



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:	HP Vertica
Date recorded:	September 12, 2014
Host:	Claudia Imhoff , President, BBBT
Guest(s):	Steve Sarsfield , Product Marketing Manager William Cairns , Lead Data Scientists
Run time:	00:13:23
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 Will Cairns and Steve Sarsfield. Will is the Chief Data Scientist and Steve is the Product Marketing Manager for HP Vertica. Welcome to you both.

SS: Claudia, thank you for having us. It's great to be here.

WC: Thanks, Claudia, excited to be here.

CI: All right, well, Steve, let me start with you. Most people are familiar with Vertica, but let's talk a little bit about the differences between legacy architectures, the new NoSQL ones, and maybe where Vertica fits into this paradigm that you've got.

SS: Yeah, good question. If you look across the big data solutions, I think you'll find that the solutions across the market for big data analytics are really full of compromise. You have your traditional enterprise databases, and I don't think the original developers of the SQL databases really envisioned the data volumes that we have today.

Then there are the new Hadoop based solutions, they're pretty good. They handle the data volumes they have, they have clustering, the ability to handle those. But handling the traditional analytics that were out there are very difficult for them. They're sort of scrambling to try to put SQL on top of these new platforms.

Really, what the difference between sort of those platforms and Vertica are that we have the ability to handle big data, petabyte scale big data, huge volumes of data. We can deliver very fast analytics. We can work with legacy tools, based on SQL, visualization tools and ETL. We have



some of the new types of analytics, so Python and R, and some of the new ways to do analytics are available to you.

You can also do advanced analytics with Vertica. We think there's a big difference in being able to handle all the problems that you have with big data analytics.

CI: Interesting perspective. Now, Will, let me bring you into the conversation, because you did have a slide claiming to be insanely fast. I think I have to ask you, what is the secret sauce? What's under the covers that makes Vertica so very good and so very fast?

WC: We certainly have survived and have matured in the marketplace by just marketing ourselves as being fast. For a large portion of our history, that was a very successful way of broadening our reach in the marketplace. Yes, we came on, we were a young company, we were maturing. We were orders of magnitude faster.

But I think in the most recent, say, year or two, other database vendors have taken notice of that. Right? Just simply marketing yourself as fast is not good enough and I think we are doing things on the analytics side and we'll talk more about those on some of the follow ups here.

But getting back to kind of the secret sauce, Vertica built, of course, on column store architecture, MPP hardware. Our customers can select the SLA that best suits their needs. If you have 1,000 users, maybe you'll need a 12 node cluster or a 15 node cluster. If you have a smaller user base, you can size that down. If you're going to do simultaneous query and load, there's an architecture to fit that use case.

CI: Well, you also have some, you have distributed queries, you've got projections. Maybe you could spend just a second talking about those a little bit?

WC: Yeah. Certainly. In Vertica, there's not a leader node of any kind. You can instantiate a query to the DNS name of the cluster, and it's going to grab any node within the cluster. That node becomes the initiator and the query gets distributed to the executor nodes. Because our data is



segmented across the cluster, every node has the ability to contribute to the query output.

CI: All right, well, let's move onto something else that I really liked. It was titled, "From Hindsight to Insight to Foresight." Explain what you mean by that slide.

WC: Long before, I think, organizations started to think about hiring data scientists, what we deemed analytics was largely backward looking pictures of building a dashboard or generating a report. Then trying to generate business decisions based off of that report, that daily activity report, or that monthly activity report.

CI: Very descriptive in nature. Here's what happened.

WC: Very, very... not very rigorous. Is a delta of a quarter of a percent significant enough that half the organization runs around and scrambles, or is that just some casual noise?

Or trying to spend a lot of time explaining for some phenomena that happened in the middle of the night. I can't tell you how many times, working in different companies that I experienced this. It was, go back and try to explain what happened last night.

Where I like to see things go, where I like to see organizations progress towards is this prescriptive, predictive analytics, mature, robust statistics. But don't do it in the offline traditional sense that maybe, say, like insurance companies did with actuarial sciences, it doesn't belong in a back room. It belongs at the forefront.

You need smart people who ask the right questions, who have the right rigor to apply the techniques and then the ability to broadcast that out to the organization, present the ideas and come up with a solution.

CI: Well, and the thing that I think is fascinating is the ability to, we've always talked about how you have to be able to, act on the intelligence that you've been given.

