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:	**DataHero**
Date recorded:	**August 27, 2014**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Chris Neumann,** CEO and Co-Founder
	Jeff Zabel, Chief Product Officer and Co-Founder
Run time:	**00:17:55**
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 Chris Neumann and Jeff Zabel. Chris is the CEO and co-founder and Jeff is the co-founder, also, and Chief Product Officer for DataHero. So, welcome to you both.

Chris Neumann: Thank you for having us.

Jeff Zabel: Thank you.

CI: Chris, let me start with you. DataHero's a relatively new company. It's about three years old. Why don't we start with just an overview of the company itself?

CN: Sure. As you said, DataHero's about three-years-old. We were founded in 2011. We're based in San Francisco, California. We are a new type of company doing self-service, cloud BI. Our goal is to empower business users to be able to work with, and get answers from, the data in the cloud that matters to them.

CI: Wonderful. All right. The cloud market has certainly taken off like a rocket. It's expanding tremendously in recent years.

I think it has forever changed that idea of everything has to be held tightly, has to be centralized on premises. Why do you think that the cloud has become so very popular?

CN: It's a great question. I think there's really two things that we're seeing today in the cloud market. The first, as you noted, is for the first time ever, we're seeing a net outflow of data from the walls of the organization.

Secondly, business users are able to pick and choose. It's really the great equalizer in the sense that business users across the organization can decide what services to use, and it doesn't always have to be the set that's pre-determined based on a traditional procurement process, where there's one piece of software for the whole organization.

CI: One of the things that we did talk about in terms of the great equalizer, too, is that it is for all sizes of companies. It's not just the very large ones or the very small ones. In fact, your software would work on any size, right?

CN: Absolutely. I think that's a really good point. Previously, we've always had this notion of software that's designed for enterprises and software that's designed for small businesses. But with the cloud, we're seeing that line go away. Business users in large organizations are leveraging cloud software that was designed for small businesses, and vice versa. You see small companies using Salesforce and large companies using software like SurveyMonkey.

CI: Yeah. I like it. I like that great equalizer being anybody, anytime, anywhere.

What do you say, though, to people who worry about the fracturing of the data, now that it's escaped the four walls of the organization? Some people worry, "Well, gee, I've got some of my BI capabilities in the cloud, I've got some on premises. It's fracturing my central data warehouse," if you will. What do you say to people like that who truly believe in the single version of the data?

CN: I don't think that we're losing the single version of the data from moving to the cloud. We're simply moving to multiple repositories. If you think about it, we have some data where the single version resides in an enterprise data warehouse, we have some data where the single version resides in Salesforce, or Dropbox or Marketo or whatever the cloud service may be.

As long as the BI tools are respecting those data sources as the single source of the data, I don't think we're going to see any significant shift in those models. We just need to make sure that we have the correct governance and administrative capabilities on the cloud services to match those that we have for traditional on-premises stores.

CI: What I like about that, I'm very glad to hear you say that, is that basically you're building the data warehouse for the cloud, as opposed to the data warehouse for on-premises. Yes?

CN: That's a great way of putting it. We're building the store where individual users can connect and leverage the governance and the administrative capabilities in services that IT has already blessed.

The IT department has set up these services. They've set up the various permissions and everything to make them comfortable with Salesforce, or Marketo, or whatever are being used by the business users. We're building a service that provides data warehouse-like capabilities while leveraging all of those good governance practices that are already there.

CI: Yeah, I like that. You're piggy-backing, basically, off of the IT security situation already.

CN: Exactly.

CI: All right. Well, let's change direction and get into DataHero a little bit. We're going to go over what you call the optimized data ingest framework for DataHero.

I really did like what I saw, especially the security part. You also have connectors and data replication, and so forth and so on. Why don't you talk a little bit about the ingest itself?

CN: Yeah, absolutely. As we said earlier, our view is that the cloud services are the single source of data and we need to respect that.

In traditional infrastructures, one of the questions we have is, if I pull data from a cloud service, I'll use Salesforce, again, as an example, into, whether it's on-premises or not, a larger data warehouse, what we lose is the permissions around that data.

