Original Article

SAP Datasphere – Business User-Oriented Approach to Cloud Data Warehousing

Baljit Singh

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

Abstract - All businesses today focus on making quick and accurate business decisions. They use data generated by IT systems to analyze and make business decisions. ERP systems generate data from various business functions in an enterprise and provide very valuable data. SAP is one of the market leaders in ERP, and businesses from all sectors have implemented SAP as their ERP system. Currently, S/4 HANA is the ERP system offered by SAP. Datawarehouse has become a very important tool for storing enterprise data and providing valuable insights to businesses. Traditionally data warehouse was offered as an on-premises solution for business analytics, e.g. Teradata and Oracle BI. SAP also offered SAP BW as an on-premise data warehouse.

With the increase in data footprint, the cost and performance of the data warehouse became key pain points. In memory computing and cloud computing were the technologies which addressed these pain points. Various cloud computing vendors such as Amazon, Google, Microsoft, and many more started providing cloud base data warehouse solutions. SAP also developed SAP HANA for in-memory computing and provided SAP BW and SAP Enterprise HANA as Datawarehouse on the cloud but not Datawarehouse as a service. SAP introduced the Datawarehouse cloud in 2019, which was rebranded as SAP Datasphere in 2023 and is a solution by SAP which is offered as Datawarehouse as a service.

This paper evaluates SAP Datasphere as a cloud Datawarehouse that can store and analyze data in the cloud. One big differentiator for SAP Datasphere is that it is a solution that is business user-oriented and enables business experts to make decisions. The positioning of SAP Datasphere as compared to other SAP analytics solutions is also explained in this paper.

Keywords - SAP Datasphere, Datawarehouse, Analytics, Cloud Computing, SAP HANA, Business Reporting.

1. Introduction

Organizations are using analytics for better decision-making. The data warehouse is a key tool that has been used for more than 20 years to cater to business needs. SAP also introduced Datawarehouse in 1998, named BIW, which was named BW in 2001. SAP introduced an in-memory database named SAP HANA to address performance issues due to a high data footprint. SAP then introduced SAP BW on HANA with an in-memory database. In 2016 SAP completely rebuilt its SAP BW with HANA-optimized code and introduced a newer version of SAP called SAP BW4HANA. SAP BW4HANA is now the flagship data warehouse solution which can be deployed on-premise or on the cloud. With the advent of cloud computing, many Datawarehouse solutions on the public cloud are offered by vendors like Amazon, Google, Microsoft, Snowflake, etc. offer various Datawarehouse capabilities and systems maintained by cloud providers. SAP also introduced a new solution called data warehouse cloud in 2019 which was available on the public cloud. SAP rebranded the Datawarehouse cloud as Datasphere in March 2023.

SAP Datasphere is a modern approach to data warehousing with an attempt to unify SAP Transaction Data and other disparate data in an organization's IT ecosystem. This product also provides business users to look at data cohesively and try to find the relations between data which never existed before. SAP Datasphere unleashes the power of an open data ecosystem (Collibra, Confluent, Databricks, Datarobot), allowing integration with all non-SAP systems.

2. Literature Review

The pace of business has recently increased due to various factors like increased processing power, cheaper storage, virtualization of machines, in-memory computing, cloud computing, machine learning, AI, and many more. Business users are asking for systems that can help them with valuable data insights to help them make better and quicker decisions. Focus is very high on the ability to make decisions quickly. Due to these requirements, data warehousing has undergone a seaside change now. New-age data warehouse systems are able to provide reporting features like real-time reporting, analyzing high volumes of data on

the fly, self-service reporting, and intuitive user experience. Cloud computing has also made data warehousing very accessible and cost-effective. Cloud computing has also made use of machine learning and artificial intelligence to get more value from data. SAP is the largest ERP software company had also been providing analytics solutions for the last 20-plus years. SAP offers BW4HANA, Enterprise HANA, Embedded Analytics, and SAP Datasphere as different data warehousing/reporting options. BW4HANA, Enterprise HANA, and Embedded Analytics are IT-oriented analytical solutions requiring technical/coding skills, and organizations shall be completely aware of the pros and cons of using these solutions. SAP Datasphere is a business-oriented system where business users can work with enterprise data fabric (Both SAP and Non-SAP) and can get more insights into data. No coding knowledge is required to build data models in the Datasphere by business users, and it empowers business users. SAP Datasphere is a solution on the public cloud managed by SAP and is a futuristic solution for reporting and analytics.

