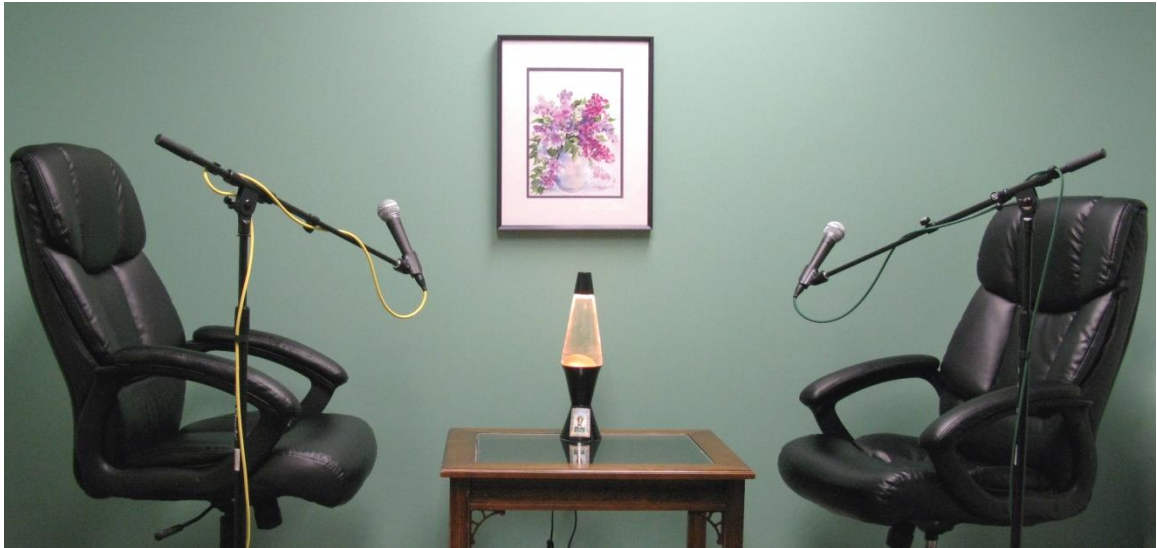




BBBT Podcast Transcript



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:	WhereScape
Date recorded:	October 25, 2013
Host:	Claudia Imhoff , President, BBBT
Guest(s):	Michael Whitehead , Founder and President
Run time:	00:14:00
Audio link:	<u>Podcast</u>
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 good friend. He's Michael Whitehead. Michael is the founder and CEO for WhereScape. Welcome, Michael.

Michael Whitehead: Thanks for having us here in fantastic Boulder with beautiful weather.

CI: It is gorgeous here, isn't it, today? Good day to be here. Let's start off with your morning "shot across the bow". You started out by saying that the data warehouse landscape has changed, and it has changed dramatically. Why don't you just briefly walk me through what was, and what is today, the data warehouse landscape?

MW: Sure. When we started off, the data warehouses were defined by their database in a lot of ways. You'd talk to an organization, and they'd say, "Well, I have a Teradata data warehouse," or, "an Oracle data warehouse," or, "a SQL Server data warehouse," or whatever.

CI: It was pretty much a single entity.

MW: Yeah. I think Microsoft probably changed the rules early on without anyone realizing it when they added cubes in, which is essentially a different database. It was still a Microsoft data warehouse. We've really seen it take off in the industry, where now we see people with mixes of technology and specialist technology for special things.

You'll see an Oracle shop with a Netezza box wield, and a Teradata shop with an Aster. We've had to really look at it, and go OK, that's a different way of viewing a data warehouse.



CI: Colin and I call it the extended data warehouse architecture. Basically, you've still got a core data warehouse somewhere, in some technology, but there are so many, now, additions to it, like you said, Aster Data, a data warehouse supply, a Hadoop thingy, something that you hang off of this environment that does specialized things for you.

MW: As that matures and becomes the norm, you then get a different set of problems. How are you going to build it? Are you going to have a different team that is doing Hadoop, a different team that's doing Oracle, a different team that's doing Netezza?

Is that really feasible? Are you going to double up on those people? What if they're away? Where have you put your best team? For us, that looked like an opportunity. Why can't we have a development environment that sits across all of them, and you can just choose where you deploy?

CI: That's where you come in, because, quite honestly, you also stated that even though you've all these wonderful new technologies, it's still really hard to build the data warehouse, and in particular, to maintain it, to change it, enhance it. Why is that? I would have thought all these wonderful technologies would help us out.

MW: I think that data warehousing is fundamentally difficult. We're trying to solve a hard problem. It is. It's hard. When you look at what we're trying to do, we're trying to take a business idea and trying to support someone. We're trying to understand what they want and provide data to support that.

We've got no place to hide. If you look at the transaction processing people, they can create an application. They can hide the data, and put screens between it. We have no place to hide. We're taking that data and make it available to people. We have to understand what they want to do with it. We have to understand the business rules. We have to understand the definitions. A lot of these things organizations don't understand themselves, or have never been forced to go through the discipline. Because it's obvious what profit means if you're sitting in one department, versus another department, versus the CEO, versus a...

CI: A line manager or something?



MW: Yeah, because it has context around it. When you're dealing with data, you don't have that context. It's hard.

CI: It is, and I like your comment. Operations can hide the data, mostly because, you're right, there is a buffer between the data and the human being that's using it. It's a very rigid buffer. They enter orders the same way every time. In the data warehousing environment, I never know what questions are going to come down the pike. There is a no buffer. You're right. There is no standardization, in terms of the queries or the questions that a person can ask. How do we fix that problem then?