We're trusting in the data warehouse or the BI team to ensure that the right people get the data. For DataHero, we're trying to empower any user to be able to get access to the data that matters to them and that they want to work with in a way that also respects that governance and that permissions model.

We've built our framework in a way that leverages the security practices of each one of the cloud services that we support. When a user connects to Salesforce, for example, they authenticate with their Salesforce credentials and they only see the exact data that the Salesforce administrator has given them access to, no more, no less. The same for Dropbox, the same for Marketo and any other product that they use.

When they're pulling data into DataHero, and they're starting to create visualizations and analytics and combined data, and so on and so forth, all of that is sitting on top of and respecting the existing security infrastructure so that that user is leveraging their data and their data alone.

CI: I think it's important to understand that you do pull the data. You do make a copy of their data and you bring it into your environment. That's the data replication piece of it, right?

CN: Yes, correct. For DataHero to work, we pull the data into our data warehouse, but that data is in a multi-tenant environment where each piece of data is built and respecting the security infrastructure of the various services that it comes from.

Then, when users are working with that data or updating that data, each time it's going back to the single source of the data and saying, what is my data look like now, what do I have access to, how has it changed, while always sitting on top of that general governance and security model.

CI: Now, one of the problems that we ran into early on with the data warehouse, of course, was the impact on the operational systems.

You are dealing with operational systems. All of these, Salesforce, and so forth, are operational systems. How do you minimize the impact that you might have on these very critical systems when you do your polling and your pulling of data?

CN: That's a great question. Every time we work with a new partner, we work with them quite closely to be able to take advantage of interfaces that are designed specifically for that.

Most cloud services, the interfaces that they provide to the outside world are transactional in nature, very similar to our transactional databases. As you alluded to, we saw similar issues with the emergence of big data and data warehouses in, how do we these complex, analytical workloads without impacting the transactional systems?

Some of our partners have APIs and interfaces that they provide access to publicly so that you can pull large volumes of data and you can synchronize data without impacting the transactional workload.

Others, we work very closely with the partners to co-develop interfaces so that we can get the data that we need access to with minimal impact to their ongoing, very valuable, transactional systems. In all cases, the connectors that we're building and the way in which we're leveraging our connection framework, is done so that we can get the data while respecting the security models, while also having the minimal invasive impact on the ongoing operations of our partners.

CI: Brilliant. All right, Jeff. Let's bring you into the conversation a little bit here. I saw some very interesting aspects to your software today, mostly, around your computation and recommendation engine. If you don't mind, you're the interface guy.

JZ: Great. Thank you. Yeah.

CI: Tell me a little bit about the interface and what's new, what's different from others?

JZ: I think the important part here is where we started, and where we started with the end user and going out to individuals and organizations and looking at how they used data, how that was changing and where they're moving to. On the first path of that was making it super simple to get the data into DataHero.

We're seeing data moving outside the company's walls and to services, so making it very easy to import that data and making that very apparent.

People have spreadsheets, files, et cetera, that they had access to and making sure it was super easy to get those into DataHero, and correctly

recognized. There's a lot of intricacies in the system on how it imports it, and what it does as it does with the data decoder and everything else.

That's really the foundation for everything. That was the important first step for us to get done. Then beyond that, we wanted to make sure it was simple for a user of any degree of expertise. We used drag and drop, you don't have to know code, you don't even have to know if something is a dimension or a measure. In fact, a lot of our users probably don't know the difference, but they do know the output they want at the end.

We tried to design that learning curve very easily so you can get to a state of, I think this is what I want. I'm seeing something. Maybe I want to see it in a different format, or all of a sudden, it sparks a new question, or I want to remove some information, and then, that aspect of the interface should then be apparent. OK, how do I change this and how do I get to the answer I'm looking for.

CI: What I found incredibly good about your software is something I have asked BI vendors for years to do and that's to bring a human factors person in. It's obvious to me that you did. That to me is such a key to how very pretty your interface is, how easy it is to use. I can do something in about two or three clicks that takes other BI tools -- I don't even know how many -- but many more than two or three. How did you come to that wonderful brilliance there?

JZ: A lot of user testing. Like we said, we started very early on testing with end users and watching, observing them. We hold constant interaction with all our users throughout the course of them using the product actively, or people we interface with to get more insights from.