2. Datasphere Architecture

SAP Datasphere is a cloud data warehouse on the public cloud offered as a data warehouse as a service like services offered by Google, Amazon, and Microsoft. The below figure depicts the architecture of Datasphere.
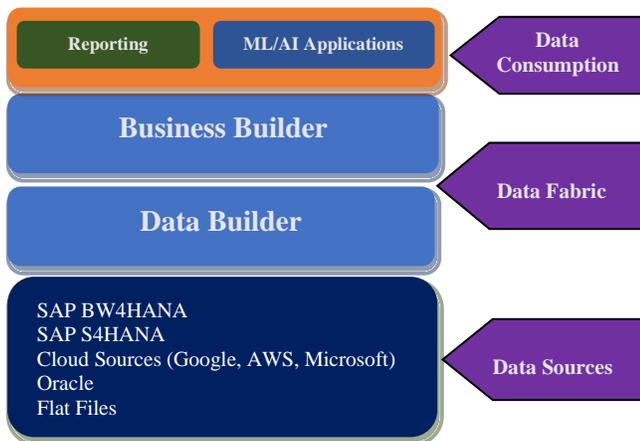

Fig. 1 Datasphere Architecture

The main components of the Datasphere are:
2.1. Space Management

Space is like a mini Datawarehouse in SAP Datasphere, which can be for a specific business function or set of users, which the organization's strategy can define. Administrators can assign the resources like disc space and memory to space. Space is the starting point for all the modeling activities. All the models are created in the space. Data can be corroborated by users and shared across different spaces,

and it avoids redundancy of objects in the system. The key thing to be noted here is that only the administrator can create and manage the spaces created.

2.2. Data Builder

Data builder in SAP datasphere is the data layer where the modeler can do the basic modeling on underlying tables and views from various sources. The objects created in the data builder can be shared across spaces and set for consumption to be used in other applications such as reporting tools, ML applications, etc. All objects modeled in the data builder use drag and drop or SQL. No specific technical skill is required. The key thing to note is that in the data layer IT team or Data Engineers will model the objects using the semantic approach, focusing on data consolidation and data provisioning. The different artifacts types in data builder are as follows:

2.2.1. Tables

These local tables within SAP Datasphere serve as building blocks for other data models.

2.2.2. Views

These are virtual database models created by joining tables in Datasphere. They serve a very important purpose in creating data artifacts to build objects with meaningful data for specific business functions.

2.2.3. ER Models

ER models can be created in the data builder to explain interrelated objects of interest in specific business domains. They provide very key insights to business users when they are creating their objects in data builders.

2.2.4. Task Chain

The task chain is very similar in functionality to the SAP BW process chain. A task chain helps in scheduling multiple tasks in a pre-determined sequence. Task chains can be run on demand and scheduled periodically as well.

2.3. Business Builder

A business builder is a business layer where business users can define objects with a business approach. The focus of business users is to model data to optimize models based on business needs. Data models in business are more stable and less volatile as the business definitions will not change over a period of time, whereas in the data layer, data models are dependent on underlying systems, and their modeling can change over a period of time. Business Builder provides an interface for business users to create objects on top of the data layer. The below figure explains how business users will do the modeling based on data objects and not on technical tables in the underlying system.
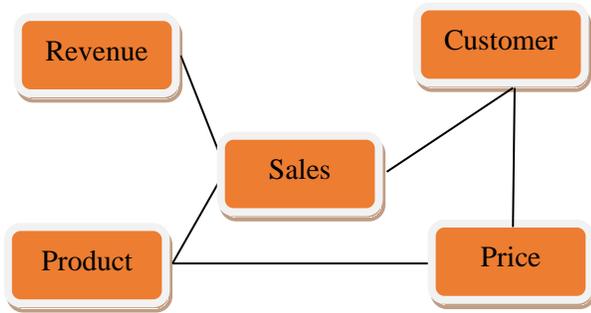
Fig. 2 Modeling in business builder

Below are the key object types in the business builder.

2.3.1. Business Entities

Business entities are based on views created in the data layer. Business entities contain information about business functions which can be measures or attributes. If a business entity contains only attributes, then the entity is of a type dimension. If a business entity contains both measures and attributes, then the entity is a type of analytical dataset.

2.3.2. Fact Models

Business entities are based on views created in the data layer. Business entities contain information about business functions which can be measured and attributes. If a business entity contains only attributes, then the entity is of a type dimension. If a business entity contains both measures and attributes, then the entity is a type of analytical dataset.

2.3.3. Consumption Models

Consumption models help you combine business entities and fact models to create a dataset that provides meaningful information for an identified business domain.

2.3.4. Authorization Scenarios

Authorization Scenario helps users create data access control, providing row-level security to objects in data builder or business builder. One data access control can be assigned to multiple objects. These are like the analysis authorizations we have in BW and BW4HANA.

2.3.5. Perspectives

Perspectives are a subset of a consumption model's measures and attributes, which can be used by visualization and other consumption tools. One consumption model can have multiple perspectives.

2.4. Data Integration and Data Flow

SAP Datasphere can get data from SAP and Non SAP systems which are available both on-premise and on the cloud. It has inbuilt integration capabilities with SAP solutions and all cloud providers like GCP, AWS, MS Azure, etc. All connections in the datasphere will have one of the below features.

