



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:	HortonWorks
Date recorded:	August 28, 2015
Host:	Claudia Imhoff , President, BBBT
Guest(s):	Shaun Connolly , VP Corporate Strategy Matt Morgan , VP Product and Alliance Marketing
Run time:	00:20:11
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 guests today. They are Matt Morgan and Shaun Connolly. Matt is the Vice President of Product and Alliance marketing, and Shaun is the Vice President of Strategy for Hortonworks. Welcome to you both.

Shaun Connolly: Thanks for having us.

Matt Morgan: Thanks for having us.

CI: It's great having you both here. Shaun, let me start with you. It's been a year or so since you last presented to the BBBT. What's been going on? How's the company doing?

SC: It's been a very interesting year. During the course of that, in December for instance, we went public. The IPO was a branding event for the company. It really helps us project our Open Enterprise Hadoop story, and our next generation Data platform and architecture story, out to the broader market.

I think we're roughly around 745 employees worldwide in 17 countries. We are, I think when the last earnings call, at 556 customers with well over 40% of those customers coming in just the last few quarters. The business is accelerating very quickly. The technology clearly over that last year has been accelerating fairly dramatically.

The use cases and the richness of the stories, which is really what I like, has been solidifying and making it clear what this crazy "animals in the zoo" called Hadoop and Pig and Hive and all this stuff... How do they come together to actually help me solve a real problem?



CI: Yeah, it's no longer an elephant. It's more of a rocket ship, don't you think? Just to keep with the analogy. Matt, let me turn to you. You mentioned the Hortonworks vision is to enable Apache Hadoop to be the enterprise data platform. What is the strategy at Hortonworks to achieve that enterprise platform status?

MM: If you talk to any industry, any vertical, the need to maximize data is now one of their top imperatives. We believe the truly innovative way to do that, is to give folks the capability to create, organize, and manage data through a central data lake. Our point of view is Hadoop solves that problem exceptionally well, but doing it in the enterprise takes a very unique approach to do it right.

We think that approach is called "Open Enterprise Hadoop." It's a categorical differentiation in terms of taking an approach to get Hadoop operating at enterprise scale. It has four basic attributes.

The first one, we operate 100% in the open source community. Every component is open source.

CI: That's an important differentiator, is it not? Aren't you about the only company that does that?

MM: We are the only provider that operates 100% in the open. Folks need to understand the difference, there is pretty significant. Organizations that leverage a 100% open source solution have the capability to capture all of the innovation that's happening in this robust community.

Also, the speed of release... It's like nothing I've ever experienced in my career. The open source community is innovating in Hadoop in so many different directions that an organization truly has the opportunity to take and select these areas that match to their business value.

We also believe that influencing that community needs to be a core opportunity for our customers, so Hortonworks does that through our committers. We have more committers than any other Hadoop distribution provider, all working in the open source community to help influence the direction. We've seen the benefit of that work.



That's just one of the differentiations of Open Enterprise Hadoop. The other three include centralization, interoperability, and enterprise readiness.

Let me spend a moment on centralization. When you start bringing together the data into a single repository, accessing and managing that data gets really, really complex. You need to be able to have specific routines and policies that you set up from a security point of view, from a governance point of view, but more importantly you need to have the overall accessibility of that data in a multitenant, multifaceted heterogeneously provided way.

The reason that's so important is because all of the data needs to be available through the analytics that are required for the data at that specific time, for that specific business case. Having a central architecture, and we have standardized all of our technology around something called YARN, a data operating system, it allows you to truly build that central data lake.

Folks may start off where everything's not centralized. That's OK. YARN works in small data ponds, if you will, but if folks want to evolve to that central data lake, having that centralized approach is unique and our approach to do that with YARN is also unique. It's a core component of Open Enterprise Hadoop.

The third area, interoperability. This is an obvious one. By having a 100% open source infrastructure, components, software, technology, we have the opportunity to provide the broadest Open Enterprise Hadoop ecosystem. We've also standardized and partnered with specific organizations around their engineering roadmap to give our customers the most options.

What's also important is Hortonworks is a founding member of ODP, or the Open Data Platform consortium. With our other members of ODP, we have standardized around a common core of technology to eliminate the combinatorial explosion of complexity that could exist if you try to go it on your own. Interoperability is also important.



Finally is the enterprise readiness. All of this stuff would be great, interesting, but not usable if you didn't deal with the enterprise hardened requirements.

We've been working diligently with the community and our partners to provide that technology, whether you're looking at it from a governance point of view, from a security point of view, or you need single pane of glass operations, your operator can truly move quickly to procure and spin up and spin down Hadoop clusters whether on premises or in the cloud, we give them all of that technology.

All of our enterprise readiness, as I mentioned earlier, run on top of that YARN data operating system. You have the capability to truly scale the individual components. This is our unique point of view in the market. Again, it's called Open Enterprise Hadoop.

