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:	**Denodo Technology**
Date recorded:	**December 5, 2014**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Paul Moxon**, Senior Director of Product Management
Run time:	**00:17:37**
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 friend today. He is Paul Moxon. Paul is the Senior Director of Product Management for Denodo Technology so welcome, Paul.

Paul Moxon: Thanks. Nice to be here again.

CI: Lovely to have you here in Boulder. Let's start off with something that was said early on in the event this morning, and that was that now is the time for data virtualization. It just seems like peoples' view, opinion, whatever of data virtualization is changing. Why don't you tell me a little bit about that?

PM: Sure. The old view, the original view, of data virtualization was very much based on the data federation technology we had 5, 10 years ago. People are now realizing, somewhat like the service-oriented architecture concepts, that abstraction is very valuable in itself. You're hiding the technology from the consuming applications.

If you think about the rate of change of technology -- whether it's big data, Internet of things, streaming data is the next wave of things -- people, IT departments, are struggling to keep up with how do I integrate this new technology.

If you hide it behind the abstraction layer, the data virtualization layer, it becomes a lot easier. It's given the agility and stability that IT departments need in what is now a very exciting and very rapidly changing time for data integration.

Data integration used to be the orphaned step-child that nobody spoke about. Everyone was on about SOA, but now data integration, big data, Internet of things, have put it in the center, and IT departments are struggling to be agile as demanded by the business.

CI: It just removes the whole complexity of having to interface with the myriad of technologies that are out there. The go-to-market strategy has also changed. It's become much more mature than it was even a year ago. It's amazing how you've changed. Why don't we talk about that a little bit? What is the market strategy?

As an adjunct to that, we have to bring in data consolidation or ETL, the physicalization as opposed to the virtualization. I do want to hear your thoughts on where that stands today, as well.

PM: Let's address that one first. That's the one most people are probably waiting to hear. We believe that data virtualization is part of your information architecture. It is a key component and sometimes the center of that architecture, but it is not the only component in the architecture. There will still be the need for ETL. You're still going to want to consolidate results.

Historical information is a prime example. If you're doing regulatory reporting, you're going to need to reproduce those reports in a year to, I think up to, seven years' time or something for the SEC. You have to have a consolidation of that historical type of data so there will always be that need.

There are also the times that you want to apply some quality processes to the data and you don't want to do that every time you make a query. Sometimes you will consolidate that data. We have customers who are doing a combination of data virtualization and consolidation in the same information architecture. That's quite common, and we just accept that's going to happen.

CI: The go-to-marketing strategy, then?

PM: Oh yes, that bit. We actually have four prongs to our go-to-market strategy. Part of it is the product. We believe we've got the best product,

and it's evolving, it's innovative, it's maturing. We're going to continue working on that.

In the post-sale side of things, it's that engagement with the customers, especially new territories. We're seeing emerging territories like Australia, Latin America, Africa, as well. Working with the early adopters there, making them successful, and leveraging their success to expand in those markets.

Partners. Denodo is not an IBM. We don't have an office in every city around the world. We use a partner network so that we can spread the knowledge, spread the message. Also, through Denodo Express we give people the ability to actually access a data virtualization platform and use it themselves, educate themselves, try it themselves. They're the main prongs that we have.

CI: Excellent, alright let talk about... There is still a little bit of confusion about when and where do I use virtualization versus ETL or data consolidation, or even just plain, old data replication. Why don't we talk a little bit about the usage patterns for data virtualization, and, if you don't mind, sprinkle in an occasional customer example of a usage pattern. That would be wonderful.

PM: We've looked quite extensively at usage patterns for data virtualization. Traditionally, people used to think of it for the agile BI reporting use cases, projects. We're now seeing people moving beyond that. What we call broad spectrum data virtualization, to give its buzzword.

We see it as categories of analytical and informational, which is the traditional agile BI, but also includes big data. How people are integrating their big data sources, their Hadoop clusters, with their other enterprise data -- operational data, historical data, et cetera. That's one side of it, and we've got customers doing quite a lot of that.

