IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 9, Issue 1, September 2021

Enhancing Biotech Data Management through SAP: A Comprehensive Review of Data Ingestion in SAP BW/4HANA

Mahesh Babu Munjala
Sr. Business System Architect, CSL

Abstract: *The importance of data-centric decision-making has amplified the role of enterprise data warehousing solutions, leading to an increased demand for the ingestion of data from diverse sources in recent times. SAP BW/4HANA, an on-premises packaged data warehousing solutionfrom SAP offers advanced and flexible data ingestion capabilities to integrate heterogeneous data sources into a unified repository. This paper explores the various data ingestion mechanisms available in SAP BW/4HANA. It reviews the native SAP HANA-based Operational Data Provisioning framework that enhances extraction efficiency. The paper also examines the different source system types supported, including Operational Data Provisioning, SAP HANA, Big Data, and flat files. Additionally, it analyses the batch, real-time and hybrid ingestion patterns facilitated by SAP BW/4HANA. Data provisioning directly with SAP HANA database using tools like SAP Data Services is also discussed. Recommendations are provided for efficient data ingestion based on real-world implementation experience. The insights presented will equip technical teams with comprehensive knowledge regarding ingestion architectures in SAP BW/4HANA aligned to their integration and analytics requirements. By leveraging appropriate ingestion strategies, organizations can optimize the performance of their SAP BW/4HANA implementation and accelerate data-driven business innovation.*

Keywords: SAP BW/4HANA, Data Ingestion, SAP ERP Extraction,SAP ODP Extraction, HANA Data Provisioning

I. INTRODUCTION

The growth in data volume and sources in our digital era underscores the critical role of data ingestion and integration in managing enterprise data. Data warehouses are instrumental in consolidating and structuring organizational data, fostering business intelligence and analytical capabilities. SAP BW/4HANA[1], a leading data warehousing solution, stands out for its extensive and adaptable data integration mechanisms, tailored to ingest data from diverse sources. The paper delves into SAP BW/4HANA's data ingestion frameworks, exploring native SAP HANA-based mechanisms, source system categories, and ingestion modes. The focus extends to SAP HANA native data provisioning approaches, offering recommendations and best practices for optimal data ingestion aligned with real-world implementation experiences.

This paper aims to provide readers with a comprehensive understanding of data ingestion frameworks and strategies within SAP BW/4HANA. The discussion serves as a guide for technical teams and architects in crafting efficient data integration architectures tailored to specific data warehousing goals and business requirements. By optimizing data ingestion capabilities, organizations can fully leverage the accelerated analytics and real-time reporting potential offered by SAP BW/4HANA.

II. DATA ACQUISITION

Data acquisition in SAP BW/4HANA is a crucial process involving the extraction, transformation, and loading of data from various sources into the system. This flexible process can be customized to align with specific data warehousing strategies for different application scenarios. The integration of sources into SAP BW/4HANA is facilitated using source systems, which provide interfaces for extracting and transferring data from numerous sources. A DataSource in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 9, Issue 1, September 2021

SAP BW/4HANA is required to transfer the data effectively. The loading of data into SAP BW/4HANA utilizes the Data Transfer Process (DTP), incorporating transformations and filters. Transformations play a key role in consolidating, cleansing, and integrating the data during this process[2].

To establish integration for data transfer in SAP BW/4HANA, source systems connection is required. Fig -1 shows all the available source systems for ingesting data to BW/4HANA. The following are the broad categories:

- Operational Data Provisioning (ODP) source systems
- SAP HANA source systems
- Big Data source systems
- Flat files

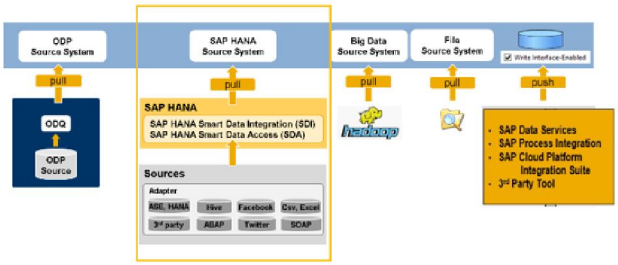

Fig -1: BW/4HANA Source Systems

2.1 Operational Data Provisioning (ODP) Source Systems

For SAP BW/4HANA, there is a native, SAP proprietary and more efficient data loading technology called Operational Data Provisioning (ODP) that replaces many of the data loading techniques used by legacy SAP BW solution.

