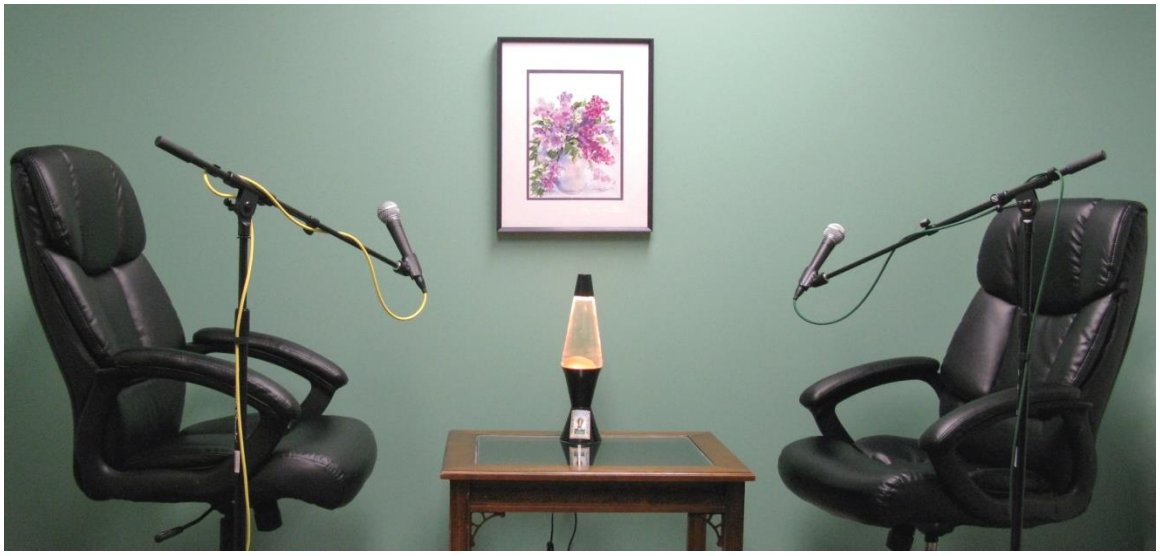
BBBT Podcast Transcript



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:	**Kognitio**
Date recorded:	**August 28, 2013**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Michael Hiskey**, Vice President Marketing, Product Evangelist
Run time:	**00:15:51**
Audio link:	Kognitio Podcast
Transcript:	[See next page]
Transcribed By:	CastingWords

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.

CI: I'm very pleased today to introduce my guest. He's Michael Hiskey. Michael is the Vice President of Marketing and a product evangelist for Kognitio. Welcome, Michael.

Michael Hiskey: Thanks a lot, Claudia.

CI: Well, let's start off with Kognitio's past year. It's been a pretty banner year for you guys. Why don't you tell me a little bit about it?

MH: It's been a really great year because of the growth in big data. Everybody is talking about big data and we've had a long history of being able to do things with it. I'm very proud that we've doubled the company from year to year from our fiscal year 2012 to fiscal year 2013, and we're hoping to double again in our fiscal year 2014 that ends in March.

CI: Some of the things you've done, you've, as you said, doubled your revenue, you've brought in new technology. I mean, it's really been a very exciting year.

MH: We really started on this idea about two years ago, that what's interesting in the market is an analytical platform. We came up with the reference architecture and product development has been delivering on functionality that meets those needs. The sales team and the marketing team have picked up with how those needs fit what clients are doing, what prospective clients would like to do in the marketplace. That's helped us really shift more of a cloud focus away from an appliance focus, it's helped us double bookings, it's helped us grow revenue. It's helped us get into some interesting new markets, for example, companies that provide analytical information services are a big hit for us.

CI: Yeah. Now, one thing that you mentioned was the extended data warehouse architecture, that's something that Colin White and I have developed and it's still evolving. Basically, it consists of three different environments. There's the operational environment, where we might have streaming analytics, for example, as well as our operational systems. It also has the traditional data warehouse environment which consists of an enterprise data warehouse and various data marts and so forth. Then, there's a third area that is the investigative computing platform, which is really where Hadoop or NoSQL types of technologies would reside.

You took it to heart, which I'm very pleased with. You placed Kognitio where you thought it would shine in these three different environments. Why don't you tell me a little bit about that?

MH: Yeah, Claudia. I think the important thing that all of the companies are trying to do today is understand in their extended data warehouse infrastructure and how they look at their information management architecture overall, how can they use the various different platforms that are most efficient? How could they most efficiently store data in a range of platforms? How could they most efficiently process data or analyze data in another range of platforms? What we're seeing is the disaggregation of how you store data from how you analyze it.

