Company Participants

- Ashish Nadkarni, Program Director, Computing Platforms
- Mandeep Singh, Software Analyst, Bloomberg Intelligence

Presentation

Mandeep Singh {BIO 15014535 <GO>}

Good morning. Welcome to the BI Analyst Briefing with IDC on the Big Data market outlook. My name is Mandeep Singh and I'm the software analyst at Bloomberg Intelligence. I'm delighted to have Ashish Nadkarni, Program Director, Computing Platforms from IDC to talk with me about the Big Data market today.

Couple of housekeeping notes; today's presentation will be recorded and available for playback. At the bottom of the slide window you will notice that you can adjust volume and maximize your screen. You're free to ask a question by submitting one to the right of the slides, we will address questions at the conclusion of the presentation.

Little bit about BI, Bloomberg Intelligence provides in-depth analysis and data sets on industries, companies and government, ESG, credit, economic and litigation factors that impact business decisions. BI Research is backed by 300 plus third-party data providers, and is delivered with no buy or sell recommendations. Our analyst team averages 20 plus years of buy and sell side research experience. BI is exclusively provided to clients who subscribe to the Bloomberg Professional Service and can be accessed at BI GO.

Let me pass it on to Ashish to go with the presentation. Ashish?

Ashish Nadkarni {BIO 18264499 <GO>}

Thank you, Mandeep. Hello, everyone, and thank you for joining today's webcast on the Big Data market out. During the presentation, I will be talking about Big Data and Business Analytics slightly interchangeably, but I'll set the context as I get through it. But my name is Ashish Nadkarni, I'm a Program Director at IDC, I cover our Computing Platforms Research and I also co-lead our Big Data initiatives, which have evolved over the last four or five years as -- it has evolved.

And the main reason for Big Data to evolve is that as companies embrace digital transformation, business analytics and Big Data and Business Analytics are to some of -- are the pivotal pillars on which people rest their businesses, their business models and so the goal is to create greater operational efficiency, building deep relationships with their customers and then creating new revenue streams, which are based on technology enabled products and services, and the underpinning of all of

this is data and so many times in the past, a lot of the business decisions were made based on surveys or user sentiment. These days much of that is based on data and the data is either collected in a telemetric fashion through devices or it's collected through user sentiment, it's collected through websites, much of that data is diverse, it's desperate and bringing it all together in the form of collation and then analyzing it is what people do.

And we believe that by 2018, a third of the leaders in every industry will be disrupted by the third platform competitors and all of these third platform competitors whether it is Uber or whether it is Tesla or whoever you look at as a disruptor in their respective market segments would be using a data-centric approach, a Big Data based approach to solving their business problems.

So with that sort of just talking about where we see IT going and much of it fueled by other reasons, but one of the things we're seeing in a Big Data world, if you will, is this notion of a distributed IT model. Now whether you talk about Internet of Things or whether you talk about connected cars or whether you talk about autonomous devices, drones, what have you, we are in a world where our data center is no longer the compute unit, it's no longer the restraining factor where all of the analytics happens.

We are in a world where much of the analytics is a distributed entity. So there is analytics on the endpoint, for example, your mobile phone might be doing some level of preliminary analytics given the compute capabilities in the phone itself. There are edge point -- edge devices that are located in different places where there is good enough compute capacity to do some preliminary training. And then finally, you have the cloud where people can store a lot of their data in an amalgamous [ph] fashion, and in addition to their core, which is basically where their business applications reside.

And so we live in this world where IT needs to manage everything, so there is a lot of emphasis on data locality, security making sure Asset Management is all in control. Why am I saying all this? Because this is sort of also a side problem to Big Data. So when we look at the market outlook, we need to make sure that all of these issues are addressed looking at the solutions that are introduced by vendors.

So looking at Big Data and it's important to note that Big Data doesn't happen in a vacuum, it's important to know that analytics is really all about these six things that I have mentioned here. The first one is data persistence. What that means is, where the data lives is important. How it leaves is important. The manner in which it is stored is important. So whether it is talking about block storage, traditional block storage in the data center or it is talking about Amazon S3 or whether it is talking about Azure or it's talking about file storage. It is important to know that Big Data requires data persistence, because that's where the data lives. We are not all talking about streaming data, so there is data in transit, which is never captured -- never stored, it's always analyzed in transit. But then there is the data in storage, which is where data persistence comes in.

