



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:	Pentaho
Date recorded:	September 5, 2013
Host:	Claudia Imhoff , President, BBBT
Guest(s):	Jake Cornelius , Senior VP, Products Evangelist Donna Pritch , Senior Director, Product Marketing
Run time:	00:19:54
Audio link:	Pentaho Podcast
Transcript:	[See next page]
Transcribed By:	CastingWords



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.

CI: I'm pleased to introduce my guests today. They are Donna Prlich and Jake Cornelius. Donna is the Senior Director of Product Marketing, and Jake is the Senior Vice President of Products for Pentaho. Welcome to you both!

Donna Prlich: Thanks, Claudia, thanks so much for having us today. We're excited to be here.

Jake Cornelius: Yes, thank you, Claudia, we're really excited to be back at the BBBT, updating you on Pentaho version 5.

CI: It's wonderful to hear your voices, as well. All right. Well, let me start with you, Donna, a little bit. Pentaho has had a pretty remarkable last year, I'd say. You gave us some NDA material, but, I guess bottom line I'd like to know, what do you attribute all of this growth and new customers and so forth, what do you attribute that to?

DP: Yeah, that's a good question, Claudia. It has been quite a year for Pentaho. But yeah, if I really think about it, what's really been driving our growth is Pentaho's been in this business analytic space for some time, but it's really the growth of big data in that market. Our customer base has just grown so significantly and we're addressing so many more big enterprise customers, and really kind of helping to solve some of the difficulties of bringing some of these new technologies in. Big data is really pushing a lot of that growth.

Embedded, too. So a lot of our embedded business and the need for the ability to sort of get analytics in everywhere and anyhow. That's been another key point.



CI: Yeah, it's interesting. You're certainly in good company. Big data has just pushed everybody along. To that point, a very exciting thing that we're now able to talk about is the announcement on September 12th of the new version of Pentaho, 5.0. Donna, I'm going to go to you again. There were some significant drivers behind this version, certainly, big data being one of them. But why don't you tell me about the other drivers?

DP: Absolutely. We really have seen that. The two things that we saw, along with big data, the other one was this concept of being able to address the broad set of users that might be involved in a business analytics deployment. What we did is we stepped back and we really spent some time looking at the user experience and what's involved there. A lot of what you'll see built into this 5.0 platform is a simplified analytics experience where, whether I'm an administrator or a business end user who just wants to create a dashboard, my experience with the product and being there is really simplified for what I need to do and being able to access the data I need to access, the tools I need to access, in a really familiar, comfortable environment.

CI: Let's dive a little bit deeper into 5.0. It's been an exciting time. I learned so much today about what you're doing. Jake, let me turn to you about this. You said that it was architected for the future, and you gave us four significant areas that you focused on. Tell me about these four areas in a little more detail.

JC: Absolutely. The first one is beyond just being able to connect to big data, which are table stakes in today's big data integration world, really delivering the notion of blended big data, for either analytic use cases or even ETL use cases. With this release, we extended upon some of our previous investments by being able to take Pentaho data integration as a platform that's ideal for taking different formats of data, different volumes of data, and architecting how those can come together to answer a question that may span those different areas.

We're really excited to present in this release a way where end users, or, let's say, content creators, can use the language of data, which is SQL, directly against blended datasets facilitated by the data integration platform.



That's one really exciting area, that notion of blending big data. The second, as Donna mentioned, is the profile of Pentaho customers is changing every week. We're seeing more and more large enterprises adopt and roll out Pentaho at scale.

A lot of the enhancements in this 5.0 release are really to tease out the specific user audiences within large organizations and provide a platform that lets the BI administrator set up the right set of capabilities for the right users, and then, again, focusing on that end user's productivity and the ease of use for them adopting Pentaho for their analytics needs.

Thirdly was further evolution in our OEM embedding capabilities. Really excited, this has always been a strength of Pentaho's as an open source platform. But in this release, we took a step back and we made major upgrades to the architecture of the business analytics platform to deliver that same level of ease of use for developers.

That comes in the form of a single set of simple REST services, and a wealth of out of the box documentation examples that help those developers get started with embedding better data integration or better analytics into their host applications.

Finally, an ongoing theme for Pentaho is our innovation in big data integration and big data analytics. With this release, we've dramatically expanded our platform support for new Hadoop distributions.

We've driven up the ease of use for leveraging NoSQL platforms, either for ETL or business analytics purposes. We've also introduced some new big data platform support, probably most notably integration with Splunk for pushing and pulling data into the Splunk platform.