CI: Yeah, It's really interesting. I want to dive into the partnerships a little bit more. Shaun let me go back to you. One of the things that I think is so impressive about Hortonworks is their joint engineering projects. You put up a slide that had just a laundry list of mostly big name companies that you are partnering with on these joint engineering projects. Why is that important?

SC: Sure. The interoperability facet is a key element to the strategy. I'm a technologist. I love what goes in our green boxes, I like to say, but if it's not well plumbed with the rest of the enterprise, then it's going to be a silo where it's going to have very limited utility.

The two-pronged approach was really make the tech great and enterprise grade, but integrate it with what folks have in the data centers as well as other emerging technologies that are coming out.

That takes engineering work. It takes time and effort to integrate it into the Microsoft ecosystem, the EMC ecosystem, the HP, SAP, SAS, right? They're analytics, they have a great community of users that want to have access to this rich lake of data, but in a familiar way with their tools. There's a raft of SQL users out there. How do you enable them in a way that they can use their familiar tools and gain access to it?



It was a founding principle of the company back in 2011. Our first two partners we actually started the journey on, they were picked for very specific reasons. One was Teradata, because they were the big data provider and continue to service large operational deployments. Top of the pyramid, so to speak. The largest of the large enterprises.

Microsoft was the other one around... their Azure strategy, Azure HDInsight is the Hadoop-as-a-service offering that's powered by HDP. It enabled us to go after this in a very large enterprise, whatever the enterprise needs, and then Microsoft's great at making technology easy to use and approachable by billions of people. That was really the rationale there.

We rounded it out with dozens and dozens of partners that we spend a lot of time, effort, and technology development and joint roadmap planning with.

CI: Let me ask you a little bit more of a probing question. The ODP is an interesting group. If you don't mind, just spend 30 seconds and explain what it is and who's in it.

SC: Sure. ODP was launched earlier in 2015. Ourselves, Pivotal, IBM, GE, and a range of others. SAS is one of the members.

CI: IBM, I think.

SC: Yeah, clearly IBM is in there. Now there's a few dozen folks involved. Altiscale, who's a Hadoop-as-a-service company, their CEO is Raymie Stata. Interestingly, he was the CTO at Yahoo when they chose to create this thing called Hadoop and donate it to Apache.

It's a good mix of vendors and end users, and really the goal simply in my mind is not to create a new Hadoop distribution. That's not interesting. It's to solve the problem for the ISVs and the system integrators on, there are so many different versions and permutations that you get into this, "If I create a solution, I might be halted at procurement because I'm not compatible with a certain platform."

Whether it's Pivotal, IBM, Hortonworks, or whoever else is compliant in ODP, the goal there is to enable the raft of new solutions to be able to be



on-boarded quickly. From my perspective, that brings more processing and more data under management, which I think is a good thing and fits with our strategy.

It is a very real problem. We hear it from the enterprises. We hear it from the software vendors. We're really trying to figure out how we go and drive compatibility into the market, in a way where it's not really about building technology in Apache... that goes very well. It's really about driving compatibility.

CI: I think it's a terrific initiative. I commend you for being a part of it. All right Matt let me go back to you. You mentioned "customer use cases." They fell into about six different categories, I guess. If you don't mind, fairly quickly, give me the six categories and maybe you can talk a little bit about some of your real customer case studies.

MM: The customer journey with Open Enterprise Hadoop is one of these variable journeys where customers can select an entry point that makes sense for them. As you had mentioned, we really find that there's six clusters of use cases that help articulate the business value and the cost savings value that they would receive from deploying Open Enterprise Hadoop.

We've organized this into what we internally call a placemat, which helps our customers understand what they can do with this technology. Across the business outcome axes, we have a total of 53 different use cases, but they're organized into three categories.

Data discovery, which is about collecting massive sets of data and tearing loose the data scientists, letting them go wild and come back with insights.

Single view, this is an exceptionally popular use case. Organizations create a 360 degree view of, frankly, an entity. It could be a customer, it could be a plant, it could be equipment. Anything that's associated with a single entity that's interesting to that business, you obviously can leverage Open Enterprise Hadoop for that.



Then there's predictive analytics. Once you start collecting this information, once you have single views of your entities, you can start understanding and seeing trend lines. With trend lines, you can see that if 90% of, let's say a customer does one, two, and three, and they also do four, you can start suggesting to the customers that have already completed one, two, and three, that, "Here's four for you." Right?

These are the things that enable an organization to operate more efficiently, but they also give an organization the opportunity to provide a better user experience.

The business outcome side of the use cases as I had mentioned in our chat earlier, it is driving the lion's share of the conversations now. It's amazing how fast these use cases have moved into providence. They are the ones that people want to start with and have conversations about.