Second is data management, so being able to glue all this data together, whether it's in the form of a database or whether it's in the form of a data fabric bringing the data together in a sort of a (Technical Difficulty) for analytics is important.

Next-Gen apps is important, the third one is because it is all about making sure that the application touches the pulse of the user. If I am using a mobile application, that mobile application is to capture my sentiment in a way that is important to me. So that the next time I try to use the app, it gives me relevant information. Accelerated computing and this is sort of -- it's an umbrella term, for things like cognitive, artificial intelligence, machine learning, what that means is the use of right technologies to improve and accelerate outcomes, whether it is cognitive functions, whether it is machine learning, it is to accelerate outcomes using certain types of technologies like GP, GPUs.

The fifth is analytics, again (Technical Difficulty) Big Data, you know, there is the data storage piece, which I talked about earlier. But then there is also the analytics and the analytics here is it has to be meaningful, it has to be done in a timely fashion and it has to be done in a fashion that is relevant to the business. We are not talking about iterative science experiments here. We are talking about analytics that is directly impacting the business.

And then finally edge workloads, which is like what I'd mentioned earlier, analyzing the data where it lives, we don't have the luxury of centralizing all the data, because much of the data might be old (Technical Difficulty) time. So you need to be able to analyze the data, whether it is Twitter feeds, or whether it is social media interactions, whether it is clickstream data, that data has to be analyzed as and where it's generated, so that you don't lose out on important information that is contained in that data. So those are -- these are the six tenants of Next-Gen IT and Big Data sort of goes across all six of them.

We are looking at a taxonomy slide here and I won't bore you in the details, but note that we look at Big Data as a multi-stack model and then multi-stack model pretty much goes across our entire software spectrum and at IDC we have a fairly robust, one of the industry's best software taxonomy is out there. And as you can see from the slides we look at Big Data that goes -- pretty much spans both the vertical and horizontal spectrum of this -- of this taxonomical deck that you see.

And you'll notice that there is talk of Enterprise Performance Management, there is talk of production planning, there is talk of cognitive AI [ph], there is talk of continuous analytics, you get the idea, I mean Big Data is not a one size fits all, it's not a single entity, it is actually a collective or a conglomerate of multiples that are spread out across the entire software spectrum.

And how do we classify Big Data? So one of the things we need to note is that we need to define Big Data, there are a lot of industry terms out there, that talk about Big Data in different ways, but for us, really it is the volume variety and velocity. And volume means that the data is collected in excess of 100 terabytes [ph], it is

generated and growing more than 60% and/or it is received through high-speed streaming and then it is deployed on dynamically adaptable infrastructure. This is key because a lot of those static applications that are used in traditional business analytics use cases. Much of the infrastructure is pretty static and rigid. For Big Data, you cannot have that structure because the data is changing so rapidly, so you need agility, you need elasticity.

And so ultimately, those are the three key principles; volume, variety and velocity that provide you Big Data. And the one we that I did not mention in this structure is value, ultimately the value of the information that is generated has to be relevant to the business, if the business doesn't find value in that information, then the Big Data Analytics (Technical Difficulty) no meaning. And so this is important to know that it's not an iterative cycle to find things that you don't know, it is an iterative cycles to answer questions that you're asking.

This year talks about the Big Data workflow and so you can think of Big Data as a big sieve of through which the information or data -- raw data passes, gets processed and then ultimately gets applied to the base. Again it's an eye chart, so I'm not going to bore you with all the information in these icons. But note that it's data created, data acquired, information processed and then ultimately fed into the business process.

And where this becomes important is, again it's a loop, business has to know what they're getting to generate questions to ask more data to be generated, so it's feedback cycle and it's a continuous cycle that goes on in many businesses, in fact, their entire businesses runs on this. So again data creation, data acquisition, machine processing and then ultimately business process.

