Podcast: Ask HANA Questions, Get Answers - Three SAP Mentors Talk to the SAP HANA Team

"An
informal back-and-fort where some burning HANA questions get answered - in detail"

Podcast Interview Date: June 16, 2011

{enclose pcast_hanaqandajonerp.mp3}

At Sapphire Now 2011, the central importance of HANA to SAP's vision was made abundantly clear. But as a result, that made the questions raised by HANA's ascendancy that much more important to answer. After Sapphire Now, there was enough lingering confusion about HANA to create a need for greater communication between SAP and its constituents. In this case, this podcast came about as a result of a burst of Tweets between several SAP Mentors and SAP. The SAP Blogger Relations team proposed the idea of doing a HANA podcast with me as the moderator and referree and I agreed.

During this 40 minute podcast, you'll hear SAP Mentors John Appleby and Ethan Jewett pressing SAP on some of the key HANA questions they have been hearing from SAP customers. I jump in once or twice with thoughts of my own - otherwise, after my intro, this is a free-flowing discussion with SAP's Jake Klein and Thomas Zurek providing in-depth answers and a good deal of useful context to the HANA discussion. Topics such as the future of BW and data warehousing at SAP, the need for HANA benchmarking, and the reality of current HANA use cases are covered at length. Klein and Zurek acknowledge that HANA is still in a relatively early phase and explain what they are doing to bring SAP customers from Proof of Concept (POC) to Go-Live.

With the HANA General Availability announced on June 20, this is a timely podcast. We didn't get all our HANA questions answered - time ran out on the taping first - but listen in to find out what we did learn.

Editor's note- podcast links: Here's SAP's In-Memory Home Page on SCN. Thomas Zurek is one of the most active SAP internal voices blogging on HANA themes. You can view the Hasso Plattner and Vishal Sikka Sapphire Now keynotes, both of which were heavily centered on HANA, with a free log in. John Appleby has blogged in detail on HANA on his Bluefin Solutions blog and also on his blog on SCN. Ethan Jewett has done the same with his own blog as well as his SCN blog. Jon Reed also covers the range of reaction to HANA at Sapphire Now in his Enterprise Irregular blog post: Analyzing the Real News Stories of Sapphire Now 2011, Part One: The Impact of HANA. Other links: SAP Mentor Initiative. Also recommended: SAP Mentors Vijay Vijayasankar and Vitaily Rudnytskiy have both been blogging in detail on HANA. Special thanks to Craig Cmehil of SAP Blogger Relations and the other influencer relations folks at SAP who helped make this podcast happen.

If you want to subscribe to JonERP podcasts, get the The JonERP Master Blog and Podcast Feed. Or find Jon on his @jonerp Twitter feed. These podcasts are also included in the JonERP iTunes podcast feed.

Podcast Highlights

1:46 John Appleby to Jake/Thomas: What are the great use cases for HANA, and what are some use cases where HANA isn't the most appropriate and BW and/or Sybase IQ might be better? Jake: HANA is not just an application or a capability; it's the kernel of SAP's next generation platform to deliver many of our applications in the future, built around in-memory technology. HANA itself is the in-memory engine, which includes in-memory database capability as well as an integrated calculation and aggregation engine. In the 1,0 version which will be in GA next Monday, HANA is targeted for operational data mart capabilities primarily; those are scenarios where customers would like to run analytics on top of data that is sourced from their SAP or non-SAP ERP systems, and do those analytics in real time. It scales extremely well, customers can ask whatever they want without having to worry about the impact on core ERP systems databases. With HANA, they don't have to aggregate data and do database tuning, etc for operational BI to work effectively.

What is it not? HANA is the first app that we have via in-memory tech, there are many more we will be delivering on top of this platform. So how does HANA relate to BW and to Sybase IQ and other data warehousing technology? The way SAP views BW is as an application that sits on top of the database. The plan with BW is to, over time, deliver BW on top of HANA and to selectively and progressively move and shift the performance intensive parts of BW into the HANA in memory engine starting in Q4 this year. In its current version, HANA is a data mart, not yet a data warehouse system. It's not designed for a large scale EDW with content lifecycle management, data management, object management and other robust EDW capabilities - those will be moved from the BW app layer into the HANA engine over time, and we'll see the first in-memory data warehousing offering later this year: BW 7.30 that is running and optimized for HANA.

Thomas: I would not call HANA an application, but it's the core technology, and we will build a huge number of apps on top of it, with BW being one of the first applications that sits on that technology. I was confronted with that type of question in a number of customer conversations, and one good comparison that came out was that HANA is the engine of the car and BW is the body of the car around it. It's confusing to customers as they go to SAP as the would the show room of the car dealer, and what they get shown is the pure engine with not a lot of body around it, and that creates some confusion. We are building a number of applications on top of HANA, some are simple, some involve Business Objects tools. We're looking in the EPM area for a number of apps, such as strategic workforce management, then there are more generic apps, I would call them implementation of best practice data warehouse management. In the past, these tools have been built on the assumptions of traditional database technology, now the rules are changing significantly for in-memory, so we optimize it in to the in-memory technology world.

