International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-3 Issue-5, June 2014

Overview of Sap Hana: In-Memory Computing Technology and ITS Applications

Srikantarao S, Radhika K R

Abstract— Traditional database management systems are designed for optimizing performance on hardware with constrained main memory. Disk I/O was the main bottleneck. The focus was on optimizing disk access, where by minimizing the number of disk pages to be read into main memory during processing. The SAP HANA database is designed from the ground up around the idea that memory is available in abundance. Instead of optimizing I/O hard disk access, SAP HANA optimizes memory access between the CPU cache and main memory. SAP HANA is a massively parallel (distributed) data management system that runs fully in main memory, allowing for row- and column based storage options, and supporting built-in multitenancy.

Index Terms—High-Performance Analytic Appliance (HANA), Structured query language (SQL), multidimensional expression (MDX), Online transaction processing (OTLP), Online analytical processing (OLAP).

I. INTRODUCTION

HANA, stands for "High-Performance Analytic Appliance" is an in-memory, column-oriented, relational database management system developed and marketed by SAP AG. This can be deployed on premise or on demand cloud product. Today's business users need to react much more quickly to changing customer and market environments. They demand dynamic access to raw data in real time. SAP HANA empowers users with flexible, on-the-fly data modeling functionality by providing nonmaterialized views directly on detailed information. SAP HANA liberates users from the wait time for data model changes and database administration tasks, as well as from the latency required to load the redundant data storage required by traditional databases. The elimination of aggregates and relational table indices and the associated maintenance can greatly reduce the total cost of ownership.

SAP HANA Platform Overview

SAP HANA is a modern platform for real-time analytics and applications (see Figure 1). It enables organizations to analyze business operations based on a large volume and variety of detailed data in real time at very high speed. Initial deployments of SAP applications on SAP HANA have shown that business users can act on shorter system response times, which opens the door to application possibilities that may not yet have been imagined. The platform can be deployed as an appliance or delivered via a cloud. SAP in-memory computing is the core technology underlying the SAP HANA platform.

Manuscript published on 30 June 2014.
* Correspondence Author (s)
Srikantarao S, Information Science, BMSCE, Bangalore, India.
Dr. Radhika K R, Information Science, BMSCE, Bangalore, India.

Combining Online analytical processing (OLAP) and Online transaction processing (OTLP) into one database, it creates a unified view on data from transaction, analysis, decision and planning systems. Real-time analytics and transactional applications, including planning, can run in mixed operations. Multidimensional Expressions (MDX) is a query language for OLAP databases, much like SQL is a query language for relational databases. It is also a calculation language, with syntax similar to spreadsheet formulas.

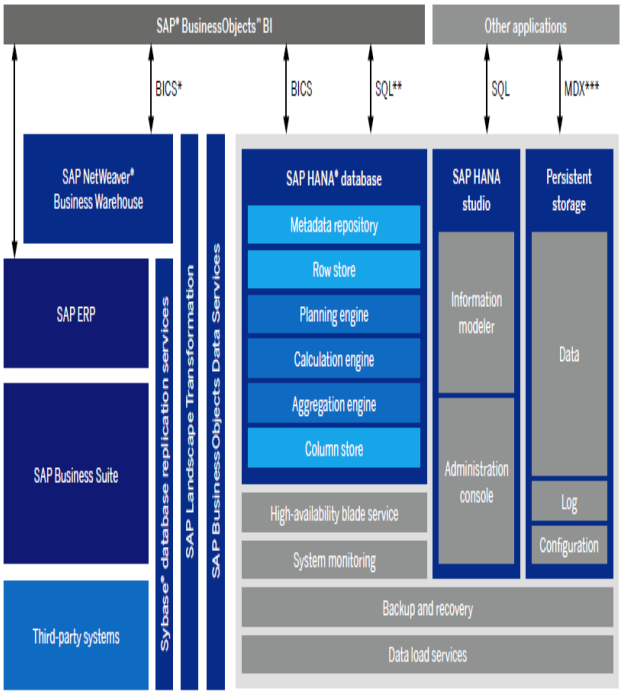

Figure 1: SAP HANA Platform Overview

II. EXISTING SYSTEM

A database table is conceptually a two-dimensional data structure organized in rows and columns. A row-oriented organization stores a table as a sequence of records which has less data compression ratio. For faster search operations, it may needs to create many index structures which are also overhead to the system. With row-oriented storage, the operations on single columns, such as searching or aggregations would be much slower because data of the same column is distributed across memory and the CPU is slowed down by cache misses. Some solutions offer columnar storage on traditional hard-disk technology, while other platforms offer the option of storing data on solid state disks (SSD). Although these disks have no moving parts and access data much more rapidly than hard disks, they are still slower than in-memory access.

