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.boulderbibraintrust.org

Vendor:	**Actian**
Date recorded:	**April 12, 2013**
Host:	**Claudia Imhoff**, President, BBBT
Guest(s):	**Fred Gallagher**, General Manager
	Mark Van de Wiel, Director, Product Management
Run time:	**00:17:48**
Audio link:	Actian Podcast
Transcript:	[See next page]

Claudia Imhoff (CI):   Hello and welcome to this edition of the Boulder BI Brain Trust, or the BBBT. We're a gathering of leading consultants, analysts, and experts in business intelligence who meet with interesting and innovative companies here in a beautiful, little bit chilly, 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 advise them on their marketing direction and messaging. The BBBT podcasts are produced by my company, Intelligent Solutions. I'm Claudia Imhoff and I'm pleased to introduce my guests today, Fred Gallagher and Mark Van de Wiel. Fred is the General Manager and Mark is the Director of Product Management for Actian Corporation. Welcome to you both.

Mark Van de Wiel (MV:): Thank you, Claudia. Great to be in Boulder.

Fred Gallagher (FG):   Yeah, good morning Claudia. It's great to be in chilly Boulder this morning.

CI: A little chilly, Fred. Alright, Fred, let's stay with you. First of all, a little bit of history about Actian Corporation. You've had some very exciting news in the last, what, 24 hours. A little bit of history and maybe a little bit about the exciting news?

FG: Yeah, sure. Action is a data management company. We have a very rich history in data management. We own the Ingres database, and about two years ago we, re-branded our company and started moving in another direction, that we called action apps, which are really about how people bring data to make decisions and automate those decisions. When we started looking at what was required to execute on our ideas around action apps, what we needed was a few other capabilities around connecting to data sources, as well as transforming and cleansing that data, and then adding it into our database technology. It was a natural fit for us to look to at a company like Pervasive, which is now Actian.

CI: You're over 400 employees now, which is kind of nice. Let's keep going with what you do have. We had a little bit of a discussion this morning about one of the claims that you made, and that was that you have the fastest big data analytic relational data base, Vectorwise. That

generated a whole bunch of discussion. I think one of them was, go ahead and justify that. So I'm going to ask you the same question. How do you justify that?

FG: Sure. I think the best way to justify any statement like that is for folks to use the product on their data themselves. That's really where the proof is in the pudding, is take your data, your queries, and bring Vectorwise into the equation. That's where we see great results and comments by customers. We have run industry benchmarks with the product, and so that's why we can make that claim. There's a test, such as the TPC, where we're at least 2X, if not much higher, in terms of the results that we deliver. That's why we make that claim, but really the proof is in the pudding, and when you eat the pudding. That is what I believe is really what people need to look to.

CI: I think that's probably a good way to look at that. I think "fastest in the world", or "fastest database" is a difficult claim. I like your idea of, "Look just use this, just see how it works." And I think that's a good one. Mark, let me bring you into the conversation, because we heard a pretty good history of databases from you -- all about the improvements that have occurred in CPU performance in, what, the last 20 years or so. You went through a very nice list of what caused these improvements, and how they improved CPU performance. Why don't you discuss that with me a little bit, or tell me about it.

MV: Sure. I think most people in the industry are familiar with Moore's Law. Moore's Law was originally about the number of transistors that would fit on a single chip. The industry, already a while ago, actually achieved the ceiling of how many transistors would potentially fit on a chip. So Moore's law has gradually morphed into more of a, "Well performance doubles roughly every two years, or every 18 months." In order to be able to keep up with that newer definition of Moore's law, CPU manufacturers have introduced sophisticated performance capabilities in CPU's, that software applications could now take advantage of. Some of the newer capabilities include SIMD instruction, Single Instruction on Multiple Data, to dramatically speed up data processing.

Includes chip caches, where now with multi-core environments, main memory gets congested, as multiple cores access the same memory pool. Chip caches are designed to minimize the overhead of congestion. We've seen hyper-pipelining capabilities. We've seen super-scalar functions inside the CPU, just in order to improve performance. Vectorwise is unique in taking advantage of those capabilities as it processes data at ultra-high performance.