MW: I think that the problem exists, and I'm not sure we can fix the problem. I think what we can do is help alleviate and stop it from defining a data warehouse. It shouldn't be defined as being difficult and hard. I think we can look to automate functions of it that we can say, "Yeah, this is difficult."

Let's spend our time on the things that are actually different, that are actually important, that are actually going to add some value. Let's not spend our time on the things that are exactly the same, that we solved the problem, years and years ago. Let's not repeat that every single time we do a project.

CI: Let's get into automation. You actually gave us a very clear example of what can be automated and that was a developer saying, "I need to develop," I'm going to test your memory here, "I need to develop a changing dimension, a slowly changing dimension. What do I need to do to do that?"

You had a whole list, a whole raft, of activities that this person would do every time they wanted to create a slowly changing dimension. They would first do this, then this, then this, and there was a whole list of them. What you said was, "Do they really need to do all of these things, or are there parts of this that we can automate?" I'm going to test you memory. Go for it.

MW: We did talk about that, and all those things do need to be done, but the question is whether the developer needs to do them. Because if you think about what's really different about any dimension, you're talking about what data is changing, what's the business key associated with it.



Then, when you end up coding it, and end up putting it into the database, and have it actually working, there's a whole lot of other things you've got to do. You've got to get the keys, and you've got to do indexing, and you've got to write all the code.

CI: Pretty much routine processes that you do every single time.

MW: Yeah. Exactly. What we see far too often is people doing these every single time, and individuals will, obviously, automate it to a minor extent. They will cut and paste the previous one they did. Most of them will, anyway. Coders never start from scratch.

You've got that code, and then you're modifying and changing that code. You're putting in the fields, because you're typing in the field names and all these things, and then you're modifying it to go out. Maybe you've got lots of different ones or lots of different types of dimensions that you're cutting and pasting from.

We can do that for you. That's easy.

CI: Why not just push a button?

MW: Why don't you just push a button?

CI: Just tell it what the attribute is and what the key is, and boom, off they go.

MW: Then you're using the smarts that the data warehouse person has got. They're figuring out what needs to be done, and then you're just pushing back to software.

CI: Not how it's being done.

MW: Yeah.

CI: I love that example. I think it's very clear. It's very obvious what can be automated. You go beyond that, though. You talk about a number of things that can be automated. I'm going to test your memory again. What else is available to be automated?

MW: It's really interesting, because we're exploring that. There are obviously a series of automation points, of things we can do. One of the ones that we



used as an example was building a data vault from a normalized data warehouse. There are a set of rules associated with data vaulting, very strict rules associated with that.

Rules are good. In this case, it means that if we follow those rules, we can automate those rules and push them through. Some of the things really interest us are moving it up from those point solutions to processes, though. What processes can we automate? How can we string those things together?

That's when you start to get some real value out of it. That's, see, for us, the really exciting part of automation.

CI: Yeah, I liked the list that you had. You had profiling. You had just any number of things that were pretty convoluted processes.

MW: Let's take a real trivial one, but one we use as an example a lot, is documentation.

CI: Oh, yes. Yes. Oh, my goodness, yes.

MW: That's the reaction of very single data warehouse person that you go to. We can automate the documentation, and they're like, "Wow!" We should be able to. We know what you have done, so with the WhereScape software products, with Red, in particular, we're taking note of everything that you're doing. We're storing the metadata about it, just so we can automate the documentation.

When you have got it automated, you get some real advantages out of it. In the documentation's case, you can document on demand. So documentation is no longer a one off thing that you do in a point in time. You can do it any time you like.

CI: It's a living and breathing document.

MW: Yeah! The concept of data warehouse documentation being the one that was done on the first project that was done really really well, and signed off on, and beautifully formatted. To figure out what's going on, you take that and then you layer on top the slightly dodgier changes that have been made.



You have to go in the right order, so you understand, because the last person just came in and did this, and all they documented was the change that they made onto it. It gets really difficult to understand. The idea that you can walk in, document on demand, have a look at it.

Impact analysis, that's another one. You want to be able to go in and look and see if I change this source system, what is going to change? Where did this data come from? The lineage, that's a classic for automation.

CI: Yeah, and it's so needed in our environments.

MW: It is. Given the conclusion that we came to very early on that change is inevitable, and should be catered for from day one. When you have that as a mindset, you just approach the problem a little bit differently.

CI: I'm going to change directions on you a little bit. Let's talk about something that's very exciting. I'm going to lead you into it, because I want to know how the vendors, the other data warehouse vendors, have reacted.

MW: I don't know if you saw that last week we signed a reseller agreement with Teradata.

CI: I did indeed.

MW: Look, we're really thrilled with that. A lot of it is the view from the market, and that Teradata saw that data warehouse automation is a valid technology. It's not just another form of ETL.

CI: It's not a nice to have. It is now a mandatory part.

MW: Yeah. As we have more and more projects that are taking longer and longer, and more and more demanding business users, that the idea of automating it is just a logical step. That was fantastic.

We're also seeing a lot of our competitors starting to talk about data warehouse automation as well. We've been circling it and using different words to describe the same thing, and we're seeing people now starting to talk about data warehouse automation. We're going to see TDWI is running a data warehouse automation course for the first time in Orlando. It's really a concept whose time has come.



CI: Yeah. I'm thrilled for you. It is absolutely a validation of your direction. It's also a tremendous boost to a little company to have a big company like Teradata take you on.

MW: We were very pleased.

CI: I'll bet you were. Unfortunately we're out of time. That's it for the BBBT podcast. I'm Claudia Imhoff, and it's always a great pleasure to speak with Michael Whitehead of WhereScape. Thanks again, Michael.

MW: Thank you.

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