We also find, and I don't want to discount too much, are the cost savings use cases. Many times our customers will use cost savings use cases to fund some of the innovations that we just talked about.

The three groupings include active archive, or taking offline data that would typically be committed to an offline server or a tape drive, and keeping it online for pennies on the dollars in cost.

ETL offload, helping ensure that the broad value ETL work that doesn't have time to be leveraged inside of your enterprise data warehouse, can be leveraged, but leveraged in Hadoop.

Then of course there's data enrichment. Taking an enterprise data warehouse and adding unstructured sensor data, variable type data, and extending the value of that work.

These cost savings areas can be significant. All of these six areas of use cases, these six groupings, are presented again in that placemat so that organizations can jump around as required.

In our talk earlier, we talked a little bit about this. We also talked about specific case studies. Mercy for example. This is a single-view case study... Amazing case study that looks at how they have taken three disparate



Epic databases and brought them all together to provide predictive insights and to give a single view of their customer. "One Customer, One Record," I think is the name of the initiative.

Another one is Symantec, trying to reduce the time it takes for them to understand and communicate the value of what hackers are doing to their endpoints so that they can protect against attack.

We also looked at other use cases that focus on predictive analytics using Internet of Things sensor data to draw trend lines.

CI: Yeah, they're really interesting, and people can find videos of these if they want more information at your website, right?

MM: Yeah. There's a collection of case studies that we've completed that include a write-up of the value that they've received and videos of the customers themselves. You can go to hortonworks.com and click on the "customer" tab. We also have some videos from the last summit, "Hadoop Summit 2015." A lot of the keynote sessions were talking about this type of transformational value.

You'll see broader case studies including the industrial Internet, and how Hadoop can power that. I would encourage everyone to check that out.

CI: Matt, let me stay with you. We only have about a minute left, so we're running out of time, but let me stick with you for a second. If you don't mind, tell me what's new with Hortonworks Data Platform 2.3 in a very short amount of time.

MM: We're super excited about 2.3.

2.3 was announced last month. It had three major themes. The first one is around the user experience. Any of you who have worked with Hadoop knew that the exposed working environment could be complex. It could be command line driven.

Basically we've delivered what has been largely called a new face of Hadoop in this release. We call it a breakthrough user experience. It's a more polished term, but it's a dramatic improvement in the user



experience, especially for the developer and the operator... for setting up, procuring, expanding Hadoop, for leveraging common development tasks around SQL.

You now have full blown graphical user experiences that are exceptionally modern. I think, Shaun, you like to say... "in some cases it's as easy as setting up a 401k portfolio now."

The other theme is enhanced security in governance. I think that there's been expansion of our capabilities on the encryption side, encrypting data both at rest and in flight. This is an exciting capability, one that's been requested for a variety of use cases. We've taken Apache Atlas as a top line project now, expanding the data governance initiative.

We've also introduced something totally new that we're super excited about. It's called, "Hortonworks SmartSense."

SmartSense is a proactive support solution. It's included in our subscription. It instruments our customers' clusters and provides the capability to understand their current status and health and predict and proactively recommend solutions for them before it becomes an issue.

The idea is to provide the shortest time possible to solution. Our proactive support in SmartSense is an opt-in solution, you're not required to use it, but we are finding that customers are super excited about it. It gives them much broader control and insights into what's happening, and using our self-service portal they don't even have to talk to folks to be able to see some of the value.

CI: Yeah, I really liked the SmartSense piece of the presentation. It was quite good.

All right, Shaun let me end with you if you don't mind. Now tell me what's really new.

You recently did acquire a company called Onyara, the Apache NiFi company. Tell me what that's all about.



SC: Sure. The news was around the acquisition of Onyara, they're a very early stage company. They incorporated in December 2014 and opened their doors in March 2015. Haven't been around a while but the technology was originally called Niagara files and was built up over the last eight years within the US government space. The NSA spun it out as part of their technology transfer program as Apache NiFi.

It's an Apache Foundation project. Very robust set of capabilities. That'll provide the core, if you will, of our new subscription offering, called Hortonworks DataFlow.

It's a complement to our Hortonworks Data Platform offering, but it's really about gathering the Internet of Anything data as close to the point of inception as possible and orchestrating that data through various paths in the pipeline, transforming it along the way to get it to the right analytic or right end user at the right time.

We'll be having a webinar, I think at September 23rd, where we'll talk more about DataFlow, NiFi, and the road ahead. I encourage people to sign up for that and learn a little bit more.

CI: Excellent. It sounds like the road ahead is very exciting for you folks. Unfortunately, we're out of time so that's it for this edition of the BBBT Podcast. Again, I'm Claudia Imhoff. It's been a great pleasure to speak with Matt Morgan and Sean Connolly of Hortonworks today. Thank you both.

SC: Thank you. Thank you for having us.

MM: Thank you.

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!