



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: Tamr
Date recorded: January 8, 2016
Host: Claudia Imhoff, President, BBBT
Guest(s): Shobhit Chugh, Product Manager
Run time: 00:12:45



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 guest today. He is Shobhit Chugh. Shobhit works in product management for Tamr. Welcome, Shobhit.

Shobhit Chugh: Thank you very much for having me here today.

Claudia: It's wonderful to have you with us. It was a very interesting session, and I really enjoyed it. Let's start off with a little bit of the examples that you gave us of the types of questions that your technology can answer. If you don't mind, what are some of the types of things that you're focusing on?

Shobhit: Yes, we have answered a variety of questions for our customers. For example, in procurement, we've answered, "What are the top suppliers for a particular part? How do I get a part at the best price?" In customer data, we've answered questions around, "Which are the right customers to cross-line and upsell to?"

The problem is not really with the questions themselves, but with the fact that data is silo-ed across different parts of the enterprise. What we help companies do is unify this data from all these different parts of the enterprise.

Claudia: That's been a problem for 25 years, as long as I've been in this business. It seems like we've got silos of data all over the place. The whole focus of Tamr, as you say, is to tame the data, is to unify the data so that we can then produce the analytics from it?



Shobhit: Absolutely.

Claudia: All right, let's talk about the difference in your approach to this data unification or data integration, if you will. What sets you apart from the other folks that do that?

Shobhit: Our main goal is to improve the productivity of people doing this integration. Traditional approaches rely on rules that take different parts of the data and then somehow try to unify them.

Our approach is: Use a combination of machine learning and then input from people. We are able to create a very small training set, ask questions of people, and from that... learn and then automate a lot of these processes, resulting in a huge productivity improvement.

Claudia: I think that was when the light bulb went off in my head. If we look at an example, and you're going to tell me about some of your customer examples, but they may have millions of records from a very small training set.

Then Tamr can take that intelligence, if you will, and apply it to the entire set of data, getting a pretty darn good hit rate of being able to determine distinct and matched records and that sort of thing... Leaving just a small amount of records for a human to come in and say, "OK, there's something really strange with these records, and now I have to step in." Is that a good explanation?

Shobhit: That's a very good explanation. Very often, customers only need to answer a few questions a day to resolve those conflicts, and keep all the data unified.

Claudia: That's what you have as the user interface, right? Why don't you explain the user interface there?

Shobhit: Absolutely. The user interface does a combination of things. When you are bringing in data sets and training them, there is a workflow around bringing in a new source, training Tamr to use that source. Then on an ongoing basis, there's a more of a workflow



management system, which is around, "Hey, these are the few things that were not characterized properly."

Another big part of this is this information does not exist in a single person's head. It's spread across the enterprise. There's various people that are experts in different parts of the data. A big part of Tamer is collecting this feedback from these different experts, learning which experts know more about certain kinds of data, and then directing appropriate questions to them based on that knowledge.

Claudia: I think that's a really important thing to keep in mind, is that the machine learning can handle an enormous number of records, but there are certain ones where these data experts - and there may be many -- have to jump in and say, "OK, this is what this record is," or, "Let me clean up the field or fix the problem," or whatever it is.

That's that interplay between what does the machine learning do versus what the human does that I think differentiates your product. Would that be accurate?

Shobhit: That's a very accurate statement.

Claudia: Let's talk about your solution a little bit, because the platform consists of three functions. One is Catalog, another one is Connect, and the third one is Consume, and if you don't mind, tell me a little bit about each of these three areas.

Shobhit: Absolutely. Catalog is a way to organize metadata about all the data sources that exist within the enterprise. It came from a question that we had from customers, that, "Hey, you know, I understand unifying these data sources, but really, I don't even know where all these data sources are, and I don't have enough knowledge about it."

Catalog is a collaborative tool that helps various people add in information about these sources, discuss, and build this knowledge about data sources over time.



Claudia: Before you move on to the next one, does the Catalog then read the data dictionary, and at least give someone a leg up in terms of what the data looks like?

Shobhit: Absolutely. We look at data, we can optionally profile it, and expose what is within it. Then it uses the data dictionaries that might exist, for example, in Hive or whatnot.

Claudia: OK, and then the second one is Connect. Why don't you explain a little bit about that?