8:00 Addressing BW confusion: Ethan: what does this do to customers' overall perception of the BW/BI architecture? How will HANA fit in? What is the end game? My understanding is eventually you'll put all your apps on this platform, including the Business Suite. But for now, what do we do about the BI/data warehousing strategy, do we keep that layer in? Thomas: The layering is mainly motivated by the way you manage and orchestrate data in a data warehouse. Some features in there are motivated by optimizing query performance - those pieces we need to rethink and rearchitect, but then there are a lot of other pieces that are not technology motivated.

One example: there was a company talking about sales order management, they had an operational database for each database, for online sales they had a database, they had databases in the various shops, and they had some more databases. Customer would order something online, get the product and return the product in the shop. But you have a consistency problem. You want the online customer to be treated as if you had the product in the shop, you need to synchronize and keep those two databases consistent. Once you move the data to a data warehouse, you need to harmonize the data between the two databases. This is not a performance problem, it's a semantical problem that needs rules, constraints, and validations and this will not go away with HANA, this will increase in the same way as in the past. However, the indexing of data will completely change. So we need to look at what this new technology can solve and what it can't.

12:15 Jake: If you look at a typical large enterprise customer's data landscape, for managing data for operational and BI, you see operational databases and the vast majority have deployed EDWs on top of those applications for data quality and corporate memory and BI, and many have gone further and deployed data marts on top of those. Why do those exist? Some for business reasons, but many are limited by technology constraints. One primary reason customers have built these EDW tools on top of their OLTP systems is to avoid a performance hit, so the data is replicated, transformed, and moved out. Over time, individual lines of business need to be agile and respond to changes in their business quickly. A lot of times it takes too long for IT to update this line of business information, so they end up with data marts that are geographically or line of business focus. You can have 15 EDWs and dozens if not hundreds data marts.

With HANA, there is no need to move the data our of the OLTP system for performance reasons, you can simply apply the analytics tools/views on top of the operational data once that data is n-memory, and performance is not an issue. It's no longer required to invest in aggregation, so IT doesn't have to be involved as much to provision the data for the lines of business. The business can model and provision the data themselves. 3-5 years out, this will result in massive data infrastructure consolidations. We've done test cases where huge global ERP and EDW systems can be consolidated onto single HANA instances with the LOB and analytics views implemented as productions on top of physical tables. There will always be regulatory and network performance reasons to have multiple instances, but this should lead to significant data infrastructure consolidation.

15:30 John to Jake: For customers that have deployed large scale BW without the BW Accelerator - some of those data mart scenarios might be applied via HANA instead. However, there are some scenarios where HANA does provide performance problems, especially with large scale ad hoc scenarios using the BusinessObjects suite. What is the team doing to work on that? Jake: full transparency: the product is in ramp up - we have made a lot of progress in the last seven months. There are situations where customers look at HANA and see that they can deploy apps on top of an in-memory app. But the primary use case today is not to replace the database options that customers can use today; it's also not an open database platform for all kinds of database application development. We haven't optimized all those query scenarios yet.

We advise customers beyond GA to look at our programming guide for HANA and adhere to those guidelines. We believe in our core fundamental technology but we need to work with customers for the types of scenarios they might use HANA for as we gain deeper HANA experience. We need to work with customers around the types of scenarios they implement. For SAP applications that we deliver on top, including BusinessObjects, we believe the performance is astounding, but some queries may require custom applications.

18:30 John to Jake: But from what I can see, the problems occur when customers ask ad hoc questions. Other data warehousing vendors have similar issues, and they seem to be investing in map reduction algorithms to enhance performance and other ways of addressing spanning tree problems, what is SAP doing? Jake: The answer is yes. We do have algorithms to distribute queries across multiple nodes. That's one of HANA's strengths - to scale across deployment. For exploration and ad-hoc queries, we're investing heavily in that area and we see that as an important area for us to excel in.

20:28 Ethan to the guys: What kinds of actions are you taking to help customers optimize their applications for HANA as you go into GA, to work with customers who aren't part o the ramp-up program? Jake: it starts with training, documentation, and field services enablement. We're committed to providing our customers with clear guidance on how to build applications on top of HANA. We're invested heavily in providing packaged content for customers that alleviates the need to do a lot of modeling for core ERP functions. We expect to have that content release available after June 20.

Thomas: What Jake just said is really significant: at SAP, business content is one of the areas we know best; we understand the tables in an ERP system and the underlying characteristics of HANA. That's why it's so important to not only deliver a piece of technology, but a type of content, I would consider this content the groundwork for customer-built apps on the customer side. We need to provide more apps like brining BPC on top of HANA, bringing BW on top of HANA, these are tools that allow the customers to build applications and they already abstract a little bit from the underlying technology. I can talk for BW because that's what I know best, but we changed the underlying physics of cubes and data store objects. Customers can continue to use these, but we have done the tough work of mapping that to HANA.

