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:	**Cloudera**
Date recorded:	**October 30, 2015**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Clarke Patterson**, Senior Director of Product Marketing
Run time:	**00:19:00**
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. He is Clarke Patterson. Clarke is the senior director of product marketing for Cloudera.

Welcome, Clarke.

Clarke Patterson: Thanks, Claudia. Really happy to be here.

CI: I was so happy to have you guys here. You did a great job I thought.

But, first of all, you started off by saying that our relationship, that is, customer's… company's relationships with data is changing. What did you mean by that?

CP: It's really about every organization starting to think data first.

If you look at a lot of the emerging organizations… we talked about Uber, Airbnb… all these companies, in a lot of ways, the problems that they're solving are ones that have been addressed by other organizations for a really long time. What separates them, though, is they're very data-centric, very heavily focused on personalization.

That's where every organization has to get to. There's no excuse these days with the technology that's available to everyone to not be thinking about collecting as much information… structured, unstructured… and using that to optimize your business.

When we say that the relationship with data is changing, it really is all about helping every organization, whether they're 5 years old or 100 years old, understand that a data-outwards view is the key to success moving forward.

CI: I thought what was particularly important in the companies that you mentioned, in Facebook and Alibaba were two others. They basically have no physical assets. They are just data. Airbnb doesn't own the rooms. Uber doesn't own the cars. It was a fascinating trend that said that data was king, right?

CP: Absolutely, absolutely.

If you look at Airbnb, there's a huge focus on the relationship between the renter and the property. If they can really make that relationship and establish that, the rest of it's kind of irrelevant to them.

They don't have to own the properties. They're just acting as that service to connect those two people together and do it in a way that's so personalized and almost intimate in a way.

It changes the dynamics of that business altogether. It is a very, very interesting trend that they don't have walls like we're sitting in here.

CI: Right. Let's talk about a few of your customers if you don't mind. How are they using the massive volumes of data that they've collected?

CP: There's a range of ways. We have companies like British Telecom and AMEA. Telecommunications is a pretty big industry for Hadoop. In the BT sense, they're really focused on the complete view of the customer with a focus on personalization.

You'll hear this a lot with Hadoop and big data generally. Personalization, using as much information as we can, is really where people are trying to get to. Customer 360, single view of customer we've been talking about these things forever. It's not a new topic. But you can get far more individual interaction with more information predicting what people are going to say and what they're going to do and all these other types of things.

BT's really focused on solving that problem, taking the, I think they have 1,600 different systems, collapsing all those things together. Pulling in new sources of information from sensor-driven devices and things of that nature to understand in a geospatial sense, what is this customer doing

right in the location that they're at right in that point in time and how can I improve the customer experience right at that point in time?

At the same time, from an efficiency standpoint in how they operate their business, there's a lot of knowledge that's being captured there as well that allows them to understand how they can deliver a better service.

If they can use the delivery of their network bandwidth and optimize that for a certain customer set, I know there's a big football game this weekend. You get all these people converging on the stadium, and you can tell that there's a whole bunch of your customers that are going to be there. You can get out in front of that problem right out of the gate.

CI: There's going to be a bottleneck at some point.

CP: Exactly, exactly. So knowing that and being able to take those steps really helps not only a company like BT to drive efficiency across their operations but, in so doing, deliver better customer service.

There's an enormous amount of what I call "Feel good stories" that we have in our customer base these days. I think you were in New York a couple weeks ago. We heard from one of our customers called Thorn.

They are combating child trafficking. It's a complex process of text analytics, doing some natural language processing, getting some machine learning, and things of that nature where they scan Craigslist posts and Facebook posts and all these different types of things. They can detect that advertisements on some of these channels are actually for kids.

They're able to notify authorities. They've seen a massive, like 40 percent, reduction in incidences as a result of this.

There's all those types of examples that are starting to emerge. It's no longer just about efficiency. It's really great to see where this platform is going.

CI: It is. I like the feel good stories I think better than selling more shoes to women.

CP: Ad placement, exactly.

CI: Yeah, that kind of thing. Well, let's transition to your product. First of all, what is Cloudera's strategy for its products?

CP: We have what we call a hybrid open source model.

At the core of our business is a focus on open source software development on Apache Hadoop. Apache Hadoop has many meanings these days, but we look at the broader ecosystem. Anything that allows us to integrate, store, process, and serve data up to an end consumer is where we focus our open source efforts.

But we realized that, as this technology becomes more mainstream, there's an enterprise need and an enterprise set of requirements that have to come along with that. You're not going to adopt some technology that is Wild West-oriented and bank your business on it.