So that said, I would sort of jump right into the numbers. And so what we are seeing here is that in 2017 we would expect Big Data to cross 32 billion, now 32 billion includes software, hardware and services. And as you can see, a big chunk of the market comprises of infrastructure, which is hardware and the reason for that is because ultimately as you can see on the right, the pie chart shows that a lot of the Big Data for various reasons, governance, regulatory, compliance is stored, it stored for much longer than one would expect, then it stored only (Technical Difficulty) people don't want to believe this data only to know that we had some necessary information contained.

So people hang on to their data a lot more. Computer is important again, because people analyze the data and then as you can see networking and infrastructure -- other infrastructure form a smaller portion of the market. And then similarly, you will notice that the business analytics market is going to be a $153 billion market in 2017, again, growing much faster. But as you can see here, the compute side of business analytics forms a bigger portion of the infrastructure market.

Now here's where I need to explain why and how Big Data and business analytics are interchangeable or not. So business analytics is the Uber market, it's the bigger

market and we think of Big Data as a portion or a segment of the business analytics of the overall business analytics market. So what you're seeing here in the $153 billion market, a 100% of that or -- I'm sorry, 100% of the Big Data market is a sub-segment of this bigger -- of the business analytics market.

Now, going back to the previous slide, and this slide you'll notice that business analytics is not all about Big Data sets, business analytics can also be about small data sets. Many of the business analytics functions don't operate on terabyte scale data sets. They operate on small data sets. And so, you know, where the differentiation comes in is the three slides ago where I talked about how a Big Data workloads are classified.

So when the classified Big Data workloads as such, everything else becomes business analytics workloads. So it's important to note that Big Data is not anything foreign, it is a portion of the business analytics market. In 2017, we have made some taxonomy changes that make it a 100% segment of business analytics.

So how do we compare? So as I talked about previously, so you could look at the market in this fashion, you can look at it as Big Data and then rest of analytics. And the rest of analytics as you can see is growing, but it's growing at a much smaller rate than the Big Data market, which is growing at a lot faster rate. And the reason for that is because many of the data sets that people operate on today require a different form of processing, those data sets are becoming larger and larger.

So the luxury of dealing with small data sets is no longer there. Many of these data sets increase in size, just by sheer quantity of the collection mechanism. So for example, if you're dealing with mobile data, then just by the sheer number of smartphones out there, your mobile data is extremely high. Or to give you another example, when we talk about aircraft engines generating data, just a single flight given the amount of sensors that are present in different aspects of the aircraft or the engine itself. The total flight time generates a whole lot of data that needs to be analyzed. So the luxury of dealing with smaller data sets and the luxury of dealing with static data sets is no longer there, it's becoming a niche more and more.

So, sort of wanted to end then [ph] with some key market drivers and inhibitors. So what are the key market drivers? And I'm not going to go through each of these. But the key things here are, there is an increased belief in data driven decision making and so much of the sea-level executives are convinced that their decisions cannot be (Technical Difficulty) nature, they have to be driven by data.

There used to be a portion of executives [ph] who were driven by their own gut or they were driven by their own market sentiment, but these days data driven emphasis is so much that it's pretty big risk, do not base your decisions on data.

The second is availability of data, there is a lot of democratization -- democratization of data and what that means is many companies can get access to social media data, they can get access to anonymized machine generated data, anonymized mobile

data, anonymized financial data and they can use that to better inform themselves. So the availability of data makes it easier for your decision making to be superior versus having (Technical Difficulty) make decisions based on the data you have at your disposal.

And then the third one is the demand for self-service. We are in a era where people don't want to wait, people want to be able to service their own requests directly and so there is a lot more emphasis on automation behind the scenes, in data integration and automation and that results in emphasis based on automated analytics or cognitive analytics or AI or machine learning. So lot of these companies are focused on a self-service for a Big Data and that results in a lot more investment in the software's segment or even in the services segment.

What are some of the inhibitors? So the main two inhibitors that I wanted to talk about to you are macroeconomic factor, so clearly we are in a place where much of the buyer sentiment is on the economic situation of the country, or the region or the industry. We are seeing an ebb and flow there, so I'm not -- you guys are the experts, so I'm not going to talk more about that. But know that market sentiment is directly impacted by macro-economic factors.