I'm frustrated, because so many organizations get stuck in those descriptive analytics, the historical backward looking reports and metrics



and so forth. They don't even know enough to ask the next question, which is the diagnostic one, of why? Why did this happen?

How do you see moving companies into these higher levels of thinking?

WC: Yeah, they tend to silo things. Not only does the data wind up being siloed, but the analysis and the outcomes tend to be siloed. Maybe that's just a factor of how organizations get territorial or how they structure themselves.

But when we get to the point where you can break down the silos, when you can share marketing data with the network group, or when you can share customer call center data with the network group. This is where, really, organizations start to leverage analytics in a meaningful way. The traditional silos need to go away.

CI: Yeah. I think we need to break those walls down.

All right, well, let's talk about some of the analytical features then in Vertica, if you don't mind, a brief overview.

WC: Yeah, sure. Interfacing with Vertica is traditionally a SQL based interface. Of course, we have some programming languages that can talk to Vertica as well.

But analytical functions in Vertica are extensions of SQL, some are compliant, some are extensions that we've come up with on our own. Things like gap filling on a time series.

If I expect, say, a beater measurement come in every five minutes and all of a sudden, I've had a problem where I missed a few measurements, maybe an upstream problem, whatever, using Vertica SQL I can quickly gap fill. I can interpolate across those missed measurements, so that I can still understand and still compute what's going on.

Sessionization, so, I can describe within data something like the progression or the click progression from, say, a home page to a product page to a checkout page. If I'm trying to understand something like abandoned shopping carts, it's a really useful feature.



It's simple SQL code and we write a few lines and we study that and then we look for, within the data set, other cases that fit that same pattern. Time series analytics, regression. You can do regression right from the SQL command line, so.

CI: Yeah, a lot of pretty sophisticated cause and effect, I suspect.

WC: As far as you want to take it. As far as you want to take it, literally. Then of course, extending it up in-database R. Take the open source R packages that are out there and could be classification, clustering algorithms, all of that is available to you.

CI: Yeah, that's great. The other thing that you've got that HP has, not necessarily Vertica alone, you have HP Labs. They've been around for a number of years, many, many years. They've also been kind of the premier think tank of HP.

First of all, what are they? Do they instill some of their innovations into Vertica?

WC: Yeah, so, I myself, I get out there at least once a month. There's really this fantastic feature of going to HP Labs. Whenever I'm with somebody new, I like to take them there.

The offices of our founders, Bill and Dave, still exist in their original form from the 1960s when they were both present, the two offices were just combined by a washroom in the middle. Right on their desks, everything from the position of their pens to novelty items on the desk to picture on the wall with former presidents. Then walk out of those 1960s offices and into HP Labs, where some of the most cutting edge research is going on today.

We go down and we take customers there, we do customer advisory boards there. We show them some of the features of what's going on there. It's an amazing group of guys there. Really great stuff coming out.

CI: They contribute to Vertica, overall.

WC: They do, they do. Kind of the flow of things is, if you're a researcher there in Labs, you might have a project. Then at a certain point throughout our



innovations, we promote projects into product. That's really where it becomes part of Vertica.

Obviously, things like distributed R that we talked about, that was a project out of HP Labs, that eventually becomes part of the product. Super impressed by that.

CI: Yeah, brilliant. Absolutely brilliant. All right, Steve, let me bring you back in. I'd like to get back to Vertica, and if you don't mind, describe the different deployment options that are available.

SS: Yeah. Vertica has traditionally been software that you can download if you want, you can go to our website and get a version that you can use for up to a terabyte and up to three nodes on a cluster. But you can also buy Vertica and use the Enterprise Edition.

We also have deployments available so you can buy either an Amazon or HP Cloud version of Vertica, which is easily deployed in the cloud. Then finally, there are appliances that you can buy, so Vertica installed on an appliance and ready for it to sort of drop in your enterprise. We have all those deployment options and we have customers in all of those areas.

CI: Wonderful. Well, it's a very exciting time for HP, so what's in the future?

SS: Well, I think Vertica's going to continue doing what they do best, and that is making very fast analytics. We're going to try to make analytics faster, better, bigger, with bigger data. I think we are looking constantly at the use cases for analytics. Could be anything from customer analytics to security analytics. We're trying to look at ways that we can make performance better. Yeah, just bigger, better, faster.

CI: Always good to have. All right, well, unfortunately, 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 Will Cairns and Steve Sarsfield of HP Vertica today. Thank you both for being here.

SS: Thank you, Claudia.

WC: Thanks very much.



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!