



## BBBT Podcast Transcript



### About the BBT

The Boulder Business Intelligence Brain Trust, or BBT, 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 BBT provides a variety of services, centered around vendor presentations.

For more, see: [www.bbbt.us](http://www.bbbt.us).

<b>Vendor:</b>	<b>Trifacta</b>
<b>Date recorded:</b>	<b>March 13, 2015</b>
<b>Host:</b>	<b>Claudia Imhoff</b> , President, BBT
<b>Guest(s):</b>	<b>Adam Wilson</b> , Chief Executive Officer <b>Wei Zheng</b> , Vice President of Products <b>Michael Hiskey</b> , Head of Marketing
<b>Run time:</b>	<b>00:20:37</b>
<b>Audio link:</b>	<a href="#">Podcast</a>
<b>Transcript:</b>	[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 Adam Wilson, Wei Zheng and Michael Hiskey. Adam is the CEO. Wei is the Vice-president of products, and Michael is the current Head of Marketing for Trifacta. Welcome to you all.

**Adam Wilson:** Nice to be here. Thanks, Claudia.

**Michael Hiskey:** Nice to be here.

**Wei Zheng:** Nice to be here. Thank you.

**CI:** All right, Adam. Let me start with you. Let's start off with an overview of the company. It's a little bit over two years old. Why don't you give me a little bit about a short history, and why it came into being?

**AW:** Sure. Trifacta really started as a joint research project between Cal and Stanford. One of our co-founders, Joe Hellerstein, who is a world renowned expert in database has got together with Jeffrey here, who is a world renowned expert in human computer interaction. They happened to have a shared grad student named Sean Kendall, who is doing a lot of research and into trying to figure out a way to bring people more into the process of understanding and manipulating data.

There was a general sense that over the past 10 or 20 years, that advances in computation, and from a lot more data out in the world, that somehow, the human aspect of what people were trying to do with data had been lost. They really set about focusing on this as an initial research project and developed a prototype that was affectionately referred to as Stanford Wrangler.



The terminology, or the name wrangler, actually came from the end users that were interviewed as part of the PhD research. They referred to the work that they did as wrangling, munging or manipulating data. Therein was born the prototype. Within about six months, there were 30,000 people using this prototype.

Jeff, Jill and Sean stepped back and scratched their heads and said, "We might actually have a company here." We went out and raised some financing and several years later, 42 million dollars in the bank. We launched a commercial product a year ago at Strata in San Jose.

Now, I have been taking this into the market and really helping people who understand the data best do the data work and the data manipulation. It's been a really exciting couple of years. A lot of fun and seeing great customers come on board, start to make use of this to solve their data wrangling problem.

CI: All those academicians' data wrangling. Let's pursue that a little bit. If you don't mind, I know there are several misconceptions about what data wrangling means. Why don't you, if you don't mind, give me a good, solid definition of what you mean by data wrangling? Then, we'll get in to the self-service data wrangling in a moment.

AW: The main point is we're solving a lot of problems that have been around for a long time, so whether that's cleansing data, structuring data, distilling data, augmenting data. These are problems that have plagued people that are trying to make better decisions and to gain insights about their business for decades. What we're seeing is for Trifacta, it's not so much a matter of changing the problem we're solving, as much as changing the person that we're trying to enable to solve it.

We're really focusing on the people who are the decision-makers, who have the context in their heads. They're able to bring that to the process, and we're enabling them to do work that traditionally had been done by much more technical users, typically writing code or using legacy ETL technologies to do the work.

We feel that if we can enable people who know the data best to do the preparation, to do the cleansing, to do the transformation that that really



allows them to iterate a lot faster. It removes a lot of cost from the equation, and ultimately helps people get to insight more quickly, which is something we're focused on as we deal with more and more data, and recognize how data-driven our organizations have become.

CI: Certainly, that's the benefits from the businesses' standpoint for data wrangling and this idea of self-service. I can do my own data prep. I don't have to have IT in the picture. IT takes a little bit of an affront from that, or worries a little bit. Aren't there advantages to IT if we remove a lot of the data prep from their backs?

AW: Yeah, absolutely. We don't look at this as an IT versus the business kind of situation. We really look at this as a collaboration between the two. I was out on a meeting with a customer recently, who said that a big part of their job is to help democratize data within their enterprise. They're doing a lot of what they consider to be data janitorial work, which is better left to people who understand the data best.

If they can set up an environment and bring to bear a set of tools, processes and governance techniques that enable the end users to do, to share some of these work, do more of these work, to be a full participant in that collaboration, then that benefits IT and it benefits the business. We look at it as a coming together of these worlds in a more profound way as opposed to something where this isn't shadow IT. This isn't trying to work around IT, but it's really trying to take a lot of the data assets that IT has been curation, and make those more generally accessible to the business, to again do a lot of the work that they need to do in a quicker fashion without having so many hand-offs, so many different tools, so many different technologies and architectures to work across.

CI: Yeah, it's a terrific idea. You mentioned the Trifacta approach to data wrangling. It's a little bit different perhaps. Can you go into a little more detail of what that is?

