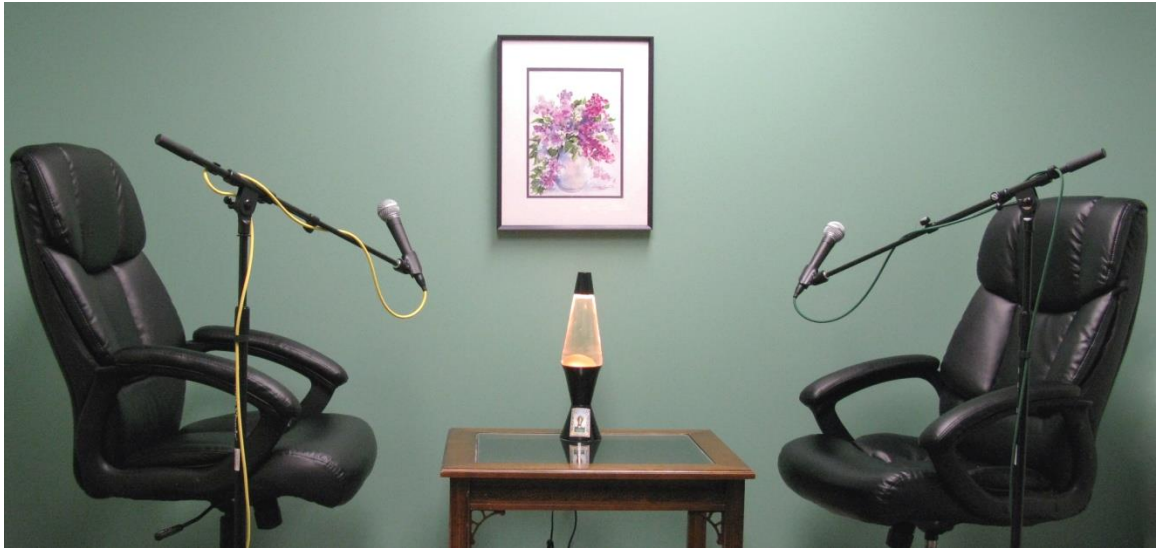




BBBT Podcast Transcript



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
Date recorded: December 13, 2013
Host: Claudia Imhoff, President, BBBT
Guest(s): Suresh Chandrasekaran, Senior VP, North America
Paul Moxon, Senior Director, Product Management



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 Suresh Chandrasekaran and Paul Moxon. Suresh is the Senior Vice President of North America and Paul is the Senior Director of Product Management for Denodo. So welcome to you both.

Suresh Chandrasekaran: Thank you, Claudia. It's pleasure to be here again. Thank you.

Paul Moxon: Thank you.

CI: Let's start with you, Suresh. You started off the BBB Team with a few messages from Denodo. Would you go over these for me, please?

Suresh Chandrasekaran: Sure. Denodo today is the only company focused exclusively on data virtualization. We have a strong commitment to it. We also came out here to emphasize that we have a broader view. We call it the, "Broad Spectrum Data Virtualization."

What we mean by that is that we are trying to, and we have always wanted to expose all information whether its data structured, unstructured, internal, external, created enterprise, or outside. And take that, many sources and expose it to as many users as possible so at the top end it's not only BI and Analytics but it's also application development, migration, mobile cloud, etc.

Our key message today was that Denodo, being exclusively focused on data virtualization, and the product differentiation, the customer deployments, and the go to market partnerships that are developing, to make broad spectrum data virtualization overtake the world and be a good thing.



CI: Yeah, you've really broadened out the definition. You've really brought many new capabilities into it. Paul, let me turn to you because that leads into my next question of what exactly is data virtualization today? Then you gave us a definition. If you don't mind, go over that again.

PM: Sure. We simplify it basically down to three words, connect, combine and publish. So that's connect and virtualize your data sources. Combine and integrate them together to produce your canonical business entities, and then publish it in the way that consumers want it, whether this is SQL for your BI tools, et cetera, or web services, restful SOAP, et cetera, for mobile apps, integration middleware, and so on.

CI: All right, and what were the key tenants then behind this definition?

PM: I think we gave 4 key tenants. One was, realize all of the value from all your data, so this is structured, unstructured, it's stuff outside your traditional databases, internal data and external data, so going out to the Web as well. Integrating things like Hadoop and realizing the value from your big data, your no SQL, as well as the more traditional enterprise data sources, that's one.