Caterpillar is one example that we talked about earlier. Integrating information they're storing in Hadoop with internal data sources and then feeding that back out to customers so that they can do predictive maintenance on their equipment.

We're also seeing operational use cases, operational data sources. A classic example of this is a single view of a customer within a call center. One of our customers, division of France Telecom, has got a call center with 3,000 call center agents, and we're feeding up customer information from multiple data sources -- billing systems, their support systems, instant management systems, CRM systems -- to the call center agents so they have one, single view of the customer in front of them when they're on the call.

That's actually reducing call times by about 60 percent, just the fact the information's there and the call center agent isn't doing the famous swivel-chair integration.

CI: Yes, indeed.

PM: They're good examples of how we're seeing the scope of data virtualization moving away from the classical analytics and reporting into other use cases.

CI: Excellent and I'm sure there are others. People can find some of this information on your website, as well, I assume. You mentioned Denodo Express, and I was very happy to hear what it is and what people can do with it so why don't you explain what Denodo Express really is?

PM: Denodo Express is a fully-functioning data virtualization platform. It's free to download, it is free to use. You can use it for self-education, for pilot projects, for POCs, even for small, departmental projects. It is not enterprise scale. We're still a commercial company. We'd like to sell you something. There are some limitations on scalability in terms of the number of rows in the result set or no clustering support, et cetera.

It's fully supported through a community, and we've got a very active community where existing customers are actually responding to questions coming in the community. It's not just us feeding the answers.

The reason we actually came up with Denodo Express was we're finding that there's so much noise and hype, in some ways, about data virtualization. People don't really know what it is, what it does, and how applicable it is for them.

CI: They tend to get stuck in their ways, I think. That's why I think this is so brilliant. They're so used to, perhaps, using ETL technology that they don't even see a need for virtualization.

PM: That is very true. That is usually one of the nuts we have to crack, and hopefully Denodo Express will make this easier. Also, people hear about data virtualization, and is it right for me? We're spending a lot of time in, actually, the sales cycle educating customers. We think let them self-educate. They prefer that. They prefer to find out by themselves rather than being fed by a sales person or a sales engineer.

They can do this, they can download it, they can play with it, they can see if it's right for them. Like I said, pilot projects, small, departmental projects. Community projects. We actually have one of our customers, one of the engineers, is actually doing something for his church. He's taken Denodo Express and he has built a little project, a little data integration project for his church.

That's free of charge. He can do that as much as he likes.

CI: Perfect.

You are a commercial enterprise so let's talk about the other side of the coin, and that's the Denodo platform. What's new, what databases can you hook it up to, and so forth?

PM: The databases, the data source, are forever growing as people come out with new technologies. The latest thing that we've been touching on is streaming data -- IBM streams, Apache Kafka, et cetera. We've got various connections to Hadoop, et cetera. This is almost like table stakes. It's what you have to do.

CI: Almost every database, every version of that database, and so on.

PM: We're just adding new ones when they come up and customers request them, come up in sales situations. Similarly, optimizations, we've done lots of work on query optimization. That, again, is something that's ongoing. You'll never, ever finish optimization. There'll always be another technique, another data source.

One of the biggest things we've changed in the Denodo platform, Denodo 5.5 which was released just a month or so ago, is what we call our global search capability. This allows people from outside the Denodo platform, people who don't have access to the design and development tooling, to actually browse and search for data assets.

Look at what data is available through the data virtualization platform, look at the metadata, look at the schema of a table, even go and look at the data themselves to make sure it's the data they want before they start building their reports or building their applications.

We provide a form of this, a browser-based form, but, more importantly, we provide the APIs that customers can build their own search capabilities on their own catalog search, their own data search capabilities, and feed that to their internal staff, usually through an enterprise-wide portal or something like that.

CI: Let's dive in a little bit more on the optimization because, obviously, one of the fears that people would have about data virtualization is, if I hook it up to my operational environment, am I going to degrade its performance? Now, I've got these unusual queries, unusual users, I don't know what they're doing. They might actually pull the performance down, and I can't have that in my operational environment.