But the second one is where I can talk at length, and what that is, is the impact of source. We are in a world, thanks to Linux and thanks to Apache and all of these open-source foundations. We are in a world where software is becoming increasingly a -- an open-source entity where people do not want to spend on software, if they can get an open-source based variant available. Developers are there becoming a lot more powerful in a company, because these developers can get access to snippets of code directly from the open-source world and then they can construct their own stack without any commercial -- need to buy any commercial off-the-shelf software.

So as you can see in our slides for the numbers, the hardware segments growing because, obviously, that's where they want to run their analytics. The services segments are growing because people obviously don't want to give up on that turnkey element. But the software revenue segments are not growing as much because a lot of the revenue is now getting lost to open-source alternatives if you will. And there is no dearth of open-source alternatives in every segment of the software spectrum, whether it is for cognitive, whether it is for machine learning, whether it is for any kind of analytics like Hadoop. You will see that there is no dearth of open-source alternative.

So if you are, in fact, the company with a culture of doing it all in-house, then your developers don't need to go by software, they can just use open-source.

And finally, I wanted to end a note with some main market trends; first is that public cloud providers like Amazon and Google and Microsoft are continuing to aggressively invest in Big Data and Analytics, they have sort of building their value

proposition beyond just the infrastructure service and therefore they are going after the Enterprise segment.

It is -- Big Data is and will be an ecosystem play, there won't be a single company that can dominate it. The ecosystem will be, you know, so to give you an example, Hadoop and Spark are some of the ecosystems. There is a sort of a growing ecosystem in the IoT space. There is an ecosystem growing in the cognitive space. So ecosystems is what will dominate the Big Data space only because it allows people to take a modular approach to -- approach to building their stack.

Infrastructure vendors, I think we see them largely giving up on their own software efforts and going with partnerships. So it's going to be partnerships with Hadoop distributors, Spark distributors and really it is all about making sure that they focus on enabling the software versus developing their own software.

We see cognitive artificial intelligence and machine learnings to be initially as offshoots, and if you look at our taxonomy, they are clearly offshoots of the bigger analytics market, but they eventually will become standalone markets as they grow up and become more mature.

And then lastly, money will follow open-source and so Hadoop is the best example I can give you, Linux is a good example, you know, companies like Red Hat have millions [ph] in commercializing Linux and SUSE is following it, Canonical is following it. So money will flow into the open-source ecosystem, which will in turn then flow into the open-source community in some shape or form and then give rise to vendors that are building their models based on or building their business models based on open-source --commercializing open source.

And so with that, I wanted to end the note and open up the call for Q&A. So, back to you, Mandeep.

Questions And Answers

A - Mandeep Singh {BIO 15014535 <GO>}

Yeah, thank you, Ashish. That was very, very comprehensive and I would encourage everybody on the call to send questions, if they have any. In the meantime, I have a few questions for you, Ashish. And I've been following the Big Data market for a very long time, and the focus has shifted from structured, unstructured data to predictive analytics, recommendation engines. It's kind of evolving, you get the sense that the ROI isn't obvious when it comes to Big Data projects and sometimes it can take a very long investment cycle to really say, okay, this is what I achieve using my Big Data system.

So would you say that standalone Big Data is that kind of the future or does it have to overlay AI and cloud to really drive that ROI?

A - Ashish Nadkarni {BIO 18264499 <GO>}

So I think it's the latter, and that's a great question, because this is basically one of the trends that Big Data has been seeing, because of how it -- of the technology going from a nascent, very primitive technology to being a more mature market. And so what people are realizing is you cannot create [ph] Big Data as a science experiment, you cannot create Big Data as a iterative model where people are just sitting in a lab somewhere and just generating data and data and just analyzing and analyzing it.

It has to be tied to a business outcome and that business outcome as people evolve into things like Internet of Things or being able to deliver products that are entirely based on market data collection, they are overlaying the cognitive and machine learning, in fact I was at a conference last week where IBM has now attached Watson to Internet of things. So they have this Watson IoT as a platform that they have developed in Munich, where it's a services play where companies are encouraged to collocate their developers with IBM, so that they can tie their analytics to the Internet of Things. And so, the whole thing then become a focus on how do we gather data from our market devices and how do we analyze it and how do we make it.

