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:	**Revolution Analytics**
Date recorded:	**November 14, 2014**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Bill Jacobs**, Director of Product Marketing
Run time:	**00:15:27**
Audio link:	**Podcast**
Transcript:	[See next page]

Page 1 of 9

Claudia Imhoff: Hello, and welcome to this addition 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 very pleased to introduce my friend. He is Bill Jacobs. Bill is the Director of Product Marketing and the Field CTO for Revolution Analytics, so welcome, Bill.

Bill Jacobs: Hey, good to see you again, Claudia. Great to join you again on this chilly Colorado day.

CI: We're into double digits today. I'm thrilled.

BJ: Feels like a heat wave after hitting -6 the other night.

CI: It's been a cold one. Let's get to business here. First of all, let's start with an overview of Revolution Analytics. It's got about an eight-year history, so tell me a little bit about it.

BJ: Revolution is about eight years old, and the company began with the same mission it essentially has today, which is to promote the adoption and use of the R language in the world. Maybe a little more focused these days, is we're focusing on the needs of the enterprise clients who are trying to leverage R as a means to grow what they can do in big data analytics, and other kinds of analytics for the organization.

CI: Let's jump right into the products a little bit. You did just announce the availability of open source R from your company. It's kind of a change, so why don't we talk about why the company felt that was necessary?

BJ: The company, for the first four or five years, focused on work station and desktop users of R, and at a point about three to three and a half years ago, began to see the beginnings of the big data wave. Now, of course,

we're looking back saying, "Gee, big data's an awfully tired term." Three years ago it was all the rage, and has been up until now.

The company focused on creating a new capability to extend the capabilities of the R language, so that we could grapple with a terabyte-sized file, and we could re-model 30 models on a billion customers in four hours, before the business day every morning.

That's a very important accomplishment, but the fact that we discovered was that even though we were able to solve those high end problems, it was probably five percent of the R user base that were actually encountering them, so if our mission was adoption of R broadly, then we were only serving a tiny fraction.

We pivoted the company to grow beyond the high end product to also support two new distributions of R, called Revolution R Plus, which is a supported and indemnified version, and Revolution R Open, which is a freely downloadable, come and get it. It's on our website, you don't even have to give your name, version of R that has been validated and tested, and has some additional repository capabilities, but is open source R through and through.

CI: You now have the three products. You have R Open, R Plus, and the R Enterprise, correct?

BJ: That's correct.

CI: If I understand correctly, only the first one is the free version. The other two are for sale?

BJ: Right, and there are reasons for that, of course, because we have to maintain a workable business. We provide indemnification of the open source code in both of those commercial editions, and this is a key capability.

A very large number of American corporations are a little spooked about open source. They're wary of it, and they should be, because open source can be a means to expose a company to some challenges on a legal basis.

With the Revolution R Plus product, as we have historically with the Revolution R Enterprise product, we offer indemnification as part of the product subscription license.

If you are challenged by someone who claims that you are using a package that you do not have the right to use, and you got it from RR Plus or RRE, we will remedy the situation and hold you essentially harmless, up to some fairly large limits with about seven digits in the financial numbers. That is a critical capability for deploying R in a production context for big corporations.

CI: It sure is. Let's go back to kind of the trends today. Yes, big data's getting maybe a little bit overused, but it's still a pretty interesting area. People are beginning to realize the size of the data is irrelevant. It just keeps going up and up and up. There's no upper limit. There's no lower limit. It's really the complexity of the analytics that you want to run against it that become quite interesting.

Why don't you tell me a little bit about what you see as far as the trends today? You had a wonderful story about a rail situation, so if you don't mind, tell me what you see as the trends, and give us that example, as well.

BJ: There's a couple of things. We could joke here and say big data is dead, long live big data.

CI: Right.

BJ: We grew the term big data on the backs of the evolution of datasets that were flying off in huge volumes off of web servers, clickstream data and that sort of thing. Probably the most interesting and compelling trend is the next wave of even bigger data that some call the Internet of Things, sensor-based data, these sorts of things.

What's notable about those is they tend to produce even larger data than we've seen today, and data that comes in a continuous stream from a sensor that is autonomous, and with all of that, if someone can deduce information and insight from that data, great business value comes out.

The example I used was one that we learned of through a prospective client, and through their sub-client, in the rail industry. The wheels on the train are critical elements. If the wheel fails, the train derails. When the train derails, it destroys the track, and it can kill people, and has routinely done that.

It can explode incendiary cargoes and burn out entire neighborhoods. We've all seen those news clips. There's an extreme value in the transportation industry, not just in railroads, but in trucks and airplanes and everything else, to identify patterns that predict failure so they can be remedied before they get to cataclysmic failure.

The particular rail example is a combination of efforts on the part of the owners of the rail, because the rail lines are separate from those who build the cars, to be able to measure key parameters from the wheels of the train as it goes flying by.

There was actually a PBS segment, I believe, or a Discovery Channel segment on this recently, that exemplified. You can go look it up, but they basically want to be able to measure the weight each wheel is bearing, the side loads it's putting on the track, the temperature of the bearing in that wheel assembly, and any vibration coming from that bearing, because vibration is a great indicator of impending failure.

That's only part of it. They're also now talking about at some point in the future being able to X-ray the wheels as they go by, because when wheels fail, they are castings. The failure begins as a microscopic crack and grows under repeated stress to where it becomes visible, and then it becomes catastrophic.

If they can catch it at the visible stage, the cost of repairing that wheel is tiny compared to the cost of replacing a derailed train, replacing the freight that's destroyed, compensating the rail line, compensating the families who may have lost loved ones.