Shobhit: When I mention unification, Connect is the core component that's doing that. It's connecting these different sources, different records within the sources, and getting you to your unified view. Then Consume is, "Now how do you consume this data?"

What we found interesting is one, not just a variety of targets and how this data is used, but also, there's different views of the data that are applicable for different purposes. That's a huge part of Consume.

Claudia: The other thing is, you finally have this wonderful repository of data. Do we only use your tool to analyze it, or do you open it up to any BI and analytic or data visualization capability?

Shobhit: Great question. Tamr on its own, we don't attempt to build another business intelligence reporting tool. We have partnerships with Tableau, with QlikView, and so on and so forth, and also make this information available via APIs, so that it can be put as part of the data pipeline that already exists at several of our customers, and used in the appropriate systems of their choice.

We don't force anyone to use Tamr for the analysis. That's often integration with a third-party app.

Claudia: I think that's smart, sticking to your knitting there.

Shobhit: That's what we try to do.



Claudia: Let's talk a little bit about a couple of your customers and how they're using Tamr. First of all, tell us what the problem was and how you solved it.

Shobhit: I'll talk about a couple of cases. There's a large auto manufacturer that's using us for customer data integration. The problem they were faced with is that they have information about customers in dealership systems, in customer relationship management systems, in several old databases.

They wanted a single view of customers so that they can do things such as figure out, "Who do I cross and upsell to? What sort of treatment do I give to each customer? When do I know when not to email a person because they've opted out?"

By using Tamr, they started to create both country-specific and then also an overall view of customers across different countries. They're using it for analytics as well as for understanding how they email customers, and then also present a view for their salespeople, their dealers, to interact with their customers knowing everything they know about them.

Claudia: They got a full, maybe not a complete 360, but pretty darn close to 360 view of the customer and everything about them. Is it as current as it can possibly be?

Shobhit: It is, and the "current" part is extremely important. People would move over time, get married, change names, and Tamr is able to link these customers together, even if these things were not 100 percent accurate or completely mashed up. That was a huge part of the value for them.

Another interesting thing they want to do, on a very real-time basis, if somebody comes to their website, understand, "Have we seen this customer before? What do we know about them? What is the relevant information we can surface about this customer?" A big part of the value proposition was this real-time matching that they're using.



Claudia: Excellent. What was the other case study?

Shobhit: Another case study is in the clinical data space. We have several customers that partake in clinical data that's been collected over the years for different studies they've done. They are unifying it for the purposes of, one, reporting to the FDA, getting quicker approvals, and second, to use it for analytics in the data warehouse.

The scale of this problem is really what is interesting. They have, in several cases, thousands of sources with a hundred to a thousand attributes each, each recorded in its own way. You can imagine how fitting it to a standard is a fairly tough problem.

What we help them do is, once they've mapped a few of these sources, we're able to recognize the patterns and learn from the decisions the user has made through machine learning, and then automate a lot of the subsequent work, dramatically reducing how much effort they need to put into unifying all this data together.

Claudia: It seems like also, from the case study that you presented, it was remarkable how much you alleviated the burden on the data scientists. They were freed more to analyze the data instead of try to bring it all together and then analyze it. You lifted that off their shoulders.

Shobhit: Right, and I think we've all heard this statistic that 80 percent of time is spent in data preparation, and how do you shorten that? Not just shorten that one time, but keep that updated so that you can use the latest data in your analysis at any point. That's a big part of Tamr.

Claudia: That was a huge ROI as well. Last question, and we've got a couple of minutes left. If you don't mind, what's the future hold for Tamr? Where are you going with this?

Shobhit: One of the most interesting things that is emerging is our effort on spend analytics. We've seen a lot of engagement from both procurement organizations and CFOs who want to understand how



they can save money and use spend analytics as a way to kick-start that.

We're coming up with a solution for that, and specifically works well in companies that have gone through several mergers and acquisitions, and have this data spread across the enterprise, and just find it hard to generate the savings they're on line for. This is one of the coolest things we're working on.

Claudia: All right, sounds terrific. Unfortunately, though, we're out of time, so 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 Shobhit Chugh of Tamr today. Thank you so much for speaking with me.

Shobhit: Thank you so much for having me here.

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!