So then ROI then becomes almost a mandate people have to say, all right, if I'm investing x number of dollars in enabling -- Internet enabling my washing machines as an example. Then I need to be able to --what is per dollar -- per unit cost and what is my ROI on it. Am I reducing the number of customer support calls, am I reducing the number of failure components so on and so forth, so ROI then becomes the overall objective.

A - Mandeep Singh {BIO 15014535 <GO>}

So any IoT system or let's say a driverless car use case or pick anything in healthcare are these going to be Big Data use cases or -- and what else do you need along with the standard Big Data technologies like Hadoop or as Spark based environment. What else do you need to make this project successful like, how should companies think about, kind of implementing their own Big Data Systems?

A - Ashish Nadkarni {BIO 18264499 <GO>}

Sure. So sort of, let me start by making a sort of a more general statement, I think the term Big Data is slowly melting or blending into the whole business analytics thing. And we have sort of started to use the term Big Data Analytics as the overall market. The reason why we do that is because if you unpack Big Data, there is a data storage component, which the Big Data part of it and then there is the analytics, part of it which is slowly getting attached to an industry or a use case.

So you mentioned autonomous vehicles, you mentioned healthcare, every so the data storage part, which is you talked about the structured and unstructured piece in the previous question, the data storage part has not been moving so much meaning that the format of the data has not been changing so much.

What has been changing is the analytics piece. And the analytics piece is slowly becoming more specialized, more industry centric, more use case driven, more if then, there, that kind of driven. And that's where things like cognitive and artificial intelligence and (Technical Difficulty) learning, which all three are different by the way, get -- become a lot more prevalent. And so, if you call it a Big Data and Analytics as the general term then know that the Big Data piece is that static or I should say the data storage piece is the less changing of the two -- when the analytics piece is the more dynamic of the two and that's where you get the industry use cases and you get the models that data (Technical Difficulty) for example, very specialized based on the outcome that kind of stuff.

A - Mandeep Singh {BIO 15014535 <GO>}

That is a really important distinction, I think and it really helps us understand in a much better way what part of Big Data market is actually moving faster versus what you -- you said the data storage part, which is more static.

So on that note, which were the companies which you think are applying that analytics piece in a very use case oriented way and are able to establish a clear-cutting edge or a use case, are there any vendors that come to your mind?

A - Ashish Nadkarni {BIO 18264499 <GO>}

So, let me kind of take it industry or segment by segment because it will be easier to talk about it that way. Apart from the basics side, right? So there are companies like NVIDIA and Intel, who are focusing a lot more on the hardware aspect of accelerated computing. So there's a lot more focus on GP, GPU's, FPGAs things that enhance the capabilities of what we would call an industry standard.

So when we started doing research on Big Data, it was all about how can we beat up the CPU in a server to analyze as much as data we can. But now there is an increasing emphasis on, hey, the CPU can't do everything, the memory in the server can't do everything, we need to be able to offload that onto (Technical Difficulty) GP, GPUs or FPGAs. And so, these companies like NVIDIA, AMD, Intel, they are all focusing on this accelerated computing.

There is the whole, all flash array market, which is alive and well where there is a lot more focus on how can we analyze large quantities of data without beating upon spinning desk.

There is the whole memory segment, which is companies like SanDisk or Western Digital now. So there is all these component vendors who are coming up with superior silicon and superior memory to hold large quantities of data. In fact, I read a comment the other day that, Intel was thinking that, in 2017, the memory market itself is going to grow and that's because a lot of these newer Big Data technologies use recomputing for, SAP HANA being one of them. So that's kind of you can see as the infrastructure market has reached the plateau of speeds and feeds, they are now branching off into different types of compute models, if you will, or compute and storage models.

And then now let's shift into the software (Technical Difficulty) now, I'm not a software expert, but from what I can see, there are companies that are becoming more and more specialized on industry specific use cases. So to give you an example, the whole Watson, IBM Watson IoT, which was all driven by applying a cognitive layer on to an Internet of Things layer.

