



## BBBT Podcast Transcript



### About the BBT

The Boulder Business Intelligence Brain Trust, or BBT, 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 BBT provides a variety of services, centered around vendor presentations.

For more, see [www.bbt.us](http://www.bbt.us)

<b>Vendor:</b>	<b>Denodo</b>
<b>Date recorded:</b>	<b>Feb. 26, 2016</b>
<b>Host:</b>	<b>Claudia Imhoff</b> , Founder, BBT
<b>Guest(s):</b>	<b>Ravi Shankar</b> , Chief Marketing Officer
<b>Run time:</b>	<b>00:17:25</b>



---

Claudia: Hello, and welcome to this edition of the Boulder BI Brain Trust, or the BBT. We're a gathering of international consultants, analysts, and experts in business intelligence, who meet with interesting and innovative BI and analytics 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 BBT podcasts are produced by my company, Intelligent Solutions.

I'm pleased to introduce my guest today. He is Ravi Shankar. Ravi is the chief marketing officer for Denodo. Welcome, Ravi.

Ravi: Good to be here, Claudia.

Claudia: It was fun having you here. What an exciting day this was.

Ravi: It was. Thank you very much.

Claudia: For those that may not know a lot about Denodo, why don't you give me a really quick review of what Denodo is all about?

Ravi: Denodo is one of the pure play data virtualization vendors. We differentiate ourselves as being the best data virtualization product out in the marketplace, and we have been recognized as a leader in this space. We have been awarded a number of awards from analysts and other firms for our product.

We have very large customers, Fortune 500 customers that are using our technology for a variety of use cases across the enterprise and in departments for very serious use cases, I would say. We have more than 250 customers at this particular point. Many of them, as is talked about, [have] more than a billion dollars in revenue, are actively using our product.

We are very excited about our product, our wins in the marketplace, and our customers using it for really serious business value.



---

Claudia: Well, For those who may not know what data virtualization is all about -- It's hard to believe anyone wouldn't know, but there are a few -- why don't you explain what it means briefly? What is the business need, and what are the challenges to data virtualization? And then we'll get into what is Denodo's solution to these problems.

Ravi: Sure. As I mentioned, a large number of our customers have over a billion dollars in revenue, which means they are really global companies with multiple different departments and hence, different systems across the place.

Business users, like in sales, marketing, customer service and so on, spend a disproportionate [amount of] their time going through these different data sources to piece together the information needed to do their day-to-day business operations. So, what data virtualization does is, it abstracts them from having to go and tap into these data sources, which could have a variety of different formats, [a] variety of ways to access them, [a] variety of data quality levels, and so on.

By positioning it as a virtual data repository, as a single bell, if you will, you just ask the data virtualization layer, "I need information about customers, what products did they buy, which channels did they buy [from], did they come through the web and buy, did they come through the brick and mortar store to buy it, and give me the sales in the last year, two years, nine months."

Data virtualization will take care of getting that information for you from the sources in real-time, as the data is up-to-date right now. You are completely shielded from going to the data sources to get your questions answered.

Claudia: I look at it as kind of a view into data, but the view can be across multiple sources of data. That's what I find so exciting about virtualization is that it can pick pieces of data out of different systems -- out of an enterprise data warehouse, out of a Hadoop cluster, and out of an operational system -- and bring them together as if they physically were but in fact, they're not. We're not moving



---

the data at all. It is simply brining it together in a virtual fashion. Is that correct?

Ravi: That's exactly right. That's why customers love our technology, because there is no physical movement of data, there is no repository to enter the repository, to hold the data. It saves on the repository costs, saves on the systems that need to hold all the data. It does this all on-demand, in real time, across any type of data source.

It can even access your own Excel or Word files within your laptop and integrate that with your CRM and ERP systems, and provide a holistic view of the data.

Claudia: The big news, of course, is that Denodo is about to launch version 6.0, and that I found fascinating. If you don't mind, just a brief overview of what 6.0 is all about, and then we'll go into some more detail.

Ravi: Sire. With version 6.0, we are accelerating fast data strategy for a number of companies. In all the cases that I talked about, the companies are really suffering from slow-moving technologies to get the data and results as fast as possible. We are providing the acceleration for fast data strategy for them through three key innovative features that we're bringing out in 6.0.