They don't necessarily have to be in the same platform, and in a big data environment, they almost necessarily have to be in discreet platforms. So we looked at the extended data warehouse architecture and said this is exactly what the clients were speaking to or trying to do.

Where the Kognitio platform fits in is that it's an in-memory layer that rests on top of your various persistence and storage layers whether they be Hadoop or your enterprise data warehouse and feeds into the tools that your business users are very happy with because changing the tools that business users touch every day is probably harder than changing their religion.

CI: Exactly. So you basically play the role of an accelerator, and it doesn't really matter what you're accelerating, what database, whether it's an EDW or a Hadoop platform or whatever. You just accelerate it.

MH: Very much that case. In a number of our clients, they might have an existing data warehouse. They might be looking at Hadoop or have Hadoop production environments. But either way, we want to make the provisioning between Hadoop and the data warehouse an IT decision and leave the business decisions with the line of business users that continue to use their tools which happen to write a lot of SQL.

CI: Yeah, smart. Now, something else that you, Kognitio, have come up with I found quite fascinating, and that was an analytics center of excellence. Basically, it's a data science lab. Why don't you tell our audience about that?

MH: Well, the data science lab takes two facets. The Kognitio Center of Excellence is our internal organization that's led by our head data scientist. It carries two or three PhDs, some interns, a couple of physicists, people who are working with our core clients. They're also working with core universities that we interact with and developing use cases for how to use some of the really advanced analytics because what we've seen is it's coming back to math. It's advanced analytical algorithms that happen to run against really large sets of data. The people doing that -- the statisticians, analysts, data scientists -- don't really care whether the data sits in Hadoop or an EDW. That's irrelevant to them. What they care about is bringing big complex algorithms in line with the data.

So the data science lab takes two instantiations. Companies are building their own data science lab functionality where they put a handful of data scientists in a room together, and they're supposed to come up with things that help the business. So they experiment. They develop. They put things together.

But their goal needs to be to graduate those things to the rest of the business because, regardless of how many data scientists we have, we'll probably have more business analysts. So the data scientists need to be fueling the business analysts. The other facet of the data science lab is actually an educational go-to center of excellence, and that's what we've created internally to be a resource for our clients as well as the organization at large.

CI: I'm fascinated by it. I think it's a great idea. Now, the other thing you have, of course, is your analytical platform reference architecture. Different layers, different functions in those layers. Briefly just talk about the layers first, and then we'll get into the next part of the question.

MH: Absolutely. At the basic layer, the bottom of the platform, is however you choose to persist your data. You could use data warehouse or Hadoop. You might use the free storage environment that Kognitio gives away, which is a disk based storage. That's the base level, and that remains an IT decision. On the other end, at the top of the platform are your BI tools, your Excel, or your proprietary applications, what the users touch. We don't want to move those either. So running Kognitio in between those two layers means I don't have to change the BI layer, the application layer, and I don't have to change the persistence layer. In fact, I could add more layers of persistence or more different types of persistent data storage.

You can continue to use Kognitio as a universal adapter in the middle. In this way, my business tools don't have to write pairwise connections to every different data source. They can just write SQL queries that fire at Kognitio, satisfying them in memory, having taken the data from whatever persistence storage it needs to.

CI: So that middle layer is all memory, right?

MH: It is. Kognitio built an in-memory analytical platform before it was cool. We added a disk subsystem later, but the idea was how could we do -- the initial client was actually for risk analytics -- very fast analytical processes, effectively a data warehouse, that would run entirely in memory and provide that with very fast answers that could be queried on the fly for ad hoc type reporting.

CI: Everybody and their cousin seem to have now added in-memory to their database offerings and so forth. How does Kognitio differentiate itself from all these other companies?

MH: We call it in-memory washing. It was cloud washing a few years ago. This is in-memory washing. All of a sudden, everybody is talking about in-memory. I'm a fan of saying that a database that can run in memory is

not an in-memory database. Something that's designed from the ground up as Kognitio is as an in-memory database has a completely different design principle. What I mean to say is that on the inner loop of every query in a conventional data system has to ask, "Is my data in memory? Where is my data? Can I get the data, bring it into memory, then do the process on it?" That's the inner circle in every query. It might run thousands of times throughout a query. An in-memory database doesn't have that. It knows all the data's in memory. That in and of itself makes the code path perhaps 100 times shorter.

