



BBBT Podcast Transcript



About the BBBT

The Boulder Business Intelligence Brain Trust, or BBBT, was founded in 2006 by Claudia Imhoff. Its mission is to leverage business intelligence for industry vendors, for its members, who are independent analysts and experts, and for its subscribers, who are practitioners. To accomplish this mission, the BBBT provides a variety of services, centered around vendor presentations.

For more, see: www.bbbt.us.

Vendor:	IBM
Date recorded:	June 10, 2013
Host:	Claudia Imhoff , President, BBBT
Guest(s):	Reed Meseck , Senior Competitive Executive Tracey Mustacchio , Director of Product Marketing, Big Data
Run time:	00:21:13
Audio link:	IBM Podcast
Transcript:	[See next page]
Transcribed By:	CastingWords



Claudia Imhoff: Hello and welcome to this edition of the Boulder BI Brain Trust, or the BBT. We're a gathering of international consultants, analysts, and experts in business intelligence, who meet with interesting and innovative BI companies here in beautiful Boulder, Colorado. We not only get briefed on the latest news and releases, but we share our ideas with the vendor on where the BI industry is going, and help them with their technological directions and marketing messages. I'm Claudia Imhoff and the BBT podcasts are produced by my company, Intelligent Solutions.

I'm pleased to introduce my guests today. They are Reed Meseck and Tracey Mustacchio. Reed is the Senior Competitive Executive, and Tracey is the Director of Product Marketing for Big Data. Welcome to you both.

Tracey Mustacchio: Thank you.

Reed Messick: Thank you.

CI: All right, Reed, let's start with you. There have been three really big announcements from IBM regarding Big Data. Let's cover those. What are they?

RM: Well, we announced DB2 With BLU Acceleration For Speed Of Thought Analytics. We also announced our enhancements to our Big Data platform to advance consume building in the performance of Big Data to help people basically get started more quickly. We also, along that same vein, introduced our PureData System for Hadoop, an appliance for Big Data and Hadoop solutions that really helps people explore and analyze more data with the simplicity of an appliance.

CI: One of the things you said in our briefing, and I thought it was interesting, you said that Big Data is all data. A lot of people have mistaken the Big Data push or trend to be just social media. What do you mean by Big Data is all data?

RM: Well, if you look at the data landscape, there has been this desire by some people to link Big Data to, exclusively, things like social data or newer forms of data, but Big Data is really about all data. It's about the fact that, even under your own feet, things like transactions have become



richer. If you're a bank, for example, it used to be that the only things you had were check numbers and amounts at the end of the month. Now, there's all sorts of rich information in those transactions.

Because we increasingly live in an electronic world, when we interact with that world through purchasing or just moving from point A to point B, there's a whole lot of information that's generated there, but a lot of that ends up in our traditional systems of record. Our systems of record have gotten richer. Our new forms of data have just increased, as well. You really have to look at this in a holistic manner and look at all data. Every piece of data has value in different ways to different organizations.

CI: A really interesting observation. I actually like it very much. I agree with you. I think it's all data. Tracey, let me bring you into the conversation a little bit. You talked about five business driven use cases. What are these?

TM: Sure. Well, what we've done is we've polled our customers. We've done thousands of engagements. Of course, there are many more than five use cases where Big Data adds a lot of value, but what we've tried to do is simplify the message to the customer, to the prospect to help them get started by identifying the top five, across industries where we're seeing customers get the most value.

So, the five that we've identified are, first, we call the first one Big Data exploration. And that's really the ability to discover, navigate, visualize, really to understand all of the data. So, Reed had mentioned, we don't want to just look at things in silos, not just one piece or one section of data. We want to understand all data inside and outside your organization, and once you can do this, it will help you with improved decision making.

The second use case that we identified is around a 360 degree view of the customer. And of course, this isn't a new idea. Many different products and capabilities have been trying to get a 360 degree view of the customer. So, what we do is utilize what organizations have already pulled together on their customers, and we marry that information with all other available data. So, in real time, we can incorporate not only MDM information, CRM information, ERP information, the social data that you



talked about, unstructured data that might be included in note fields or so forth. So, all of this information can be pulled together. We call that the enhanced 360 degree view of the customer.

And the third use case is really around security and intelligence extension. Similarly to the customer 360, folks have been pulling together security information for years. But it's not a situation where all that security information is ever going to be in one data source, in one place. So, if you want a full view of your security and intelligence information, you have to tie together all data. And this helps folks lower risk. You can do things like detect fraud, monitor cyber security, and again, in real time. You can pull this all together in real time.

