INFORMATION FRAMEWORKS

THE SAP BUSINESS INTELLIGENCE ACCELERATOR

TURBO CHARGER FOR SAP NETWEAVER™ BUSINESS INTELLIGENCE

Naeem Hashmi
Chief Strategy Officer
Information Frameworks
May 19, 2006

Table of Contents

THE BUSINESS INTELLIGENCE ACCELERATOR

For SAP NetWeaver™ Business Intelligence

The Business Intelligence Accelerator (**BI accelerator**) is good news for SAP customers. As long as I have been engaged in the SAP BI design since its conception, I see what Shai Agassi demonstrated at the SAPPHIRE'06 in Orlando is a giant performance leap forward in end-user access to the analytical data. One of the main pain points for SAP NetWeaver BI (**SAP BI**) has been its data access performance due to its sophisticated relational OLAP model. That model works well for small to mid sized business intelligence operations. Now with the BI accelerator, large SAP BI implementations will enjoy super-high data access speed that is necessary to implement 'real-time' embedded-analytics for intelligent business solutions, a key to SAP Enterprise Services Architecture (ESA) based future business solutions. To prove the super high performance gains, Shai demonstrated an analytical xApp that accessed large data volumes in SAP BI through BI accelerator at a lightening speed.

Shai stated in his keynote:

> *'…One of the biggest issues in a data warehousing is that you have to figure out all the queries you are going to get to build cubes ahead of time. You have to almost think what your users will think, and pre-think to design the architecture of a data warehouse. So what we [SAP] have done with BI accelerator is effectively removed the need to build the cubes…. **No more Cubes**…. You put the data, stored in the BI accelerator, it stores it in different way than before, but it stores it on-line, immediately accessible, and you **don't need to pre-think of any question that any user will ask**…. So No Cubes…"*
>
> *Shai Agassi, Keynote at Sahhpire'06, Orlando, May 17, 2006.*

What caught my attention were the two key points emphasized during his keynote. '*No-More Cubes'* and "*you don't need to pre-think of any question that any user will ask*".

Humm… 'No more Cubes'… perhaps this needs clarification on what Shai meant which confused quite a few SAP BI customers who approached me after his keynote. Let me explain why such confusion.

To best of my knowledge, BI accelerator is an 'Appliance' that speeds up Data Access (queries) task of SAP BI. The BI accelerator sits on top (or on the side) of a SAP BI instance. Data access is only one important piece of a data warehouse's overall performance equation. BI accelerator addresses data access performance, scalability and to some extent data load-performance issues (as no aggregates are required anymore). Period.

HOW SAP NETWEAVER MAKES ANALYTICAL DATA AVAILABLE TO ITS USERS?

In SAP BI (as well as in former SAP BW), you design Cubes (InfoCubes) based on '*end-user requirements*'. Each Cube has one Fact table (usually very large, stores only measures) that is linked to many Dimensions tables (relatively small compared to the Fact table). When a user queries against a Cube in SAP BI using BEx (or any other access tool), the relevant data is fetched for InfoCubes (physically a set of relational tables organized in a star schema) from database server up in the application server cache in the form of a virtual multidimensional structure (to support navigation). Subsequent queries will be served from this cache (at application server level) and when further data is needed, the SAP BI analytic engine will request the Data Manager to fetch more data from the database server.

As Cube volume grows or when dimensions complexities increase (snow-flaked, timed based, multiple hierarchies, etc.), the data access performance starts to degrade rapidly. When this happens, you build aggregates in SAP BI based on small *user data navigation views* – the frequent queries. Next time, when user issues a query, the analytic engine uses these aggregates to fulfill user requests without accessing the larger Cube. This speeds up data access because Data Manager has to do little work to look up data from aggregates instead of the full cube scan.

SAP BI supports two directly access-able data stores for reporting and analysis purpose - Cubes (InfoCubes) and Operational Data Store (ODS Objects – now renamed as DSO, Data Store Objects in SAP NetWeaver 2004s). ODS objects are flat database tables without dimensional structures. The ODS objects can contain any type of data, structured, unstructured or blobs, while Cubes hold only numeric data (measures) in its Fact tables.

