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:	**SAS**
Date recorded:	**July 31, 2015**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Mike Frost**, Senior Project Manager
	Wayne Thompson, Chief Data Scientist
Run time:	**00:13:26**
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 so happy to introduce my guests today. They are Wayne Thompson and Mike Frost. Wayne is the chief data scientist and Mike is the senior product manager for SAS Institute. So welcome to you both.

Wayne Thompson: Thank you, Claudia.

Mike Frost: Thank you, Claudia.

CI: It's good to see both of you. Wayne, let's start with you. SAS has got a long and really innovative history. For 40 years you've been doing analytics and all kinds of things like that. If you don't mind, give me a quick overview of these 40 years, or almost 40 years.

WT: I've been there 23 years myself, Claudia. But SAS is a great place to work, first and foremost. We're really centered around analytics and data management.

I think what you'll see, through the progression of our company, is that we've moved more from technologies into a solution based delivery. We have, now, tools for fraud, also for cyber security, retail, credit risk, and so-on.

Of course, we've been around for a very long time and we have customers throughout the globe. I think that you'll see that one thing that's happening with SAS is it's moving away from more of a closed system to being much more open and interfacing with platforms like Hadoop, and having open APIs until you can actually use SAS on the cloud.

The last thing I'll mention is that I think, too, what's really exciting about SAS today is we have a much more renewed interest in academics. We have our SAS Academic University Program. SAS is free for anybody that wants it, and we're really working closely with the universities to get more SAS being used.

CI: Perfect. Wonderful. 23 years. Wow. Mike, let me bring you into the conversation because the whole purpose of today's session was to talk about SAS and Hadoop. The first question I have for you is: "How does SAS fit within the Hadoop ecosystem?"

In other words, there are some major players, distribution players, like Hortonworks and Cloudera and MapR and so forth. The question is: "Do you play with all of them or have you picked out your favorite child?"

MF: That's a great question. We actually do play with, certainly, all of the major players and we'll keep an eye on any up-and-comers. We have business as well as technical partnerships with folks like Cloudera and Hortonworks and MapR. We're looking to expand those partnerships that we also have with pivotal and support distributions like BigInsights from IBM.

What that gives us is the ability to make sure that no matter which Hadoop distribution our customers choose, regardless of what the business reasons for doing so, they can expect their SAS applications and solutions will certainly work, and work well, and they're designed, from the ground up, to work with those environments.

CI: Let's dig into it a little bit more. Let's talk about the SAS architecture on Hadoop. How about if you describe it for our audience and some of the advantages that you see with this relationship?

MF: Sure. I mentioned earlier that we have these technical relationships. Why that's important is because SAS's architecture is to have a bit of what we would think of as a SAS shell. We call it the embedded process. The embedded process actually installs and runs across all of the nodes of a Hadoop cluster.

It's that piece that allows us to do things like submit SAS code and have that SAS code run within your Hadoop cluster. As your cluster expands and takes on more information, you're able to expand and have the compute capability also expand with it. We're able to augment the capabilities that you would get with the Hadoop cluster through technologies like MapReduce and Hive with capabilities that aren't native to that environment, but are something that most SAS customers would know and are familiar with.

CI: What I find interesting, and it's such an important relationship, is the performance that Hadoop gives you. It was almost like it was made for SAS. Would you agree?

MF: I would say, certainly, the Hadoop architecture is uniquely suited for the kinds of data that customers are now looking to process. We've made sure that you can leverage the performance benefits of that architecture and get that delivered and exposed in part of the SAS environment as well. Yes.

CI: Let's talk about the theme of today's Boulder BI Brain Trust. That was that Hadoop brings the big data but SAS brings everything else. I really like that. It's a wonderful tagline, by the way.

Why don't you walk me through what SAS does? Let's start with SAS Data Loader. You can go through the various and sundry pieces that lead, ultimately, to the end goal which, of course, is the analytic results.

MF: Data Loader, despite its name, does a lot more than simply load data. You can access data that does not exist outside of Hadoop, as well as within your Hadoop environment; bring data into Hadoop, move it off of Hadoop, transform it in place in Hadoop… So your data integration types of capabilities to subset, filter, do those types of things, merge.

You can also augment data quality capabilities that come as part of the general SAS platform and get that within your environment.

Again, all these capabilities are running down inside of Hadoop so you get to take advantage of all those performance capabilities and expose those to users that may not necessarily be ETL experts.

Now that you've got your data loaded, you want to make it available for exploration and visualization. You now use SAS Visual Analytics. The product has been around for a while. What you can do is load the data directly from your Hadoop cluster into memory, so that you can begin your visualization, start doing your explorations, and look for issues that would then inform your initial attempts to start doing the types of statistical analysis that folks like Wayne would do using the visual statistics capability that's part of that environment.

Then, being able to take and operationalize that process using Factory Miner, so that you're starting to expand the number of models that you can build in a fast, efficient manner, and roll those out and start to score them in real time against data, either at rest in your Hadoop cluster or, for that matter, before it even streams on to your Hadoop cluster up in memory using something like event stream processing which he illustrated for us today as well.

