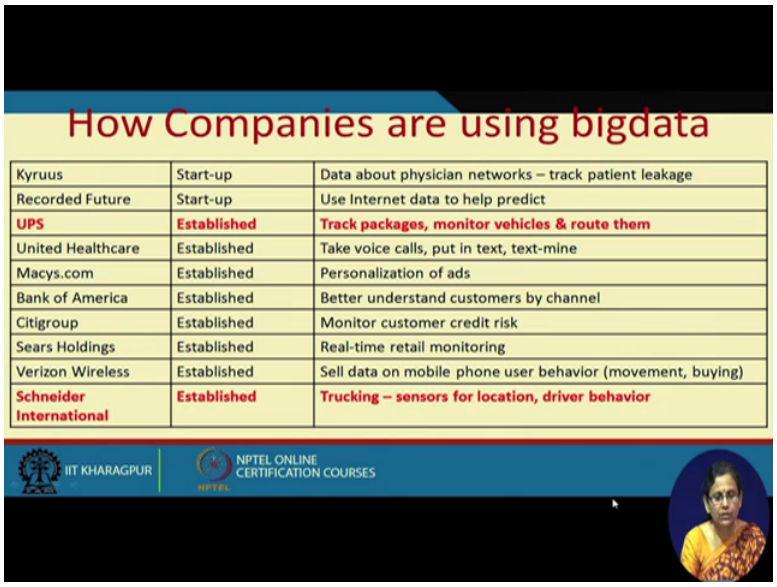
Now in the supply chain literature, big data is coming from many sources, while tracking the products from GPS and RFID, in manufacturing, in manufacturing so far in we have not have any decision on this. The data, the parts themselves are talking about, parts are now talking to each other and they are sending data to some centralised server, these are called cyber physical systems. In fact Industry 4.0 is all about it, this is about machine to machine communication, as a result huge amount of data is getting generated.

Next is prevailing sales and marketing while analysing weblogs, understanding customer reviews, again you encounter big data. This is a more detailed example of various, during various datatypes and based on this volume, velocity and variety. For example sales, you have more detailed price, quantity, item, time of the day a date, customer, where the sales is actually happening. This is velocities from monthly, weekly to daily and sometimes it is hourly. Variety, direct sales data, distribution, recruited field data, Internet sales data, International sales, competitors sales, data is coming from various sources, this is variety.

In case of consumer, you have more details about the items browsed, bought, frequency, dollar value and timing. Then velocity is coming from the click through to the card users, credit card users. Then you have variety, which includes software identification, emotion detection like his likes, tweets, etc., his correct reviews and so on. Inventory, perpetual inventory by style, color, size, this is can be, velocity can be found monthly update to hourly, monthly update to hourly update, data is coming from warehouses, food, Internet store, when the inventories and so on.

Location and time data, sensor data to detect locations, better inventory, control and all. Velocity is, frequent updates within the store, in transit and so on. It is not only where but what is chosen, who moved it, path, future path and so on.

(Refer Slide Time: 22:57)

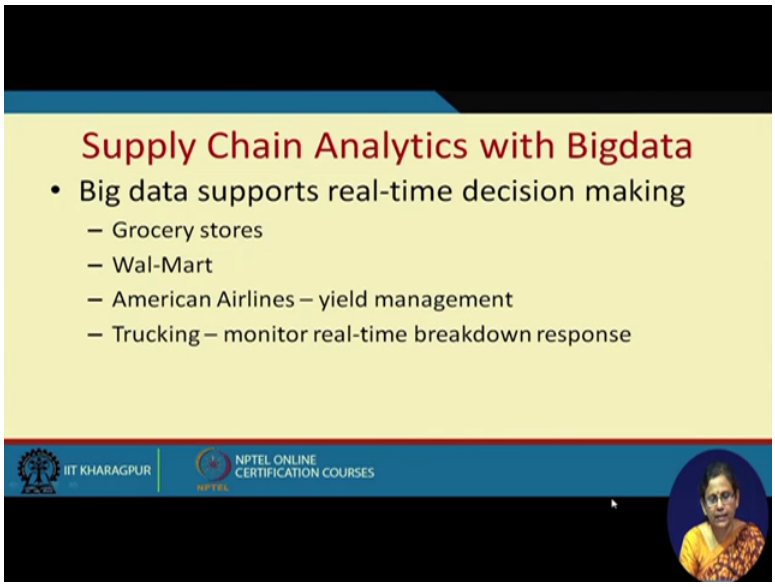


Then this big data provides various opportunities to improve the supply chain operations for demand forecasting by linking the real-time sensors to machine learning sensors to machine learning algorithms, barcode checkouts and in case of Walmart RFID chips already exists. Then this enables real-time responses about the demand. Warehouse design and location, you can design system for optimality, it is a classical operations research trouble of course but it can use network analysis to be more complete. Supplier evaluation selection, it can consider more factors and more up-to-date data.

Selection of transportation nodes, real-time truck and rail assignment, so such systems already exists and as more data comes in, but algorithms are warranted. Companies are using this big data for their advantage. Both, it includes both start-ups and established companies.

For example UPS, it tracks packages, avoidance vehicles and tours them. Then Schneider International for trucking, for trucking it gets the sensor locations, driver behaviour, etc. So these are some of the uses the companies are already making.

(Refer Slide Time: 24:39)



In grocery stores, in stores like that of, big stores like that of Walmart, by airline companies, by trucking companies, are all happening or getting revolutionised because of this big data. So with this we finished this lecture, from next class onwards we are going to learn about the decision support modules in e-business, thank you very much.