HOW SAP BI ACCELERATOR WORKS?

Simply said, the BI accelerator is an appliance – pre-installed 64-bit Intel® Xeon® processor-based blade servers from HP and IBM, running Linux Operation System. The TREX (text retrieval and classification) engine with in-memory attribute search engine for structured data is installed on each blade (say, 4 blades, 4 server instances, 4 TREX instances) to support parallel search capability. It has file system for storage (for Indexes/data) but no database engine.

At present, you establish an RFC connection between one SAP BI instance to one BI accelerator. Then in each SAP BI instance, define remote links to the BI accelerator instance.   In other words, an SAP BI accelerator can act as a unified enterprise wide business intelligence hub by sourcing raw data from several SAP BI instances.

HOW DO SAP BI AND BI ACCELERATOR WORK TOGATHER?

In an SAP NetWeaver 2004s BI instance you define an RFC link for the BI accelerator instance. When you 'ROLLUP' a cubes in SAP BI, you elect the Cube to be rolled-up (aggregates) in the connected BI accelerator instance. Cube Dimensions are evenly stored on *individual* blades. On the other hand, the Fact table is horizontally partitioned and stored *across* all blades, whereby columns can be accessed individually (vertical decomposition)

Both, indexes and data is moved to the BI accelerator file system. This helps the BI accelerator server on each blade to not only search the indexes but also the associated data as well at lightening speed – all in parallel using all blades.

When a user issues a BEx query or access through BAPI, or API based application, the analytic engine (earlier BW OLAP Processor) knows that a BI accelerator is attached to the SAP BI instance. The analytic engine simply routes the query to the BI accelerator by bypassing the relational store (skipping aggregates or Cube tables in SAP BI). The analytic engine formats the user query in a BI accelerator-proprietary language and passes it along to the BI accelerator where BIA servers running on each blade runs at full throttle to search what user has requested and hands over content back to analytic engine which internally maps back to user requested format. The relational database is completely bypassed in this situation (Fact/Dimension Tables).

Originally, TREX search engine employed techniques such as KNN, Centroid and SVM classification on top of a vector space model for classification of content (used for unstructured content – documents; SAP NetWeaver Portal uses these search algorithms) but for BI accelerator, SAP has now implemented an in-memory attribute search engine for structured data that does ad hoc aggregation at a lightening speed. In other words, you build aggregates on the fly in BI accelerator based on user query. This is the reason why the very first time a query is some what slower because BI accelerator needs to build the aggregate in its memory, like OLAP Cache in SAP BI Application Server.

The BI accelerator literally acts as an in-memory database for complete SAP BI Cubes providing access through the BIA engine instead of SQL. In SAP NetWeaver vocabulary, BI accelerator serves as a cache at the SAP BI application server and is at the same time a full Cube /Aggregate.

Also notice here that SAP BI supports remote InfoCubes as well as Multi-Providers (union of multiple physical InfoCubes). A large number of SAP BI customers have implemented Multi-Providers. Notice that, today BI accelerator does not support Multi-Providers, but I assume one can simulate multi-provider like model (union of Cubes) programmatically through APIs against underlying aggregated InfoCubes in BI Accelerator (and much faster).

Note here that at present, you can not use BI accelerator for ODS tables. It works only for Cubes defined in SAP BI. Perhaps future versions of BI accelerator will incorporate other SAP NetWeaver data stores as well. However, today, you have to have Cubes in SAP BI before you can use BI accelerator. Period.

ANALYSIS

The BI accelerator is simply an appliance for SAP BI. You have to have SAP BI. In SAP BI, you model and construct Cubes – Meaning, you have to get user requirements. You have to define Facts. You have to define Dimensions. You have to define complete Data Warehouse Architecture – starting from User Requirements – Extraction – Data Transformation – Data Mapping – Data Quality and Distribution and Data Access Model and data load strategy in Cubes before you touch BI accelerator.

