International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V71I1P107

Volume 71 Issue 1, 45-49, January 2023

Original Article

Enterprise Reporting on SAP S/4HANA using Snowflake as Cloud Datawarehouse

Baljit Singh

SAP Analytics Architect, IBM Corporation, 600 Anton Blvd, Costa Mesa, CA, USA

Received: 09 December 2022 Revised: 12 January 2023 Accepted: 22 January 2023 Published: 31 January 2023

Abstract *- Today, if a claim is made that Data is a much more valuable resource than oil or gold, all professionals working in various technology fields will agree to it. Data is the most critical tool to enable business enterprises to take quick decisions. All successful companies are using ERP software to integrate business processes across their enterprise to run their business operations seamlessly. SAP is the market leader in ERP software, and most companies are using SAP as enterprise software. SAP S/4 HANA is the flagship product of SAP in the ERP market space, and it stores key enterprise data in its in-memory database called SAP HANA. SAP S/4 HANA becomes a key and pivotal source of information for enterprises as it stores tightly integrated data across all business functions. Analytical reporting on S4 HANA helps organizations make key decisions quickly. Enterprises must choose a strong tool to implement analytics to cater to their analytical needs quickly and cost-effectively. SAP provides various enterprise analytics software, but they require specific technical skills and hardware and software costs are involved in using these solutions. There are various cloud based on demand analytics solutions available that are very cost-effective and require lesser technical skills than SAP analytics solutions. This paper aims to evaluate snowflake, a cloud-based reporting solution that can store and analyze data within a Snowflake. Snowflake's three-tier architecture, data modeling approach for SAP S/4 HANA enterprise data in snowflake, data loading in Snowflake from S/4HANA, and comparison of using snowflake in place of SAP Analytics solutions have been done in detail.*

Keywords *- SAP S4 HANA, Business Intelligence, Enterprise Reporting, Analytics Cloud, Snowflake Data cloud, Data warehousing.*

1. Introduction

With the emergence of cloud computing analytics in the last 7 years, the analytics space has been revolutionized with the advent of new generation cloud computing solutions. Many cloud computing analytics solutions are available in the market, e.g. RedShift by amazon web services, Synapse analytics by Microsoft Azure, and Big Query by Google cloud. Snowflake is also one such solution that provides cloud-based data storage and service termed "Data as a Service". Snowflake company was founded in 2012 and publicly launched listed in 2014. It has been running on Amazon S3 since 2014, Microsoft Azure since 2018, and Google Cloud Platform since 2019. Data is loaded in snowflake using ETL tools, modeled in snowflake, and integrated with all widely used visualization tools like Power BI and Tableau.

2. Literature Review

Cloud computing is the on-demand availability of resources (hardware and software), especially data storage and computing power which can be increased and decreased on demand without any significant investment in managing resources. There are many cloud providers available in the market. The three biggest cloud providers in the market are amazon web services, Microsoft Azure and Google cloud platform. They provide both infrastructure as a service and software as services based on different product offerings. On the other hand, the data warehouse is also known as an enterprise data warehouse which is used for enterprise-wise data reporting and analysis. With the advent of cloud computing, there are many data warehouse services provided by cloud providers. Snowflake is one such product offering that can have storage based on amazon web services, Microsoft Azure and Google cloud platform. With so many data warehouse solutions available in the market, both on-premise and on the cloud, it becomes imperative for organizations to evaluate which is the best available solution to adopt. With that in mind, organizations evaluate these data warehousing options on key parameters such as cost to acquire, cost to maintain, skills required to implement and sustain the solution, data warehouse performance and ease of implementation. Well-informed decisions in choosing the best data warehouse will enable organizations to become intelligent enterprises.

3. Snowflake: Enterprise Reporting on SAP S/4 HANA

3.1. Architecture of Snowflake with SAP S4 HANA

Snowflake has a three-layered architecture. Below are the three layers of Snowflake architecture.

3.1.1. Storage

Data is stored not in the snowflake but in the storage layer of the snowflake. Organizations have the option to choose the storage layer of AWS, Microsoft Azure, or Google cloud platform. Data is stored in hybrid columnar storage in the form of blobs which is different from the traditional data warehousing approach where data is stored in the form of rows. Hybrid columnar storage is used in most of the big data solutions, and this hybrid columnar approach is more efficient and provides better performance while querying data from Snowflake storage.