The fourth use case is operations analysis. And this is where we really talk about and bring in the concept of analyzing machine generated data. And again, the idea here is not to analyze data for the sake of analyzing data. It's to analyze data so that we can help customers improve their business results. This machine data can come from anything. IT networks, meter sensors, medical devices, the list goes on and on. There's endless possibilities across multiple industries.

Then, finally the last one, or one of the more popular ones that we've identified, is data warehouse augmentation. This use case really integrates Big Data with your existing data warehouse capabilities. It lets you increase overall operational efficiency and, really, improves your ability to discover and analyze all the data you have more deeply.

I don't want to leave you with the impression that there's only five use cases. There's many across multiple industries, but what we're seeing is this pattern where these are the most popular use cases for folks to get started with. They are also where people see the most value quickly.

CI: Yeah, and I think that's what's important is that it's the value here. They really are able to use one of these five use cases to garner some significant value in a short amount of time, right?

TM: Right. Exactly.



CI: All right. Well, Reed, let me return back to you. If you don't mind, describe the new IBM architecture for data and analytics a little bit. You've simplified it. "The whole is great than the sum of the parts," I think is the quote that I have from you. Why don't you go ahead and describe it to us.

RM: Well, in many respects, this isn't so much IBM defining new a architecture as capturing it, in the sense of looking at what has been happening in our customer environments and you start to see patterns emerge, basically. A lot of those are driven by things like the outcomes that people want to see. The results people want to see. They want to see things like real time intelligence and analysis. They want to have better access and broader decision management with more sources of data. They want to be looking at, more and more at, predictive analytics. They also want to be able to look at navigation and discovery, where they pull together a bunch of data sources to help them discover new opportunities.

They also have to do this at the same time as dealing with, not just data at rest, which has been there all along, but they have to deal with data in motion. Then, data in all these new and different structured and unstructured forms. As a result, what we've really seen developing here is a real time analytic zone, an informational ingestion and operational information zone, an exploration and integrated warehouse and data mart zone, and a landing area which is an analytic zone and archive.

Around these areas, they support different forms and types of data and different types of analysis and exploration. For example, the real time analytics zone is really where you want the data in motion flowing directly into, and you're going to operate it on things like our Strings technology, for example. IBM Infrastructure Strings.

Data in the landing zone is, a lot of times... The traditional ingest that we've done in data integration unfortunately strips off some of the interesting information. It's about landing things like raw data there and being able to look at it with new forms of analysis and new tools, and then be able to basically find new value in that.



All this sits on top, of course, a layer of information governance, security, and business continuity, because more and more these things are connected. If one piece fails, or if you've got a leak somewhere from a security perspective, or if you're not paying attention to data being fresh or stale, you've got a major problem.

It's all getting combined together and it's all interrelated. That's why we've seen this kind of thing emerge, and so what we've tried to do is document that and communicate that to our clients.

CI: All right, well let's dive into a little more detail on the three big announcements. The first one I mentioned, or you mentioned, is DB2 with BLU Acceleration. It's an interesting change, an interesting set of enhancements. One of the things it can do, of course, is improve the OLTP performance, the transactional processing side of the house, certainly. I got from IBM that the main emphasis is really on improving the analytic performance. Is that right?

RM: Yes, it is. Basically, what we have with BLU Acceleration is a way to greatly accelerate your reporting and analytics. It's a new, dynamic, in memory columnar, technology that we've spent years developing. It's actually a third generation technology that has seen various iterations in various IBM products and now is available as a software only product in DB2 10.5.

It uses dynamic in memory columnar processing, actionable compression, parallel vector processing, and data skipping to really bring together a cohesive solution to the deep analytic problems that people have. Because of the fact that they want to analyze, and not pick and choose, the piece of data that they might put into some form of an accelerator, they don't know what the question is. They need to be able to look at a broad set of data.

Bringing all of these technologies together really allows us to have something which allows people to look at a broad range of data and not have to predefine things like indexes and aggregates and so forth, and still get 8 to 25 times faster reporting and analytics, and even some queries up to 1,000 times faster. By the way, you get a whole bunch of storage savings as well.



Things like actionable compression that we used to do this is about being able to act on the data all the way from the disk, to memory, to the processors, and have it in a form that we don't have to decompress to do things like joins, for example. This is extremely important because a lot of other technologies out there from competitors, they can only work with the data so far in the process. Then they've got to do a lot of work to actually match the data up. This is something that we're very proud of, and the performance results have been just absolutely astounding.

The other part about this is that, as I mentioned, it's part of DB2. It's part of DB2 10.5. It has all the same great capabilities that you're used to in DB2, with the great security, and all the great memory management that we've had in the past. You don't have to buy a football field of memory to use this. You can work on very large data sets, again, without having to have it all in memory. For us, it's been the combination of having something that is a huge innovation, but, at the same time, it's incredibly easy to adopt. To us, that's really the magic behind it.