This lack of clarity resulted in confusion among SAP Customers when Shai emphasized repeatedly 'No More Cubes'.

To use BI accelerator, today, you have to have Cubes in SAP BI. When Shai emphasized twice in his keynote 'No-More-Cubes' is not what it means rather a very forward-looking statement. With BI accelerator in place, eliminates the need for aggregates in SAP BI. Also, data modeling may be simplified, as many restraining performance considerations are no longer valid.

Moving forward, SAP intends to make InfoCubes optional in the process, and then indeed it will conceptually be perfectly fine to have 'no more Cubes'. Of course, that does not mean that business process- or data modeling won't be required anymore. Data models can be kept more generic, as the data modeler will no longer required to extensively 'pre-think' (or anticipate) future requirements but such broader general purpose model and lack of 'end-users pre-thinking' will add more complexity in secured access protocols especially when information has to be aggregated from several layers supporting different generic models. Such approach raises several privacy and security issues (typical issues with search oriented models). In nutshell, BI accelerator may simply a few areas in implementing SAP BI but at the same time will open new challenges on the design-run-management-time practices and established business intelligence best practices and architectures.

And to define a cube (to be physical instance or an aggregated cube in BIA) you must also have to gather **user business and access requirements** and think of all possible ways to architect a high performance Business Intelligence solution. When making available all sort of corporate information without end-user access and content requirements, you will end-up with an environment like the Internet. Tons of quickly accessible information but not relevant what the user is looking for – no context. So business access requirements have to be at the hart of any data/information access strategy to architect an viable ecosystem best aligned with the business and their end-user needs – both structured & unstructured content.

For Data Warehouse pros, the concept of BI accelerator is similar to good old HOLAP, although the technology and approach is radically different. Meaning, the content is transformed into proprietary structures in another layer on top of Relational-OLAP implementation. User access layer sends incoming queries to HOLAP for quick access/navigation instead of Relational-OLAP. The only difference here is that BI accelerator uses powerful search engine technology, transparent to traditional data warehouse end users.

The BI accelerator cost (as mentioned in the keynote) is quite appealing to SAP BI customers. I see more applications of BI accelerator in other intense data access environments. For example, the master data management implementation is only beneficial when mission critical applications use a central master data management infrastructure. I see next SAP NetWeaver MDM implementation to use BI accelerator like appliance for high performance master data lookups. I can think of several innovative implementation of BI Accelerator. I will discuss a few such scenarios of BI Accelerator in my upcoming book 'Analytics Design Strategies' to be published late September 2006.

When looking at the BI accelerator pricing, it takes me back to Shai's keynote at the SAPPHIRE. Shai stated.

> *"… One of their customers was considering two options to build a high performance data access environment. The two options plan to accelerate their Data Warehouse Program – One, the Highly SMP solution with man many CPUs on vary very large box with a database from …a database vendor… [and another option the SAP BI accelerator]. We put comparison on what happened between the two options.* **3 Millions Euros** *worth of Hardware with* **½ a Million Euros** *worth of license on the Data Warehouse from them. Compared to* **150 Thousands Euros** *blades based IBM box, Intel based processor with many many processors but still only 150 Thousands Euros [BI accelerator] option. Now the interesting thing is not just how much it costs us to acquire the hardware and software, it is how much it costs you to manage and operate these two environments [One being Database Vendor Option vs. SAP BI accelerator]. If you look how many DBAs it needs to maintain and operate on stored data warehouse of this magnitude vs. what you need to do with the box that literally plugs in. Whirlpool plugged in this box [BI accelerator] within one week and live and productive in one week. That what you get in comparison....* --- *No more DB, no More Admin, No more Cost…"*
>
> *Shai Agassi, Keynote at Sahhpire'06, Orlando, May 17, 2006.*

Humm … are we comparing Apples and Oranges here? We are talking about an Appliance vs. an Entire Data Warehouse Infrastructure. It is correct to say that BI accelerator is an appliance that Turbo Charges SAP BI data access and that's all. In other words we are comparing cost of "high octane gas' or 'Turbo Charger' for a Car to perform better vs. a brand new high-end race CAR. Not a right comparison. The best comparison would be to add SAP BI costs (HW/SW/Support/DBA) plus BI accelerator and then compared with the new option for a Data Warehouse Solution from another vendor(s). That would have been a right comparison. There too, SAP solution will relatively win but this will be a right way to justify and compare.