III. PROPOSED SYSTEM

SAP HANA runs on multi-core CPUs with fast communication between processor cores and containing terabytes of main memory. Here all the data is available in main memory, which avoids the performance penalty of disk I/O. Either disk or solid-state drives are still required for permanent persistency in the event of a power failure or some other catastrophe. This does not slow down performance, however, because the required backup operations to disk can take place asynchronously as a background task.

IV. IMPLEMENTATION

In column storage the entries of a column are stored in contiguous memory locations. SAP HANA supports both, but is particularly optimized for column-order storage. Columnar data storage allows highly efficient compression. If a column is sorted, often there are repeated adjacent values. SAP HANA employs highly efficient compression methods, such as run-length encoding, cluster coding and dictionary coding. With dictionary encoding, columns are stored as sequences of bit-coded integers. That means that a check for equality can be executed on the integers; for example, during scans or join operations. This is much faster than comparing. Storing data in columns is functionally similar to having a built-in index for each column. The column scanning speed of the in-memory column store and the compression mechanisms – especially dictionary compression – allow read operations with very high performance. In many cases, it is not required to have additional indexes. Eliminating additional indexes reduces complexity and eliminates the effort of defining and maintaining metadata. SAP HANA Database High-Level Architecture is shown by the figure 2.

A running SAP HANA system consists of multiple communicating processes (services). The following figure 2 shows the main SAP HANA database services in a classical application context. The main SAP HANA database management component is known as the index server, which contains the actual data stores and the engines for processing the data. The index server processes incoming SQL or MDX statements in the context of authenticated sessions and transactions. The index server uses the preprocessor server for analyzing text data and extracting the information on which the text search capabilities are based. The name server owns the information about the topology of SAP HANA system. In a distributed system, the name server knows where the components are running and which data is located on which server. The SAP HANA database has its own scripting language named SQLScript. SQLScript embeds data-intensive application logic into the database. Classical applications tend to offload only very limited functionality into the database using SQL. This results in extensive copying of data to and from the database and in programs that slowly iterate over huge data loops that are hard to optimize and parallelize. SQLScript is based on side-effect free functions that operate on tables using SQL queries for set processing, and is therefore parallelizable over multiple processors. The database persistence layer is responsible for durability and atomicity of transactions. It ensures that the database can be restored to the most recent committed state after a restart and that transactions are either completely executed or completely undone.

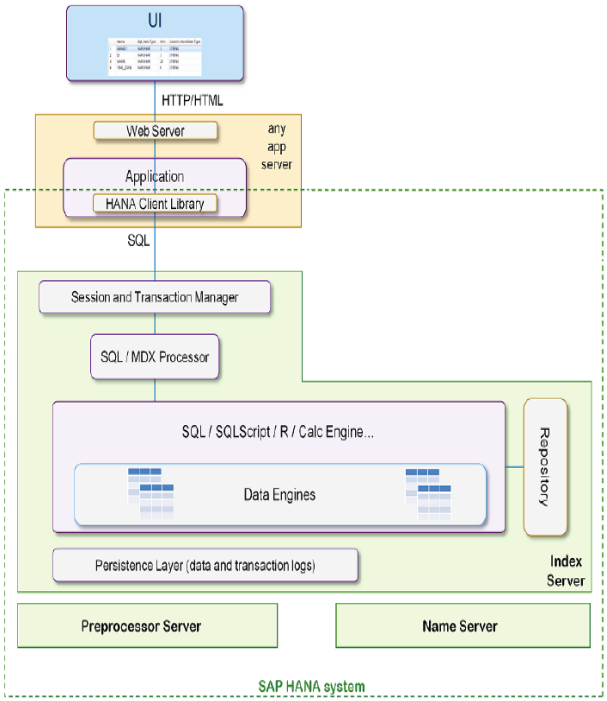

Figure 2: SAP HANA Database High-Level Architecture

V. APPLICATIONS

a) Recently SAP has been integrated their ERP solution in HANA called Business ByDesign to get the full use of the capabilities of current hardware to increase application performance. It is fully integrated on-demand enterprise resource planning and business management software product designed for small and medium sized enterprises. Business ByDesign solution is a type of software as a service (SaaS) offering which can be run on a PC with an Internet connection and a web browser, while the software and data are stored on the host servers. Business applications are delivered as an on-demand service via a secure Internet connection and a standard Web browser. Like most SaaS solutions, SAP Business ByDesign has a pay-per-use fee instead of an up-front investment [5].

The SAP Business ByDesign solution is designed to track end-to-end business processes across the following modules.