The next thing that's important is that this isn't just a cache. This is actually satisfying all the processes in memory. All the RAM storage, all the tables, everything the system needs is all in memory. There literally is no reliance on disk, and that's very important.

Then last but not least, once you have data in memory, that is necessary but not sufficient. Memory is just another storage area, just another parking spot for my data. What becomes important is how many CPUs can I drive? You have to think about the amount of data you have per CPU because the goal is to keep those CPUs 100 percent busy for 100 percent of the query as often as possible because it's only the CPUs that give you processing power. Always is, always was.

CI: All right. Let's talk about some of the new features coming down the pike for Kognitio, things like MPP in-memory code execution. Let's start there.

MH: This is the most exciting thing that Kognitio has ever done. In-memory code execution in a massively parallel format means that any standard binary, any standard programming language, effectively anything that could run in Linux and take a stood-in/stood-out line of data or set of data into it can run within Kognitio in an MPP fashion. Now, what's important to note here is we haven't created a translator for R, or for Sass, or for Python, Pro, or anything. What we've done is we've just created a little hole, something that allows the system to run, let's say, thousands of R scripts across hundreds of nodes all in parallel all at the same time and run that within the database as an organizing principle.

Finally, the thing that's really important in that vein is this is all wrappered in SQL. So there was a SQL and no-SQL war. I think SQL won because it's what business users have and it's what the organizations have.

By wrappering things like complex R algorithms inside of SQL, it means I can call full functions, full product forecasts, that are literally thousands and thousands of lines of code. I can call with a simple SQL statement that could be running in, for example, MicroStrategy, Logi Analytics, Cognos, or any BI tool.

That means a standard business user can effectively be doing high-end complex analytics, which were previously only the prelature of data scientists from within their standard BI tools with no understanding of big data. Perhaps the data sits in Hadoop or even the R algorithms underneath it. The statisticians and data scientists can update those algorithms, load them into a system, and the business users can use them as they will.

CI: Yeah, which is exactly what you want, isn't it?

MH: It's really power to the business users.

CI: Yeah. Well, now the other big news is that you're about to launch version 8.1 -- 8 just came out. Now, you're about to come out with 8.1, so what are the features here?

MH: This is a release that was so big. Version 8 is the most important release in our history. We had to split it into two. In the first release, version 8.0 came out in accordance with the Hadoop Summit, Hadoop connectivity that a lot of clients have been using for over a year now. There were external tables so I can see data that sits in Hadoop or other systems, high-speed imports, high-speed exports, et cetera. As we moved into version 8.1, the release that comes out in September, we started doing a lot more performance optimizations. We added compression in memory, and we did some tricks that give us speed. Now on the TPCDS benchmark version 8.1 will be twice as fast as our old version 7 was just last year, that it gets up by doing advanced partitioning, things like RAM prejoins, RAM presorts. It comes from having a lot of experience.

And make no mistake. It took us years, in some cases a decade, to perfect some of the things we do with RAM because it's a lot of work. It's very intricate and very involved. So using all of that maturity and using all of that that we've learned over those years, we've been able to build that into version 8.1 and continue the good work we started in version 8.0 to make this just a blockbuster couple of releases for us.

CI: All right. Well, last thing, you mentioned that the appliance part of your business is OK, but what's really taken off is the cloud side. You have two ways of implementing the cloud. Let's end on that.

MH: It's really two flavors, so Kognitio cloud has been available in some way, shape, or form since 1993. The ability to run as a platform as a service in the cloud is engineered into our core product. With the advent and really uptake of Amazon Web Services and Amazon Marketplace, there are more technical users than ever experimenting and playing with software. So it was important for us to get that out into the market as the tip of the arrow. What we find quite frankly is that a lot of clients are playing with a POC or even a pre-POC. They're doing a self-run POC -- we would call it -- in the Amazon environment, running it, using it, and then coming to us and saying, "We like that. We'd like to actually run that in something that's a little more secure. We're a little concerned about Amazon. Great service gets expensive after a while. Do you have something else?"

For them, we have a private cloud platform as a service with tier three data centers, and that's been an effective solution for a number of clients who use that and one more security. Then still in some cases clients really rather prefer to buy just the software or an appliance, and we support those people as well. The interesting shift is that the cloud is the tip of the arrow, and it's where more businesses are looking every day.

CI: It certainly has gotten traction lately. 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 my friend, Michael Hiskey, of Kognitio. Thanks so much, Michael.

MH: Thanks a lot. Happy to be here.

CI: I hope you enjoyed today's podcast. You'll find more podcasts from other vendors at our web site, 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!