Operational Data Provisioning provides a unified framework for data provisioning and consumption. ODP supports extraction and replication scenarios for various target applications, with delta mechanisms in place. It acts as the central infrastructure for data extraction and replication from SAP ABAP applications into SAP BW/4HANA. ODP offers features such as timestamp-based recovery mechanism, configurable data retention periods, intelligent parallelization options, and highly efficient compression mechanism with up to 90% compression rates. Compared to traditional methods like Info Packages, using ODP reduces runtime by over 40%, making the Persistent Staging Area obsolete in SAP BW/4HANA[4]. Fig -2 demonstrates the various ODP contexts available for consumption with BW/4HANA.

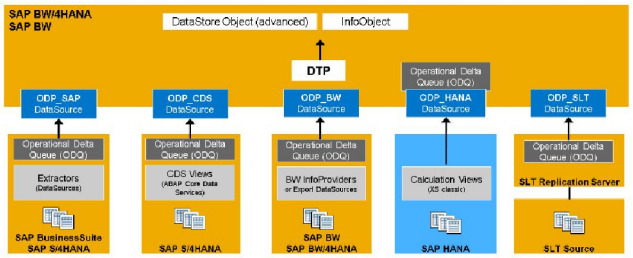

Fig -2: ODP Context [4]

The available ODP source system types for SAP BW/4HANA system as a destination are as follows:

2.1.1 ODP Contexts

SAP BW/4HANA Data Sources can function as Operational Data Providers, which are populated by extractors retrieving data from ABAP systems. The ODP framework supports released Data Sources without requiring additional configuration. These sources are mainly part of the SAP ERP systems and can be accessed by consumers through the Remote Function Call implementation of the ODP consumer interface.

ODP - ABAP CDS Views: ABAP CDS views in the ODP context are used to provide corresponding analytics data sources for use in SAP BW/4HANA. This enables access to extraction annotated ABAP CDS views from SAP S/4HANA and SAP S/4HANA Cloud, allowing full or delta extraction and data access as needed.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 9, Issue 1, September 2021

ODP - BW: SAP BW/4HANA provides data throughOperational Data Providers for extraction, in cases where there are multiple BW/4HANA systems involved in the landscape. All SAP BW/4HANA Info Providers aresupported as a source, accessible via the RFC implementation of the ODPconsumer interface.

ODP - SAP HANA Information Views: SAP HANA views can serve as Operational Data Providers, facilitating data transfer to other warehouse systems and allowing for the direct creation of BW/4HANA objects. This facilitates pushing the business logic from ABAP-based layers to the SAP HANA database, encompassing key figure calculations, currency conversions, joins, and unions specified in SAP HANA views for better performance.

ODP - SLT: The ODP infrastructure, in combination with trigger-based replication in SAP Landscape Transformation Replication Server (SLT), allows for real-time data transfer from multiple source systems to one or more target systems, including SAP BW/4HANA. The SLT Replication Server acts as a provider for the ODP infrastructure by making tables from supported database sources accessible as delta queues.Fig-3 illustrates the different producers and consumers available in general as part of the ODP framework offering.