You have to have security. You have to have governance and be able to weave it into those types of policies. It has to be easy to manage. So we really focus on "fast, easy, secure" as the three tenets of our overarching product strategy and building those capabilities on top of the Apache Hadoop ecosystem so that it can be adopted with confidence.

CI: And we'll get to the security in a moment because I think that is still a concern somewhat. But let's talk about Hadoop a little bit and what exactly Hadoop is.

You showed us a terrific slide just of the myriad products that are coming out that do different aspects of this called "Hadoop ecosystem." Why don't we start? It is still somewhat complex, but it is still very confusing to many enterprises to even try to get their heads around what "Hadoop" is. We'll use the air quotes around Hadoop.

How does Cloudera then help these companies incorporate this ecosystem into their environments?

CP: A lot of it is, simple education of what all the individual projects are. It's interesting, the air quotes aspect of it, because, depending on who you ask, you're going to get a different answer in many cases.

If you go by the official definition of what Apache Hadoop is, there's three central things to it. It's HDFS, it's MapReduce, and it's YARN. As we know, there's a whole bunch of other things. We showed that chart. There's an enormous amount of activity that's going on.

But what we spend a lot of time with our customers on is trying to understand what the requirements are, first and foremost, and helping them understand that, while there's many projects across this ecosystem, they all vary in terms of maturity for one but, more importantly, the applicability to what the customer's trying to do.

I hear it all the time where organizations are saying, "We need to do Hadoop. But you're showing me this chart, and it's got all these projects."

CI: Scary.

CP: "I don't even know how I'm going to start with this." They don't understand that maybe five of them apply to their initial use case.

There's a, I'll say consultative, an engagement conversation that we have to say, "What are you trying to do? Here's what these things are," and educate them on what the various pieces are so that they can understand how they can apply those to their current problems. Then, of course, educate them on the additional ones as it applies to ongoing evolution of it.

CI: What I thought was really good was the fact that Cloudera is kind of packaging these things up in a way that do meet a particular customer's needs. You don't have to cherry-pick through all of these things. Cloudera will do it for you. Right?

CP: Definitely. If you look at Cloudera Enterprise, the product that we sell, there's three primary ways that it can be adopted. We have what's called Basic Edition, we have what's called Flex Edition, and what's called Data Hub Edition in support of our enterprise data hub vision.

All three of those are designed specifically to help organizations navigate through this complexity of what applies to them. The name should say that Basic Edition is really that core base case.

We talked about data engineering today. A lot of the capabilities that you need within Hadoop come in that particular offering. That helps the decision process for anybody that's thinking about this really get down that path.

Then as you go up in maturity, obviously there's different ways to try and adopt it. You can layer on the capabilities based on what your needs are, your skill sets, all these other things that you need to consider.

CI: I think that's the scary part for a lot of companies is they don't have the skill set. They don't have the right people to put it all together again. They really are looking to companies like Cloudera to give them advice, tell them, "Here's what you need."

CP: Yep. It's a big investment for us as an organization. There's three central pillars for us.

One is, call it public training, where we actually bring events to regions. We'll come out here to Denver. We'll sit here for a week and say, "You want to be a Hadoop developer? Come and sit down. You can do certification at the end of the class."

We do private training directly for companies as well when they say, "Look, we're going to adopt Spark perhaps. We need help getting X number of people up to speed on what this is." We've got a whole program designed around that. We have that public aspect of it.

The second piece is just getting training or Hadoop knowledge generally out into the market at large. We partner with Udacity, is one of those other curation vehicles where they take our curriculum, it gets pushed out, and more and more people, I think the last number I saw was 40,000 people or something like that worldwide has gone through training through that venue in fashion or another. There's a scale aspect of it in trying to just make it consumable more broadly.

Then the third piece of it is partnering with universities. We have what we call a Cloudera academic program. We cover over 100 universities around the globe today. It's a constant focus for us.

The idea is if we can infuse a lot of our training knowledge into curriculums, we're creating a groundswell of people coming out of school that are going to be ready for this stuff because it's not going away. There's admittedly a shortage of talent. We're really trying to fill that gap.

CI: I think that is absolutely admirable. Let's get back to the product. There are different deployment options, right?

CP: Yep. We focus on customer flexibility at the end of the day… so on-prem, obviously, cloud, through all the major providers, engineered systems…so we have a Teradata appliance now, for example, one with Oracle as well. Depending on what the customer's looking for, lots of different deployment options.

CI: Almost anything, hybrids, mixtures, whatever it is.

Of course, security, we wanted to circle back around to that. Security always seems to be top of mind for many organizations. How does Cloudera respond to this concern?

CP: We've seen security as a big deal for a long time now. We've made engineering investments, we've made acquisition investments, and we've really been trying to get around this for well over two-plus years.

