SAP ERP with Big Data and Cloud Computing

Rishi Reddy Cheruku
Department of Computer Science, University Of Akron
rc81@zips.uakron.edu

Abstract - *Enterprise Resource Planning (ERP) systems are currently involved in every part of organization as they provide a suite of integrated applications to meet the organization needs. SAP is evolving it's in-memory technology with introduction of SAP High- Performance Analytic Appliance (SAP HANA) software, which is the core of SAP's new data management platform. The goal of SAP HANA database is to provide a powerful system for both transactional and analytical query scenarios, on the same data representation within a highly scalable execution environment. Within this paper, we start by giving an overview of distributed query processing in SAP HANA and it is features, then we give an brief overview on cloud computing and its advantages of implementing SAP ERP on cloud and we continue further by looking into real time analytics that the Big Data provide to the ERP systems. In summary , the paper aims at illustrating how the SAP ERP is evolving with Big Data and Cloud Computing to expand their market share.*

I. Introduction

In today's challenging business environment, best-run companies need to compete against more efficient competitors. In order to stay ahead they need to get the real time view of business process, by using many kinds of software which might not interact with each other. Customization of these software's may be difficult in few cases which will reduce the optimized functioning of companies business activities. The Company will be facing many problems in many areas of its functioning like Invoices regarding material purchases, salaries. All these things change when a ERP (Enterprise Resource planning) system is implemented. ERP is comprehensive suite of commercially available integrated modules which provide end-to-end support for statewide administrative. ERP allows you by controlling all parts of your business in one software and gives you the real-time view of its core business process such as production, order processing, and inventory management [1]. Though there are many ERP vendors available in the market SAP is leading enterprise application software provider.

SAP Enterprise resource planning (SAP ERP, formerly SAP R/3) application is an integrated backend application with tens of thousands of installations worldwide designed for tracking and managing business processes in midsize and large enterprises [2]. The application is built on software integration platform that gives the application to control concurrency between database server and application server. Data management for enterprise applications have changed significantly over time, we no longer distinguish between transactional and analytical access patterns. In a nutshell applications demand a consistent and detailed view of its business processes ready for querying and analytics. SAP has introduced SAP High-Performance Analytic Appliance (SAP HANA), a platform which has the business logic and data base in an in-memory database architecture. The SAP HANA comprises replication and data transformation services to easily move SAP and non-SAP data into HANA system. Everyday huge amount of transactions take place around the world, this data is pretty much important from company point of view. In order to capture the entire data store it and process it we call this as Big Data. Big Data refers to the challenges of capturing, storing, managing, and analyzing large volumes of various types of data with great velocity. Most of the systems work for a specific domain and provide ubiquitous access

Thus, this paper provides an overview on SAP ERP integration with Big Data to improve the real time data analysis and will the scalability improve by implementing it on Cloud. Additionally , SAP HANA in-memory capabilities would benefit SAP ERP application.

II. Literature Review

According to several Research papers, we can divide it into three parts : SAP HANA architecture and implementing to business applications, advantages of moving business applications on cloud and Integrating

Business application with Big Data analytics. First, we discuss on SAP HANA architecture and integrating with business applications, Analytics by integrating SAP ERP with Hadoop [5]. SAP HANA database is positioned as the core of the SAP HANA Appliance to support complex business analytical processes in combination with transaction-ally consistent operational workloads [6]. Emphasizing the distinctive features that differentiate the SAP HANA database from other classical. From a more application-oriented perspective, we outline the specific support provided by the SAP HANA database of multiple domain-specific languages with a built-in set of natively implemented business functions.

Next, we focus on Implementing business application on Cloud. Cloud storage mode helps enterprises transform existing resources to deal with large-scale data storage in high-growth business and high-speed data processing services [4]. With the prevalence of cloud computing, more and more modern services are deployed in cloud infrastructures to provide rich functionalities [3]. We focus on the challenges and advantages faced by moving the business application on cloud.