CONCLUSIONS

In my opinion, SAP BI with BI accelerator option will be a win-win proposition for high-end SAP BI customers, not necessary from HW/SW/Support costs angles but from what business value SAP BI will add naturally to your SAP investment such as pre-defined business content (the most complex and time consuming task), analytics, meta-model and continued future built-in capabilities that will come naturally with SAP BI as SAP adds other business solutions and most important value of SAP BI being an integral part of NetWeaver to support its ESA driven intelligence business solution.

I strongly agree that the BI accelerator is a great option for SAP BI customer with large SAP BI data warehouses. Even with BI accelerator, SAP BI customers need to remember that they need to be very careful in architecting their enterprise business intelligence solution that is aligned with emerging business and technology architectures (Service–Oriented) which requires embedded-analytics rather typical reporting and analysis model. The key is to listen

to user's needs. How they want to use information to make right decisions. And remember; only a handful of users will require detailed data but most will be just as happy to get minimum data points to meet their needs. And if you plan to use BI accelerator, you still have to maintain your SAP BI infrastructure – Cubes (for now), ODS implementations, ETL, Data Loads, Maintenance etc. And most important fact to remember is that you need to be quite careful how you model objects that to meet your needs without BI accelerator and with BI accelerator. Otherwise, you will end-up redesigning core SAP BI objects when you plan to implement BI accelerator. So you have to think of possibility of BI accelerator implementation at the design time of SAP BI objects.

In nutshell, when Shai stated *'No More Cubes'*, perhaps he was referring to 'no more aggregates in SAP BI' and for *'you do not need to pre-think of any question that any user will ask'*, perhaps, he meant 'no user run-time ad hoc queries definition'.

About The Author: **Naeem Hashmi, Chief Research Officer, Information Frameworks**

Naeem is a world-renowned thought leader, author, speaker, and expert on emerging Information Technologies with 25+ years of experience in Enterprise Architectures, Distributed Service-Oriented Business applications design, ERP application Integration, Information Delivery Architectures, Data Warehousing; Data Mining; CRM, Management Consulting, Knowledge and Technology Transfer and visualizing new products. Current research work includes Information Visualization, Social Networks, Information Psyche Mining, and the Enterprise Architectures.

Naeem is Founder, Chief Research Officer for Information Frameworks Naeem advises IT vendors on innovative technologies and product He authored/co-authored four books; the best selling book 'Business *Information Warehouse for SAP'*, a contributing author for upcoming book 'Technologies *for Government Transformation: ERP Systems and Beyond* edited by Shayne Kavanagh et.al. To be published by the Government Finance Officers Association. Naeem is a frequent speaker at national and international IT conferences. He has authored and published well over 100+ articles on Information Technologies and has authored several cover stories and feature for the Intelligent Enterprise Magazine, Intelligent ERP Magazine, Manufacturing Systems International and the CIO magazine. Naeem served a distinguished judge for Intelligent Enterprise Award for IT Excellent 2005 in Oct 2005.

Upcoming Books:
Analytics Design Strategies. To be published: Sept. 2006
Enterprise Architectures and Governance Models. To be published. Jan. 2007

Naeem can be reached at nhashmi@infoframeworks.com   T: 603-432-4550 M: 603-661-6820
Web Site http://infoframeworks.com

ACKNOWLEDGMENTS

Disclaimer:
As SAP NetWeaver Business Intelligence Accelerator is still evolving, this paper analysis is based on best information available for the present and near state of Business Intelligence Accelerator as of May 19th, 2006.