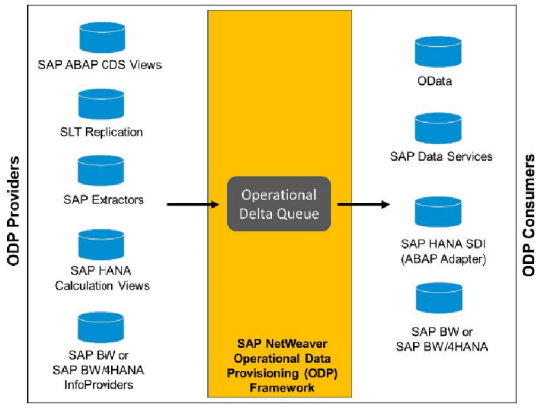

Fig -3: ODP Framework [4]

2.2 SAP HANA Source Systems

SAP HANA source systems provide unified access to sources through SAP HANA smart data integration (SDI) or smart data access (SDA). It also provides access to objects from the local SAP HANA database or an SAP HANA tenant database. The data can be accessed virtually, or persisted into SAP BW/4HANA, in batch mode, or real-time[2]. Below are three different types of connections available with this source system:

Local Database Schema: Provides access to non-BW external schemas on the same SAP HANA database on which the SAP BW/4HANA application is deployed.

Smart Data Access (SDA): Leverages the customer-defined remote connections based on either SDA or SDI to access data on remote sources outside the local SAP HANA database.

Tenant Database Schema: Grants access to additional schemas on different tenants within the SAP HANA database for multi-tenant deployment.

2.3 Big Data Source Systems

Big Data source systems offer centralized access to sources via SAP HANA Smart Data Access using SPARK SQL or VORA adapters[2]. These source systems facilitate the integration of big data sources into the SAP BW/4HANA environment, enhancing its analytical and data processing capabilities with the ability to handle large volumes of diverse data. Big Data source systems support both real-time and batch load ingestion patterns, providing users with the flexibility to ingest data based on their specific requirements. Prerequisites for Big Data source systems include the setup and configuration of SAP HANA Smart Data Access, ensuring that the necessary connections and adapters are in place for seamless data integration.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 9, Issue 1, September 2021

2.4 Flat File Source Systems

Flat file source systems enable the transfer of data from flat files in ASCII or CSV format, allowing users to load data from external sources, such as Microsoft Excel, into the system for analytical purposes. These source systems support batch load ingestion patterns, providing a straightforward method for loading data from flat files into SAP BW/4HANA. Flat file source systems do not have specific prerequisites, as they offer a simple and direct approach to data ingestion from flat file sources, making them accessible for users without complex setup requirements.The SAP BW/4HANA web cockpit also offers a feature for uploading csv and excel files directly into the ADSOs, which improves the self-service capabilities for business users.

III. INGESTION PATTERNS IN BW/4HANA

Ingestion patterns in SAP BW/4HANA refer to the methods used to acquire data from various source systems into the BW/4HANA system. SAP BW/4HANA distinguishes itself from other data warehouse solutions by providing extensive pre-built content for extracting data from SAP ERP systems. SAP has a rich repository of standard data models, extractors, and transformation rules that can be used for data ingestion with its tight integration from SAP ERP to BW/4HANA. These pre-built content offerings in SAP BW/4HANA ensure faster and more efficient data ingestion processes. These patterns can be broadly classified into three categories: batch, real-time, and hybrid.

3.1 Batch Ingestion

Batch ingestion is a traditional method where data is imported in batches at scheduled intervals. This method is suitable for large volumes of data where real-time analysis is not required. The information is extracted and transferred to the BW/4HANA system during times of low usage to reduce any negative effects on system performance.

SAP BW/4HANA supports batch ingestion through various mechanisms, including: ODP based extraction, SAP HANA source, File-based extraction, schedule-based Data Transfer Processes triggered using a process chain. The process chain executes a series of predefined data transfer and transformation steps with increased flexibility and programmable capabilities.

3.2 Real-time Ingestion

Real-time ingestion is a method where data is imported into the BW/4HANA system immediately as it becomes available in the source. This method is suitable for scenarios where real-time analysis and reporting are required. SAP BW/4HANA provides several options for real-time data ingestion, including: SLT, Smart Data Integration, Data Services, ODP Real-Time Delta Extraction and streaming process chain.