AW: Fundamentally, the idea's to get eyes on data as fast as possible. What we're trying to do is to really invert the traditional paradigm. In the past, specification was written. It was handed to individuals who were highly technical, who would write code or build data pipelines, in order to get



data in a specific shape. Then eventually, a sample which usually run and then some visualizations were created. There would be a round trip back to the end user, who generated the specification to say, "Is this what you wanted?"

That process, depending on where you worked, would take weeks, months, quarters or even years. Invariably, what we found is by the time you went back and actually showed the output, there were questions, issues or concerns, or maybe it had taken so long that even the problem that we were trying to solve had changed, or the questions that were being asked had evolved.

The idea was, through Trifacto, let people get eyes on the data very early in the process, start to sample the data, start to work with the data, using then, predictive transformation technology, which allows us to say based on how the human or user interacts with the data. We can make specific suggestions as to how they might want to transform the data, so that they're not having to invent all this.

Every day is not a new day. We've seen these data sets before. We understand the types of things people want to do with data. If we can put the machines to work, to help suggest and recommend the work that needs to be done, it automates a lot of the things that are reasonably manual and reasonably tedious. But we also recognize that in the end, there are decisions that a human being is going to make.

That's an important part of the process, and important to make sure we don't lose sight of the fact that the context that's in someone's head eventually has to be translated into the decisions that build up the logic that will then compile down to whatever the appropriate processing environment is.

CI: It's a very mature approach I think given that you're only a baby company, two years old. Michael, let me bring you into the conversation a little bit. This inversion, flipping or reversing of the process itself is actually a pretty big paradigm shift. How do people grasp that? How do they go about understanding where we're going here?



MH: Flipping the paradigm is an important way to think about big data. When I look at data that's beyond the scope and scale of what I can consume visually in rows and columns, I have to start approaching that visually, and I have to do it in an interactive way. This ability to have a sampling paradigm that scales up to a big paradigm is very important.

We've done a lot of studies of how people interact with the data. As much as we're a young company, there are decades-long experience that come from the founders in academic research to understand how human computer interaction works and specifically, what that means for data.

That interactive capability is very important. Putting the visual in front of the data, so that when I unboxed that data, I get a visual right away of some representation, I have the ability to highlight things that are interesting, and see how that relates to that visual, make a big difference for how people can add their value. Because as I've alluded to, these are the users with the greatest context to the data that can easily give their intelligence and create a useful end product.

CI: I like that. That's actually a very nice way of looking at it. By visualizing the data you put it unboxing it, we can see almost immediately, there are some very clever parts to Trifacta...We'll get into that with you Wei in just a moment...that allow you as you put it to see the data, to understand what was in the box, and then we'll manipulate it.

On that note, let me bring Wei into the conversation because you did first of all, a fabulous job with the demo. It really made everything very clear and clean. I liked it very much. It also demonstrated not just the tool, but how Trifacta is different. Its differentiators from, let's say, the traditional ETL or BI Tool that does data blending or whatever something along those lines. If you don't mind, I want to spend a few minutes with you. Let's talk about these differentiators.

WZ: A lot of differentiators actually come from some of the characteristics that Michael and Adam were talking about. One big difference is the user that we're targeting. We are talking about self-service data prep. For that user, who understands the data the best, usually from a domain perspective,



---

from the research that the founders did, one of the things that they noticed right away is this user does not perform a linear processing flow.

They actually look at the data. They examine the information. A lot of them say that, "I don't really know what I'm looking for. I really need to get my eyes on the data before I can make the decision of what I want to do with it." One of the differentiators for Trifacta is that we provide a natural iterative workflow for the user.

Starting with visualizing the data, looking at the data, and then wrangle, the term that we used to talk about cleaning, standardizing, augmenting. But, it doesn't just stop there. The user has the ability to go back, and look at the result as its being run on the entire data set and really be able to iterate. That's one differentiator.

The second one that I'll talk a little bit about is the power of getting your eyeballs on the data. One of the things that we do is what we call interactive or integrative visual profiling. This capability is different from traditional data quality or data integration tools that when you are actually getting your eyes on the data and start interacting with the pattern or histogram that you're seeing, the system provides transformation suggestion that's related to those things that you're looking at.

From that perspective, it's not just about looking at charts and graphs to get a better understanding, but it's also allowing you take some remedial steps or actions that you would do to make the data into the shape and quality that you like. That's a second differentiator.

The last one I would say is that Trifacta really was built with big data in mind. When I say big data, I do mean all aspects of big data, I mean the volume, variety, and then in terms of the way where people are looking at those type of information. Scaling in an interactive way is hard. One of the things that we've done is again watching how the users do it today, even though they're doing it manually.

They subset the data. They go through this iterative process. They come back, retest the hypothesis over and over again, and then they scale to take advantage of Hadoop or Spark or anyone of these paradigms. For Trifacta, we knew early on, even though we were going to go to the



Hadoop market, we knew early on that the execution paradigm may shift as new technology comes on board. Things are running faster with a bigger volume. We want to be able to leverage that. The whole flow of the exploration and all of that is independent and abstracted away from the execution paradigm.

