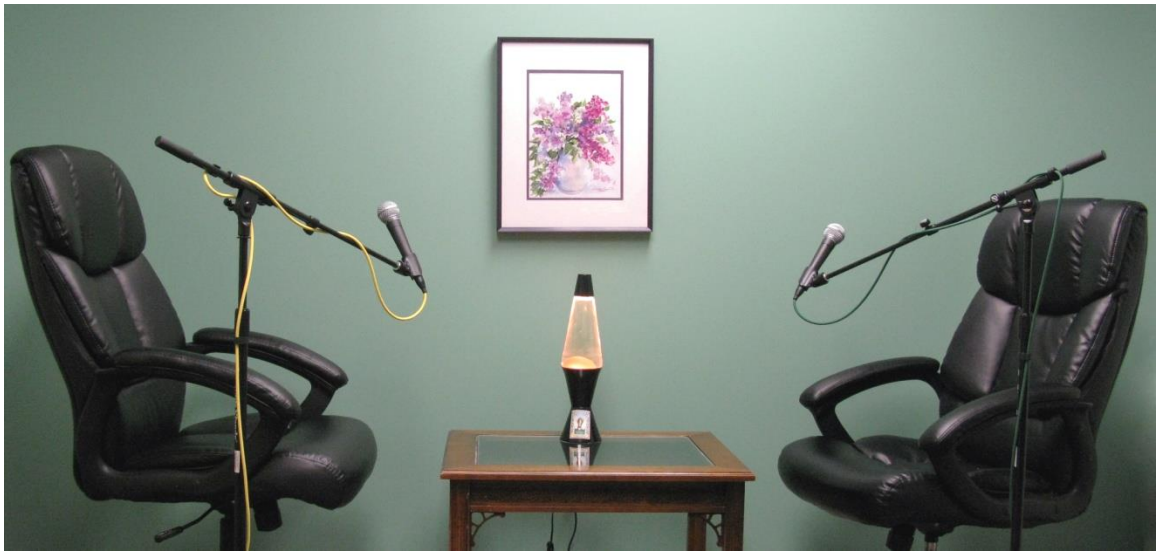
BBBT Podcast Transcript



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:	**Teradata**
Date recorded:	**July 18, 2014**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Chris Twogood**, VP Product and Services Marketing
	Dan Graham, Director of Technical Marketing for Teradata
Run time:	**00:14:11**
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 Chris Twogood and Dan Graham. Chris is the Vice President of Product and Services Marketing and Dan is the Director of Technical Marketing for Teradata. So, welcome to you both.

Chris Twogood: Thanks, happy to be here.

CI: All right. Chris, let me start with you. Teradata has been like a laser, focused on big data analytics, but there are significant challenges for companies trying to analyze big data, so why don't we start there.

First of all, what are some of these challenges and how does Teradata Aster fit into this emerging area?

CT: That's a great question, Claudia. I think what we see a lot when people are doing big data analytics is they're trying to do things in terms of a data warehouse infrastructure and a relational structure with SQL.

They're also trying to do stuff specifically around stats, and file structures, and systems. They're trying to do stuff in Hadoop, with MapReduce, and some of the new technologies like Spark and Storm. They're also trying to do things in graph engines -- storing things as a triple store.

So, the problem with all of this is you have this proliferation of all these different analytic tools, fragmentation of data. You've got to have different skill sets that can interact with all of those different technologies. A lot of times, it takes a longer time to value to bring it together. What we really do with Aster is we deliver what we call, "an integrated discovery platform."

This platform really takes all of those different technologies and puts them as first class engines inside of a single infrastructure. So you can do stats, and MapReduce, and path and pattern analysis, text analysis, graph function, SQL all into a single platform. But the key about it is, there's a SQL framework, which is the integration layer to all of that.

So, mere mortals, as I talk about, can really invoke it through SQL, and then we have a pre-built set of analytic functions -- over a hundred of them now -- that sit on top of that, that really make it easy for a broader set of users to deploy it. That's really the value we bring from an Aster perspective.

CI: Well, it's fascinating. Let me turn to you, Dan. Let's go into a little more detail here.

