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:	**Alation**
Date recorded:	**04-29-2016**
Host:	**Claudia Imhoff**, Founder, BBBT
Guest(s):	**Satyen Sangani**, CEO
	Stephanie McReynolds, VP Marketing
Run time:	**00:16:44**

Copyright © 2016 BBBT - All Rights Reserved

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 Satyen Sangani and Stephanie McReynolds. Satyen is the CEO, and Stephanie is the vice president of marketing for Alation Welcome to you both.

Satyen: Thank you.

Stephanie: Great to be here.

Claudia: It's lovely to have both of you again. It was a great session. Let me start with you, Satyen. You started by describing the problem today as, tribal knowledge is an inefficient foundation for self-service data. What do you mean by that?

Satyen: To use data is not, unfortunately, a simple exercise. You need a lot of information about really trivial and very minute things like what the field name means, and what the values represent, and which database to go to, and how to log in, but you also have to have a lot of business context around which tables to use, and what the business process happens to be, and what the definitions of churn are.

Understanding all of that information, understanding all of that tribal knowledge is not something that is easily documented today, in any complicated enterprise. As a consequence of that, people have to rediscover and reinvent the information over, and over, and over again.

What people do today is, they ask people. Sometimes they read documentation, but sometimes that documentation is stale, and

Copyright © 2016 BBBT - All Rights Reserved

old, and it doesn't work, and it's inappropriate, and it can't be found easily, so people ask fewer questions, because it's just too expensive to ask questions. It's too much time.

Either they ask fewer questions, or the questions they do ask, they ask inappropriately, or they get the incorrect answer. We really see that there is a huge problem in learning about data, and as a result, people can learn about their enterprises and their businesses much slower than they otherwise would.

Claudia: The solution that you suggest is something called a data catalog. How would you describe, or define what a data catalog is?

Satyen: We use catalogs in our everyday lives, in a variety of different contexts. One great example of a catalog is AirBNB. People crowdsource and volunteer information about space that they have in their homes, that other people can rent and use for a night, or a couple of nights, or a couple of weeks.

Amazon is another catalog for products on the web. People can buy products on Amazon, and you can see any product, from anywhere. LinkedIn is a catalog for your professional network, and Google is a catalog for web pages.

What we thought needed to exist was a catalog for enterprise data, a catalog that would allow people to surface all of the knowledge in a single place, so that they can go to a single place to answer all of the questions that otherwise might have taken them days and weeks of research with lots of emails, and lots of back and forth.

Go to a single place, build a catalog, allow people to use it very flexibly by finding the information that they need, and using that same catalog to understand the information they find.

Claudia: What do you consider to be the best type of catalog, then?

Satyen: I think a catalog has to be specific to the domain that it's applied to. In the case of enterprise data, you have to have all the relevant

Copyright © 2016 BBBT - All Rights Reserved

questions about that enterprise data that you're asking. Which reports exist out there? What questions do those reports answer? Who has used those reports in the past? What is the lineage of that particular data? What are the definitions around those data sets?

You have to basically answer all of the relevant questions that somebody might ask of that particular thing. Whether it's a professional network and you want to know where somebody's worked before, in the context of data, you want to know all of the context, all of the metadata that's relevant to consuming that information.

Claudia: All right, Stephanie, let me bring you into the conversation. For a modern data catalog, what are the critical techniques that you believe are needed?

Stephanie: IT teams have been creating data inventories for a very long time, and so the idea of creating a list, and descriptors, and metadata about the data that's available in an organization is not a new concept. What is new about the modern data catalog is that we are automating the creation of that inventory.

If you think about it, inventories are really for suppliers. The ability to go out, and crawl, and parse, and index data to create an inventory is a very powerful, modern approach to understanding what data exists in the organization.

I think more interestingly, in the modern data catalog, is the collaboration between what machines can automate, and how humans can collaborate with those machines to create a really useful catalog, not just an inventory for suppliers.

We want to create catalogs for consumers that can encourage more self-service consumption of data in the organization, and to do that in a trustworthy and accurate way, and that's where technologies like AI, inference, machine learning, natural language processing, as well as techniques for gathering human information,

Copyright © 2016 BBBT - All Rights Reserved

which is things like crowdsourcing, and things like expert sourcing and data curation all come into play.