CI: Yeah, these are terrific and they're spot on, aren't they, for the future? Yeah, there are also some really interesting new features, and I want to spend the rest of our time together on those features. Let's drill into those in a bit more detail. The first one that I thought was quite interesting is, you mentioned blending at the glass. Again, it's that data blending idea. What are the new features that support this idea?



JC: We break the data blending into two categories. When we talk about blending at the glass, we see a number of products on the market that talk about empowering end users to blend data. There are use cases where that makes sense -- if the data is naturally conforming, if there're common actions you can join on. But at the same time, there're risks associated with that. If the data is not of the same quality, or if, let's say there's a difference in granularity, then sometimes putting that power in the hands of end users can result in inaccurate results or unintended results. One of the big drivers in this release that I'd mention is that notion of an architected blend, or blending with big data, and even non big data data sources.

We feel very strongly that Pentaho data integration is one of the best platforms available on the market today because of its breadth and the type of data sources that it can access as well as the ease of use in the design environment. To accommodate for some of those differences that need to be accounted for when you're taking data, let's say, from a documented oriented store NoSQL, but trying to answer a question that leveraged that information along with maybe some data from a social media outlet or from my enterprise data warehouse.

In Pentaho, data integration makes it really easy to describe and account for how that data gets blended. Then in this release, we then expose that to a much broader audience by letting them use a simple JDBC driver and SQL queries against that ETL engine at run time. It really simplifies, then, the delivery -- lets you create accurate blended results and then push those results out to a really broad audience.

CI: It certainly is a distinctive feature of this new version, as well. It's something that kind of sets you apart a bit, doesn't it?

JC: It does. It's a concept that we've been kicking around for about 18 months. Would this even be possible? It's definitely one of those areas that, architecturally, having that data machine, that 100 percent Java based, embeddable data machine, empowers these really interesting use cases and so we're really excited. It's something that we haven't seen in the market. We think it's going to be really beneficial as IT organizations change and expand and set about building their next generation of



application infrastructure. We see time and time again these use cases where they still need to ask questions and have answers against data that blends information from all these multiple different data sources.

CI: Yeah, it's really nice. Well, another feature, or another area that you spent a great deal of effort in is the user console features. Let's talk about those, touch on them at a high level, but you have a whole bunch of new features there.

JC: We really do. We really do. We hate to come to market with, "Oh, it's super real, it's really easy. It's really simple." We don't treat that as a headlining feature, but you're absolutely right. From the ground up, with this release, what we looked at is particularly, again, back to that enterprise audience, looking at the profile of users within the organization, looking at the common workflows, whether that's, "How do I deploy the system and manage it?" to, as an end user, "How do I go through common workflows like I have a work and I want to make sure that it's scheduled and delivered on a weekly basis?" And then diving deep on the usability of all those workflows.

From your first impression of Pentaho 5.0 you'll see a fresh, new, modern look and feel, but then as you use the product you'll find that all those common paths have been carefully thought through to make sure that users get the highest amount of productivity, the highest amount of ease of use, and really lessen that burden for new users coming into the system, maybe using Pentaho for the first time.

CI: One of the features really caught my eye. I've been doing a lot of writing and talking about self-service BI. One of the things I have insisted upon is that IT can't control the assets that their business users use. We've never been able to control the information assets. But IT does need to at least be able to monitor what assets are being used. Who is using them, how often are they using them, and, more importantly, for what purposes. If they're using, for example, an ungoverned source, like some kind of Hadoop database, or whatever it is, for a compliance report, boy, that ought to send up red flags all over the place.



One of the things that caught my attention this morning was the fact that you have added the ability to really monitor the usage of the environment, of the different sources of data. Where did they come from, who blended them together, and ultimately what BI component did they go into? I love that. Explain a little bit more about why you guys came up with that.

JC: I'm really glad you keyed in on that. As a business analytics company we do tend to focus a lot of our time and effort around that business analyst or that information consumer. But Pentaho IT has always had a fondness and a love for IT and administrators. They're often times the folks that are downloading Pentaho, and installing it, and testing it, and see if it's the right fit for their organization. As we were developing 5.0 and really focusing on those user profiles, we did say, "What can we do for that administrator to make his life easier in that deployment administration of the console?", and "Why not, as a business analytics company, take advantage of the same power that we put in the hands of those end users, and allow those administrators to take advantage of that for doing their work as stewards of the system?"