You talked about the SNAP Architecture. It was defined… I actually looked it up because I wanted to see what it was all about. I looked it up in a recent blog, a Teradata blog. It was defined as the new spine of the Teradata Aster platform, namely, a unifying discovery framework we call SNAP.

So, what exactly is SNAP? What does it do?

Dan Graham: SNAP has two primary components. One is just the recognition that there are multiple file systems all under one umbrella. So, you have a row store, a common store, you have a file store, where you can copy in Hadoop data directly without any effort. It's pretty simple to do.

So, having multiple file stores under a relational database is something new under the sun. Probably more important, however, is the optimizer. The optimizer is actually optimizing across multiple workloads, so those workloads would of course include SQL.

But it also includes MapReduce workloads and graph workloads and R workloads. Having an optimizer that spans all of those in one request is kind of new as well.

CI: All right, Chris, let me go back to you. You mentioned earlier in the presentation today that open-source R is not really built for big data. So, I guess my first question about that is, "Why?" Let's just start with that.

CT: We can start there. R is a great open-source movement. It's free, you've got lots of the academia behind it. You've got lots of users adopting it. In things we've seen, almost 70 percent of the companies out there are using R inside of their organizations.

Plus new algorithms are coming to market every single day. There's a lot of stuff that's really great about it, but the challenges about it, and why I talk about "not really designed for big data" is if you want to be able to analyze and use any of those algorithms at scale, it quickly runs out of memory. It quickly runs out of processing power because quite often they're taking sample sets and moving it down directly into their server or their laptop and running the different algorithms and they can run out of memory. Or it just takes hours and hours and hours to run.

So, to your point, what we've done with our Teradata Aster R is really make open-source R massively scalable and powerful. So, really taking the foundation of what R is... and we have three components to it. The first component is our Aster R library. This enables you to invoke a lot of the open-source R that's commonly used in a parallel function across all the data and you don't have to sample anymore.

Plus we've taken a lot of our core Aster functions, like nPath, and we've enabled them through R.

We also have introduced the parallel constructor, the Aster R parallel constructor, and what this is, is that our programmers can take any open-source technology and push it up into a parallel framework inside of the Aster R, and then we've done the integration tightly into what Dan was talking about earlier with the SNAP framework and integrating it tightly with the optimizer.

CI: Certainly, if people want to see or read more about it, you do have a June 26th announcement that goes into much more detail about that.

CT: We have a June 26th announcement. We've had a lot of great coverage about it. You can also go to Teradata.com and search on "Aster R" and find out a lot more about the product.

CI: Yeah. Very Good.

All right, Dan, back to you then. First of all, you did mention collaborative planning. Let's talk about that a little bit more. Why don't you describe, at a high level, just exactly what it is, how it works and maybe some of the benefits of this function?

DG: Correct. This is part of the SNAP Optimizer, the part that runs the queries and runs the complex things inside the system. It's not two people collaborating and getting together over some plan.

CI: Clever.

DG: Basically, the collaboration here is that it's easy enough for Teradata to optimize SQL queries. We've been good at this a long time. Aster has benefited from that.

However, what happens when you... in your SQL query you are invoking a MapReduce function, let's say you're doing some kind of complex algorithm like nPath, or Bayesian analytics, or you're doing any number of text analytics, or maybe you're doing a graph analytics. You're looking at things that are purely not SQL and not relational by their very nature. They're not considered something you'd see in rows and tables in the old thinking.

The consequence here is the collaboration basically says, the optimizer talks to the function and exchanges information, and comes back with a plan that meets with this non-SQL function does.

CI: Well let's continue. Chris, let me go back to you. Let's continue a little bit with these Aster graph analytics. A lot of people are confused about them. Myself, I have to admit I'm in that group. Why don't you tell me a little bit about what is Aster SQL graph?

CT: Before I do that, let me take it up just a level.

CI: Thank you!

CT: This is actually what we call connection analytics, because we believe everything is connected. People are connected to people. People are connected to products. Actually, products are connected to products. Machines are connected to machines.

Machines are connected to products. But, what I mean by they're connected, is there's relationships. and there's influence that gets driven across these different connections.

