BBBT Podcast Transcript

Vendor:	**Informatica**
Date recorded:	**Mar. 04, 2016**
Host:	**Claudia Imhoff**, Founder, BBBT
Guest(s):	**Dave Lyle**, Vice President of Business Transformation Services
Run time:	**00:18:22**

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 and analytics 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 Dave Lyle. Dave is a vice president of business transformation services for Informatica. Welcome, Dave.

Dave: Hi, Claudia. Thanks for inviting me.

Claudia: It's great having you. Let's talk a little bit about Informatica, a little overview of where Informatica is today, especially after the return to kind of a private status.

David: Informatica has been around for 20 years. We grew to the point of getting over $1 billion a year. We were taken private last August. I think this has been a very good thing for the company. As you know from the industry, the balance of software that's being bought on a subscription basis through the cloud is increasing. Everything is changing in our business.

We just found that it would be better for Informatica to be thinking medium term and long term than some of the quarterly short-term approaches that you take as a public company. We had some very long term thinking private equity investors that bought us with the idea of taking us to not just $2 billion, but also changing the mix of our subscription to enterprise license approach and doing this in an intelligent way that benefits our customers.

Then someday, probably, hopefully, going back public again. Anyway, it's a very exciting time. We really have the same energy or

the same company. I'm still here. I'm still happy doing what we're doing. It's a really exciting time to be in the data world.

Claudia: You're still standing! Let's talk a little bit about Vibe. I found Vibe just very interesting. It's a new platform, if you will, a data integration platform. Tell us a little bit about what the purpose is of Vibe, and how is it different from traditional data integration platforms.

David: The way traditional data integration platforms have worked is you go into this user experience where you define your logic in some graphical way. Actually, that logic is, to a certain extent, kind of WYSIWYG, what you see is what you get.

What we did, starting with version nine – this actually isn't that new anymore; it's been around for seven years or so – we separated that definition of what the user wanted done from the how it actually gets executed.

Now, we've created, in between, an optimizer that you can use to help decide whether I want to push this work down to Hadoop. Do I want this to be done in Teradata? Do you want this to be done or need this to be done maybe by the Informatica engine?

Of course, each of those can be split up into many different parallel pieces, and so forth. It really becomes an incredibly powerful way to not just execute things but the most important part of this is, it becomes an insurance policy, going forward.

If you think about it, we didn't know about Hadoop or weren't working with Hadoop five years ago. Now, oh my gosh, Spark is here. We've been joking today about, "Have you heard about Wolverine?" I just made that up.

The point is, maybe I wasn't making that up. There is going to be something tomorrow that is going to come along, and our point with Vibe is that we're going to be able to help you take advantage of that new power without having to hire really expensive programmers, who are already knowledgeable in that, that are really hard to find.

There are all sorts of really powerful, under-the-covers benefit to the way Vibe helps you take advantage of the data Renaissance, kind of, that's going on there.

Claudia: I think a good way of looking at it is that it's portable. Today, I may be on "fill in the blank" relational DBMS. Tomorrow, I may be on a different relational DBMS or on Hadoop, Map Reduce or Spark or, like you said, Wolverine, whatever that is.

David: Absolutely, or the different document no SQL systems or the different key value store systems or the graph databases. We got all these different tools that are all really good at their own things. The problem is, they each have their different syntax, actually.

They each do a different get input and they don't all talk SQL, and they don't all do these things. We hide that from you, allowing you to really talk to just about everything, using a SQL-ized kind of way of working that you're used to. That portability of skills is more important, almost, than the portability of code.

Claudia: Briefly, why don't you tell me a bit about the components of the platform? What is it made up of?

David: It starts with the connectivity, going from main frame, all the way through to all the new stuff, and being able to connect to things in both a batch oriented way, a parallel way, a high speed ultra streaming kind of way.

You need to be able to connect to things. Then have that pipe or that either data input or output appear to you in a common way so that you can see that structure, that hierarchy, that thing in a way that you can then connect up to other things.

Once you have that, then you're going to want to apply logic to that. You may want to cleanse it or conform it in some way. Some of the things we've done in the intelligent data platform is be able to suggest things to you where we can automatically recognize different data types and apply or suggest certain common things that other people have applied to a similar situation.

We also give you the ability to see what the data would look like, to profile not just data that may be sitting somewhere in, say, a relational database or Hadoop or wherever it might be but also profile the data as it might look, virtualized as it might be traveling from one place to another.

This is an incredibly powerful agility-increasing way of debugging and building what you're doing so that you can prototype things and even create ODBC objects of these things that you can then view through different BI tools to put in front of your business to get feedback as early in a project as you can possibly get.

That becomes a really powerful way of putting these different pieces together. Unfortunately, I have to take another breath, because Informatica has grown so much that that's not even it. That's where we came from, but now we also have some of the strongest, most mature master data management capabilities that allow you to also think of it as a multi-domain problem.

You typically don't just have products. You also have suppliers and customers and GL accounts or whatever it might be. All those things need to be mastered. Then you need to be able to serve that up or publish that and make that available.

Security is also one of the really big investment areas Informatica has focused on, adding abilities to mask data, to archive data, to create test data systems that are subsets of other systems. Informatica's portfolio has grown greatly. Of all the areas that are growing, it's actually the cloud availability of these services, integration platform as a service, or IPAS, that's probably the fastest growing area for Informatica.

Claudia: We had a number of questions about the organization of a critical component of this overall environment, and that's metadata. It's the glue, if you will. It's what keeps everything harmonized together. If you don't mind then, go over Informatica's strategy and the technological support it has for metadata.