Next we discuss on Big Data analytics were we analyze the feasibility and advantages of storing mass data and data storage flexible methods based on hadoop. The emerging large-scale service-oriented systems often involve a large number of services with complex structures. The big data generated from these systems are typically heterogeneous, of multiple data types, and highly dynamic [5]. Due to the fast increase of system size and the associated massive volume of service-generated data, creating value in the presence of massive system and data becomes an inevitable challenge. We discuss how the real time analytics improve by adding the Big Data to the business applications. Adding the cloud based service to Big Data will improve the ERP systems.

III. Problem Formulation

SAP is the world's largest business software company which introduced SAP HANA which is an in-memory framework. SAP HANA comprises of a multi-engine query processing environment. We start by giving an overview of distributed query processing in SAP HANA and it is features. Then we give an brief overview on cloud computing and its advantages of implementing SAP ERP on cloud and we continue further by looking into real time analytics that the Big Data provide to the ERP systems.

IV. SAP ERP Architecture

The SAP ERP mainly consist of three layers. The Presentation layer which deals with how the business objects are displayed to users of the software. This contains the components that implement and display the user interface and manage user interaction. The application server has the ERP suite and servers. The database server has the database physically located. But Unlike traditional database the application server is built into SAP HANA itself. This is application server is called SAP HANA Extended Applications Services (SAP HANA XS) [12]. This increases the performance as most of the calculation is inside the database, which requires very less processing.

The programming model when working directly with XS dictates that applications are designed in the following manner: front-end processing should generally be delegated to the browser (or mobile device) using HTML5 and Client-Side JavaScript. By far the vast majority of application logic, business logic, calculations, or any data-intensive operations should be implemented using SQL, SqlScript, or Calculation Engine (CE) functions [12]. The Database used in SAP HANA is an in-memory database which mean all the write , read operations take place in memory which will optimize the software of database engine which increases the memory access. The SAP HANA database can also store data in row format and in a column format. When the data is stored in column format the data compress much better and can store more data in the database.

V. Distributed Query Processing In SAP HANA

The SAP HANA database is the core component of SAP's HANA roadmap playing the foundation to efficiently support all SAP and non-SAP business processes from a data management perspective [7]. The SAP HANA database takes a different approach to provide support for a wide range of data management tasks. Data is organized along its life cycle either in column or row format, providing the best performance for different workload characteristics. Update rate and point queries are routed against a row store; analytical workloads with range scans over large datasets are supported by column oriented data

structures. In addition to a high scan performance over columns, the column-oriented representation offers an extremely high potential for compression making it possible to store even large datasets within the main memory of the database server [6]. The goal of the SAP HANA database approach consists in scaling over a reasonably large number of nodes without sacrificing overall system performance and all well known transactional guarantees, i.e., Atomicity, Consistency, Isolation, Durability (ACID) properties [8]. The SAP HANA database is designed for scalability in three different directions:

- Scale Up: Due to main memory requirements SAP HANA was designed to run on "big machines" offering multiple CPUs and a fairly large number of threads [6].
- Scale Out: The SAP HANA database runs in a multi-node environment to balance the need of CPU power and main memory capacity providing the same level of transactional guarantees like in a single node scenario [6].
- Scale In: Scale in typically denotes multi-tenancy support and therefore ability to host multiple logical databases within a single physical instance offering a certain level of schema and data sharing [6].

The core database challenges can be classified as Deployment Schemes and Data Distribution, Distributed Transaction Management and Distributed Metadata Management.