The modern data catalog contains those five categories of things to really encourage, not only an inventory for suppliers, but broader based and more accurate consumption of data by data consumers.

Claudia: Excellent. Staying with you, you also introduced a new term, at least new to me, data curator. Very interesting term, but what's the difference between a data curator and perhaps a data integrator, or a data preparer?

Stephanie: It's interesting, because the concept of curating something has been a concept around for a long time. Humans have curated news stories in newspapers, and have curated art in museums. Curation is really a process that can be a very social process.

It's about sharing perspective and context on something with a broader audience. If you think about it, the individuals who are best positioned in our data and analytic organizations to curate data are really the users of the data themselves, whereby, if you think about data integration, or development of code, you really want specialized experts doing that because of the need for the accuracy of that code, and the specific skill set needed.

On the flip side, with data curation, anyone who's able to apply data to a business problem, and explain a particular perspective or context should be able to share their opinion, and curate data, just like we now curate news through social media platforms like Twitter. We curate ideas, and photos, and concepts through Pinterest.

There's more modern examples of how individual users of anything, but in our case, data, can actually be curators and share their own context on the data with others.

Claudia: Satyen, let me come back to you. We've had, I thought, an excellent question come in from one of our BBBT members about how far along a company should be with its fundamental data

Copyright © 2016 BBBT - All Rights Reserved

processing maturity level before it actually makes sense to use Alation to go to the next level. What's your response to that?

Satyen: It's interesting. When we first started the company, we didn't really have a sense for that question. We didn't really have an answer around it, because we didn't know how people used it, we hadn't had a lot of experience in understanding what people were doing with it.

What we've found was that, interestingly enough, people who are highly mature with their data can obviously get value out of Alation because we can unearth very complicated, very well-established data architectures, and we can learn from them.

We can allow companies with 800,000 tables in their databases, with 1,500 analysts who are querying those databases every day to unearth insights inside of those very complicated, very well-established data architectures.

On the flip side, we also have customers that have silo data where they're trying to embark on a project, where they want to re-architect, and re-create their data architectures. They'd like to be able to learn from those different systems, and those different silos.

They're at the very beginning of a journey or perhaps the beginning of a journey that they'd want to be able to start, to re-architect their data. Alation, in that case, becomes a tool for people to discover what questions are people asking? What people are doing with the data, how they're using it, where they're using it.

In that case, you can see a very low maturity, relatively speaking, but in both cases, you can really start to capture knowledge, and at the end of the day, what we're trying to do is capture the knowledge as it lies, so that people can then figure out where they want to go.

Figure out where you are, and then based upon where you are, take assessment, and then figure out where you can go, and learn

Copyright © 2016 BBBT - All Rights Reserved

about what's important, and learn about what people are actually doing, so that you can make sure you can hit your goals accordingly.

Claudia: I found it most interesting, because in some ways, it's similar to data quality technology. How do you justify bringing a data quality tool into your organization? All you have to do is do a profile of the data, and you can see immediately, you've got data quality problems.

What your tool does is basically the same thing, in terms of the metadata, the cataloging, if you will, of the data. All you have to do is show, look at the redundancy you've got in your data, and people reinventing the wheel. It seems like it's a no-brainer at that point, right?

Satyen: Yeah, I think it's a little ironic, because we're building data systems, which allow us to measure our businesses, but on the flip side, we of course have these data systems, and we don't measure the success of those data systems.

The irony here is that we're not really analytical about our analytical systems, and we're not analytical about how we use data, which is fascinating, but I think that's going to change. I think as we all start asking questions at a faster rate, as there's all of this data that exists out there, we're not going to be able to do anything but become analytical, because unless you understand where the ball is going, you're not going to be able to solve the problems.

Faced with limited options, I think being smarter about what we do is going to be the only way to proceed.

Claudia: All right, Stephanie, back to you. You gave us a number of really good customer use cases. If you don't mind, talk about a few of these.

Stephanie: Our customer base has grown tremendously in the past year. We have grown the customer base more than five times, and that's

Copyright © 2016 BBBT - All Rights Reserved

been across a variety of different industries, from retail, to insurance, to public sector with the city of San Diego, to pharmaceuticals.