3.1.2. Query Processing

This is the second layer of Snowflake architecture, where data computing is accomplished in snowflake. Virtual warehouses are created in this layer where virtual compute resources are used for query processing. Virtual warehouse shall not be confused with traditional Datawarehouse. These virtual warehouses are virtual servers that can perform massively parallel processing when the amount of data is very high and complex calculations are required to be computed on large data volumes. Organizations have the flexibility to choose the computing power of resources and the number of servers required to scale out. This layer works as the muscle of the system. Virtual warehouse comes in different sizes, the bigger the size more computing resources are available, and more will be the cost. The below small table depicts the number of servers available in different sizes.

Table 1. Virtual Warehouse Sizes

Virtual Warehouse Size	Number of Servers
XS	1
S	2
M	4
L	8
XL	16
2XL	32
3XL	64
4XL	128

3.1.3. Cloud Services

This layer works as the brain of the system where we manage the infrastructure, manage the security, optimize the queries, manage metadata, etc. The below figure explains the architecture used with SAP S/4 HANA for enterprise reporting. Data can be extracted from SAP using any ETL tool and loaded into the Snowflake storage layer. Snowflake has good partnerships with various companies of ETL, Business Intelligence, and Data Integration tools.

Snowflake has partnerships with many ETL tools like Informatica, Qlik, Snap Logic, Talend, Matillion, etc. After loading data into a snowflake, it can be modeled in a snowflake for enterprise reporting. Once the data has been modeled in snowflake, it can be used by reporting and visual tools such as Powerbi and Tableau. Snowflake also integrates with leading machine learning and data science platforms like Alteryx, H2O, Hex, Zepl, etc. as well.

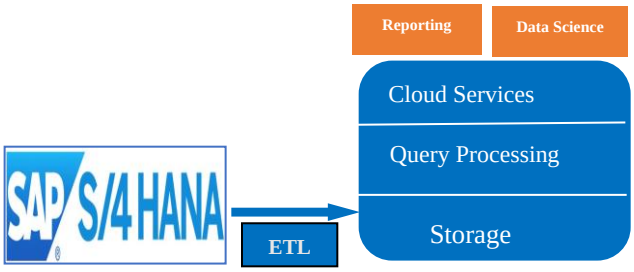

Fig. 1 Snowflake Architecture with SAP S/4HANA

3.2. Data Loading in Snowflake

SAP S4HANA stores data from all business functions such as finance, sales, purchasing, inventory, and other functions based on SAP functionality implemented in columnar tables in the HANA database. Data can be loaded into Snowflake storage using ETLs like HVR or any other ETL chosen by the organization. One of the key concepts in snowflake for loading data into tables and databases in snowflake is the stage. A stage is a database object containing the location where data can be loaded. The stage has some properties like location, URL, credentials etc., of data. There are two types of stages external and internal. The external stage is the more common type which can be blob storage provided by cloud providers Amazon, Microsoft, or Google.

3.2.1. Bulk Loading

Bulk loading uses using compute power of a virtual warehouse. This is a more frequently used method. At the same time, loading data into Snowflake transformation is possible while loading data in snowflake. Data is loaded into tables using copy command.

3.2.2. Continuous Loading

This method is not to handle large volumes of data but small volumes of data. Data is automatically loaded from stage to snowflake if there is any change in data. This method provides the feature to provide the latest data for analysis at the earliest. Snowflake provides a tool called Snow Pipe, a serverless feature to load continuous data from the stage. Snowpipe works on event notifications which are configured in our cloud provider for storage. The event is triggered in storage and triggers a copy command to load data into tables in snowflake continuously.

3.3. Modelling of SAP Data in Snowflake

Once loaded in storage, data from SAP must be ingested in snowflake and modelled. It is recommended to follow the layered data approach. Below are the four layers which are recommended to model SAP Data in Snowflake.

3.3.1. Physical Tables

This is the first layer for SAP data in snowflake. In this layer are physical tables, e.g., the MARA table for material master. This layer will contain the Raw data exactly as it exists in SAP. This layer contains only the technical details, such as metadata and technical names of fields that business users cannot understand.

