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:	**07-08-2016**
Host:	**Claudia Imhoff**, Founder, BBBT
Guest(s):	**Michael Whitehead**, CEO
	Neil Barton, Senior Architect
Run time:	**00:14:50**

Claudia:	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 Michael Whitehead, and Neil Barton. Michael is the CEO, and Neil is the senior architect for WhereScape, so welcome to you both.
Michael:	Thanks for having us here, Claudia.
Neil:	Thank you. It's nice to be here.
Claudia:	All right, Michael, let me start with you. What's happened in the last five years with WhereScape, and what remained constant? You mentioned a whole bunch of changes that have twisted our business, and made new terms, new ways of doing analytics, but some things have remained constant, so let's talk about those to begin with.
Michael:	Yes, we were first here five years ago, and one of the biggest changes is actually terminology. We've talked all about data warehouses. Now we talk about data warehouses, data hubs, data lakes, data vaults, data marts. It's not just semantic. There actually are some differences.
	Also data – data warehouses were structured, and now we're in unstructured and semi-structured. We've got usage changes where the data warehouse used to be just a system of record, now it's used for lots of different things. Power base has moved from IT to users, and style, there's now a lot more self-service.

Some of the things that stayed the same? These systems still take too long to build, and they're still too hard to change, and governed data is still really important to keep people out of prison.

Claudia: That last one's pretty darn important. How have these things impacted the direction of WhereScape, then?

Michael: We've kept to our core principles, and our core ideas about delivering value quicker. We used to talk a lot about delivering warehouses faster. Now we talk about delivering value much faster. That's quite a big change. It's meant that we've taken a far wider view of the world, and freed us up to go different directions in the product.

Claudia: I think the other part of that is that it's no longer the data warehouse. You now have to automate the construction of, let's say, a data lake and a data warehouse, or a data mart and a data lake, or something like that. It's a very fluid environment, in some respects.

Michael: Yeah, and we have to be very aware that there are multiple silos out there. The transformations of data are no longer done just that the data warehouse layer. It's also done in self-service tools. It's also done in really powerful front-end tools like Tableau.

Claudia: Let's put some boundaries around data warehouse automation. What do you think data warehouse automation is not?

Michael: That's a good question. Data warehouse automation isn't the same as ETL, and it isn't the same as ELT. Extract, transform, load, that's only one part of the process. When you look at data warehouse automation, it's all about building things out, stage tables and load tables, and building something that people can use. It's not just about the data aspect of it. While ETL, ELT is a component, data warehouse automation goes much, much further.

It's also not agile, although in my opinion, you'd be crazy to do an agile project without an automation product, because an

automation product enables you to do so much more per sprint, and not have to spend sprints on going back to fix up infrastructure.

It's also not a case tool. A lot of people view it through a case lens, of defining what you're going to build, and then using a software product to produce that, to generate the code and things. There are some similarities, but there's not the religion associated with it. We're not trying to create something that you can just build in the metadata layer.

Claudia: I think it's important to go back to: it's the process. The whole process. ETL is certainly a part of it. You don't ignore that. If they have an ETL tool, fine, the more, the merrier, right?

On the other hand, there's so much more than just extract, transform, and load. You have to design the environment. You have to understand the metadata behind what's going on, and sometimes that's left out. That whole documentation of the process is missing, and unfortunately, people have a really difficult time if a developer leaves, or if a DBA leaves. What was he doing? What did he do in this environment? That's where you fit in.

Michael: Yeah, in a lot of ways, data warehouse automation keeps you honest, that when you look at it through an ETL lens, it becomes a very technical part of it. I have just got to move data from here to here, and you miss the why, and what am I trying to achieve, and how am I getting to value part?

The data warehouse automation is all about, how can I deliver value faster? It does bring in that user aspect to it and it brings in that real, hanging over your shoulder, am I doing the right work? Am I spending the time on the right places? I've only got a finite amount of time or resource to spend on this. Am I using it correctly?

Claudia: It really puts a bright spotlight on value. What kinds of customers does automation resonate with? You talked about a number of case studies, I guess. Tell me about those.

Michael:	We've got a lot of really good case studies on our website. There're probably six standard usages for data warehouse automation. The first is building a new data warehouse, and that's great. We don't get the opportunity to do that very often these days, because most organizations have an existing warehouse, but sometimes it's just better to start again. So building a new one is one use case.
	Incorporating big data into an existing data warehouse is something that's coming out, that's more and more common, where the warehouse can deliver a lot of value, but be aimed solely at structured data. Sometimes, that's using big data techniques like offloading to big data to get more performance. Sometimes, it's bringing in different sorts of data.
	Another one is migrating an existing data warehouse. Regardless of how you view them, if you turned off a data warehouse inside an organization, you often have a problem. Quite often, people need to migrate and change it while still using it. That's a really tough problem. It's like remodeling a house, changing your roof, and your foundations, while still living in the place. It's tough.
	The last couple we see as common is, doing rapid prototyping. Putting something in front of someone, using the value of automation for how fast you can build something, to build multiple prototypes.
	Finally, quite a simple one is, sometimes organizations have got a warehouse that they like, but it's just too hard to maintain, and too hard to change. How can I take all that manual, hand coded ETL and just put it into a managed environment? That's a common automation use case, as well.
Claudia:	And a really important one. Why don't we talk about what differentiates a data warehouse automation technology?
Michael:	The first thing that differentiates automation is that we view everything through a time lens. We look at everything in terms of, how fast can I do something? If I spend time on it now, will it save

me time later? That's in the back of our heads whenever we're doing R&D, whenever we're making changes to the product; it's how it's going to impact time.