So IBM is an example, there is a Company called Cognitive Scale, which is based on Watson that does a lot of healthcare specific stuff. Then there is the, there are companies like Splunk who are more into the operational analytics, then of course there is a lot of emphasis on Spark investments. So IBM committed like a couple of billion dollar (Technical Difficulty) Spark. So there is a lot of companies who are trying to make Spark a standard overlay on top of the Hadoop model, Cloudera and Hortonworks and MapR have their own variances, if you will of Spark.

Traditional companies like Oracle and Teradata are now getting into the market. And underlying themes that there is a lot and lot of start-ups. So while the hardware, the infrastructure space might have become a lot more cooled off a little bit, I would say, minus the comments I made earlier about these GPUs and such, the software space seems to be very much alive and well. And there's a lot of activity there on building industry specific use case.

One thing that I did not mention and this sort of straddles both software and hardware is blockchain. Now you will notice IDC focusing a lot on blockchain this year. I think blockchain in the context of Big Data and Analytics is going to be huge, primarily because it drives a lot of security. And it is a lot more focused on secured transactions and where does blockchain become important is, blockchain is a very compute-intensive ledger system.

So there is a lot more analytics behind (Technical Difficulty) that people are working on to make sure blockchain transactions can be optimized and such. And then of course, but let's not forget services, all of these companies that were in the BPO market earlier, Tech Mahindra is an example, are becoming more and more focused on analytics as a service, Big Data as a service, public cloud service providers are getting into it only because a lot more -- many companies don't know how to get there. So they are making it easier for others to get there.

A - Mandeep Singh {BIO 15014535 <GO>}

Great. We have a question from the audience, Steve, is asking, are the Big Data open source software as powerful as commercially available software?

A - Ashish Nadkarni {BIO 18264499 <GO>}

So that depends on how you see the whole stack, so what -- let's take Vanilla Hadoop as an example, a few years ago, when Hadoop became known to the enterprise a lot of the companies shows to stick with commercial plugins if you will. And so, when we did a survey three years ago, the amount of -- the number of companies going with Vanilla Hadoop, the whole Hadoop distribution were in a limited minority, because Vanilla Hadoop had a lot of problems, it had a lot of single points of failure, it wasn't stable.

Fast forward three years later, many of those things [ph] have been ironed out, many of the issues have been sorted out, Vanilla Hadoop is a lot more stable, it's a lot more robust. And so, when something starts off in the open source, it is never as good as the commercial variant or the commercial (Technical Difficulty) but developers, many of whom work at these companies and are contributing code both into the commercial space, as well as into the open source space are very much in tune with the challenges that they have to address.

And over a period of time, those issues do get sorted out. Additionally, so take Spark as an example. Now if Spark were (Technical Difficulty) in an Apache Foundation as a open source project, it would have probably seen a slower trajectory in terms of the commercialization or the enterprise readiness, if you will. But there are companies like IBM and Cloudera and others who are making Spark a priority. And so, today, this year, we have a Spark Conference. And so, a lot of these commercial vendors are going to put their might behind Spark. That accelerates the adoption, that accelerates the development effort.

And so, overall you will start to see a lot more open source project become a lot -- give commercial variance a run for their money in a shorter amount of time.

A - Mandeep Singh {BIO 15014535 <GO>}

And I want to piggyback on that, so basically you have a lot of the legacy data warehousing companies, which have a huge installed base and they are kind of dipping their toes in the new Big Data solutions that are based on Hadoop and Spark what you mentioned. Do -- which is the way the industry is headed? Are they going to change their data warehousing solutions to this new kind of evolving technology in the Big Data realm or they're going to continue with what they have. How do you see that data warehousing market evolving?

A - Ashish Nadkarni {BIO 18264499 <GO>}

So, I think in the short to medium term and perhaps even in the long run, the traditional data warehousing market is probably going to -- okay and the only reason for that is because it directly is fed from traditional business analytics and business intelligence applications. So you will see that all of these enterprises, whether it's retail or whether it is healthcare or whoever else has got these traditional business analytics applications are going to continue to keep their proprietary enterprise data warehouses, keep investing in them because of the one to one correlation there.

However, what we see is they are going to start to supplement those EDWs with these next-gen data warehouses that are going to be based on Spark or based on Hadoop and based on, that's where their next-gen application development data is going to be fed.