We've seen a lot of different use cases. This is a generalized technology in terms of how to catalog your data. I think some of the more interesting results that are coming out of our customer base, one example is at Albertson Safeway, where Christie Allen was under the gun to prove that personalization, when it comes to customer loyalty, really is a more impactful marketing strategy than offline fliers.

She had about four days to make that argument to her management. She did that through the power of Alation, in terms of being able to define the business definitions, and metadata that were necessary, and make that available to an analytics team that was moving very rapidly, under the gun, to get to some answers.

I think that's one example of how a small team can really collaborate very effectively, with the help of some technology that increases their productivity, to get a very meaningful answer, which was, "Hey, let's invest more in continuing personalization of outreach to customers, rather than the standard, traditional marketing flyers that don't have as much return."

Claudia: She also is a good example of your data curator, isn't she?

Stephanie: She is a good example of a data curator, as well. I think this is how we're changing the definition of, what does it mean to be a curator? There certainly is always going to be a role for expert curators in the organization. We have customers like Citrix, that has a very active expert curation program, and eBay also has a very advanced data stewardship program.

There's also a need for every analyst to become a curator, to share their perspective on the data, and to help bring others into the fold in terms of problem-solving around data. I think this is a discipline that really comes out of a perspective on data science, where

Copyright © 2016 BBBT - All Rights Reserved

individuals, any individual can hypothesis test, and put an idea on the board, and ideate about what might be possible with data.

That allows the experts to then dig into, what is an appropriate fit between the data and the algorithm, to test out that experiment. We see usage of our product by both expert data curators, as well as individuals who use data on a day-to-day basis.

Claudia: Tell me another case study.

Stephanie: Another interesting case study is actually the city of San Diego. I think this is particularly interesting, because of the challenge that you have in local government of really proving what the impact has been of some of the decisions that the government is taking.

There was a project Alation was used for to identify if the mayor had met his express goal of paving the streets of San Diego within a five-year period. It was very difficult to pull together all of the data sets to prove that.

An inventorying approach to pulling together that data was the first solution that the team thought of, and then they realized that the way that you calculate what percentage of a street is paved is not, maybe, what you would think. If a street is double-wide, you have to account for that in your calculations.

Only through pulling all of the data together in Alation, and starting to uncover the details, did the team realize that they could've made a dramatic mistake in communicating to the public just where they were in that initiative. It opened a lot of conversations internally about how you need to be literate, not only about the data itself, but about the algorithm that is being used in order to apply that data in order to have accuracy as well as speed in evaluation.

Claudia: What an excellent story. I would never have guessed it was so complex.

Copyright © 2016 BBBT - All Rights Reserved

Stephanie: Very complex. Unfortunately, or fortunately, it's a very interesting part of the public sector that many of us don't get to see every day, and I have a lot of respect for those individuals that are in data science, or chief data officers in that domain.

Claudia: Satyen, let me give you the last question. How does a company get started, then? Is this a very complex project, or is this something that you can ease into?

Satyen: Certainly, we designed, and I think design is the key word, Alation to be relatively easy to consume, not only for the initial users when they start using the product, but for the organization when they get on boarded.

I think the reason for this is because the value proposition is one of productivity. People want to be able to see productivity gains, not after two years of investment, but after 5 to 10 days of work, and being able to roll out a tool, and to see if they like it.

We've made it relatively that easy. What that means is that an organization will typically purchase the product, and they will put it on their servers, and they will install a product on, typically, a native VM, and they will install this, and configure it relative to other databases, which means they'll connect using ODBC or JDBC.

All of that work will be done within a matter of days. Generally, by the time the training's done, which can be in a week, people are up and running, and using the product, and they decide whether they like it, and they decide whether or not they're getting value out of it.

Luckily, I'd say in more than 90 percent of the cases, our customers do. There are rare exceptions, but most of the time people say this is exactly what they want. We do make sure that the product can address use cases where the customers have problems before we engage, but that's exactly how it works.

Claudia: Wonderful. Unfortunately, we're out of time, or we could talk about that some more, I'm sure. That's it for this edition of the BBBT

Copyright © 2016 BBBT - All Rights Reserved

podcast. Again, I'm Claudia Imhoff, and it's been such a pleasure to speak with Satyen Sangani and Stephanie McReynolds of Alation today. Thank you both.

Satyen: Thank you.

Stephanie: Pleasure.

Claudia: 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!

Copyright © 2016 BBBT - All Rights Reserved