David: Absolutely. This is an area close to my heart. Informatica has obviously run itself on metadata over the past several decades. It's an important part of what we do, but that recognition about what can be done with metadata has really exploded in the last several years. It is the secret behind the concept that Informatica has an intelligent data platform.

The intelligence comes from an analysis of mostly the metadata and the ability to be able to recognize things that we've done before, to be able to suggest better ways of doing things based on what people have done before.

There are all sorts of different layers to this. Informatica has acquired many different companies over the last 10 years, 15 years. We've really, in many cases, completely integrated that metadata into our core metadata repository.

Even so, we have several others, and it's our responsibility to ensure that a common metadata is portrayed to all the roles who use our product, regardless of whether we physically integrated that metadata ourself.

This should actually sound familiar to everybody, because that's what we all do with data at our different client or home organization sites – to try and make data appear integrated – even though we will never achieve a single Uber business database for any given company.

Anyway, what we're doing now with live data map is taking this to a whole another level. We've got a high speed graph database, creating the common patterns that allow us to on board new metadata from different places in a very fast pattern-oriented way, which I'll talk about, coming up, as well as get metadata out of here.

The most important part of this is the metadata visualization of what is in the data lake, what is valuable. Based on what I've looked at before, what might I be interested in today, what are data sets that

look like they might go well together, what quality or confident scores can we attribute to the data sets that are out there?

What do I need, for instance...Metadata provides a foundation for this... What do I need to continually improve the quality of the contents of the data lake in my area?

Claudia: You put up a slide of the hairball, if you will, of most systems' architectures. What do you suggest a company facing this kind of hairball do to mitigate the costs and the risks, really, of creating an analytics environment?

David: Whether a company has actually gone to the work of drawing that hairball or not, we kind of all laugh at it. We know that it exists. Like they always say in management, you can't really manage what you can't measure. You can't really manage what you can't see, in this case.

Of course, this comes back to metadata. We need our application portfolio understood. We need to understand how those applications are talking to each other. We need to do an inventory of what's going on. We need to understand who's doing it.

I guess you could say is, that in my travels of the world of integration competency centers or centers of excellence for integration or whatever that might be, the one thing that separates the most efficient groups from the least efficient comes down to metadata and their use some competency and maturity with that.

That really is your ability to be able to expose the hairball. Typically, when folks do the first time, they'll find that at least half of what's out there is duplication and redundant or not needed anymore.

Claudia: One of the other things that you suggested too was for people to begin creating patterns of data, if you will. How does that help?

Dave: At first, I was doing a lot of different similar data marts and data warehouses a long time ago and found that we end up working

with master data, reference data, cross reference data, and actual transaction data, whatever it might be.

Based on those types of data, we do the same thing. Those are patterns. We thought that if we had something that allowed us to work within a pattern but generate the logic that may be appropriate to sales orders or inventory movements or something, that we'd be able to do it faster.

It turned out that, yeah, you got faster when you work with patterns or templates. The bigger benefit was building in quality and the speed that that would give you, an agility that would give you, over the lifetime of the logic you build. When you do that, you're building in the best practices that your integration architects have found to be the best way of working.

What's the alternative, right? You don't give the person a pattern. You give them a blank white sheet of paper. You know what happens when that happens. You're going to get chaos within a week. Everybody is going to end up doing the same thing.

In fact, if I had a nickel for all the people that can bring the skill that they already had to the table using any product, let alone Informatica, they'll end up, for instance, using Informatica as a glorified stored procedure scheduler. No. Focus on patterns and templates, and you'll be building in quality.

Claudia: I agree with you. The shiny thing in the data integration area is this technology or this technique called data preparation, if you will, or data wrangling, data munging. It has a number of terms that I'm not real wild about. The idea is that it's kind of a data integration light, if you will.

It doesn't do the heavy lifting of ETL, the good old extract, transform, and load, or it doesn't do a lot of data quality. What is Informatica's response to this data prep movement?

David: In effect, we've been doing the same thing in parallel with something we used to call spring bock and we know call Rev. I

really suggest to folks listening, go to rev.informatica.com and try this out for free. It's really easy. Just grab a data set and play with it. You'll really like it. I actually end up using it quite frequently.

The way I like describing it is – a data integration tool guides you towards laying out, graphical logic on a canvas. If you want to see the data, you have to click into that graphical logic to bring up a grid of data. In other words, you see the mapping first and the data's behind it.

Let's turn that around. This is really the way Rev works. It's also the way some other data preparation tools work. You've got your grid of data that looks like Excel, but the overlays of data changes and alteration and logic that you're using to prepare and clean up that data is creating logic behind the scenes. It's creating a mapping back there.

The difference is, in the way that, I'd say, that our tool is a bit different than some of the lightweight things that are still useful but out there, is that the mappings that you're creating using Rev behind the scenes are actually creating reusable logic that you can operationalize in the power center, Informatica intelligent data platform domain.

What you find is you end up doing many of the same things over and over again, as well as all our different data preparation products are doing things where we're automatically recognizing things like addresses or different data types.

That's nice. There are so many other things that humans are adding their value to the clean up and you'd like to be able to incorporate, operationalize, and reuse that logic again. That's what we provide by making Rev really a part of the entire Informatica platform.

Claudia: That's really nice. What's really good about it is it's free.

David: Now you can get more usage. Hey, it's something that anybody can start playing with.

Claudia: Absolutely. Unfortunately, we're out of time, so I have to say that that's it for this edition of the BBBT podcast. Again, I'm Claudia Imhoff, and it's been such a pleasure to speak with Dave Lyle of Informatica. Thank you so much, Dave.

David: Thanks, Claudia.

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!