A. Deployment Schemes and Data Distribution

Every single component of a database system has to be "distribution enabled", i.e., not only working correctly but also efficiently in a distributed environment. From that perspective, the fact of distribution affects functional as well as non-functional service primitives ranging from distributed (multi-node) query processing to caching strategies of metadata repositories [6]. A single table can be split into multiple partitions using hash, round-robin, range partitioning strategies. Individual partitions are then allocated at different nodes pursuing two different strategies. For example, a single landscape may consist of one very large machine and multiple smaller nodes as shown in Figure 1 [6]. The large machine node will then host all "transactionally hot" tables or partition of tables avoiding distributed transactions with network traffic and protocol delay. More analytically oriented applications targeting multiple partitions of historical data or databases coming from external data sources will then hit the parallel nodes to improve query performance. Since all datasets are part of one single SAP HANA landscape, the database system is able to run cross-joins within multi-node transactions, if the query demands it—the allocation and deployment scheme just tries to reduce the communication within the cluster.

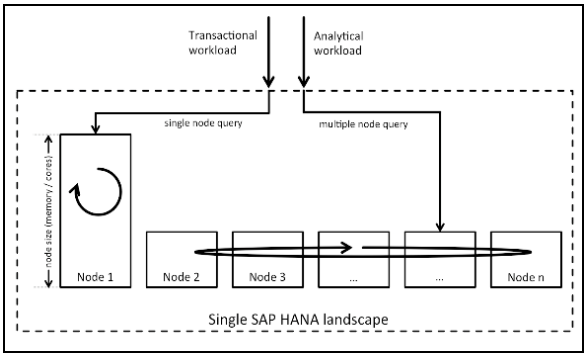

Figure 1. Asymmetric deployment of an SAP HANA landscape

B. Distributed Transaction Management

In opposite to scale-out solutions like Hadoop, SAP HANA follows the traditional semantics of providing full ACID support. In order to make good the promise of supporting both Online transaction processing (OLTP) and online analytical processing (OLAP) style query processing within a single platform, the SAP HANA database relaxes neither any consistency constraints nor any degree of atomicity or durability. The deployment of the system usually reflects the intended use in order to have a benefit of a large node for heavy transaction processing and a number of usually smaller nodes for analytical workloads where the additional overhead of distributed synchronization reflects a relatively small portion of the overall query runtime [6].

C. Distributed Metadata Management

Within an SAP HANA database landscape, a coordinator node stores and manages all the persistent metadata such as table/view schema, user information, privileges on DB objects, etc. To satisfy requirements for consistent metadata access, the metadata object container provides both Mulitversion Concurrency control (MVCC) based access and transactional update (ACID) on its contents. It also provides index-based fast object lookup. In order to illustrate the key concepts of distributed query processing within SAP HANA, we will follow a query in different scenarios. Within a single node setup, the client connects to a

particular server and starts the query compilation process [6].

VI. SAP ERP On Cloud

To meet ever changing business needs organizations need to invest time and budget to scale up there IT infrastructure such hardware , software and services. However with on premises IT infrastructure the scaling process can be slow. so, the organization are frequently unable to achieve optimal utilization of IT infrastructure. Cloud Computing is paradigm shift that provides computing over internet. It consists of highly optimized computing data centers that consists of various Hardware, software and Information resources to use when needed. Organizations can simple connect to cloud and use the available resources as pay as peruse bases which reduces the organizations capital expenditure and additional on premises Infrastructure resources and instantly scale up or scale down according to the business requirements.

In today's challenging world SAP ERP has been providing many solutions to optimize the core business process that every business need to do. Cloud Service is a hardware or software provided to you as a service and accessible through internet. There three main cloud services that we can distinguish. The first one is the infrastructure-as-a-service model, where you get things like virtual machines, servers, you get storage. Then we have software as a service, where you get tools like e-mail, like a Customer Relationship Management (CRM) system, or like virtual desktops, online games. And the third one is the platform-as-a-service model. This model actually looks to provide services for developers. It provides things like execution runtimes, like databases, like development tools, Web service. This is where typical Cloud SAP ERP system or ERP in the Cloud system resides [3].

