BBBT®
Boulder BI Brain Trust

Leveraging
Business
Intelligence

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:	**August 21, 2015**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Harriet Fryman**, IBM Analytics VP of Marketing
Run time:	**00:13:22**
Audio link:	**Podcast**
Transcript:	[See next page]

Claudia Imhoff: Hello, and welcome to this edition of the Boulder BI Brain Trust, or the BBBT. 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 BBBT podcasts are produced by my company, Intelligent Solutions.

I'm pleased to introduce my guest today. She is Harriet Fryman. Harriet is the Vice President of Marketing for IBM analytics platform, of course for the company, IBM. Welcome, Harriet.

Harriet Fryman: Thank you very much, Claudia.

It's great to be talking to you again.

CI: It's lovely having you here. I'm just so thrilled that you gave the presentation for IBM today, but let's talk about what's going on. IBM announced in June its support for Apache Spark. What was the reason for this?

HF: The announcement around Spark was a really major commitment for IBM. When we look at our mission with analytics to build intelligence into everything, we're really looking for a way or a technology that can access data wherever it resides, enable data scientists to build models and work with the data engineers, and then work with application developers to implement those insights into the everyday business.

When we look at where analytics is really going to give a return for our customers, it's putting it into applications so that we can make smarter applications, and it's Spark that has the technology that enables us to do that.

CI: You mentioned in the presentation today about the insight economy as being the ultimate goal of analytics. What are some of the examples of this, and what is its purpose?

HF: We definitely believe that we have now entered the decade of data where everything around us is digitized and captured in data. We know that we have the analytics and technology available to analyze that data and also do it at a cost-price-performance that means we can do analytics in far more parts of businesses than we did before.

We really see that as driving the next economy, this new economy based on insights, so the insight economy. Where we see front runners already gaining value out of analytics is when we look at our data which is 69% of leaders say they have a significant positive impact on business outcomes when they use insight.

We see 60% created significant positive impact on revenues, and 53% created a significant competitive advantage. That is really where analytics is uncovering new business opportunities, new business models, and in fact, whole new businesses in different industries. That's why we really believe we're entering this decade based on data and insight.

CI: The next slide, I believe it was the next slide that you showed us, was also a maturity scale, if you will, of companies that started first with just bare bones, operational systems just so they could get up and running. Then we moved into the era of the data warehouse.

If you don't mind, talk a little bit about that slide, because it really was the last two levels of maturity where we really got into the value of analytics. It's not just performing the analytic, or creating the analytic. It's doing something with it, right?

HF: Yes. It's almost like we intellectually know the value of having analytic applications where we look at data, we understand customer churn, we understand supply chain, demand, forecasting. Those are applications where analytics has a great play today.

The next wave of innovation and indeed the value that's going to transform businesses is when we put that insight actually into applications and everything around us... Where cars can know based on analytics to break when there's a dangerous condition, when thermometers know to

adjust the temperature of heating based on the ambient temperature in the world around us.

When we have the inside actually built into applications and everything is smarter because they all have a little bit of DNA of analytics in them. That's where the next decade of value is going to come from in analytics.

CI: Yeah, it is, and it's almost a real time analytic now. We've got to make decisions within seconds, right?

HF: We need analytics to be real time if we're really going to drive a smarter world around us and Internet of Things, mobile interactions, people engagement.

CI: Let's get into a little about Hadoop. There were lots of questions about the relationship of IBM and Hadoop, IBM and Spark, and IBM and Spark and Hadoop.

CI: You had a slide up that said that Spark was complementary to Hadoop. Now, a lot of people have said that Spark is replacing Hadoop. What do you say about that?

HF: We have to look at Hadoop as having a great role to play in managing masses of data in a file system, and that ability to manage the volume of data will stay. Spark is not for that big data management problem.

Where Spark comes in is it can actually help extract the value out of Hadoop by providing the way to gain insights, and run real time analytics against the Hadoop file system. We definitely see it as complementary to Hadoop.

There are some vendors that would say Spark is for Hadoop, but I think that's too narrow.

The value that Spark can provide is a broader analytic operating system across multiple data sources and for multiple purposes.

CI: Let's talk about the June announcement a little more. Big commitment to Spark. It was actually surprising to me how much IBM has poured its

resources into the Spark community. Tell me and our audience a little bit about what IBM is doing.

HF: If we would stay in the context of the decade of data, we see Spark as possibly the most significant open source project for data and analytics. Therefore, we've made a serious commitment to Spark, and I would summarize it in three main areas.

We have provided an open source of our SystemML technology system, machine learning technology, to the open source community, with the intent that we provide it as a committed part of Apache Spark.