A couple of other things that are really important to the automation land: one thing is, if you can't do it manually, we can't automate it. If you bring us a problem as we're working with customers, we're all about, "How are you doing that now, and can we make that process more efficient?"

Claudia: In other words, the process has to exist for you to automate it.

Michael: Yeah, and that's really where we fit in, as well. The automation tools are generally used when you want to do something on an on-going basis. It's not for a one-off thing. It will get used for that, but the value is the ability to make changes later. You do that when a process is important, and valuable.

That's probably the third differentiation that's worth talking about, is that automation vendors are often not the first vendor to support a new, exciting use case that's come out. We're not the new, shiny object. We're the, "How do I make use of this object to deliver value to people?"

Ask us about technology and we'll go, "OK, how is this being used?" It's that question of, "How are you getting value from it?" Once we understand the value of that, we go, "Right, let's now automate that."

Claudia: All right, the last question for you, and then I'm going to turn to Neil. You had a big announcement at the BBBT. I'm so thrilled that you chose us to announce it. What was it?

Michael: This is really exciting for us. We've announced a strategic partnership with Dan Linstedt, who is the inventor of data vaults. This is just massively exciting for us, because we see so much demand for data vaults. We've got customers all over the world, US, Europe, Asia-Pac, building data vaults now.

Page 6 of 10

In some ways, it's been there for a while, but I think with data vault 2.0, it's really just taken off. What we've announced today is that we're going to be working with Dan. He's going to be helping us on the engineering side, so to make sure that we've got the best product for data vault. That's our challenge, and with Dan's input, we're going to give it a go.

Claudia: Very good news. All right, Neil, let me bring you into the conversation now. You now support any methodology and data modeling technique, for example, third normal form, star schema, and now, of course, data vault – any of those used to create the analytics repository. Tell me about the two products that you currently offer.

Neil: Absolutely. We have our 3D product. 3D stands for our data-driven design tool, and we really use that for doing discovery of source systems, to understand what their structure is, and what the attributes are. We also can, with 3D, profile the data within those source systems, to look for dirty data, understand unique values, null values, and any other potential issues we may have when we start to load data into the data warehouse.

From there, we can then build models, be it a star schema, or a logical model, a third normal form, a data vault, whatever you want. We can build those models from those designs, and then we can document those models, and also generate and publish those models into our other product, WhereScape RED, which is what we use for the code generation, and the deployment on the data warehouse platform itself.

RED's responsible for managing all of the schema objects. It will create, manage, alter, drop tables, views, indexes, all of the objects within the data warehouse. It's also responsible for the code generation. It'll generate the procedural code within the database engine, to do all of the data processing.

From there, we can then operationalize the data warehouse through our scheduler. We can run our jobs on a frequency,

whether it be nightly, or hourly, or every 15 minutes. Load the data in, we do full auditing and logging of those processes. If we have any issues, we can go back through all the logs, resolve the issues, and restart the jobs.

Some of the overlooked features of data warehouse development, documentation and impact analysis are the last two components of the RED product. We can, with literally the click of a button, generate a full documentation of our data warehouse, including all of the metadata we've captured, the descriptions, all of that useful business information around what's in the data warehouse.

Then, from a second order cost perspective, if I want to change the data warehouse, what's my impact? We can do diagrams, and charts showing track back and track forward diagrams, impact analysis. For example, if we're changing the size of a column on a source system, we can understand what all of the objects in the data warehouse that will be impacted, and use those to scope the effort to make those changes, and understand which tables need to be modified, and which code needs to be rebuilt.

Claudia: Absolutely brilliant. I love asking this of the guys that do the demos. What's your favorite set of features?

Neil: For me, I've got a few favorite features, but I think the big ones for me are the fact that we now have templates within our product, so we can generate code for our standard platforms. We can also generate code for new platforms, using templates which allows us to more easily enhance and add new features as those new platforms come out. As those platforms add new features, we can use our templates to generate the code, automatically.

The two big ones for me are really the documentation and the impact analysis, always overlooked in data warehouse design, but critical if you want to maintain a data warehouse over its full lifespan.

Claudia:	I think the one that I was just blown away with, of course, is the lineage tracking across all of the data repositories. We talked earlier – it could be a data warehouse, a data lake, a data mart. You track across all of that, right?
Neil:	Absolutely. Gone are the days of a single database for a data warehouse. We now have heterogeneous platforms, whether we're loading data from, using an example from earlier today, from Oracle into Hive. As data processes through that Hadoop ecosystem, maybe it goes into a data warehouse running on SQL server, or Teradata. We can point to any object anywhere in that environment, and actually do lineage across all of those platforms, so we really understand the full impact, as we move data between the various platforms.
Claudia:	Nice. Last question, what does the future hold for WhereScape?
Neil:	For us, it's about adding more metadata, about more products, and also understanding and integrating with other products as we have a more diverse ecosystem now, for these data warehouse environments. The last one would be the automation capabilities in the big data space. That's certainly our focus going forward.
Claudia:	Yeah, focusing a lot on big data.
Neil:	Absolutely.
Claudia:	Unfortunately, we're out of time, so that's it for this edition of the BBBT podcast. Again, I'm Claudia Imhoff. It's been such a pleasure to speak to Michael Whitehead and Neil Barton of WhereScape today, so thank you both for speaking with me.
Michael:	Thanks, Claudia, and thank you for putting on beautiful weather.
Neil:	Thank you, it was a great time.
Claudia:	The weather was all my doing. You're welcome.
	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!

Page 10 of 10