



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:	Infobright
Date recorded:	October 11, 2013
Host:	Claudia Imhoff , President, BBBT
Guest(s):	Don DeLoach , President & CEO Graham Toppin , CTO
Run time:	00:19:01
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 Don DeLoach and Graham Toppin. Don is the CEO and President and Graham is the CTO for Infobright. Welcome to you both.

Don DeLoach: Thanks. It's great to be here.

Graham Toppin: Thanks, Claudia. Great to be here.

CI: It's always good to have you guys here. You always bring up such interesting things, and we'll get to that in just a moment. However, the last time you were at the BBBT was a little over a year ago. Don, I'm going to start with you. What's happened?

DD: A pretty good amount. It's been a busy year. Last year when we were here, we talked about us introducing a new product called the Infoplance, talked about some of our direction in terms of where we were bringing the product. We talked about really looking at what was going on in the markets and making sure that we were responding in kind. Since that time, we've really fairly dramatically grown our customer base. We now have the Infoplance out and in production, and our experience with that has been very positive. It's delivering the type of results that we had hoped it would, so that's good.

Our growth, especially in the OEM markets, they are our key providers that are servicing the mobile network operators. At the start of this year, we set as a goal that we would be doing business with all of the major suppliers into that space. At this stage of the game, we've either covered that off in its entirety or at least are very, very close with a lot of the household names that you can imagine. That's been good.



We've also filed a number of patents. As we discussed today, we are going down the path of what we call scalable knowledge, which we find very exciting. We've grown the employee base.

All in all, good momentum. I think our investors are happy. Certainly our customers seem to be very happy, and we're going down a good path.

CI: Yeah, it does sound like it. You've also changed your path a little bit, or you've brought in some new terms, at least this time around. Let's talk about these. You said that Infobright is now focusing on machine generated data and this new thing called the "Internet of Things." Why don't we start with those two? First of all, machine generated data, what is it? Give me some examples. Then, if you don't mind, a little primer on the Internet of Things.

DD: Sure. Our focus on machine generated data is not new. I think that we began to narrow our scope to machine generated data in mid to late 2010, and have been building on that ever since. The thing that's new is really not a function of us, so much, as it is the market and the growth of what people are characterizing, largely, as the Internet of Things. Cisco calls it the Internet of Everything. Ericsson calls it the Networked Society. Fundamentally we're talking about a connected world, where the increase in the sophistication of sensor technology, combined with the increase in the capabilities and the sophistication of communication technology, is allowing people to connect a world of both human and non-human devices, to enable things like smart grids and smart cars and smart cities and mobile health and a raft of new use cases that, really, will change life on Earth at its most basic level. It's interesting, and we see that the fuel for all of this is machine generated data, hence our very very keen interest in this trend.

CI: The idea, again, machine generated. It could be sensors. It could be anything. It could be RFID tags. There are so many things that are spewing information at us, right and left. Our cell phones, everything on the Earth, literally, is now capable of sending back signals to the mother ship. A lot of it is real time. A lot of it requires real time sensitivity, if you will, to what's going on, but a lot of it is also based on historical patterns, historical trends that we've seen in the past. Where do you see Infobright fitting into this Internet of Things, then?



DD: Infobright is, basically, focused on machine generated data as a persisted store for that data. So all of the accumulation and analysis of very very vast amounts of this data, works very well with us. We don't try to be all things to all people. Our approach is not as a general purpose data warehouse, but is really strictly focused on storing and analyzing machine generated data. Whether it's call data records, or sensor data, or network events, or Web logs, or IT logs, or even things like financial prints, or historical transaction records. All of these tend to go into a structure that works very very well with the underlying mathematics that drives what we do.

We have some unique capabilities. Our customers seem to like it a lot. It's exceptionally well suited for our OEM customers that are solution providers to things like the mobile network operators. When we look forward into the Internet of Things, the very things that we're doing right now, seem to be at the forefront of where that whole market is going, so we're excited about servicing the needs of that market.

CI: Excellent. All right. Well, I get the benefits. I see them. I understand them. I also think that there are a lot of challenges for the first one, certainly. You are now drawing information or data from a multiple, or a myriad, of sources. There's always the problem of integration. There's always the problem of the quality of the data itself. First of all, what kinds of challenges, other than those, are you seeing, and how are companies overcoming these rather large hurdles?