And so it became really important for us to say, "How can we create an infrastructure and a foundation that, in a very scalable way, can help analyze these different connections?" So, with Aster 6, we brought out the Aster graph, or SQL-GR.

What this does is really provides a framework that fits within the whole Aster discovery platform for being able to run graph analytics. There are a number of pre-packaged functions that we've built, so that with just the same framework with SQL, users can invoke this and start to understand the relationship between all of those different connections of multiple products, machines, and people.

CI: Alright. Well Dan, let me turn to you, then. Part of the confusion is that we've been able to do the types of analyses that these graph databases do, but it was very difficult, right? They make it easier to do these multiple connections and that sort of thing. Is that a correct way of saying it?

DG: That would be correct. We've had a number of telecommunications customers using the Teradata database to do multiple self-joins, and there's the trick right there. Multiple self-joins take a long time.

You can't do these things with SQL, and you can also spend a lot of energy and time doing them. When you switch over to a graph engine, the thing dramatically drops in elapsed time, the performance is very nice, and allows you to do some things that you weren't able to do with all those self-joins.

CI: Where do you see graph databases fitting into the analytics landscape itself?

DG: There's fundamentally a couple things going on. There are graph engines and graph databases in the marketplace. They're radically different. They do two different tasks. They overlap somewhat, but these do not replace your statistics.

Don't give up on SAS analytics, right now, not a good plan. Also, you're not going to be able to move forward as effectively, unless you have something like Aster SQL-MapReduce. So, being able to apply Python or Pearl, or complicated algorithms like nPath, you still need those things, but the graph engine is a new capability that allows you to look at the relationships between things in a new way, as Chris said, the connections.

Instead of focusing on the people, let's focus on the connections. Instead of thinking about the airplanes, let's look at the relationship between them and the airports themselves. Let's see what happens as they move through this process.

CI: All right, Dan. Why don't you give me some graph engine use case examples?

DG: First, you have to get this notion of a ripple effect, that if you touch some part of the graph, some person, some entity, some area of the graph, it emanates an influence to those around them. This is something I call the ripple effect. We know this in things like pricing, for example. Pricing is a very obvious example.

Anybody that's sold a house knows that if the neighbor down the street sells a foreclosure, it affects your price of your house, your value goes down. Same thing happens in the stock market every day. That's what you hear about -- there's sympathy pricing going on where certain stocks go up and down in sympathy to one another.

This is what we're looking for in Graph Analytics – Where are the points of action, and what is the influence being exerted? And how can I turn that into business benefit? There's applications, you can find things like... Social network analysis is what everyone gravitates towards because of their experience with LinkedIn and Facebook, but there's an awful lot of other applications, such as fraud detection and healthcare.

You're seeing, as I said, pricing. There's a whole mass of areas where graph can be applied. So, the buzzword is, "Graphs are everywhere."

CI: And we will have to pick them up and use them in our analytics environment. Alright, one last question then, and that's just about the

future. Chris, let me go back to you. Where do you want to go with all of this?

CT: Well, it's interesting you know, for us, there are a lot of different engines that are coming out. Dan talked about a number of those specific to graph. Certainly MapReduce has its place, SQL has its place, NoSQL engines have their place, right.

As you start to look at all these different engines and different file stores, how can you tie it all together in a very seamless way that takes away the complexity from the end user? Because they just want to be able run analytics and get the best insight.

We look at it more about, what is the role that Teradata can play in helping to orchestrate this? Whether it's doing optimization, so that you get the fastest results across multiple dimensions, but enabling the users to have simple interfaces to data, orchestrate it throughout their overall infrastructure, and then being able to get the best value of insight from their big data.

So, that's where we see it going, and we think Teradata, with the advances in Aster 6, with what we announced earlier this year with Teradata Query grid, we're well on the way to being able to deliver this for our customers and providing those benefits.

CI: Yeah, I think you've got a pretty darned bright future coming up. Alright, well that's it for this edition of the BBBT Podcast. Again, I'm Claudia Imhoff, and it's been a great pleasure to speak to my friends, Chris Twogood and Dan Graham of Teradata today. Thank you for speaking with me.

CT: Thanks, Claudia.

DG: Thanks.

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!