3.3.2. Harmonization Layer

Tables created in the first layer have technical names for the fields. In the harmonization layer, business context is added to these tables by creating views on these tables and giving English descriptions to the fields. The intention is to make these views understandable to business users and data analysts. There are many resources which are available which can be leveraged by referring to table definition in SAP S4 Tcode Se11, Hana live, and website https://www.sapdatasheet.org

3.3.3. Propagation Layer

Now we have data with the business context in the harmonization layer. In this layer, data is transformed and enriched according to business requirements for reports to be created by creating views on the harmonization layer. Filters, calculated columns, and other transformations are applied in these views. Business logic for reporting across various functions is applied in this layer.

3.3.4. Data Mart Layer

The data mart layer provides views created over the propagation layer to consume for reporting and visualization. Dimensions and facts are processed to create a dimensional model. Dimensional models are created based on business requirements for reports required for business decision-making.

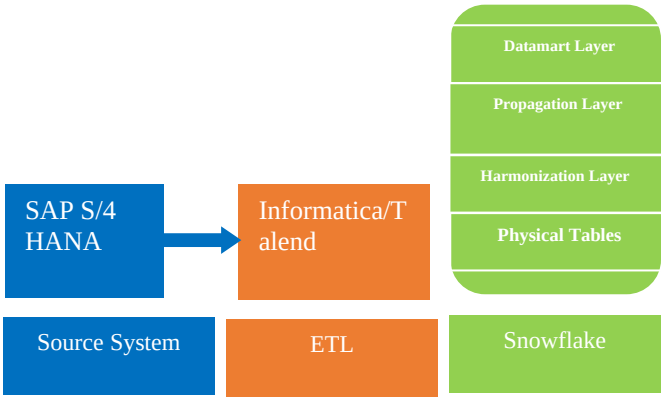

Fig. 2 Data Flow in Snowflake from SAP S/4HANA

3.4 Performance Optimization

With a significant data footprint increase, reports' performance is a key factor in choosing an analytics solution. Snowflake also provides various ways to increase the performance of data models.

3.4.1. Implement Dedicated Warehouse

Snowflake allows the creation of multiple virtual warehouses and assigns them to users. Identifying users who work on large data sets, classifying them, and assigning them a dedicated warehouse is recommended. These warehouses can be sized independently based on data volumes that users are working on. Identifying users and assigning dedicated resources can also be optimized for a minimum time when dedicated warehouses are assigned and scaled up/out. For example, Marketing and Finance users work on large data volumes and can be assigned dedicated virtual warehouses. This approach can help in managing the resources and costs of using snowflakes.

3.4.2. Scaling up

Snowflake allows changing the size of the warehouse on demand. Identify the trends when users are doing high data computations. Finance users are known to work on high data volumes during period-end close. Snowflake provides flexibility to scale up and scale down as per the organization's demand. Scaling up allows the deployment of more virtual servers and helps speed up massive data processing. As explained above, organizations can identify the pattern of users when they need more computing power. They can temporarily scale up the computing power to cater to business needs and then scale down to avoid the high costs of using more computing power for longer durations.

3.4.3. Scaling Out

Scaling out allows for an increased number of clusters of virtual warehouses on demand. If more users are using the system, the scale-out option can be used to deploy more clusters and give better performance and user experience.

3.4.4. Enable Cache Usage

Caching is an automated process to store the results of queries executed in system memory and gives very high performance if the query is executed multiple times. Data is stored in the cache for 24 hrs. if underlying data is unchanged. When the user executes, the query for the first-time system will take time to compute the results and execute the query. When the same query is executed again for the next 24 hrs., the snowflake will execute these queries very fast and give a delightful user experience. The cache is a very important feature for high performance and is automatically used by the system.

3.4.5. Create Cluster Keys for Large Tables

Cluster keys are subsets of rows to identify data in micro partitions. Cluster keys improve the scan efficiency of the

system on large tables. Cluster keys are automatically created and change over time; these should be customized to improve performance.

3.4.6. Use Materialized Views

Materialized Views are pre-computed sets of data derived from queries and stored for further use. Materialized views give very high performance for query execution but add an additional overhead of storage cost and should be used judiciously. Materialized views are mainly created on data mart views where query results are pre-calculated and stored for quicker query performance.

3.5. Comparison of Snowflake with SAP Analytics Solution

SAP also provides multiple analytics solutions that are part of the SAP products family and are very well integrated with SAP S/4 HANA. Below are the key enterprise reporting solutions provided by SAP.