All of that allows us to work with data really of all size and all shapes. I think that's a unique differentiator for Trifacta.

CI: It certainly seems to be. Michael, let me go back to you. Speaking of your customers, can you give me briefly a few examples of customers, and how they're using Trifacta?

MH: Initially, a lot of people would think of Trifacta as something that was used by the Data Science Team. On first analysis, that might make sense. But, as our data scientist friends will tell you, "Correlation does not imply causality." Those folks are just kind of ahead of the curve. They'll think about things in a forward-looking fashion.

What we're seeing now, now that we've been in market for over a year is there's this new breed of information worker or data analyst. The role name changes, but the functionality of that role is very much the same. These are business savvy people in the IT Department or IT savvy people in the line of business. They have one thing in common.

There's more and more data being pushed down into the business units. There's more and more a reality that new innovations will come through more intelligent uses of data.

That's really what the target audience is about. It's people that have the greatest context of the data. They want to be able to get their hands on the data, and you direct manipulation with it in a way that they don't need to take a two-year program, and how to code R or Python or Pearl, or anything else for that matter.

It's very much this new emerging information worker that sits at the intersection of IT and business as those lines start to blur.



---

CI: Thanks, Michael. That was an interesting answer. Adam, can you give me some specific examples of customers that you have?

AW: It's been really exciting for us to see our message and technology resonate with customers. Folks like LinkedIn, GoPro, AutoDesk, Orange, a lot of customers that are dealing with machine-generated data or social data that's highly unstructured, need new and interesting ways to really go after this in more meaningful way. Trifacta's a big part of their solution for getting that done.

One customer, particular that I just visited with recently was MarketShare. MarketShare focuses on a marketing attribution. They're a leader in that field. They use technology for competitive advantage to really help their customers get a complete 360-degree view of how effective their marketing is. For them, they have to marry the online and offline world as we all do.

They have Hadoop environment built on altiscale that they're using to understand what's going on with Clickstream Analysis and a lot of the digital presence. A lot of that information comes from sources like Double Click and others. At the same time, they have hundreds and hundreds of spreadsheets that are sent to them from all of their clients, where they're trying to measure what's going on in the physical world.

For them to get that composite view, they have to wrangle all that data together. This is really a storyline that combines big data and small data. They combine Hadoop, and non-Hadoop, and combines many different formats of data. Ultimately, provides them with an opportunity to marry that information, to cleanse and standardize that information, and then ultimately to feed a number of different downstream systems.

They may be feeding statistical models and SAS. They may be providing extracts for Tableau. They may be landing some of the data back out in Excel. Trifacta provides them with a flexibility to not just get that data together and harmonize it, but ultimately to deliver the data to the most appropriate interface, given the analysis that they're trying to do.



---

CI: Let me give the last word to Wei, if you don't mind. You guys, like I said, did a very good job of describing what Trifacta is now. I'm curious though. Where is Trifacta going for the future?

WZ: I'm really excited that there are a lot of things that we're looking at that we want to do. I think one of visions of the company, going back again to Joe, Sean and Jeff, when they founded the company, one of the visions is step one, getting the people that understand the data, to have the context, to be able to touch and work with the data.

Step two is once you get all these users using tools like the Trifacta's Wrangling Platform, you must be able to get collective intelligence from all these users. One thing that was really interesting earlier in the conversation we've had, people brought up the notion of Metadata and the notion of providence and looking at data lineage. We started with the most difficult problem, which is discovery and transformation. But, there's so much more above and beyond data wrangling.

When you think about the analyst user, he is faced with repeating the same workflow, over and over again, being able to manage the change that's coming through the data streams. As part of the next evolution of Trifacta, we are going to be taking a bigger, more holistic look at the entire life cycle of what that data analyst goes through.

I believe that the way that we disrupted traditional transformation with this visualization first approach, a lot of those same principles also apply to ingestion of data, discovery of data, looking at Metadata lineage, and how that impacts the user. Also, the monitoring and the workflow of how user tie together, all of those tasks that surround data wrangling. That's one aspect that we're really focused on.

Another thing I shall mention too is we got some great feedback from the users that talk about, "I really want you, and you talked about the vision of extending wrangling to non-technical user." Some of the things that we're working on is providing optionality in the way our user interface is shown to the user. For someone who is not completely not familiar at all with any kind of programming, we have an improved user interface that gives them a better way to work with the data.



---

That's another thing that we're working on.

CI: All right. Unfortunately, that's it for this edition of the Boulder BI Brain Trust Podcast or BBBT. Again, I'm Claudia Imhoff. It's been a great pleasure to speak with Adam Wilson, Wei Zheng and Michael Hiskey of Trifacta today. Thanks to all of you again.

WZ: Thank you.

MH: Always a pleasure.

AW: Thanks for having us.

CI: I hope you enjoyed today's podcast. You'll find more podcasts from other vendors at our web site [www.bbbt.us](http://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!