CI: All right. Well, the other thing you talked about, too, was the I/O challenge, and I agree with you there. It's always been an I/O challenge, it seems like it just seems to move in different places. But we're still there. Again, tell me what you see as the I/O challenge today, and how does Vectorwise mitigate this challenge?

MV: Right, so the I/O challenge that we see with the Vectorwise data processing engine, is related to the phenomenal fast data processing performance. To give you some indication, we've seen Vectorwise process over 1.5 gigabyte per second, per CPU core. If you think about mult-core servers these days, when it's almost impossible to get a server with fewer than 4 cores and 8, 16, even 32, 40 cores are readily available on very cost effective configurations. If you multiply 1.5 gigabytes per second by that number of cores, with Vectorwise doing parallel execution, or running multiple queries concurrently, you end up with an awful lot of I/O that would potentially need to take place, in order to simply feed the CPU's to keep them busy.

What Vectorwise does to mitigate those challenges, is it uses columnar storage to only retrieve data that's absolutely necessary. It uses storage indexes, so that we quickly filter out database blocks based on filter criteria, or based on joined conditions. Vectorwise uses aggressive compression, and it uses compression algorithms designed for ultra fast decompression, as part of Vectorwise execution.

Finally, we actually use a very significant chunk of main memory, so the RAM, as a high-speed buffer in front of disk. Then Vectorwise will, by design, pre-fetch data off of disk into main memory. A lot of the I/O bottleneck gets mitigated as data is actually processed.

We keep data still compressed in the memory buffer, so that we can make most optimum use of the total resources on the server, even though we obviously understand that decompressing data every time does take up some CPU resources. But again, the compression algorithms are designed for ultra-fast decompression.

CI: Let me talk about that just for a little bit, because you talked about -- you also are able to store data row-wise as well, in case there is a need. There are situations, the width of tables and so forth, which lead people to go down the path of row-based, as opposed to column-based. Do you have that capability?

MV: That's right. We do have that capability inside of Vectorwise, the analytical database, to store data in a row-based manner. The main reason is indeed, as you alluded to, certain use cases. I think from what we've seen in our customer base today, it's the corner cases, where you might have an extremely wide table with relatively little data, or very frequent updates across the entire table, and queries typically retrieving all columns, more so than individual columns. That's where we see row-based storage be useful for those types of implementations.

CI: On the compression front, I don't decompress all of the data, right? I'm only decompressing the bit that I'm querying at the moment?

MV: Exactly. That's correct.

CI: Again, we're trying to mitigate as much as we can. You compress it, you decompress only what you need, and then you move on, right?

MV: Exactly. That's the way it works.

CI: OK. Lots of analytic databases out there today. It seems like every time I turn around, a new one is announced. There seem to be many of them. Why don't you tell me what you see as the main differentiators for Actian versus these many different analytic databases?

MV: Actian Vectorwise is definitely unique, in fundamentally being built on top of the concepts of vector processing and taking advantage of modern server architecture and modern chips. You do see other vendors coming

out with technologies that introduce, for certain use cases, similar algorithms in order to take advantage of the improvement of CPUs, but Vectorwise is truly unique in being built from the ground up to leverage those types of technologies. If you look at existing database technologies who start to adopt these types of techniques, they create almost, if you like, a bolt-on piece to the existing database, which they almost naturally have to do, because there's a potentially large existing customer base that they need to satisfy.

To just rewrite a database execution engine from scratch to take advantage of those newer capabilities is almost impossible, thinking of the impact that that might have on existing applications that use the current technology. You get a bolt-on capability that does take advantage of similar capabilities for certain use cases, but it's not going to be as fundamentally ingrained into the technology as it is with Vectorwise.

With that, at least for some time to come, we'll be able to differentiate ourselves by seeing far more use cases that take advantage of our vector processing and our optimizations relative to other technologies. We see the technologies come around every day as well, and while I'll not say we are worried, we do realize there are organizations out there with an awful lot of financial resources who could get something going one of these days.