You're absolutely right. We now leverage our full analytics and reporting capabilities. We provide them with nice out of the box reports and dashboards to answer those common question they might want to know about what content's being accessed. Is there legacy content that needs to be cleaned up? Where might there be performance bottlenecks, or even errors taking place in the system?

Great out of the box capabilities built on the business analytics platform, but also that same level of ease of use for doing self-service, ad hoc querying against all that information being captured about usage patterns within the system. Now they can also go in there, start from scratch, ask their own new and interesting questions about activities on the business analytics server.

CI: Wonderful. Finally IT has some analytics.

Let's turn our attention to Pentaho Data Integration 5.0. Again, good features in that. Talk to me about them.



JC: Yeah, really big release for Pentaho, Data Integration 5.0. In Data Integration 5.0 alone there's over a 100 new features. I couldn't possibly describe all of them. We touched on one of the exciting things, which is this notion of a SQL access to a blended data source facilitated by Pentaho Data Integration. Beyond that a lot of the top level features within Pentaho Data Integration are reflective of our move into large enterprises.

We're seeing, more often than in the past, Pentaho Data Integration being deployed as a centralized service for large scale ETL operations. Some of the top level features include things like being able to set checkpoints within long running jobs. If a job fails, let's say it's a six hour back processing job, and three hours into it I've successfully gone through half of the process but it fails, being able to very easily go in there and restart that job from the last successful checkpoint.

Similarly, maybe controlling the behavior such that, if I didn't want a failed job to remain in that broken state, automatically being able to roll back very sophisticated job workflows back to the beginning if a failure does occur.

In addition to those, there are numerous improvements on the performance side. To be able to take advantage of your available compute resources for ETL processes. Then, again, focused on that centralized deployment of ETL as a service, we've enhanced the security model by providing greater control over different user profiles. What they can perform, what access to different common connections they have, and what they can do with those connections.

We're really excited about the features in there, particularly those that address some of the emerging needs we see in our enterprise customer base.

CI: It certainly was something that I think is very exciting, as well. The last one I want to touch on is what you're doing for your OEM vendors. Tell me a little bit about what's new there.

JC: You know we're very passionate about open source. That being in our DNA, we've always looked at ways, "How can we leverage this great



platform of technology to build custom solutions, to help others enrich their applications with better analytics, better data processing capabilities under the hood?" With this release, again, major progress by simplifying the design experience for application developers. We've rolled out a whole new set of REST based services. It's kind of an upgrade from our previous URL and SOAP based services to a very clean REST interface to the BA platform.

In addition to that, we went that extra mile to include out of the box integration samples and test harnesses around that API that help new developers working with developing platforms come up to speed very quickly with live interactive demos leveraging those new REST services.

CI: Very briefly, let me go back to Donna. I see a tremendously bright future for Pentaho. I think you guys have put together a wonderful package of features in 5.0. Donna, what do you see now in the future for Pentaho?

DP: Thanks, we're really excited about the future for Pentaho. 5.0 is really, I think, going to just allow us to continue to address a lot of the complexity and the tumultuousness of organizations trying to bring big data in. We see that that's going to be a big growth for us. The exciting thing there is that we're already so involved early on with a lot of these early deployments without big data customers. We see that as it's going to be an exciting time to help them through this process and to help evolve these applications where people are going to really start to see value from big data. We see that as a really exciting thing to just keep extending out this ability for end users.

As Jake said, we're always going to have a fondness for IT. Just really being able to address the needs, whether it's developers who need help dealing with things like MapReduce or administrators trying to manage a growing set of users, delivering blended analytics to end users so they can get to data faster. We're excited about having this platform that's just been around for a long time, purpose built, open source, ready for this new world. We're excited about that.

CI: To put a little icing on the cake, you've got Pentaho Labs, which is just a fascinating, innovative area of the company as well.



DP: I was just going to say, Jake spent a lot of time on that today, but it's exciting because that really is the source of a lot of what you're going to see from Pentaho over the next probably 6 to 12 months.

CI: All right. I think we could talk about that for another five minutes, but unfortunately, we are out of time. 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 Donna Prlich and Jake Cornelius of Pentaho today. Thanks to both of you for speaking with me.

JC: Thank you so much, Claudia. It's been a pleasure reconnecting and updating you on Pentaho 5.

DP: Thank you, Claudia.

CI: I hope you enjoyed today's podcast. You'll find more podcasts from other vendors at our web site, www.boulderbibraintrust.org. If you want to learn 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!