- SAP BW4HANA
- SAP Enterprise HANA
- SAP Embedded Analytics

Below is the comparison of the above-listed solutions provided by SAP with Snowflake.

3.5.1. Cost

Acquiring a software solution incurs costs related to license and hardware. There are costs associated with getting the license from SAP to use the solution and infrastructure costs for SAP analytics solutions, which can be cloud or on-premise costs. Since snowflake is provided as data as a service, it only incurs solution costs based on the computing power and storage required. Cost can be effectively managed in snowflake as it gives the option to increase/decrease the storage and compute power on demand. In the case of SAP, proper planning is required to identify the resources required as scale-up and scale-down are not on demand and are flexible, as in the case of a snowflake. During the specific time in a month and year, the number of users accessing the reports will be high. This means that system does not need more computing power but more parallel servers working to cater to the high number of users. Organizations can identify the trends and accordingly deploy more virtual warehouses to ensure optimal performance during those periods.

3.5.2. Technical Analytics Skills

For SAP solutions, special skills are required, which can be expensive to acquire. Skills required for SAP solutions are E-HANA Modeling, BW4HANA Modeling, SAP CDS scripting, SQL, FIORI, etc. These are very special and specific skills required for SAP solutions and have very high costs to hire or acquire. For the snowflake, no special skills are required. Only working knowledge of SQL is required to build data models in snowflake, which is relatively easier to find and less expensive to acquire,

3.5.3. Performance

SAP HANA is an in-memory platform where all the SAP solutions provide great performance and a delightful user experience. If data volumes are high in SAP, more costs can be involved to improve infrastructure. Snowflake also provides good performance; the system can be scaled up, multiple clusters can be deployed on demand, and scaled-down when not needed.

3.5.4. Realtime Reporting

With SAP HANA evolving into new technologies, all the SAP analytics solutions listed above provide realtime reporting performance. SAP BW4HANA using ODP can provide realtime performance. SAP CDS views are created directly on S/4 HANA tables and provide realtime performance. SAP Enterprise HANA has S/4 tables replicated in real time using SAP SLT, and views are created on these replicated tables, which provides real time performance. Snowflake can provide near realtime performance but not similar performance as provided by SAP solutions.

3.5.5. IT infrastructure

An infrastructure team, in-house or outsourced, is required to manage the landscape and adds to the cost of maintaining the solution. A very lean team is required to manage the Snowflake system.

3.5.6. Metadata Management

SAP analytics systems manage metadata for data in the S/4HANA system, and there is no overhead to manage metadata because source systems and analytics systems are under the SAP umbrella. Metadata management in snowflake requires additional effort.

3.5.7. SAP Delivered Content

SAP provides extensive standard reports/content that spans multiple business functions and industry sectors. Standard reports are delivered in SAP BW4HANA, termed as business content, which contains the entire data model and flows from SAP S/4HANA to BW4HANA. SAP also provided standard CDS views, which can be copied and modified. SAP HANA Live also provides pre-delivered views like SAP BW business content. This content can be leveraged to create enterprise reports and helps in shortening the content development time. Snowflake does not provide any standard content specific to SAP and will require longer development time.

3.5.8. Solution Upgrades

An organization must plan and apply the upgrades to the SAP solutions. Organizations must plan, apply, and test the upgrades, which require significant effort and cost. The organization is responsible for the application of upgrades to the SAP system. In the case of a snowflake, it is different as upgrades are applied and managed by a snowflake. There is

no overhead on the organization to apply and manage the upgrades.

4. Conclusion

With unlimited computing and storage available on Demand, Snowflake is a cloud computing analytics solution that provides a viable and futuristic solution for reporting on SAP S/4HANA. Substantial cost savings can be achieved with efficient management of Snowflake resources, easy availability of SQL skills, and a lean team for infrastructure management of snowflake. Snowflake also provides a viable source for ML/AI applications as it can store and analyze both structured and unstructured data. For organizations with diverse IT landscapes with numerous SAP and Non-SAP source systems, the snowflake is more suited as Datawarehouse because it provides better integration with Non-SAP systems than SAP Analytics systems.