We do that in a lot of ways. We'll track them, obviously, analytically to see what they're doing. We research and analyze that. We just observe. Watching five users tells us sometimes more than any of our analytics can underneath. We ask a lot of questions about how it should work.

The interesting part about data, at least for me, was what a complex product it is underneath. There's so many products that there's a couple key insights, a couple key things you can do. Data's almost a double-edged sword. There's a lot of things where we'll make something

Page 7 of 10

very easy, and then all of a sudden, realize, oh my gosh, we've broken five other use cases.

Continually putting that together, redesigning it, putting everything back on the tables is a way that I think we've, over time, have slowly progressed into something that is an easy to use business intelligence product.

CI: One of the best ones that I've seen was the recommendation engine. A lot of BI tools have that. They have some form of that, let's say. If you don't mind, tell our audience what your recommendation, if you can, it's hard to give words to a visualization, but if you don't mind, talk a little bit about your recommendation engine and what it does for the more technologically naïve user.

JZ: Actually, a lot of our product sits on top of that and it works across all aspects of it. Everywhere from the data import stage, all the way through actually producing that end visualization.

As you ingest the data, we're learning from the type of data that you're ingesting, the cardinality on individual columns, the column names, headers, what else not, and that's influencing the next import that occurs.

We're learning from, did you correct something that we thought was a state to something else. We're learning from, did you say these things are currencies, and is there something that we should have recognized if there wasn't a currency symbol in front that these in fact are currencies.

By learning from that, we start to make better suggestions down the line so the next step, of course, is we realize that lots of people bring in data and they didn't even know what they wanted. They said, I have data, give me answers. Immediately, our product gives you suggested charts.

We do those based on a number of factors, the type of data that you're bringing in, the service it's from, if we know that, basically, all these factors correlate together to come up with the best visualizations of the things we think you might be asking. That really helps guide the user that might start and say, oh that's really interesting. I'm going to start there. Then, it's just a couple clicks away to get to the actual question they had or to something else.

We use our recommendation engine throughout our product and it's constantly learning. That's the next piece to it is that as we learn more about what you like and your personal preferences, what other people that are doing with data like yours or in other instances to you, it's helping us continually tune that. It's one of those things that gets better every day.

CI: It really does. One of the things that I think is a tremendous advantage to a cloud-based implementation is that you do see every move that a business user makes, every click, every report that they generate, every visualization.

I think that's probably a huge advantage to the on-premises kinds of situations where they're completely blind to how the users are using their software. I wondered, do you actually -- do you use your own dog food here -- do you actually analyze what your business users are doing and perhaps make recommendations for things that they A, may or may not be doing efficiently, or don't even know exist?

JZ: Definitely. We do use the analytics on both sides. One, to improve our product and watch how far do people get, what are the good aspects of our interface and where are they understanding something, where are they getting stuck perhaps? Then use our one-on-ones to start to drill down at those areas and start to target them.

We also use it, of course, to enhance the experience for not only you when you come back again -- do you always want to see things via a bar chart or that are a single dimension, a single measure, or do you like to look at a pie chart?

We also have information based on services and what you're doing to know, well, if you're using Salesforce and you're a sales guy, we will show you a pie chart and it might not be the Tofta approach, but it's what our users in that domain typically like to see.

CI: Excellent. I think that's a wonderful position that you are in. I see a bright future.

You do have a freemium product, right?

CN: Yes. I think we talk about the accessibility and the aspect for any business user, and that's probably the number one thing where we start is that it is free, but more importantly, it's online. It's in the cloud, you sign up, you immediately get access. There's not someone in the middle where you need to download something, install something, you don't need an IT guy in the middle of it. You sign up, you get an email, you log in, and within seconds you're importing data. Our hope is that within a couple seconds later, you're getting answers from it via the visualizations.

CI: I sure hope so.

All right. That's it for this edition of the BBBT podcast. Unfortunately, we are out of time.

Again, I'm Claudia Imhoff, and it's been a great pleasure to speak to both of you, Chris Neumann and Jeff Zabel of DataHero today. Thanks for speaking with me.

CN: Thank you.

JZ: 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!