24:00 John to the guys: The HANA mid-term strategy looks to me like it's away from the data warehouse. You are building out data services which will allow loading from ERP to any database within reason; you are building out HANA apps with very specific use cases. So is BW and data warehousing as a concept dead? Thomas: There are two fundamental approaches in the DW world. They are independent from HANA. You can work in two directions, you can assume a physical schema of tables, and you cover those with a semantic layer such as the BusinessObjects tools. Someone designs a table schema and you put a semantic layer on top and this is the reporting layer. You have all the freedom in the world to design the schema but all the responsibility also. This is how most data warehousing databases are defined. BW has a different approach. There you define the semantic concepts, and the underlying schema is done for you based on those concepts. Other EPM tools do this as well, design the concepts first and then generate the layout underneath. We are in between these two worlds. Either way is valid. There are very large data warehouses that need to be manually designed and crafted, and there is the off the shelf DW approach which BW caters to.

Jake: DW is not dead from an SAP perspective at all. DW is a best practice and a set of capabilities that you need to have on top of your data layer. Our strategy is to continue to enable the best practices in the BW product today, running on top of HANA as an app with all the performance intensive aspects moved into HANA. Staring in Q4 this year, BW will be one of the apps that run on the in-memory platform as well as other apps such as Strategic Workforce Planning. The best practices that we have in BW that are modeled in concepts like the layer scalable architecture - we're going to keep those, though some of the implementation and management practices will change. Thomas: with HANA, we can reconcile the two different approaches. For the first version of BW running on top of HANA, we will be able to have a BW managed schema as you have known in the past, but we will also be able to access from BW an arbitrary schema where you expose tables as a cube or multi-dimensional structure. This is possible because HANA has the characteristic that you hardly have to care about query performance. That's a great aspect that relieves you from tradeoffs you had to spend a lot of time on in the past.

30:26 Jon to the guys: I had a reporter ask me at Sapphire Now, "Wait - are you trying to tell me that these customers aren't live yet?" I explained many of them were in the proof of concept (POC) phase. I talked to customers about proof of concept - one was looking at HANA to apply to a high sales volume area where the BW latency issues were hurting their sales opportunities. What is SAP doing to work with customers to move them from proof of concept to go-live? Jake: we do have a large number of POC customers we are working with. When you think about HANA, it is new technology that is not only something that is new from a product perspective, it also changes how IT can interact with the line of business. We've heard from a number of customers that HANA enables a greater degree of Line of Business self-sufficiency, so that requires some change management. Hilti's input is that this will change the way IT services the needs of the business.

Many customers want to get their feet wet with POCs to target specific scenarios. We are working with a large number of POC customers and moving them towards go live post GA. How it works is that we outline a number of substantial scenarios where the benefits of in-memory can have a real impact, we look to complete the POC in a short period of time and bring the results of the POC to senior executives along with a proposal for a much larger rollout. This can have huge implications for data landscape transformation. Customers can need some time to think through that. The POC phase is a great way to have those discussions while also moving forward with a practical use case on the ground.

34:10 Ethan to the guys: A couple months ago, there was a release of an internal benchmark SAP did with an independent auditor to verify the benchmark findings. The idea was that the internal benchmark was very impressive. But it's difficult to get a handle on what that means for individual customers without getting a handle on that benchmark. Is SAP issuing a recreatable benchmark that gives customers the data set and queries that were run so that it can be verified on a customer site? Jake: The way benchmarking works at SAP is that we have defined a large number of benchmarks for different kinds of use cases. In order to become an official SAP benchmark, the SAP Benchmarking Council has to agree and vote for it to become a benchmark that can then be released to various companies and third parties for companies to run. We're in the process of moving the "H" benchmark through the process, before we do that we can't openly share it or have third parties test against it. It's something we've spent a lot of time defining, and it is something that is based on the TPC-H queries, however the data models and data sets have been greatly extended to reflect realistic customer use cases. The performance we have achieved is astounding, we are 100 times faster than traditional database systems across the queries, and we're looking to have the council sign off on the benchmarking tests and hopefully we'll be able to share it publicly.

Ethan to the guys: Will it be possible to use that on another database, such as Max DB and Oracle? Jake: Once it's approved, we would define this as a new business intelligence benchmark, and all our OEM partners would have an opportunity to executive the tests and submit the results. That's exactly what we did internally.

38:13 Jon to the guys: Want to turn the tables and give you a chance to sound off - are there any misconceptions you've seen about HANA that drive you a little crazy? Jake: I don't think folks in the blogging community are purposely misrepresenting things, but there has been a lot of confusion around BW. To be clear: BW is not dead, data warehousing at SAP is not dead. Thomas: what I would like to do is to emphasize one thing that holds true all the time: from a technology perspective, HANA is something outstanding, but when you talk about HANA, you should keep in mind that what we are going to do is make the best fit for running SAP's applications, that's a competitive advantage that we have beyond the traditional set of working on standard databases. HANA will have application-specific algorithms that are performance-optimized within HANA.