SAP HANA: The Evolution from a Modern Main-Memory Data Platform to an Enterprise Application Platform

Vishal Sikka, Franz Färber, Anil Goel, Wolfgang Lehner
SAP AG
{vishal.sikka, franz.faerber, anil.goel, wolfgang.lehner}@sap.com

ABSTRACT

SAP HANA is a pioneering, and one of the best performing, data platform designed from the grounds up to heavily exploit modern hardware capabilities, including SIMD, and large memory and CPU footprints. As a comprehensive data management solution, SAP HANA supports the complete data life cycle encompassing modeling, provisioning, and consumption. This extended abstract outlines the vision and planned next step of the SAP HANA evolution growing from a core data platform into an innovative enterprise application platform as the foundation for current as well as novel business applications in both on-premise and on-demand scenarios. We argue that only a holistic system design rigorously applying co-design at different levels may yield a highly optimized and sustainable platform for modern enterprise applications.

1. THE BEGINNING: SAP HANA DATA PLATFORM

A comprehensive data management solution has become one of the most critical assets in large enterprises. Modern data management solutions must cover a wide spectrum of additional data structures ranging from simple key-values models to complex graph structured data sets and document-centric data stores. Complex query and manipulation patterns are issued against the database reflecting the algorithmic side of complex enterprise applications. Additionally, data consumption activities with analytical query patterns are no longer reserved for decision makers or specialized data scientists but are increasingly becoming an integral part of complex operational business processes requiring support for analytical as well as transactional workloads managed within the same system [4].

Dealing with these challenges [5] demanded a complete re-thinking of traditional database architectures and data management approaches now made possible by advances in hardware architectures. The development of SAP HANA accepted this challenge head on and started a new generation of database system design. The SAP HANA database server now comprises a centrally, and tightly, orchestrated collection of different processing capabilities, e.g., an in-memory columnar relational store, a graph engine, native support for text processing, comprehensive spatial support, etc., all running within a single system environment and, therefore, within a single transactional sphere of control without the need for data replication and synchronization [2].

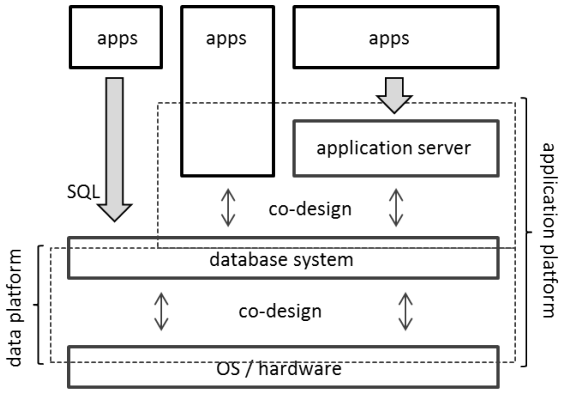

Figure 1: The SAP HANA platform

Secondly, and most importantly, SAP HANA has triggered a major shift in the database industry from the classical disk-centric database system design to a ground breaking main-memory centric system design [3]. The mainstream availability of very large main memory and CPU core footprints within single compute nodes, combined with SIMD architectures and sophisticated cluster systems based on high speed interconnects, was and remains, the central design guideline of the SAP HANA database server. SAP HANA was the first commercial system to systematically reflect, and exploit, the shift in memory hierarchies and CPU architectures in order to optimize data structures and access paths. As a result, SAP HANA has yielded orders of magnitude performance gains thereby opening up completely novel application opportunities. Most of the core design advances behind SAP HANA are now finding their way into mainstream database system research and development, thereby reflecting its pioneering role.

As a foundational tenet, we see rigorous application of *Hardware/Database co-design* principles as the main success factor to systematically exploit the underlying hardware platform: Literally every core SAP HANA data structure and routine has been systematically inspected, redesigned

Proceedings of the VLDB Endowment, Vol. 6, No. 11

and optimized for corresponding H/W primitives.

The HANA database server is now the core of the SAP HANA data platform (figure 1) extended by a comprehensive set of additional data management services, ranging from modeling tools to data loading and transformation utilities to a wide infrastructure for data consumption [1]. Additionally, the breadth of SAP HANA data platform encompasses Hadoop-based systems for low-density, high-volume data sets and includes SAP Event Stream Processor for high-velocity stream based consumption patterns.