It's a real interesting story, but there are as many of those in the aviation industry and in the automobile industry, and the chip manufacturing industry that we're working with, as well. That's a big one.

A couple of other areas that are driving our business, and actually figure in our pivot back to open source R, our goal is to drive the adoption of R. What is R good for? A lot of things, but one of the things it's very good for is allowing companies to tap a newer and lower cost workforce to do analytics.

R is taught in most mathematics and statistics programs, and as we've seen earlier today, even in some MBA programs, like here at the University of Colorado just across the road.

With that, we notice that the average age of a legacy analytics programmer, a SAS programmer or an SPS programmer, is around 50 years. That's an expensive population compared to fresh college grads, who by and large out of stats and math programs, can program in R.

The average age of R programmers is therefore down about the 26-year-, 25-year-old range. That's a very compelling number for people who are working with big workforces of analytics programmers.

CI: Other trends that you see?

BJ: There's lots more. The IOT, the aging of the traditional programming. The next big one is that analytics is moving from being an individual sport to being a team sport, and requiring collaboration.

We focus on three audiences. The first audience we focus on are certainly the data scientists. These are the R programmers, but their work is not really made of value until it is provided out through a series of business analysts who are typically working in another class of tools.

Maybe GUI-driven, drag and drop, build a flow, build an analytic application kind of a tool, and then that is then propagated out to the actual end users, who may be 5,000 call center workers in Omaha, or they may be the actual telecom client dialing in through their phone to change their rate plan.

By building a tool that serves that data scientist such that he can deliver his work easily to the other two or three audiences, we enable them to

get more done faster by amplifying their skills through those other audiences.

CI: It was an interesting discussion that we got into about that, of why is Revolution Analytics focusing so much on the IDE interface, as opposed to the GUI interface of something like Qlik or Tableau, that easy drag and drop kind of thing.

I want you to reiterate why you're doing that, because I think it is an important difference where you play in this evolution of analytics from data scientists to business analysts, to everyday Joe, call center person and so forth.

BJ: The question is best summed up by answering in this way. Our goal, our mission is to propagate the adoption of R. R is a language designed of, by, and for data scientists. It's not necessarily a good language for line of business executives to work in. It's a programming language.

What we do is we focus on building the surrounding elements that make R adoptable and usable and stable, and then we go partner with companies like QlikView and Tableau, and MicroStrategy, and Alteryx, and Predixion, and there's probably 20 more.

But the reason we do that is that they own the user experience for that line of business-oriented worker, and we make the R programmer's work directly accessible to that class of programmer, and we stay focused on R by doing that.

CI: As a small company, that's a brilliant move, stick to your knitting. Maybe sometime down the pike you might change that, but right now, you've got such wonderful partnerships. Why would you want to upset that by trying to compete with them?

BJ: That's pretty much true. The brutal reality is when you're a modest company, the most important thing is focus. We've all been in the tech industry for a while, and we've seen a lot of companies with very great promise fail.

One of the consistent failure themes in modest-sized companies is trying to do too much halfway. Do a half-something or other job of 10 things is not nearly as enduring as being the best at the thing that you know is going to last for a while. We are the best at supporting and deploying R in the corporation, and we will stay adherent to that mission.

CI: Where I grew up, it was the pig on ice syndrome. All four feet are sliding out from under him, and he's moving like crazy, but he's not going forward.

All right, let's end. We've got about three or four minutes left. I do want to end on the advantages. You gave us three very important slides at the end, the advantages of Revolution Analytics to the different audiences in the organization, so let's end on those.

BJ: As I said, the three audiences we typically focus on are the data science team, the IT team who are charged with standing up the framework that hosts big data so the data scientists can get at it and use it, and generate value from it, and the line of business guy who funds all of this because he's the one who needs an improvement in product quality, or an improvement in customer satisfaction.

To the data scientists, we provide them a language that they know and they love. It's open source, and they can gain access to a huge world of freely available and traded and exchanged intellectual property that is open sourced. The 5,000 and some odd packages that are available in R program are only one example of how serving that audience with R helps them do their job.

To the IT organization, who have largely been charged with standing up the big data infrastructure and making it usable, we provide an alternative to the traditional, I would call them the, legacy analytics vendors who, sometimes their pricing, their practices, the Balkanization of their products are exceedingly painful to their organizations. We allow them to have an alternative to build new applications atop, and not continue to be stuck with the choices they had five years ago.

To the line of business guy, it's simple. They want something that can be quickly brought to bear on a problem, and at the same time be enduring

enough that what is developed this year for automating failure detection in the NOC, network operating center, that same application will run a year from now, or two years from now, on whatever the latest, greatest platform is.

The best example would be, what follows Hadoop? What follows MapReduce? Most of us think it's Spark. If we offer a platform that gives the IT guy a chance to say to his user, "This platform is available today on MapReduce and Linux and Teradata, and tomorrow, if you want to use Spark to do your modeling, Revolution will bring you a version of Spark," then you have a continuum of consistency that means they don't have to recode and redevelop. That's good for the business guy.

CI: Kind of a buffer between the technological innovations and the ultimate analytics that they really wanted.

BJ: After all, what is the software business? We are about abstraction. We are in the abstraction business.

CI: Absolutely. Unfortunately, we're out of time, but that is it for this edition of the BBBT podcast. Again, I'm Claudia Imhoff, and it's been a great pleasure to speak to my friend, Bill Jacobs, of Revolution Analytics. Thanks so much, Bill.

BJ: Thank you, Claudia. It's been great.

CI: I hope you enjoyed today's podcast. You'll find more podcasts from other vendors at our website, www.bbbt.us. If you want to read more about today's session, please search for our hashtag on Twitter. That's #BBBT, and please join me again for another interview. Goodbye, and good business.