DD: For a start, we're at the infancy of this. I think that while there are a lot of people talking about it, and a lot more today than there were two or three years ago. If you went to Mobile World Congress two years ago, maybe a tenth of the vendors were really talking about it. One year ago, it was half of them, and most recently 90 percent of the exhibitors there are talking about the Internet of Things. You can see that it's coming on strong, but yet, what's coming on strong, as a notion of where the world is going, runs ahead of the practical implications of making it work. While people are gravitating to the concepts of the Internet of Things, they're still struggling with exactly what it is, how it's going to happen, and the practical implications of that.



To that end, the struggles, I believe, are far reaching. It's everything from data governance, privacy issues. Who owns the data? Where's the data stored? Who is the steward of the data? Is the owner of the data the steward of the data? Maybe so, maybe not. What are the security implications? It's one to hack into somebody's bank account. It's another thing to take down Cleveland, as a result of what's going on.

There are a lot of impediments, and I think that there are organizations, like the W3C, which is the Tim Berners Lee initiative that's based in Cambridge and in Europe, and there are lots of consortiums. If you look at what's going on in Europe, there are all kinds of bodies, particularly interested in the governance issue.

These struggles will work themselves out. The market acceptance will demand that these problems work themselves out. I think that we will get to a point where the Internet of Things becomes a reality, but the struggles are big and meaningful, and it's going to take some time to work through it.

CI: It seems like, you're right, first all, standardization is going to have to play a huge role. If we really want to connect different things spewing information, they're going to have to have some standards between them. The other one is that it seems like we need to start small. If someone wants to get into this environment, they're not going to take on the city of Cleveland, poor Cleveland. They're probably going to take a small section and see if it works, or a small piece, and maybe some of the technologies, not all of them. Would that be a recommendation that you would see, "Yes, I think this is a good way to get going?"

DD: Whether it's a recommendation or not, and yes, I would say that's a recommendation, I'm always a big believer in "start small and iterate." I think, in reality, that's what we're seeing. We have customers, like Fuseforward out of Vancouver or Opower in the smart meter space, that are doing things that are really really impressive in the context of advancing the notion of smart grid technology and the type of information that can be derived from, in both cases, energy usage, that people wouldn't have thought about before. If you talk to the folks at these various companies,



they'll tell you it's but a small step into a brave new world, where there's going to be all kinds of other developments as we go forward.

I use the example, in energy delivery alone there's 23 different, probably more than that now, there's 23 different unique types of sensor technology that exist between the time the power's generated until it gets into the home, that all work together to contribute a more efficient and effective delivery of energy. The amount of savings and what that represents in terms of life on Earth, really, in some ways, is profound.

It's clear that this is going to happen, but it's not going to happen without some of the obstacles that we've talked about being addressed along the way.

CI: Very exciting. A little scary, but certainly a very exciting time. Let's bring it back around to Infobright. You are about to launch a new feature. It's something called an Approximate Query. I thought it was very interesting. If you don't mind, please explain what it is.

DD: Sure. At a high level, approximate query really leverages the underlying architecture of Infobright, which is really based on rough set mathematics applied to granular computing. That's only important to the extent that we build a metadata layer using set theory, and in doing so, we create a mechanism by which you can query a vast amount of information about the data before you ever touch data.

CI: Explain what you mean, without touching the data. What are you looking at?

DD: When we ingest the data, we collect all kinds of information about the data, and we organize it in a way that is leverage-able using rough set mathematics.

CI: So it's basically the metadata about the information.

DD: It is, and it sits in memory. When we query the data, we go through a multiphase evaluation, called the rough evaluation, and the exact evaluation. The rough evaluation is basically all but instantaneous. It happens in memory against this data that's organized in a way that we can



get at it very, very, very effectively. What approximate query does is it goes through the rough evaluation, grabs information, and then applies statistical sampling to deliver an answer that's an approximate answer within a tight tolerance and a confidence factor associated with it. But we don't actually ever touch any real data. Therefore, queries that might take 30 or 45 minutes over super large data sets in complex queries might return in 5 seconds.