Tell me a little bit about the optimization techniques. Also, you showed us a slide that did a bunch of benchmarking about, if you add on Denodo, what is that going to do to performance?

PM: Yeah, optimization, like I said before, is this never-ending task. We have various levels -- various techniques, should I say -- of optimization. Query rewriting is a big one. If we can find a more optimal way of doing a particular operation we will rewrite the query on the fly behind the scenes. Our development team has got PhDs who know all about relational algebra, and they baffle me whenever I talk to them.

I asked for the idiot's guide to this, but they can't talk at that level, I'm afraid. They're the ones who come up with the optimization techniques. There's query rewriting.

Push down query delegation. As much as possible, push the query down to the underlying data sources. People like Oracle, Netezza, IBM with Netezza, et cetera, Microsoft with SQL Server, they spend millions and millions of dollars optimizing their particular databases. Even when you move to Hadoop, companies like Cloudera, Horton Works try and optimize their data sources, as well. We take advantage of their work by delegating as much as possible.

We also support things like caching. This is when you're talking about protecting operational sources. It's not just pushing all the information, all the work, down to them. It's protecting them, as well. We support things like caching. You can cache information so you're not hitting the operational data source all the time.

We have query throttling. You can actually say that I'm only going to allow so many queries or a query to have a certain lifetime before I kill it. I'm going to block all of the queries to protect the operational data sources. We have a lot of customers using operational data sources, and they wouldn't do it if they didn't feel they were protected.

CI: In terms of performance hits, you have done some work in terms of with Denodo and without Denodo. What's the bottom line there?

PM: We actually did some performance testing with IBM as part of our partnership with IBM. We're partnering with the Pure Data Systems division and, obviously, Netezza Pure Data for Analytics, as they now call it, is one of their primary products. They wanted to make sure if they entered this partnership that they wouldn't be introducing their partners to something that was not performing, should we say not performing.

CI: Detrimental.

PM: Detrimental. Thank you. Obviously, they've got a reputation to keep. They pride themselves on the performance of Netezza, of the Netezza appliances. We did some testing with them in their labs against Netezza and Big Insights. We found that having Denodo as an intermediary layer between an application and the actual underlying platform had minimal impact.

It was something like on a 90-second query which was querying over five billion records in their go sales data set, we had a 3-second performance impact over 90 seconds versus direct access to the underlying data source.

CI: Less than five percent.

PM: Yeah. We actually performed very well in that. In some cases we seemed to perform faster than the native Netezza so I'm not sure about those ones, but generally it was considered to be minimal impact. We've seen that from customers using data virtualization, the Denodo platform against their large data warehouses.

They're saying the performance impact is very small and doesn't actually get in the way of the fact that they can do so much more with data virtualization.

CI: Excellent benefit far outweighs any performance.

PM: Exactly.

CI: Interesting. We've talked about the strategy and the optimization, and we've talked about the impact. Let's end on the future. What is the future view for Denodo?

PM: We see we've got a very bright future. We hope, anyway.

CI: I would say so.

PM: Data virtualization is growing. It's becoming a mainstream technology now. Customers are adopting it. We're seeing customers coming back and expanding their footprint and use of data virtualization.

Also, as we move forward the agility it adds becomes very important. As new technology is emerging, in memory computing techniques, real-time streaming data, et cetera, cloud architectures, cloud computing, this makes the whole data virtualization layer more efficient, more performant.

All the concerns that people have about getting in the way and not being able to do what they can do with batch-driven ETL process, that

starts to diminish. It's never going to completely replace the ETL technologies, but we see it doing more and more of that work.

There will be a bigger gray area between "Do I use a traditional ETL or data quality tool?" versus "Can I do it in real time with data virtualization?" We just see its footprint, its growth, continuing. It's accelerating at a huge rate. Let's just see what happens.

CI: Life looks good.

PM: It looks very good.

CI: Unfortunately, we're out of time. That's it for this edition of the BBBT podcast. Again, I'm Claudia Imhoff, and it's been so much fun speaking to Paul Moxon of Denodo so thank you, Paul.

PM: Thank you for having me.

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!