CI: Wonderful. It certainly has been a very interesting discussion for the rest of the Boulder BI Brain Trust as well. Tracey, let me turn to you about the other two, the IBM Big Data platform and the IBM PureData System for Hadoop as well. Why don't you describe the capabilities in these two new features as well?

TM: Sure. Just as we were talking about the use cases -- and one of the reasons that we spent so much time documenting these cases and making them accessible to customers and prospects -- our whole idea was to try to make it easier for folks to get up and running with Big Data. That was also our focus with the improvements that we've made technology-wise to the Big Data platform and the new offering that you mentioned, the PureData for Hadoop.

First, I'll talk about the Big Data platform, and then we'll talk a little bit about PureData for Hadoop.

In the platform, we've made multiple improvements to help customers get up and running quickly. In our InfoSphere BigInsights product, that's our product where we have enterprise class Hadoop. It's our offering that



extends the open source Apache Hadoop distribution. What we've done is we've focused on making it more consumable and more secure so it's truly enterprise ready.

One of the improvements that we've made in this last release is we're providing full standard SQL access. Folks can leverage their existing SQL skills that they have in their organization rather than go and have to learn new ways to access the data. They can enhance and leverage the existing SQL skills and applications to be able to access Big Data. That's one thing that we've done.

Another thing in that same product, the InfoSphere BigInsights, is we're now including IBM's general parallel file processing system. That really improves compliance and security. It delivers positive compliant security and reliability, again, making BigInsights, making Hadoop enterprise class and enterprise ready. The other thing that we've focused on is making sure that we increase the availability and the performance of all MapReduce applications within BigInsights.

The second product that's part of the overall Big Data platform, Reed mentioned it earlier, is InfoSphere Streams. That's our high performance, real time ability to analyze data in motion. In the new release, what we've focused on is simplifying and enhancing the Java performance capabilities and high availability capabilities. When folks are pulling this into large enterprise environments, they're able to just get up and running more quickly, more easily. We've also added some new adapters, some new industry integration, again, just to make it easier for folks to get value quickly.

The last focus, that I'll talk about within the overall platform, is the new product that you mentioned, which is our PureData System for Hadoop. One of the things that we've focused on is to try to take away any barrier for a customer. Sometimes a barrier for a customer is just to get all the hardware and the different pieces up and running so that they can get access to the Hadoop capabilities.

With this new version, what we're doing is we're making Hadoop consumable for pretty much every enterprise. We've tied together our



BigInsights with a PureSystems management console. What we're really doing is integrating the compute, storage, and networking capabilities all together in one system. Instead of folks separately trying to pull these pieces together and having it take them quite a while to get up and running, we can get them up and running with Big Data in just a couple hours.

CI: All right. Fascinating as well. Reed, let me return back to you. We're about out of time, but I do want to touch on a problem IBM is doing a good job of handling. That is, the confusion around Big Data for a lot of BI implementers. You stated that if it were easy, everyone would be leveraging it. What is the problem with Big Data, and what's IBM doing to simplify it?

RM: One of the things is, first of all, there's been this confusion about what is Big Data. Some people, again, going back to the former question of thinking that Big Data just equals Hadoop. People run off in that direction, and that's only part of the problem. The other thing is that, in that particular area, there's a real skills shortage. What we've done to our products, for example, is we've added Big SQL to our Hadoop implementation, which is IBM BigInsights. We have created our Big Data University which has had over 70,000 enrollments. We've worked with tons and tons of customers, like, I think right now it's 1,550 customers or more.

We get down and engage with customers to help them accelerate, but we've also tried to make the technology easier to adopt so that skills gap is addressed. Things like Big Data University are places where people can go and learn about the technology, in general. Not even from an IBM centric point of view but just about Big Data in general.

Then, of course, we give the option to learn about our products, as well. We're trying to really close in that gap between learning, and help them progress along the way with services and also with engagements and fill in the gaps where they may not have the expertise under their own roof. Also, help people get out of this mindset that "I've got to keep doing things the way I've done them for the last couple decades", because as I mentioned, in the new zones that we are seeing emerge in IT organizations, it really fundamentally changes the way you manage data.



We are encouraging people to look outside the box from what they've normally been doing, and keep an open mind about this stuff, because it's going to take some change in the organization to make it effective.

CI: Yeah, I agree. I agree wholeheartedly. Unfortunately, I think we're out of time, at this point. That's it for this edition of the BBBT Podcast. Again, I'm Claudia Imhoff and it's been a great pleasure to speak with Reed Meseck and Tracey Mustacchio of IBM today, thanks to you both.

TM: Thank you.

RM: Thank you very much.