Cloud uses a shared data centre and pick up all ERP you want and only pay what you want. The cloud is safer as it as multi level data security, regular automatic battery backups and system administrator upgrades, maintenance, scalable, reliable, flexible, cost reducing. The challenges of implementing ERP on cloud are :

- In organization behind the software security, enterprises have control of their data. In cloud the organization must trust the provider. so many organizations do not want to entrust their sensitive data and their reputation to the public cloud.
- There is no well defined service level agreement by the cloud providers, like the uptime for the servers.
- We have latency and network limits as the same network speed might not be available at every place.

The main advantages of implementing SAP ERP on the cloud are :

- We can eliminate our capital expense because we don't have to purchase any hardware.
- The provider takes care of maintaining the software and keeping it up to date which even reduces the operating costs.
- These are available everyday and the provider takes care of the systems.
- Cloud services also provides flexibility, so we can either buy more or less of these services, and scale the systems up and as we need it. We only need a Internet connection and computer to access it.

Cloud computing reduces the organization costs and would rapidly scale up or down the resources as needed on pay-as-you-use model. Deploying ERP on the cloud would allow the user with an internet to use hardware and software on demand.

VII. SAP ERP With Big Data

We are getting deluge by the amount of the data we are generating every day. Every time we make a purchase or every time the light switch goes on and off, some amount of data is getting generated. According to the statistics of the World Bank, the output of modern service industry takes more than 60 percent of the world output, while the percentage in developed countries exceeds 70% [9]. Data analytics has proven its potential in providing decision support in financial, administrative, and scientific sectors by enabling complex computations to generate knowledge, insights, and experimental proofs for scientific discovery. However, the amount of data that needs to be analyzed is growing at an exponential rate and the experts on data analytics technology such as data mining and machine learning often do not have the required domain knowledge to understand the data that needs to be analyzed [3].

Big Data has emerged as a widely recognized trend, attracting attentions from government, industry and academia. Big data are high volume, high velocity, and/or high variety information assets that require new

forms of processing to enable enhanced decision making, insight discovery and process optimization [9]. high Volume posses a great challenge and opportunity as big data will allow organizations to understand people and allocate resources effectively. The Velocity of the amount data flowing into the database at tremendous rate also needs to be monitored regularly and people want to access the data real time. The Variety of the data types that should be processed is becoming increasingly diverse. Today various kinds of data like photos, audio, simulations need to be stored into the database in efficient way. The overwhelming service-generated data become too large and complex to be effectively processed by traditional approaches.

Hadoop is an Apache open source project which consists of two primarily parts: Hadoop distributed file system (HDFS) and Map Reduce programming model. HDFS is an open source version of the Google GFS implementation, as a highly fault-tolerant distributed file system, which provides high throughput data access, suitable for mass storage (PB-class) of large files (usually more than64M) [5]. Hadoop, which is an efficient distributed file system and not a database, is designed specifically for information that comes in many forms, such as server log files or personal productivity documents. Anything that can be stored as a file can be placed in a Hadoop repository [10]. It distributes storing and processing of large datasets across groups or clusters of server computers. It also detects and compensates hardware failure at the application level. This allows high level of service continuity to be delivered by clusters of individual computers each of which may be prone to failure.

Integrating SAP HANA with the Hadoop is like turning infinite storage into Instant insights. SAP HANA accelerate Big Data processing across flexible data management options to support data processing immediately. It acquires data from a variety of sources and combine structured, unstructured, machine, and human data for comprehensive insights. Big Data with analytics solutions discover new insights across the enterprise and act quickly using applications that leverage Big Data insights to ignite new revenue streams and improve operations [11].

To enable Big data Analytics Software-as-a-Service, the open source Apache Hadoop software framework is widely employed by leading companies (e.g., Yahoo!, Amazon.com, Apple, eBay, IBM, Facebook, LinkedIn, Microsoft, SAP, etc.) [9]. To enhance the process efficiency, Hadoop breaks down a task into smaller subtasks, executes the subtasks simultaneously on different computers, and finally reassembles the results. Using SAP HANA with Hadoop would turn infinite storage and get the real time analytics for the organization to better understand their customers.