References

[1] The Snowflake website, 2020. [Online]. Available: https://docs.Snowflake.com/en/?_ga=2.131591436.2086032114.1673908722-1347278072.1673734864

[2] The SAP website, 2022. [Online]. Available: https://help.sap.com/docs/SAP_HANA_PLATFORM/4505d0bdaf4948449b7f7379d24d0f0d/bc18c37df9154cb0bfacad4b045f7c49.html

[3] The SAP website, 2022. [Online]. Available: https://community.sap.com/topics/bw4-hana/hana-2-0

[4] Sikender Mohsienuddin Mohammad, "Cloud Computing in IT and How it's Going to Help United States Specifically," *International Journal of Computer Trends and Technology*, vol. 67, no. 10, 2019. *Crossref*, https://ssrn.com/abstract=3629018

[5] Haleem Khan, and Yu Jiong, "Cloud Computing Effect on Enterprises in Terms of Cost," *International Journal of Computer Trends and Technology*, vol. 67, no. 5, 2019. *Crossref*, https://doi.org/10.14445/22312803/IJCTT-V67I5P103

[6] Laxmikant S. Bhattad, and P. P. Deshmukh, "Secure Data in Cloud Computing Using Encryption Algorithms," *International Journal of P2P Network Trends and Technology*, vol. 5, no. 3, pp. 1-4, 2015.

[7] Yonah Etene, and Josphat M. Karani, "Performance in Layered Software Architectures: the Case of Customized Organizational Software," *International Journal of Computer Trends and Technology*, vol. 67, no. 12, pp. 15-19, 2019. *Crossref*, https://doi.org/10.14445/22312803/IJCTT-V67I12P104

[8] Marcel Chibuzor Amaechi, Matthias Daniel, and Bennett. E. O, "Data Storage Management in Cloud Computing Using Deduplication Technique," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 7, pp. 1-7, 2020. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V7I7P101

[9] G. Anitha et al., "A Survey of Security Issues in IIOT and Fault Identification using Predictive Analysis in Industry 4.0," *International Journal of Engineering Trends and Technology*, vol. 70, no. 12, pp. 99-108, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I12P211

[10] Twinkle Graf.F, and Mrs.Prema. P, "Secure Collaborative Privacy in Cloud Data with Advanced Symmetric Key Block Algorithm," *SSRG International Journal of Computer Science and Engineering*, vol. 2, no. 2, pp. 40-44, 2015. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V2I2P109

[11] Sidath Weerasinghe, and Indika Perera, "Taxonomical Classification and Systematic Review on Microservices," *International Journal of Engineering Trends and Technology*, vol. 70, no. 3, pp. 222-233, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I3P225

[12] Manjunatha. S, and Dr. Suresh. L, "A Study on Consolidation of Data Servers in Virtualized Cloud Atmosphere," *SSRG International Journal of Computer Science and Engineering*, vol. 6, no. 11, pp. 47-50, 2019. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V6I11P110

[13] Dr. P. K. Rai, and Rajesh Kumar Bunkar, "Architectural Data Security in Cloud Computing," *International Journal of Computer & organization Trends*, vol. 4, no. 4, pp. 23-27, 2014. *Crossref*, https://doi.org/10.14445/22492593/IJCOT-V10P306

[14] N. Madhavi Latha, and Y. Pavan Narasimha Rao, "A Novel Secured Data Transmission Model with Load Balancing for Cloud Computing," *International Journal of Computer & organization Trends*, vol. 6, no. 1, pp. 45-50, 2016. *Crossref*, https://doi.org/10.14445/22492593/IJCOT-V29P304

[15] G. Jai Arul Jose, and C. Sajeev, "Implementation of Data Security in Cloud Computing," *International Journal of P2P Network Trends and Technology*, vol. 1, no. 1, pp. 6-11, 2011.

[16] K. Phanindhar, and Ch. Suresh, "Cloud Computing: a Responsibility Data Sharing in the Cloud Computing," *International Journal of P2P Network Trends and Technology*, vol. 3, no. 4, pp. 11-16, 2013.

[17] Dr. D. Shravani, "Model Driven Architecture Based Agile Modelled Layered Security Architecture for Web Services Extended to Cloud, Big Data and IOT," *International Journal of Computer & organization Trends*, vol. 6, no. 4, pp. 55-64, 2016. *Crossref*, https://doi.org/10.14445/22492593/IJCOT-V35P304

[18] Lingaswami, and G. Avinash Reddy, "Offensive Decoy Technology for Cloud Data Attacks," *International Journal of P2P Network Trends and Technology*, vol. 3, no. 6, pp. 23-27, 2013.