Clearly companies like Teradata and Oracle have acknowledged that this is going to happen, and therefore they are trying to offer -- so if you are an EDW supplier then when it comes to investing in the next-gen data warehouse, they want to maintain that installed base. So they will -- they would rather the vendor -- the buyer come to them versus go to something else.

So they are offering connectors that allow you to offload (Technical Difficulty) the data from EDW into these next-gen data warehouses as it gets. So the initial play is going to be EDW for traditional workloads, next-gen data warehouse for next-gen workloads. And then the next-gen data warehouse for today is also going to become a tier 2 stack for archival purposes, because as everybody knows enterprise data warehouses are expensive [ph]. So if I am -- I'm not going to spend a whole lot of money in doubling my EDW capacity, I'd rather spend my money in doubling my next-gen data warehouse capacity.

So in the short to medium run, I think the installed base will probably be, okay. But they also have realized that if they don't invest in next-gen data warehouses, their honey journey is going to come towards a slow crawl, and then eventually they will be in a trouble.

A - Mandeep Singh {BIO 15014535 <GO>}

We have another audience question, he's asking, how is Big Data adoption by industries, which industries are lagging behind?

A - Ashish Nadkarni {BIO 18264499 <GO>}

You know, I think the industries that are lagging behind -- so let's, let me start by saying that, we see Big Data and Analytics pretty much go span across all industries, most industries have acknowledged that they cannot deal with traditional go-to-market mechanisms. Traditional go-to-market mechanisms, even if you take the telco industry, which has been really -- the traditional telco industry has been very much disrupted by mobile and Internet of things and all of these new, they have been upset by these newer mobile operators like Google and Facebook entering the fray as well.

So even a traditional industry, a closed industry (Technical Difficulty) has started to accelerate their Big Data initiatives. So it would be unfair for me to say, which industry is lagging, I mean, clearly there might -- the most heavily regulated industries are probably the ones who might have a pause for investing a lot more in Big Data vis-a-vis their current investment, you know, enterprise data warehouse or enterprise business analytics applications. But most of them have acknowledged that they do need to dip their toes into Big Data and start to jump on that journey, if not they are going to be eradicated in a short amount of time.

A - Mandeep Singh {BIO 15014535 <GO>}

Okay, fair enough. And do you think cloud helps facilitate that migration just in terms of from and legacy data warehouse to a next-gen warehouse, does cloud have a role to play, does it expedite that transition? And I mean there are certain companies that are still using mainframe, so obviously they haven't really moved to a Big Data environment, they aren't even close. So can cloud help these companies get there faster?

A - Ashish Nadkarni {BIO 18264499 <GO>}

So cloud definitely helps, I think helps in two ways; one is, for companies who have a very traditionally run IT, which is very rigid, very time consuming, have an alternative to go quickly jump on the cloud and start to test their waters with Big Data. The other way it helps is, as companies like Amazon invest in turnkey Big Data solutions, they can -- the companies can start to avail of those services without having to build their own stack from the ground up.

The third way the cloud can and does help is that, it offers a very quick, or if you will, to trying out new things. So what they might have otherwise been very hesitant to test, can test it now quickly without having to touch their bread and butter infrastructure. So they can do it without any risk to that. Now you mentioned something about mainframe and what we are seeing is because people are so hesitant to mainframe, IBM is in fact bringing Big Data to the mainframe. So I don't know if you were tracking IBM, but last week they announced machine learning for z Systems. And from what I'm told that machine learning is for z/OS customers, it's not for Linux customers, but it's for z/OS customers.

And so, clearly what IBM is recognizing is, if you are a mainframe (Technical Difficulty) and if you don't get the Big Data experience directly on the mainframe, then it's going to, in effect alienate that installed base. And so, they are smart in bringing Big Data to the mainframe.

A - Mandeep Singh {BIO 15014535 <GO>}

Okay, that's very interesting. I didn't know there could be Big Data overlaid on mainframe system. But I guess, IBM is making it possible, so clearly an interesting trend over there. I wanted to talk about AI a little bit more. So obviously, every company that we know often in AI has taken somewhat of their own approach when it comes to developing AI systems. Now do you think having an edge in Big Data gives you kind of more capabilities when it comes to AI or AI is more about algorithms and how you go about really crunching your data to applying machine learning and Big Data or not, doesn't really make a difference.