Streaming Process Chain is a unique feature in SAP BW/4HANA that enables frequent and rapid data updates to advanced Data Store objects. This feature allows the process chain to be started and executed very frequently, making real-time data updates possible. The streaming process chain functions with a newly implemented Queued Task Manager Framework, allowing for a variable number of "worker" tasks responsible for retrieving the next process from a queue[5]. If a second, third, or subsequent execution reaches the currently active process, these requests to start the process once more are written to a queue. Once the current operation is complete, the worker role accumulates all new initiation requests for a specific procedure and initiates the process once. As a result, the subsequent iterations in the second and third sequences are disregarded, allowing only the latest iteration to proceed. This mechanism significantly enhances the efficiency and performance of data ingestion in SAP BW/4HANA.

3.3 Hybrid Ingestion

Hybrid ingestion is a combination of batch and real-time ingestion methods. In this pattern, real-time ingestion is used for critical data that requires immediate analysis, while batch ingestion is used for less time-sensitive data. This approach provides a balance between the need for up-to-date information and system performance.

SAP BW/4HANA features the Advanced Data Store Object with Write Interface, which enables data to be transferred using tools like Data Services or through SAP Cloud Platform Integration by PUSH mechanism[3]. This open API for writing to DataStore objects can be utilized by external tools.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 9, Issue 1, September 2021

The ADSO with Write Interface is versatile and can be used for all types of standard ADSO and staging ADSO. However, it cannot be used in conjunction with Inventory enabled or Planning enabled. When data is loaded using streaming, it becomes active automatically.

Data can be loaded into the ADSO using either RFC or HTTP data. There are two possible procedures: Sending data without a request, where a new internal request is opened for every call, and sending data with a request, which follows the sequence: Open Request, Send Data, Close Request[3].

SAP BW/4HANA also offers various APIs for writing data directly into ADSO, expanding the possibilities for automating manual file loading processes. This includes the ability to automate the uploading of flat files from the application server through ABAP program and process chain setup, utilizing the APIs outlined below. This enables efficient and direct data ingestion into SAP BW/4HANA without the need for intermediate staging areas or additional data transformation steps.

The RSDSO_WRITE_API and RSDSO_WRITE_API_RFC offer functions for directly updating data into the inbound table of a standard ADSO. Activating requests in the ADSO is managed by the RSDSO_ACTIVATE_REQ_API_RFC, ensuring that the data is stored and accessible in the active table for reporting purposes. Additionally, the RSDSO_DU_WRITE/DELETE/CLEANUP APIs allow for writing, deleting, and cleaning up data within direct update type ADSOs[3].

IV. SAP HANA DATA PROVISIONING

SAP HANA data provisioning: We have seen handling data ingestion specifically related to BW/4HANA application layer, but data ingestion in SAP BW/4HANA has evolved significantly over time to provide efficient and streamlined processes. One of the key advancements in data ingestion is the use of SAP HANA as a data provisioning method. With SAP HANA data provisioning, data can be directly loaded into HANA database tables without involving BW/4HANA application. This paves the way for a plethora of native HANA capabilities comparable to any SQL data warehouse. Data provisioning in SAP HANA can be accomplished using a range of approaches, such as trigger-based replication (SLT), SAP HANA Direct Extractor Connection (DXC), ETL-based replication (SAP BODS), log-based replication (SAP Replication Server), and the use of SAP HANA Smart Data Integration and Smart Data Access[6]. Each approach offers its own benefits and is appropriate for varying scenarios.

When working directly with the HANA database, there are numerous ingestion tools in the market that can be used to push data to SAP HANA tables and complement SAP integration technologies. Any ETL tool or data integration platform supporting JDBC, ODBC, OData, or web services can be utilized for ingesting data into SAP HANA. Fig -4 illustrates the separation of BW/4HANA and Native HANA schemas for independent data ingestion mechanisms.