It's percolated from, when I started the company just over two years ago, surveys had it actually surprisingly low, Hadoop security specifically. If you look at those same surveys today, it's becoming number one, number two concern for the information getting there. So we're investing very heavily to make sure that we've got all the bases covered.

Austin, Texas, we have our office there called the Center for Security Excellence. All that team does is think about security of data that's coming into Hadoop environments and how we can make sure that it's protected and encrypted and all these other things.

Within the product itself, we take a pretty different approach to the security problem for Hadoop than you would see elsewhere. The easiest way to describe it is there's multiple different ways to get at the data. You've got Hive and Spark and MapReduce and all these different types of things.

For all intents and purposes, they all have different security touch points at the data that sits in HDFS or HBase or Kudu now in Cloudera Enterprise.

Because they're all different, they all represent multiple, different points of vulnerability. You define your policies for Hive. That's all great, but if somebody goes through in MapReduce, say, how do you reconcile what happened? Are there changes that you need to make across one of them?

We're focusing on addressing that by building security into the core of the distribution. We have this new project we announced a month ago called RecordService. It's like a unification service for all data. It sits between the storage and the consumption layer and makes sure that, regardless of how you're coming at data, you're getting served up all that information in a consistent manner… So you know if something's going to happen, your policy is defined one way, you're accessing it one way, it's getting served up one way, and it really simplifies the whole thing.

CI: It really does. It's so needed to have that centralized hold over that area.

CP: It absolutely is. It's one of these topics. We have leading financial services organizations that use our product. They refer to their deployment as a compliance hub where they literally will say it is the safest place to land, in their words, all data for their organization because they know it's secured in one way. It can be encrypted…everything else, and it's in one spot.

When you want to serve it out to your warehouse or whatever, you can go and get it and move it and whatnot, but at least they know that they've got that vault, if you will, where it sits. It makes the whole process easier for them.

CI: All right. We got about two minutes left, so let's go through this a little bit briefly. Tell me a little bit about your partners.

CP: Partner ecosystem is very important for us. Three pillars to our focus here.

One is breadth. The system makes sure that we have the coverage for the existing infrastructure that our customers have already deployed. We want to work with it so we're constantly looking at it. We're over 1,900 partners across SIs, data integration, apps, the whole nine yards. We've got that base covered.

Integration is key. So we focus heavily on partner engineering to make sure that we've got a dedicated team of people that work with our partners to make sure our products work seamlessly together.

We also want to focus on whole solutions as well. So we're starting to work across multiple parts of our partner ecosystem to see how we can come up with accelerators, if you will, to get whole solutions out to customers' hands a whole lot faster than we could before.

CI: All right. Last question then. It's kind of an odd one to end on.

Gartner came out with a survey where they said, I think some 60 percent of companies never got beyond the proof-of-concept phase for their Hadoop environment, and some 50 percent said that they really needed cultural change in order to even get the value out of their big data initiatives.

That's problematic or at least I think they see it as "This is something we need to solve." What's Cloudera's response to this?

CP: It's interesting because I think if anybody tells you that Hadoop is simple, you got to maybe shake them a little bit. This stuff is tough. It's complex technology. We acknowledge that.

What we're trying to do to alleviate that is focus on really two things. One is understanding… almost a Hadoop maturity model, kind of an adoption pattern.

There's three areas that we see… data engineering, which is your storage, your integration type of things, analytics and discovery, almost like

self-service BI and what you do there, and then building data products or data applications.

What you do with the product varies in each of those stages, how you shift from one to the other takes time, and it depends on your business requirements. We're getting really prescriptive in helping people understand that there is different ways to adopt this.

The other aspect of it is helping drive a conversation around a data strategy. All the way back to the first question, "What is the role changing with data?" We need to not only think about the technology but the people and the process aspect of this whole thing… And make it an executive level conversation and drive that downwards: "Data is important to us for these reasons. We need to think about cultural shifts. We need to think about the business processes we're trying to impact. How is the technology going to be adopted?"

Put all those pieces together. If we have that thoughtful from the onset, everyone's going to be more successful in the end. We're going to win, the customer's going to win, and everyone's going to be happy. So that's where we're really focusing our efforts.

CI: Yeah. I liked the last slide. It was basically "There's one platform but many applications that can come out of that." Let's get a little sanity back into the architecture.

CP: Exactly.

CI: Unfortunately, we're out of time for this edition of the BBBT Podcast.

        Again, I'm Claudia Imhoff. It's been such a pleasure to speak with Clarke Patterson of Cloudera today. Thanks so much, Clarke.

CP: Thank you. It's been great to be here.

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!