A - Ashish Nadkarni {BIO 18264499 <GO>}

So, let me sort of start by untangling the terms that we use, often we use interchangeably. So, (Technical Difficulty) way of looking at machine learning, artificial intelligence and cognitive is by comparing it one on one with the three different aspects of our brain. So in our brain, in the human brain we have a amphibian brain, which is sort of very trigger driven. So for example if you touch a hot stove, you take your finger away from it, because you know that it's hot and it hurts you, so you are very quick to react or someone pokes, you want to quickly react and run away or what have you. There is that amphibian part of the brain, which you could almost compare it to machine learning.

Then there is the mammalian part of the brain, which has the scale and the ability to think in some basic patterns. That's where artificial intelligence comes in, that could be compared as artificial intelligence. And then (Technical Difficulty) you could almost compare to the human, the human part of the brain, which is what makes us different from our amphibian, I am sorry, a mammalian counterparts in the kingdom.

So, so you could almost think of these three things as staggered and a stacked approach to adding more and more intelligence to your application. And so, when you look at the Big Data and Analytics, the information, the Big Data part of it could probably be the same. So if you're providing that Big Data information to the machine learning part or to the artificial intelligence part or to the cognitive part, the information itself doesn't change much, but the way you operate on it, the analytics part changes a lot.

So if the artificial intelligence, or I'm sorry, the machine learning part operates on the Big Data, it could be a very binary kind of yes or no kind of a algorithm that is processed, then you put in artificial intelligence, to the same data then you're now doing a lot more intelligent stuff autonomous cars, drones, the Alexa, the Echo, Siri all of these could be (Technical Difficulty) as artificial intelligence, where there is a lot of pattern matching going on and a lot more processing going on.

And then finally, when you apply cognitive to the same data, you are now taking that data to a very different level, the intelligence now is at a very different level. I don't think we are there yet, I mean, a lot of times people talk about cognitive, but really we are straddling between machine learning and artificial intelligence. Cognitive would be very much reading the data as a human and then being able to make advanced decisions on what to do next. I don't think we're there yet.

A - Mandeep Singh {BIO 15014535 <GO>}

Okay. Wonderful. That was very enlightening and I'll end with audience question. There is one more question around studies done on the ROI for the use case of Big Data in traditional organizations or is there a lot of hype?

A - Ashish Nadkarni {BIO 18264499 <GO>}

So would you be able to repeat that question. I am sorry.

A - Mandeep Singh {BIO 15014535 <GO>}

Are there any studies done on the ROI for the use of Big Data in traditional organizations or is still a lot of hype?

A - Ashish Nadkarni {BIO 18264499 <GO>}

So that's a very good question. I believe the answer is, there is a lot of hype. Studies are few and far between because ROI is such a interesting concept when it comes to applying it in industries and in organizations, because it is so very custom. So the ROI of even the same industry -- different companies in the same industry might be very different. It's a lot more custom. It's a lot more specific to how the company tries to mobilize its investments and try to get return on that.

So, my answer is, I think there are, the studies are emerging, there are a lot more commissioned studies out there, which have a bias towards the installed base or the bias towards the vendor. There are lot few uncommissioned and neutral agnostic

studies that would truly try to shine light on the challenges of gathering the ROI or realizing the ROI in term -- in large investments.

A - Mandeep Singh {BIO 15014535 <GO>}

Okay. Thank you so much. Ashish, for your time. It was great to have, in fact both the hardware and a software perspective, when it comes to the Big Data market. And I think some of the trends that you told us around blockchain and just kind of delineating the storage part versus the analytics part and some of the other use cases that you talked about, I think our audience would find it really useful.

Just a housekeeping note, the slides and the transcript will be available on the terminal as well as a replay of the webcast. I would also encourage everybody to take a look at the theme that we have published on Big Data on our infrastructure software dashboard, that is more focused around vendors. And I think everyone would find it interesting. Again, thank you so much, Ashish, I really enjoyed the webinar and I look forward to future webinars. So that's all, thank you.

A - Ashish Nadkarni {BIO 18264499 <GO>}

My pleasure. Thank you.