It was minimize replication. In other words, this is not a consolidation strategy, this is about accessing the data from the source, where it sits. Leave it there rather consolidating to a very expensive data warehouse.

Then we go on to services, data services. Create a data services layer to give a uniform and consistent access to the data in these canonical business entities, that I mentioned previously, through whatever means is needed by the consumers. If they want to access it at a SQL, they can do so. They can access data through web servers, or whatever. So that's the data services layer.

The final one was enterprise readiness. Many of our customer deployments deal with very large data sets, large data warehouses, big data, et cetera. Being able to perform at this scale is what's necessary. Data virtualization platform needs to have those four key attributes.

CI: Wonderful! Let me stay with you for a second. And let's talk about some use cases. You gave us six that were very clear use cases. Some people



get very confused about when do I virtualize and when do I physicalize. Your use cases brought it all home and they were pretty darn good. So why don't we talk about that a little bit.

PM: Is physicalize really a word?

CI: Well, it is now.

PM: It is an officially sanctioned BBBT word.

CI: Physicalize.

PM: There were six patterns, data virtualization patterns, that we have identified, and these are very broad patterns. There are sub patterns into some of these, Agile BI was one. And that's one of the most talked about patterns we'd see. Virtual MDM and that has a number of sub patterns about how you integrate with an MDM solution. Enterprise data warehouse extensions, this is extending the data warehouse, extending the ETL underlying that. Again, a large number of sub patterns underneath that.

Then we go on to some of the less common ones, less familiar ones. We talk about single view of the customer, single view of an entity, really it could be products, as well as customer. More recently, we're seeing the adoption of Cloud and Big Data Integration, through Data Virtualization, bringing that big data from the data science silo into the enterprise. The final one is the data services layer and again, like I said, all of these have various sub patterns underneath them.

SC: Claudia, I'd like to add something to that. So these patterns, six, are not fixed in time, they can be fluid. And as Paul was saying, I'll just expand on a couple of them. Even if you think about big data and the use of data virtualization with big data there are two or three sub patterns, one, for example, that you see the climate corporation taking, is the idea of taking very complex observational and sensor base data for weather, which they get 30 terabytes a month and do analysis on it for weather crop insurance.

Exposing that and combining that with business systems like Salesforce, NetSweep, et cetera. Here you have more of a democratization or SQLification of big data as a pattern. In other examples, like Telcos we see



the use of big data as a low cost way of storing data. For example for call records, for compliance, and then providing a data virtualization layer on top to make it very easy to access at the time of need.

If you go into another example or a broad pattern like application, modernization, and migration, we have customers like a large railway company that's migrating from legacy systems to modern applications for their whole train operating system.

We have companies migrating from, let's say, an Oracle or a PeopleSoft to a lot of cloud HR providers and yet retaining unified view of HR using data virtualization or you have yet another type of migration modernization, which is when companies merge or split, to use data virtualization as an abstraction layer. My point is that the patterns are evolving and our role is to help capture and present them back to the customers as solution use cases so they can leverage data virtualization across multiple different ways.

CI: Paul, let me go back to you very briefly if you don't mind because we're running out of time. Just talk about the architecture in terms of the three broad areas that your architecture covers.

PM: Just in the interest of time, I'll talk about five.

CI: Add two more.

PM: It's somewhat similar to what we talked about before, there is the connect layer. This is where you connect the various data sources through the connect or wrap as this we sometimes call it.

CI: You connect almost everything.

PM: Yes, we've very, very broad range of data source that we can really connect running from the traditional database, data warehouses, through the no SQL, Hadoop styles, all the way out to unstructured web style data as well, so broad spectrum.

Then we have the connection layer. That's the tool we saw in the demo, the ability to combine and integrate the various components, the various views that you've accessed. That you've created from the online data



sources and then publish them in the protocol and form that's required by consumers. So that's the traditional three layers. But then there is also a performance aspect to it. How do you guarantee performance? Through caching, and query optimization.

The fifth one is the governance, which is the critically important in any data architect. How do you ensure the veracity of the data? How do you know its data lineage? How do you know how that data is transformed between the source and the actual consumer? In the government security side of things, we provide a lot of capabilities there as well.

CI: We'll get to it, but that's a big differentiate for you folks as well. I really like the governance piece in particular.

PM: Yeah, the ability to see where the data came from, how it changed, where it's being used, whose consuming it, where are base views or drive views being used elsewhere within your system, is incredibly powerful. We see a lot of our customers liking that.