- Customer relationship management: This module supports processes that span marketing, sales and service activities.
- Financial management: This module can help provide companies with a single, up-to-date view of financial condition by integrating core business processes and financials that span financial and Management accounting and cash flow management.
- Project management: This module contains an integrated project management solution.
- Supply chain management: This module covers supply chain setup management, supply chain planning and control and manufacturing warehousing and logistics.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

- Supplier relationship management: This module focuses on relationships with suppliers, procurement processes aimed at reducing costs and to perform self-service procurement.
- Human resources management: This module spans organizational management, human resources and employee self-service.

b) One more recent application of HANA platform has been selected to develop a statistical analysis tool for IPL to assist Kolkata Knight Riders (KKR) management during the player auction at Bangalore on 12 February 2014. The real-time application named SAP Auction Analytics leverages the predictive analytics capabilities of SAP HANA which enables KKR to derive intelligence on the players to be featured in the auction. The platform provides significant insights based on various aspects of the players and a live dashboard of the performance of other teams during the bidding process. SAP Auction Analytics, capable of running on multiple devices, is custom developed for KKR. However its features are extensible and can be used by other teams. SAP Auction Analytics will deliver significant competitive edge to KKR via real-time actionable insights that can make decision-making more efficient [6].

VI. HIGH AVAILABILITY FOR SAP HANA

SAP HANA is fully designed for high availability. It supports recovery measures ranging from faults and software errors, to disasters that decommission an entire data center. High availability is achieved by eliminating single points of failure (fault tolerance) and providing the ability to rapidly resume operations after a system outage with minimal business loss (fault resilience). Fault recovery is the process of recovering and resuming operations after an outage due to a fault. Disaster recovery is the process of recovering operations after an outage due to a prolonged data center or site failure. Preparing for disasters may require backing up data across longer distances and may thus be more complex. The key to achieving high availability is redundancy, including hardware redundancy, network redundancy and data center redundancy. SAP HANA provides several levels of defense against failure-related outages:

Hardware Redundancy – SAP HANA appliance vendors offer multiple layers of redundant hardware, software and network components, such as redundant power supplies and fans, enterprise grade error-correcting memories, fully redundant network switches and routers and uninterrupted power supply (UPS). Disk storage systems use batteries to guarantee writing even in the presence of power failure and use striping and mirroring to provide redundancy for automatic recovery from disk failures. Generally speaking, all these redundancy solutions are transparent to SAP HANA's operation, but they form part of the defense against system outage due to single component failures.

Software – SAP HANA is based on SUSE Linux Enterprise 11 for SAP and includes security pre-configurations (for example, minimal network services). Additionally, it has system software also includes a watchdog function, which automatically restarts configured services (index server, name server, and so on) in case of detected stoppage (killed or crashed).

Persistence – It persists transaction logs, savepoints and snapshots to support system restart and recovery from host failures, with minimal delay and without loss of data.

Standby and Failover – Separate, dedicated standby hosts are used for failover, in case of failure of the primary, active hosts. This improves the availability by significantly reducing the recovery time from an outage. As an in-memory database, SAP HANA is not only concerned with maintaining the reliability of its data in the event of failures, but also with resuming operations with most of that data loaded back in memory as quickly as possible.

SAP HANA supports the following recovery measures from failures:

Disaster recovery support:

Backups: Periodic saving of database copies in safe place.

Storage replication: Continuous replication (mirroring) between primary storage and backup storage over a network.

System replication: Continuous update of secondary systems by primary system, including in-memory table loading [7].

Fault recovery support:

Service auto-restart: Automatic restart of stopped services on host (watchdog).

Host auto-failover: Automatic failover from crashed host to standby host in the same system.

VII. ADVANTAGES

- Greater computation power faster than disk [1]
- Faster aggregation with column store.
- Highly dependent on actual data used.
- Analysis of large data sets Complex computations.
- Flexible modeling, no data duplication.
- Fast data loads.

VIII. DATA REPLICATION TECHNOLOGIES

In-memory reporting and analyzing of business data requires the replication of the data from a source system to the SAP HANA database.

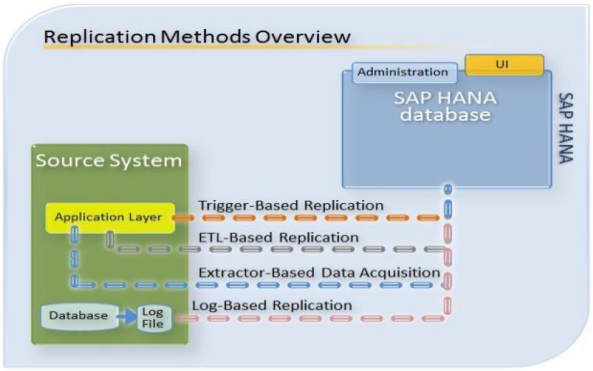

Figure 3: Replication Methods Overview

The figure-3 gives an overview of the alternative methods for data replication from a source system to the SAP HANA database. Each method handles the required data replication differently and consequently each method has different strengths [7]. Trigger-Based Replication: Trigger-Based Data Replication Using SAP Landscape Transformation (LT) Replication Server is based on capturing database changes at a high level of abstraction in the source ERP system.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

This method of replication benefits from being database-independent and can also parallelize database changes on multiple tables or by segmenting large table changes.