Let's say, for example, I'm asking for the top 10 revenue producers on my website, over the last 36 hours. The way that's derived is through a number of different calculations, and say that the data set is very very large. Well, that might be a long running query, but if I could get an approximate answer to that, my top 10 list might actually, in reality, be the first nine and then the eleventh. But that's probably good enough if I can get the answer in five seconds, to take action that I want to take. We don't always need an exact answer. Approximate query is really a breakthrough in terms of rethinking, "What data am I asking for? What price am I willing to pay? How much time am I willing to wait, and what am I going to do once I get the answer?"

CI: If the answer is, "I do need it to be exact," what happens?

DD: Then not only can you go after the exact answer, but in many cases having the benefit of what the result set of the approximate query was, can probably help you reshape your query so that when you do ask for an exact answer, you ask for one that is much more effective and efficient than what you would have asked not knowing what you were dealing with.

CI: Exactly. Perfect. All right, Graham. Let me bring you into the conversation a little bit. You talked about Infobright's alternative to the traditional deployments, and you might want to mention what some of these are. You called it the Scalable Knowledge, and I believe it is trademarked, just to make that clear. Why don't you go ahead and tell me about what this is?

GT: I think when we were imagining what our next generation of machine generated data analytics would look like and we really confronted, as Don had alluded to, the changes in scale and complexity implied by the Internet of Things. We realized that we couldn't just take the paradigm we



have today and simply extend it, naively, and achieve the types of results we wanted. A lot of the way that people think about data today, is in what we call data silos. So silo being exactly that. Walled...

CI: Isolated, walled off?

GT: Exactly. Walled off, and my reality is defined by things like ETL. The integration problem is one that has been largely skirted. In other words people try to avoid large heterogeneous landscapes. With scalable knowledge, what we wanted to do is reimagine how people thought about the data, and how people would integrate the data. Again, when we looked at it, we recognized that we weren't going to try to create a panacea, where all data searches could be seamlessly integrated, seamlessly consistent, and seamlessly queried. We also realized that that wasn't always required.

Instead, we took our core technology, which is the knowledge grid. The knowledge grid being, I guess the way that I always describe it is; the maximum amount of knowledge stored in a minimum amount of space. When we looked at the data for a given query set, our goal with the knowledge grid is always to store that maximum covering set of information, in the minimum amount of space. We said, "This would be a fantastic unit of work for a large analytical work set." That became the knowledge fabric, and how it was all stitched together, where we would either ingest data, or leave it where it was would be the data fabric. That choice is important.

CI: The data fabric itself is made up of the small knowledge grids of the data wherever it resides. It doesn't have to reside in Infobright. It could still be in its Oracle database or its whatever, DB2 or MySQL or wherever, right?

GT: Exactly. That's correct, yup.

CI: So these little knowledge grids come together as a fabric basically. You are now querying this morass of info grids, if you will, to get the answers that we just talked about for approximate queries and so forth.

GT: That's correct, and the premise is that for these large changing data sets, where we started getting into petabyte, exabyte, data sets, you can



certainly invest a ton of effort and money into large infrastructure, large teams, and there is a definite market for that. We're not trying to contest or change that. But we also recognize that there is a space that needs agility, that needs to be able to get up and running very quickly, where the models are not going to be things that I can change over the course of a week, that I need to respond very quickly. These are the businesses, as Don had alluded to, finance, ad tech, real time bidding, where we need to respond very quickly.

We're going to compete not so much on the products that we develop and the services that we deliver but on the analytics that drive them and how we shape what happens with those in the future. That's really is what scalable knowledge is geared toward.

CI: And the fabric itself. Fascinating. What I like about it is you don't have to redundantly deal with the data. You leave it in place. It's not going to have to be duplicated here, there, and all over the place. You just simply analyze it in place. That's something that a lot of companies, as they get into this Internet of Things, they are going to have to face that problem. You can't duplicate the big data all over the place.

GT: Exactly. The knowledge about the data is what's important to duplicate. The knowledge is what you want to be able to replicate and move around.

CI: Excellent.

GT: That's the key differentiator for us.

CI: Very exciting. I think the company is right on track, and I love the fact that you guys stay so focused on what your core competency is. That is your best selling point.

OK. Well, I think we could talk about this for another set of hours, but unfortunately we're done. 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 Don DeLoach and Graham Toppin of Infobright today. So thank you both very much.

DD: Thank you.



GT: Thank you, Claudia.

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!