CI: What I found interesting about that is something, I think, Wayne, you said, that you are your own best partner. A lot of analytic companies, or BI companies, would claim that they do the end-to-end, once you get it into the database then they go from there to the end product. I'm not sure that that's exactly true, or at least not to the level of sophistication that I saw today. Would you agree, Wayne?

WT: Yeah. I really feel like if you look at the analytics life cycle, we really marry all facets of that process, from data origination into the exploration. One thing I want to emphasize, even though you build and score models, exploration needs to happen in each phase and very often.

We're able to do it in an interactive way, too, that enables you to work the way you think to solve problems.

CI: What I liked was your comment about... yes, a lot of companies can do data integration, or ETL light, if you want to call it data wrangling, but once they've done that they then have to hand it off to somebody else, another company. Whereas with SAS, you hand it off to one of your own technologies. Right?

WT: Absolutely. That's, I think, really the foundation of the SAS platform. What we've done is bring these processes to the data. That's how SAS has evolved. Whether it's an in a relational database, like Teradata, or in a Hortonworks or Cloudera cluster, we work with the data with the analytics embedded.

CI: Wayne, let me stay with you. Your analytics, you commented: "Were made for the ..." I'll put it in quotes: "The citizen data scientist."

WT: Actually, Alexander Linden of Gartner came out with that term. I think you have to give the Japanese credit for that, if you think of the Citizens Watch. The concept was that watch is for everyone and everyday use, and it's easy and it's durable.

That's the way we think about, also, starting to blend our analytics into business intelligence. Making it approachable, not just saying that the business person does it by themselves. They get started and they have a platform to work in concert with the data scientist.

As you alluded to earlier, I think doing some storytelling, building some cognitive learning into some of our systems, is a way forward.

CI: I like the term too, because it's not necessarily a deeply educated statistician that we're talking about. Correct? The Citizen Data Scientist is not necessarily this ivory tower kind of statistician.

WT: Absolutely not. Tim Rey was a good friend of mine, at Dow Chemical, he was the chancellor of statistics there. I've learned, over my years, working a lot, that these people are going to do it anyway.

What we're trying to do is provide a little automation until they get good answers fast. Then, as I said, make sure that we provide the platform to do that storytelling in concert with the more advanced analysts.

CI: You also, rather enthusiastically I might add, stressed that these analytics are interactive, really interactive. Do you consider that a differentiator for SAS?

WT: Yeah, I do. I don't care how throughput is measured on a clock. I want to know how well I can get at the data and turn and do things on a dime. That's the way I solve problems every day, when I talk to a child or a family member, and we work through a problem.

Because this is interactive, the software, running in memory, I address data analysis the same way, very conversationally.

CI: The other thing, Mike, you mentioned Factory Miner and Event Processing Studio. Wayne, why don't you tell me a little bit about what these two do in a little more detail?

WT: I'm really excited about the Event Stream Processing. We brought in a guy, background at AT&T Labs, been in the business for many years. His name is Jerry Bowyer. He has done some really great work on being able to analyze and process data in motion.

If you think of the Internet of things today, whether it's your refrigerator or some big power plant, we've got to analyze and act on the data as it's being collected. That's what that product is all about.

CI: Real time analytics on real time data.

WT: Anywhere, any time.

CI: Excellent. Last question and we'll wrap it up. We've got about a minute and a half left. Let me start with you, Mike, and then, Wayne, I'd love to hear your comments as well. Where do you see the future going for SAS?

MF: That's a great question. I'll tell you, from my perspective the things that most excite me about what we're thinking about is ways to bring analytics to bear for a non-classic analytical problem. We think of data management serving analytics or visualization. We don't often think of visualization analytics serving data management.

But we're actively working on ways to do just that, to take all the things that people turn to SAS for, all the great research that goes into making our algorithms and our models and our people the best in the world, and bringing them to bear to solve those really hard problems that are going

to get even harder, as Wayne mentioned, as the Internet of things takes off. That has me very excited.

WT: I'm just happy to have our 40th birthday coming up. We're going to celebrate SAS 10 next year.

CI: That's going to be a big cake.

WT: Doctor Goodnight said we were definitely going to have a big picnic with lots of drinks. We're looking forward to that.

I think what you'll see, too, in that release is better purpose built software for the cloud. We're really excited about that. In fact, Mike Frost is going to lead up some of that effort.

We're energized. We're not going to quit. Doctor Goodnight takes a daily process in working in the business. I speak with him a lot about our products, especially analytical. The future looks bright. We'll wear shades.

CI: Excellent. That's it for this edition of the BBBT Podcast. Again, I'm Claudia Imhoff. It's been such a pleasure to speak with Wayne Thomson and Mike Frost, of SAS Institute, today. Thanks again to both of you for speaking with me.

MF: Thank you.

WT: No problem.

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!