ETL-Based Replication : Extraction-Transformation-Load (ETL) Based Data Replication uses SAP Data Services to specify and load the relevant business data in defined periods of time from an ERP system into the SAP HANA database. It can reuse the ERP application logic by reading extractors or utilizing SAP function modules. In addition, the ETL-based method offers options for the integration of third-party data providers.

Extractor-Based Data Acquisition: The SAP HANA Direct Extractor Connection (DXC) is a means for providing out-of-the-box foundational data models to SAP HANA, which are based on SAP Business Suite entities. DXC is also a data acquisition method. The rationale for DXC is essentially simple, low TCO data acquisition for SAP HANA leveraging existing delivered data models.

Log-Based Replication (for non-SAP systems) :SAP Sybase Replication Server (SRS) is a sophisticated transactional data movement product that moves and synchronizes data across the enterprise without the limitation of geographical distance, to meet demanding requirements in the enterprise such as guaranteed data delivery, real-time business intelligence and zero operational downtime. SRS facilitates this by non-intrusively handling data at the source and target, while ensuring high performance and transactional integrity.

IX. SECURITY CONCERNS IN HANA

Data Storage Security:

For maximum performance, the SAP HANA database holds the bulk of its data in memory. However, it still uses persistent storage to support system restart and recovery with minimal delay and without loss of data in the event of failure. For example, after a power failure, the database can be restarted like any disk-based database and returns to its most recent consistent state. In addition, SAP HANA provides functions for backup and recovery, as well as high availability (disaster recovery and fault recovery).

Server Side - Data at rest**:** To protect data saved to disk from unauthorized access at operating system level, the SAP HANA database supports data encryption in the persistence layer. This is referred to as *data volume encryption* [7].

Passwords: All operating system user and database user passwords are stored securely on the SAP HANA database server. In addition, credentials required by SAP HANA applications for outbound connections are stored securely in an internal credential store, which in turn is secured using an internal data encryption service. It uses the secure store in file system (SSFS) to store all internal SAP HANA encryption keys. These keys are encrypted with the SSFS master key.

Client side - The SAP HANA user store (hdbuserstore) can be used to store user logon information to allow client applications to connect to SAP HANA without having to enter a user's password explicitly. The Eclipse secure storage can be used to store user passwords in the SAP HANA studio.

X. CONCLUSION

SAP HANA can make full use of the capabilities of current hardware to increase application performance, reduce cost of ownership and enable new scenarios and applications that were not previously possible. With SAP HANA, it is possible to build applications that integrate the business control logic and the database layer with unprecedented performance. The more we do directly on the data in memory next to the CPUs, the better the application will perform.

XI. ACKNOWLEDGMENT

I am thankful to department of information science, BMS College of engineering, Bangalore. I deeply express my sincere gratitude to my teacher Dr. Radhika K R, Professor, Department of Information science and Engineering for her constant encouragement, guidance and support since beginning. I would like to thank SAP Labs India Pvt. Ltd. Bangalore for providing me an opportunity to carry out my project work. Last but not the least I would like to extend my gratitude to my teachers and friends who indirectly supported me for the success of this paper.

REFERENCES

[1] http://www.decisionfirst.com/uploads/SAP%20HANA/SAP_HANA_Technical_Overview.pdf
[2] http://help.sap.com/hana/SAP_HANA_Modeling_Guide_en.pdf
[3] http://docs.media.bitpipe.com/io_10x/io_109274/item_671528/VCE_sSAP_IO%23109274_E-Guide_042313.pdf
[4] http://help.sap.com/hana_one/SAP_HANA_Developer_Guide_en.pdf
[5] http://www.sap.com/pc/tech/cloud/software/business-management-bydesign/overview/index.html
[6] http://www.news-sap.com/indian-cricketing-franchisee-teams-sap-hana/
[7] http://help.sap.com/hana/SAP_HANA_Master_Guide_en.pdf
[8] http://help.sap.com/hana/SAP_HANA_Security_Guide_en.pdf

Srikantarao S, is a PG Scholar in Computer Networks and Engineering at B.M.S College of Engineering, Bangalore. His research areas are in-memory database and ERP solutions.

Dr. Radhika K R, is a professor at department of Information Science Engineering, BMS College of Engineering, Bangalore. She is Sr. Member at IEEE.