CI: It sounds like yours is pretty good, at the moment anyway. Let me turn to a discussion that actually heated up the BBBT, to say the least. It had to do with your claim of a paradigm shift. Just to set this up a little bit, when we started out this morning, you talked about Vectorwise being an adjunct to the EDW, to an enterprise data warehouse, and that it could be a sandbox or an experimental area or something like that. At the end of our session, we moved into the Vectorwise technology being the EDW, actually replacing the EDW, or at least that was the paradigm that I got out of it, and a lot of discussion about would that work, what does it need, what does it take, and so forth and so on. Hopefully, I described that correctly. Why don't you tell me about the paradigm shift that you see, and we'll discuss some of the points that we heard this morning?

MV: Sure. First of all, let me make a statement that definitely we would not position Vectorwise, today, as a replacement for an EDW in an existing organization who've spent 20, 30 years designing and optimizing their data infrastructure. It would be naïve for us to think that that's what the technology can do today. We see Vectorwise be successful in those types of organizations as an analytic data mart that can offload some of the heavy processing that's ongoing on an EDW, but the paradigm shift that we were talking about this morning was more related to a number of organizations that we see. Those are organizations who don't necessarily have a legacy of an existing infrastructure. It's the new media companies. It's the Internet firms who generate awful lots of data, and not necessarily an awful lot of revenue, but certainly are the younger companies.

Instead of going down the path of designing an ODS, an EDW, bringing together multiple data sources, etc, for certain use cases, we see them apply Vectorwise side-by-side with all of their data stored in a Hadoop cluster.

It's not necessarily the hard financial results that need to be reported out of that type of analysis that they want to do, but they want to do things like optimize advertising so they want to make sure that they place ads at the right spot on their web sites so that they can increase the value of the ads that they sell to their vendors.

They are trying to optimize user experiences. It's more of the softer aspect of the business where an organization has lots of data in Hadoop, all of their data, let's say click-stream for the last year or the last couple of years.

Hadoop, being very powerful processing engine, but not very fast interactive, doesn't help you inquire the data to figure out what marketing campaign am I going to run, or how should I place my ads to be successful? Then use Vectorwise as an analytic data mart to pull a subset of data, but still a large volume of data, very fast into Vectorwise, do the processing there, in order to then decide how to continue to design the website or what marketing...

CI: What I hear is that it is a relatively homogeneous set of data to begin with. It's click-stream data. We're not talking about taking data from the general ledger and from the order-entry system and from here, and from there, and all over the place, and trying to bring that into Vectorwise. You're talking about a relatively homogeneous set of data to begin with. Correct?

MV: Yes, we are indeed. In those types of use cases, we start from a single source, and it's one type of data. It's not a variety of types of data. That is definitely an important clarification. Thank you.

FG: Claudia, let me add a couple of points around this concept of a paradigm shift, really, two or three points. When we talk about paradigm shift, it's every vendor and many organizations that are investing in data see this paradigm shift. What I see it, from a customer organization perspective, is moving into what we might call the "age of data." Data is taking on a new meaning in organizations. I use this phrase, "It's a factor of production," similar to capital and labor. I think everybody has recognized that, and traditional vendors, by the way that they've revamped their product lines, are in fact saying that their older technology is not capable of addressing the new "age of data." That's my first point.

Let me summarize a couple of things that I think we were saying here, is folks that have invested in an existing data infrastructure that might include an EDW, there's still an opportunity to have faster analytics, faster reporting, to address this new "age of data." They're moving from an existing data infrastructure to one that is addressing modern needs.

There's a whole industry, an ecosystem, of what we call data service providers, especially in the digital media advertising and marketing area, that are implementing greenfield, or almost greenfield, types of applications. Those folks will start with a different starting point.

Really this paradigm shift depends upon where you're coming from and where you're going to, and so many people see the opportunity to get a competitive advantage and deliver value, both from a capitalistic

perspective or a governmental perspective, moving to this "age of data." That's really where the paradigm shift is taking place.

CI: We could talk about that for another 20 minutes, or maybe an hour-and-a-half but, unfortunately, we are out of time. So that's it for this addition of the BBBT podcast. I'm Claudia Imhoff. Thanks so very much. It was a great pleasure to speak with Fred Gallagher and Mark Van de Wiel of Actian Corporation today. Thanks so much to both of you.

FG: Thank you very much.

MV: Thank you.

CI: I hope you enjoyed today's session. If you want to read more about it please search for our hash tag on Twitter. That's #BBBT. Please join me again for another interview. Goodbye and good business.

Transcription by CastingWords