CI: Yeah, I would too. All right, Suresh let me go back to you. Part of the difficulty with Data Virtualization versus data consolidation or ETL processing, many people, many of my customers, many of the people in my classes confuse the two and say, "Why do I need one, when I got the other?" They are not the same though. You made a good point of differentiating from ETL. If you don't mind, again somewhat briefly, because we're running out of time, but do differentiate between data virtualization and ETL, the Extract, Transform and Load Processing.

SC: Sure. The first idea of data virtualization is regardless of whether you want to consolidate the data, or the federate it, or access it in real time, the idea of abstraction and de coupling is still valid. We often have ETL, which is batch accessing of virtualized data object and then ETL-ing it, if you will.

But if you come down to the fundamental differences what we're saying is not to eliminate ETL, but to minimize replication you need to on a case by case basis see which data needs accessed in real time or can be, which data needs to be cached, whether it's in a disk space or in memory caching, and which data needs to be moved.



Some of the integration functions are the same, the cleansing function are the same. Obviously, if you have multi pass you would want to lean towards ETL, because it is very process intensive and may not be suitable for real time.

We're really talking about a hybrid execution model of real time query delegation and optimization, caching and batch oriented operations. But over and above that we're talking about data virtualization being different in that is providing an abstractive view of your original data sources, and that's the key difference.

CI: All right, let's talk about some of the reasons then why companies would select Denodo over one of its competitors? Let me stay with you Suresh for this one, for a moment, and Paul, I welcome your comments as well. Why would a company choose Denodo?

SC: There are, today, maybe two or three evolved data virtualization products in the market but as a company, Denodo is the only one that's focused on it a 100 percent.

With current projects and projects in Agile BI, or Logical Data warehouse, or Virtual MDN, things like that, most of us have the functionality to support those current use cases. So the differentiation really comes from performance scalability and security. We're typically an intense POCs, we're pulling away in some of these areas. We win the largest ship maker, the largest pharma company. It's because we proved those technologies.

There are some key differences in how we do it. We have an extended relational model. We use a better query delegation and optimization. We have a very granular security model. Those are for the current reasons.

But very often, I talked about broad spectrum data virtualization that means companies are not just looking to solve today's problems, they're looking to the future, the nexus of forces, like mobile and cloud and big data and social media. They find that Denodo has the most flexibility, both on the source side, as we talked about, as well as the different ways interacting with the data.



We support not only SQL and Rest and SOAP and JSON and Portlets. We also support multiple interaction model such as starting with the search, then moving to a browse paradigm, then moving to query and analysis. Those features are not available to other federation products. That's a key reason.

CI: What's important is that you've done a really good job of putting your finger right on the pulse, on where we're going with this stuff.

SC: Right, and in fact if I were to add another key reason, is that since unlike other companies we are 100 percent focused on data virtualization. We are very interested in capturing more, and we talked about use patterns, and enabling our customers to capture and use more use patterns. We're constantly doing that feedback loop of what customers are doing and cross pollinating that, but also putting that back to the product. A lot of these features I talked about actually come from customers wanting to use data virtualization and new patterns to meet future challenges.

CI: Speaking of your customers then Suresh, why don't you quickly in about a minute or so, give me a couple of examples of how your customers are using your technology today?

SC: I'll go from a couple geographies, in Europe for example, we have some several large Telcos, Telefonica, Vodafone, rCable, JasTel. Also, some insurance companies like MAPFRE and banks that are using it, and I'll talk about rCable, using it to provide a single view of customer, they started off with data services being layer that powers their self-service customer portal and then they found that those data services could be reused for their call center and that they could be reused for externalizing it to partners. In that case we're actually being used on operational sense with the BPM layers sitting on top.

Coming over to state side, we have large companies, in the financial sectors, like Oppenheimer Funds or Nationwide Insurance. Oppenheimer Funds uses it as a data service layer to feed five or six different application development teams that consume the data services. Some of those teams are doing BI and analytics while others are building web applications using RESTful interfaces but the logical view of the data is the same in both cases.



Those are two and we have also other patterns big data use, for cloud integration, for migration, as I've talked about before.

CI: Just go to your website, I'm sure people can see them. Unfortunately we're out of time. So 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 Suresh Chandrasekaran and Paul Moxon of Denodo today, so thanks to both of you.

SC: Thank you.

PM: Thank you.

CI: I hope you enjoyed today's podcast. You'll find more podcasts from other vendors on our web site. That's www.boulderbibraintrust.org. 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!