2. THE EVOLUTION: SAP HANA ENTERPRISE APPLICATION PLATFORM

As early as 10 years ago, Jim Gray opined "We're ready for a pretty radical departure from how people currently access databases, and I don't know what it's going to be." [6]. Our vision is to provide robust enterprise solutions to support the ever increasing spectrum of (implicit) database users ranging from knowledge workers to data scientists and more. The SAP HANA data platform is being defined and delivered as *the* foundation of a next-generation application platform to build modern and large-scale business applications that allows for flexible and efficient development on the one side, and robust, non-disruptive deployments, and low TCO operations on the other side.

As illustrated in figure 1, the core strategy once again consists of applying a *Database/Application co-design* principle as rigorously as it has been executed for the HANA data platform. This co-design challenge is being tackled without compromising traditional SQL interfaces. From the application viewpoint, the SAP HANA platform supports two contrasting paradigms:

- Highly critical business object contents, logic and locking protocols may be redesigned and intertwined with features of the underlying data platform which provides comprehensive support for business function libraries tightly integrated into the core engine, e.g., into query optimization and distributed query processing, for optimal performance.
- New applications may take advantage of centrally delivered platform components natively accessible from within a tightly integrated application server as part of the SAP HANA enterprise application platform.

From the database system perspective, a large number of functional and non-functional extensions are provided to ease application development and deployment, as well as to optimize the runtime behavior of business applications. These non-traditional extensions are essential for a data platform to serve as the foundation for the SAP HANA enterprise application platform with a focus on large-scale on-premise and on-demand scenarios. Without diving into detail, we give some examples:

- **multi-tenancy support:** SAP HANA provides support to virtually isolate different tenants within a single system. Isolation may happen at different layers. For example, in traditional SAP application environments, not only database schemas but also static content of SAP deployments may be shared among different tenants without compromising individual extensibility. User transactions are guaranteed to stay within the sphere of a single tenant while administrative transactions, e.g., to upgrade an application, may cross the boundaries of multiple tenants.
- **deployment support:** In hosted environments, it is very desirable for applications to roll forward without interrupting running business landscapes. The database system is therefore required to atomically apply 1000s of schema and content changes without affecting running transactions, thus preparing them in the background and minimizing (or even avoiding any) application downtime.
- **support for rich and customizable data models:** Generally speaking, significant semantics is lost when complex business objects are mapped to the relational layer; when reconstructing the objects using complex SQL queries, the optimizer has to deduce some of the relationships based on statistics and simple constraints. In order to better support business applications, SAP HANA exhibits the ability to register complex structured business objects directly within the meta data repository. The system (esp. the optimizer) is then able to exploit this information during compile and runtime.

3. CONCLUSIONS

Hardware/Database co-design was one of the main pillars of success for the current SAP HANA data platform that has defined and delivered a new vision for data management on modern hardware. Systematically mapping database primitives to hardware features and characteristics provides a best-of-breed approach. As a next step, SAP HANA will evolve from a data platform to a comprehensive enterprise application platform for existing and new SAP and non-SAP applications. Focusing on an efficient and flexible development process, non-disruptive deployment process, and robust and low administration runtime, the SAP HANA platform again will follow *database/application co-design* approach. Only this systematic and holistic approach may yield a sustainable, well-structured, and highly performing software stack.

4. REFERENCES

[1] F. Färber, S. K. Cha, J. Primsch, C. Bornhövd, S. Sigg, and W. Lehner. SAP HANA database: data management for modern business applications. *SIGMOD Record*, 40(4), 2011.

[2] F. Färber, N. May, W. Lehner, P. Große, I. Müller, H. Rauhe, and J. Dees. The SAP HANA Database – An Architecture Overview. *IEEE Data Eng. Bull.*, 35(1), 2012.

[3] H. Plattner and A. Zeier. *In-Memory Data Management: Technology and Applications*. 2012.

[4] V. Sikka, F. Färber, W. Lehner, S. K. Cha, T. Peh, and C. Bornhövd. Efficient transaction processing in SAP HANA database: the end of a column store myth. In *SIGMOD Conference*, 2012.

[5] M. Stonebraker, S. Madden, D. J. Abadi, S. Harizopoulos, N. Hachem, and P. Helland. The End of an Architectural Era. In *VLDB*, 2007.

[6] M. Winslett. Interview with Jim Gray. *SIGMOD Record*, 32(1), 2003.