2.4.1. Remote Tables

This feature helps in deploying the table as a remote table where data can be accessed directly from the source.

2.4.2. Data Flow

The data flow artifact gives the functionality to build data flows in the dataflow editor. Modeler can add a source object from connection to a data flow to integrate and transform data.

2.4.3. Replication Flow

Replication flow is a new cloud-based replication tool which is designed to simplify data integration processes by eliminating the need to maintain additional on-premise components. Replication flow helps copy multiple data assets from the same source to the same target quickly and easily and does not require any complex projections.

2.4.4. Model Import

The model import feature helps in leveraging the metadata of the models created in the BW4HANA and S4HANA cloud as sources. The main advantage is that the objects need not be recreated manually again in the datasphere.

3. Consuming Data

SAP Datasphere is tightly integrated with SAP analytics for the cloud. Data from SAP SAC can be used for reporting and creating visual dashboards. SAP SAC also has planning capabilities. Data planning can also be done to create strategic and operational plans for an organization in different business functions. In SAP SAC, users can create planning models to acquire and process the plan data. The biggest advantage of planning with SAP Datasphere and SAC is that the entire organization's actual data is available in datasphere, which can be used to create planning models. SAP SAC also comes up with two very powerful features such as smart assistants and smart insights. These are inbuilt machine learning capabilities in SAP SAC, which business users can leverage without knowing the statistics and algorithms required for machine learning. Users can focus more on data insights and not focus their efforts on building ML Models.

SAP Datasphere also integrates with all other mainstream reporting tools, such as Powerbi and Tableau.

4. Value Proposition of SAP Datasphere

SAP Datasphere is a public cloud offering by SAP as a data warehouse as a service like other cloud providers like Google, Amazon, and Microsoft. This solution helps fix the data growth issue in all an enterprise's systems. This product helps customers create a business data fabric that provides consistency of data available for both on-premise and cloud systems. This product helps businesses build their reporting

and analytics solutions by being agnostic of source systems' underlying technology, which was earlier done by implementing complex data warehouses. SAP Datasphere also supports real-time reporting by accessing remote tables and avoiding the need to transfer data to SAP Datasphere. SAP Datasphere can handle large volumes of data and provides high-performing analytics solutions as it is built on the in-memory technology of SAP HANA. Timelines to implement SAP Datasphere are short compared to traditional warehouses because of standard business content provided by SAP across different business functions and industry sectors.

5. Conclusion

SAP Datasphere is a newer and improved version of the SAP Datawarehouse cloud. It is a public cloud offering of the data warehouse as a service, and maintenance of the system is managed by SAP and not by the customer. This solution enables business users to create their own reporting solutions, as no technical skills are required for modelling. This solution is going to be a futuristic solution provided by SAP for laying down the analytics foundation for their customers and helping them build a business data fabric.

References

[1] Sanjeev Kumar, "Data Intelligence and Planning Using AI and Machine Learning with SAP Analytics Cloud – SAC," *International Journal of Computer Trends and Technology*, vol.69, no. 2, pp. 1-4, 2021. [CrossRef] [Publisher Link]

[2] Ravi Kashyap, "Data Sharing, Disaster Management, and Security Capabilities of Snowflake a Cloud Data Warehouse," *International Journal of Computer Trends and Technology*, vol.71, no. 2, pp. 78-86, 2023. [CrossRef] [Publisher Link]

[3] The SAP Website, 2023. [Online]. Available:
https://help.sap.com/docs/sap_datasphere/c8a54ee704e94e15926551293243fd1d/5c1e3d4a49554fcd8fcf199d664d1109.html

[4] The PWC Website, 2023. [Online]. Available: https://www.pwc.com/us/en/services/alliances/library/sap-project-data-suite.html

[5] The Capgemini Website, 2023. [Online]. Available: https://www.capgemini.com/gb-en/insights/expert-perspectives/value-propositions-of-sap-datasphere-bw-bridge/

[6] The SAP Website, 2023. [Online]. Available: https://www.sap.com/products/technology-platform/datasphere.html

[7] 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, pp. 14-19, 2019. [CrossRef] [Publisher Link]

[8] 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. [CrossRef] [Publisher Link]

[9] 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] [Google Scholar] [Publisher Link]

[10] 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] [Publisher Link]

[11] 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] [Publisher Link]

[12] 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] [Publisher Link]

[13] 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] [Google Scholar] [Publisher Link]

[14] 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] [Publisher Link]

[15] 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] [Google Scholar] [Publisher Link]

[16] Dr. E. Kesavulu Reddy, "The Analytics of Clouds and Big Data Computing," *SSRG International Journal of Computer Science and Engineering* , vol. 3, no. 11, pp. 31-35, 2016. [Publisher Link]

[17] 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] [Publisher Link]

[18] 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. [Google Scholar] [Publisher Link]

[19] 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. [Publisher Link]

[20] 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.

[21] 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. [Publisher Link]