VIII. Conclusion

The SAP with its latest SAP HANA in-memory data platform is primarily designed to provide scalability. It would provide direct access to operational data without affecting the performance of SAP ERP. With its in-memory capabilities SAP HANA can improve the performance and provide a generic but powerful system for different query scenarios, both transactional and analytical, on the same data representation within a highly scalable execution environment. By moving the Entire ERP on the cloud would really improves the scaling and reduces the Hardware, software to use by Organization. The organization just need to pay as they use which further reduces the cost and Integrating SAP HANA with Hadoop can easily handle huge amount of data volume which is reliable, scalable and fault tolerant and can process data in real time.

This paper provide an overview how SAP would perform on Cloud and Bigdata. Thus Implementing SAP ERP with combination of Hadoop and SAP HANA in cloud would help make critical decisions with an organization.

IX. References

[1]. Yan Xu; Rahmati, N.; Lee, Vincent C S, "A review of literature on Enterprise Resource Planning systems," *Service Systems and Service Management, 2008 International Conference on* , vol., no., pp.1,6, June 30 2008-July 2 2008

[2]. Jerry Rolia, Giuliano Casale, Diwakar Krishnamurthy, Stephen Dawson, and Stephan Kraft. 2009. Predictive modelling of SAP ERP applications: challenges and solutions. In *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools* (VALUETOOLS '09). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, , Article 9 , 9 pages.

[3]. Zulkernine, F.; Martin, P.; Ying Zou; Bauer, M.; Gwadry-Sridhar, F.; Aboulnaga, A., "Towards Cloud-Based Analytics-as-a-Service (CLAaaS) for Big Data Analytics in the Cloud," *Big Data (BigData Congress), 2013 IEEE*

International Congress on , vol., no., pp.62,69, June 27 2013-July 2 2013

[4]. Hofmann, P.; Woods, D., "Cloud Computing: The Limits of Public Clouds for Business Applications," *Internet Computing, IEEE* , vol.14, no.6, pp.90,93, Nov.-Dec. 2010

[5]. Da-Wei Zhang; Fu-Quan Sun; Xu Cheng; Chao Liu, "Research on hadoop-based enterprise file cloud storage system," *Awareness Science and Technology (iCAST), 2011 3rd International Conference on* , vol., no., pp.434,437, 27-30 Sept. 2011

[6]. Juchang Lee; Yong Sik Kwon; Farber, F.; Muehle, M.; Chulwon Lee; Bensberg, C.; Joo Yeon Lee; Lee, A.H.; Lehner, W., "SAP HANA distributed in-memory database system: Transaction, session, and metadata management," *Data Engineering (ICDE), 2013 IEEE 29th International Conference on* , vol., no., pp.1165,1173, 8-12 April 2013

[7]. 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*, vol. 40, no. 4, pp. 45–51, 2011.

[8]. Lars Frank and A. V. Senthil Kumar. 2012. Architecture for mobile control functions in supplier deliveries for distributed integrated ERP modules. In *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication* (ICUIMC '12). ACM, New York, NY, USA, , Article 84 , 8 pages.

[9]. Zibin Zheng; Jieming Zhu; Lyu, M.R., "Service-Generated Big Data and Big Data-as-a-Service: An Overview," *Big Data (BigData Congress), 2013 IEEE International Congress on* , vol., no., pp.403,410, June 27 2013-July 2 2013

[10]. "SAP Hana Analytics with Bigdata", http://www.sap.com/pc/tech/in-memory-computing-hana/software/analytics/big-data.html (Accessed: 6 December 2013)

[11]. "SAP Bigdata", http://www.sapbigdata.com/ (Accessed: 25 November 2013)

[12]. "SAP HANA Architecture", http://www.saphana.com/ (Accessed: 1 December 2013)