The first one is the dynamic query optimizer that, during run time, figures out the best way to execute the queries and bring back the results.

The second one is, for the first time of any solution in the industry, data virtualization will be hosted on Amazon marketplace. No longer [will] you have to come to us as a vendor and buy it. You just go to Amazon, you select whatever configuration you want, and you're off and running to using it.

The third one gets the data directly to the business users. It's a self-service capability. Data virtualization holds the metadata about the sources and the kinds of data. For the first time we are



---

introducing the capability for business users to look up the data in the data sources without having to rely on IT.

The value proposition across all these three: we are accelerating the data with the dynamic query optimizer, we are accelerating your ability to get to the solution faster with the cloud version, and we are accelerating this data to the business users directly. We are totally accelerating the fast data strategy.

Claudia: Terrific. Well, let's go into – as I said – a little bit more detail on this. First of all, the dynamic query optimizer. Overall, how does it work and how do you see it benefiting corporations? You gave us a very good example. If you don't mind, telling us a little bit about how it does work.

Ravi: Sure. You will find the data virtualization products out there have a variety of different capabilities. Some, the very basic ones, rudimentarily move a lot of data from the sources into the data virtualization data layer to crunch it. That's the very basic ones, and given the volume of data, especially big data sources that run into hundreds of millions or billions of rows, it's a very inefficient way to do it.

The smarter ones actually push down a lot of the aggregation and querying into the data sources, and crunch them there; then bring back the results. In doing so, there are a lot of different techniques like partial push down optimization, full aggregation push down, that are on-the-fly data movement, partition pruning and so on.

The data virtualization engine has to determine which is the best way to do it. So far, a lot of the data virtualization technologies had to have the admin determine during design time, what is the execution plan they should use.

With 6.0, we're actually moving past this tactic into dynamic. It dynamically reads the past performance of the underlying systems, dynamically rearranges the queries to say, "The best method to use here is partial pruning, followed by on-the-fly data movement." It can do that very fast and the results that it provides is much



comparable to having a physical data warehouse. A single data warehouse.

If it takes, for example, 23 seconds to execute a query, data virtualization goes against the data warehouse, against another database, and an ERP system in our cloud system, and provides the same 23 seconds to execute the query.

So, one difference between static and dynamic. In a static sense, if you execute the query at 10 o'clock and 1 PM, you'll get different results, because it is following the same execution plan. For us, whether they execute at 10 or 1, we will get close to the same timing, because [we] dynamically which is the best way to execute.

Claudia: The way I understand it, then, is you don't have to predetermine the path, if you will, of the query. That's kind of the static way of doing things. What happens with your dynamic query optimization is that that particular capability will do that for you. Is that right?

Ravi: Exactly right. No longer you have to determine during the design time the query execution plan. We will determine that based on the performance of underlying system, so we can guarantee faster results.

Claudia: Excellent. The second part then of version 6, or a big capability that's coming in, is data virtualization in the cloud. There was so much interest in this. If you don't mind, just give me a few details on how that will work.

Ravi: Sure. We have customers who are using data virtualization on the cloud already. But what we are doing for the first time is making it available on the AWS, or Amazon Marketplace. No longer [will] you have to come to Denodo to license it. You just go to Amazon, you swipe your credit card [and] you choose your configuration. We provide a menu of different options. You can either select it by the number of sources you are connecting to, or by the number of queries or results that you need to get back.



The best part about this option is, you can start small. We have [a] number of questions that come from researchers, for example. They want to start using data virtualization, but they're not ready to buy it yet, because that has to be done through the central IT.

The researcher can go into Amazon. They can connect to one or two sources, they can start doing the data virtualization using that, and benefit from the results. They can then prove it to the central IT. "Look at the value I'm getting." Then, it becomes a much broader one. With this capability, we are allowing the users to start small, and to be able to expand, and we get scale, horizontally and vertically, to more data sources, more queries, or even adding more images for your data virtualization needs.

Claudia: Just to drill in a little bit, do all of my sources, then, have to be in the cloud for this to work, or can you actually go to on-premises sources as well as cloud ones? Make it a hybrid environment?

Ravi: That's the beauty of this. You can go beyond just the cloud data sources and also on-prem data sources within your data centers. Given the dynamic query optimizer that I talked about, we can actually do this, because you are transmitting the data over the wires. We can do the query optimization in the source and bring back the minimal set of results that will guarantee the performance.