Secondly, we have committed to educate a million data scientists and data engineers to help fill that shortfall of skills that we see in the market place to really make sure that analytics can be everywhere it should be.

Then third, we announced the formation of a Spark technology center based out of San Francisco to really drive the innovation and the adoption of Spark.

CI: Let's pick on the first one, and that was SystemML. A little bit about that. First of all, what exactly is it? You gave us a little bit of a definition, but what exactly is it? Secondly, why would you contribute SystemML to the Spark community? My final question on that is what are the benefits of that contribution, especially to IBM?

HF: SystemML follows a set of systems that have been developed out of our modern research. System S for streaming analytics, System T for text analytics and SystemML for machine learning.

You can think of it as an engine in which you can express algorithms and those algorithms are then managed in the execution by the engine. A data scientist whose expertise is in writing algorithm models doesn't have to worry about the optimization and the execution of that model. The engine's going to take care of it.

Why this is important for Spark is we really see Spark as centering on the data scientist skillset. With SystemML, we provide an engine that is almost an infinite library of possible algorithms that can be deployed today or

built tomorrow. It's really going to enable the establishment of a standard way to execute machine learning in Spark.

CI: I think that's the important part of that. That standard that you want to establish for Spark. The reason for that is what? Why do you want to establish this standard?

HF: If we truly are to get the value out of putting analytics into everything, we need to empower data scientists to be able to build and deploy those algorithms in a far bigger population than we have data scientists today.

By standardizing SystemML, we provide a way to almost on-board data scientists into Spark and then deploy those out into applications.

CI: Yeah, it's terrific. You blew me away when you said that IBM wants to educate one million data scientists and engineers on Spark. How are they going to do that?

HF: First off, we know there's pent up demand of hundreds of thousands of people that want to be able to build data projects, be able to develop their skills and data science, so there's pent up demand there.

IBM is also the largest contributor to Big Data University. There's about a quarter of a million people already on that community taking courses in big data. There are two Spark fundamental courses out there.

Just in the first two weeks, I think for the fundamental course, there were 3,500 people who took the course. We're also looking at the expansion of that university to China and to Brazil so that we can get even more people on.

We are partnering with the train in education companies out there like Galvanize and Metistream in order to really make sure that everybody can get access to the education they need. We feel very comfortable that we're going to achieve that one million target.

CI: One million, that's wonderful. Why don't you tell me a little bit about how IBM will now incorporate Apache Spark into its own products?

HF: We are committed to putting Spark into the foundation of our product strategy in four main areas. First off, we deliver Spark as a service on IBM Bluemix, our developer cloud, so you can use Spark as a developer, or data scientist, without necessarily using Hadoop.

Secondly, we distribute Spark as part of our IBM open data platform so that it can be leveraged alongside our big insights capabilities.

Third, we are building Spark into the foundation of our IBM products, products like SPSS, Commerce, and Smarter Cities Solutions.

Finally, we're looking to be a part of the community that's contributing to the core so that we can assist Spark in all it needs to be to be a production environment for our customers.

CI: All four areas. Very important. Stand alone with Hadoop doesn't matter. You're going to mix and match and be able to offer any of the combinations there. Let's end with a few examples from your vast array of Spark customers. Tell me a little bit about how these customers are using Spark.

HF: A number of our customers are using Spark with the Hadoop deployments they already have in that first use case we discussed which is the way to gain real time insights from the data held in Hadoop and Spark gives advantages to them over MapReduce. We see some people using Spark as the alternative to MapReduce to get the data in Hadoop.

What's more interesting is other companies such as NASA and the SETI Institute which are looking to outer space to identify if there are any signals, or conversations happening in outer space.

You can imagine that's analyzing terabytes of complex deep space audio signals, and they're using Spark machine learning capabilities to hunt for patterns that might portray the presence of intelligent extraterrestrial life.

I love that example as really showing, and demonstrating the volume of data that Spark can analyze as well as the sophistication of the analytics that Spark can execute.

CI: I hope they do find life. It's a pretty exciting project all onto itself. We could talk about many, many other examples. You do have so many in companies using Spark, but unfortunately, we're out of time.

That's it for this edition of the BBBT Podcast. Again, I'm Claudia Imhoff. It's been such a pleasure to speak with my friend, Harriet Fryman of IBM today. Thanks so much, Harriet.

HF: Thanks, Claudia.

CI: I hope you enjoyed today's podcast. You'll find more podcasts from other vendors at our web site www.bbbt.us. If you want to read more about today's session, please search for our hash tag on Twitter. That's #BBBT. And please join me again for another interview. Good bye and good business!