A High-Performance, Scalable Big Data Appliance

WebTech Q&A Session, February 20, 2013

1. **What is ETL?**
   Extract, transform and load (ETL) refers to a process in database usage and especially in data warehousing that involves extracting data from outside sources, transforming it to fit operational needs (which can include quality levels), and loading it into the end target database (operational data store, data mart or data warehouse).

2. **How do I size for the SAP HANA BW database?**
   The process is described in SAP Note 1514966 as well as automated through the use of the SAP Quicksizer utility. The primary factor is the amount of physical RAM required by the in-memory database. Not only does SAP HANA need to store the customer’s full data set, it also needs enough physical RAM to perform the operations traditionally performed at the application tier. The SAP manual as well as the Quicksizer output should be considered a starting point and never an official HDS recommendation. Always work with the local SAP account team and customer to establish an official recommendation prior to ordering the HANA appliance.

3. **Can you tell me more about your partnership with SAP?**
   Hitachi has a 20-year relationship with SAP. In 2011, Hitachi Data Systems grew its partnership by joining SAP’s elite partners as a Global Technology Partner. Our joint work around Big Data (http://www.news-sap.com/big-data-bundle-from-sap-capitalizes-on-power-of-hadoop-speed-of-in-memory-computing) is one example of how SAP and HDS are currently working together to bring new and more integrated solutions to market.
4. **How about your partnership with Hitachi Consulting and their expertise in this space?**
Hitachi Consulting, a longstanding SAP Global Services Partner and a close partner of Hitachi Data Systems, provides expertise in traditional SAP services as well as HANA. Hitachi Consulting offers various HANA-related services that can help customers size an appliance, integrate it into a SAP landscape (delivering also all required SAP upgrades as well as data integrations and migrations), manage the environment, and deliver measurable ROI and business value around HANA within weeks (through their Live In Five program).

5. **To we need to be concerned about about DR with scale-out and/or physical backups?**
Operations are critically important as customers embark on their HANA journey. This point is emphasized by SAP’s announcements around “Suite on HANA” – clearly customers look for the same level of protection when ECC is powered by HANA as they have today for ECC with a traditional database. As HANA is an evolving technology, customers should evaluate various HANA hardware vendors’ strategies and how their core technologies can add value to the overall solution. Hitachi Data Systems has tremendous value and unique differentiators when customers are considering an enterprise solution.

6. **For SAP HANA, is Fibre Channel SAN a feasible storage solution or can file-based storage be used?**
SAP supports both Fibre Channel and network attached storage for the HANA persistent storage layer (data and log volumes). Although some vendors use network attached storage, it is our belief (confirmed by our customers) that Fibre Channel is required to build mission-critical platforms that deliver predictable and consistent performance, scalability, reliability, serviceability and operational robustness. Also, when considering NFS- or MPFS-based systems consult SAP Note 1740136, which points to issues in those types of setups.

7. **Does a customer buy a separate Hitachi Virtual Storage Platform (VSP) for the HANA scale-out or can an existing VSP be used?**
At this point, SAP HANA certification does not allow an appliance to share storage resources with other applications or other HANA appliances. Customers will need to purchase each HANA appliance as a complete integrated system. This may change over time as the SAP certification process evolves, but the requirements are made by SAP. VSP would certainly be uniquely capable of sharing workloads.

8. **What about multi-petabyte environments; for example, 672PB growing to 1EB in 18 months?**
This may be a good use case for combining technologies like Hadoop and SAP HANA. Hadoop can perform the preprocessing and then load the HANA appliance with the relevant (and smaller) data set via data services. While HANA appliances are not limited to the 16 nodes that are required for SAP certification, 1EB is certainly a very large data volume even if one factors in compression through HANA.