Claudia: That's important in a cloud-to-on-premises environment. It's that traffic flow back and forth. You do want to minimize that.

Ravi: That's right. The constraint is on the bandwidth.

Claudia: Alright. Let's talk about the final piece, and that's self-service data discovery and search. Again, a little more detail.

Ravi: Sure. Data virtualization layers usually have a good understanding of the data sources in terms of the tables, the columns, and the rows of data that it connects. We call that the metadata. It understands the information about the data. For the first time, we're exposing the data itself through the data virtualization layer.



You are a business user. and you want to understand a particular customer. and look up all that customer's information. We have a capability that, once you see the list of customers, let's say you're looking for Mike Storm and you look up all instances of Mike Storm, you can click on Mike and you can see all the products that Mike has bought. You can see, did he buy it through the web? Did he come to the store and buy it? Do we have active warranties for it?

This is all transactional data that is sitting within the sources, and we are exposing that through the data virtualization layer. The beauty of this is, it's all Google-like search. We have a new user interface that is browser-based. You just point to that and then you query, you do your data discovery and search.

Once you like what you see from a result perspective, click a button, save it as a search. Now, it can be accessed through your reporting tool. You can run your dashboards, your reports, everything. It enables business users to discover data, search, explore, and then start using it.

Claudia: That's my favorite capability. That's terrific. You mentioned that it's very large corporations, billion-dollar corporations, who are using your technology. That also means that your environment must become quite complex. That gets to my last question here about version 6.0. What do you do to support large organizations with very complex data virtualization environments?

Ravi: Sure. So, we have many customers that are using multiple deployments for multiple use cases. We provide the ability in 6.0 release that – the new resource manager capability – that provides two things. One is, there's a diagnostic tool that visually shows you the performance in terms of CPU, the memory and all that. There are a extensive amount of log information that is available so you can actually detect where there is a slow down or point of problems, and take corrective actions.

The second biggest capability is that, if there are multiple departments that are going against the data virtualization system to



run queries and get back the results, you can assign priorities. For example, year-end, the finance department gets the priorities. The queries that come through will be given 100 percent or 80 percent capability.

You can constrain the queries coming from the marketing side so that, either you temporarily stall them and then process them later when the CPU and the memory bandwidths open up, or you can actually allow them to proceed with a certain amount of constraints. It's not running at full capacity, but it will still deliver the results.

We are providing a lot of these capabilities for these customers to manage multiple different deployments very efficiently, and process the results back to the prime need of certain departments.

Claudia: Beautiful. Alright, well drum roll... When is version 6.0 going to become available, and where would someone find out more information about it?

Ravi: Sure. On March 30th, we are unveiling the wraps to showcase the 6.0 in a virtual event. It's called the Fast Data Strategy Virtual Summit. At 9am, you will be able to see on our banner on our home page the information that you can go register. You do not have to come to the event. It's not a physical event. It's a virtual event. You can sit at your desk, log onto the browser. The sessions will start rolling at nine o'clock with keynote sessions from analysts.

We will also have sessions from the BBBT. We will have customers speaking. We will have panel sessions with partners. We're going to have 20 plus sessions on this, where you can learn about how customers use it in real time. If you're new to data virtualization, you can learn about it. If you already know data virtualization, you can get the deep dive information.

What is the best reference architecture for operation versus analytical? How do I configure my systems for performance? How do I make this work with big data systems? That will be plenty of information for everyone.



---

Claudia: That's March 30th for the US, but you also have another one for the rest of the world.

Ravi: That is correct. For Americas, for the US, it is going to be on March 30th, 9am, Pacific. And for, in the EMEA, APAC, we are rolling it at 9am GMT UK time on the 31<sup>st</sup> of March.

Claudia: Excellent. I wish you much luck with the launch. It sounds like a very interesting one, and I'm certainly very impressed with version 6. You've done well.

Unfortunately, we're out of time though. That's it for this edition of the BBBT podcast. Again, I'm Claudia Imhoff. It's such a pleasure to speak with Ravi Shankar of Denodo today. Thank you so much, Ravi.

Ravi: Thank you, Claudia. It was a pleasure being here.

Claudia: I hope you enjoyed today's podcast. You'll find more podcasts from other vendors at our web site [www.bbbt.us](http://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!