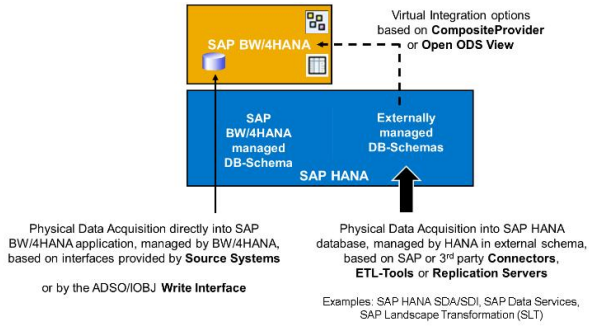

Fig -4: BW/4HANA and Native HANA Schema separation

V. BEST PRACTICES FOR DATA INGESTION

When considering data ingestion in SAP BW/4HANA, there exist numerous recommended approaches that can promote efficient performance and effectiveness. First, it is crucial to meticulously plan and devise your data ingestion strategy before commencing the process. This entails identifying data sources, comprehending data needs, and selecting the most suitable approach for ingesting data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 9, Issue 1, September 2021

SAP's best practice design for the BW/4HANA data warehouse emphasizes the importance of building an Open Operational Data Store Layer (ODS Layer)[2] where data is collected virtually or physically into the data warehouse. This approach plays a crucial role in facilitating efficient data ingestion and providing data to subsequent layers of the data warehouse. Acquiring the data to closely match the original source tables facilitates a more efficient process, easily transformable, and resilient ingestion for simplified maintenance, standardization, and integration with other data sources. For BW/4HANA managed data, implementing a field-based staging layer can also enhance data ingestion performance by enabling closer proximity error handling and validations at the source. The addition of identifiers for the source system and load timestamps to the staging layer's source tables is an often overlooked but essential technical aspect. These identifiers not only aid in tracking and managing changes to the dataset over time in a data warehouse, but also provide a way for business intelligence tools to manage delta loads.

LSA++ defines the Open Operational Data Store layer as the initial stage where data is virtually or physically acquired into the data warehouse[2]. This entails acquiring data while maintaining its semantics, syntax, and quality from the source system. The loading of data using BW/4HANA source systems ensures that SAP BW/4HANA application has full control over data management. The managed schema in which this data is stored is known as the "BW-managed schema" within the SAP HANA database.

Data integration with SAP HANA database involves managing the data in DB-tables or DB-views of customer-defined SAP HANA schemas which are distinct from the internally managed ("Externally managed") schema by SAP BW/4HANA. Such information can be modeled and integrated to any required degree into SAP BW/4HANA through use of Open ODS Views or SAP HANA Calculation Views.

One of the primary sources for data ingestion in SAP BW/4HANA is SAP ERP system. SAP ERP provides a rich source of structured transactional data in the form of standard data source extractors that can be ingested into SAP BW/4HANA for further analysis and reporting. Based on several implementation experience and best practices, it is recommended to use the SAP Business Content extractors for data ingestion from SAP ERP into SAP BW/4HANA as a primary source for data integration. Any enhancements or customizations to the standard extractors were traditionally done using the SAP ABAP programming language but with the advent of SAP ABAP Core Data Services (CDS), data ingestion from SAP ERP sources has become even more streamlined and efficient. It is recommended to utilize ABAP CDS to enhance or custom build the data sources for data ingestion from SAP ERP system into SAP BW/4HANA. SAP has implemented automatic Change Data Capture capability with ABAP CDS beginning from SAP S/4HANA Cloud 1905 and SAP S/4HANA on-premises 1909.

Housekeeping activities in SAP BW/4HANA are essential for maintaining the system's health and performance. These tasks encompass managing administrative data, requests, application logs, IDocs, and statistics. They can be classified into general monitoring, system health monitoring, performance-related monitoring, and occasional activities. Key housekeeping tasks include log management, request management, and data management. The Process and Request Status Management (RSPM) service provided by SAP BW/4HANA supports these activities to ensure the smooth running and optimal performance of the system.

VI. CASE STUDIES

6.1 Large Healthcare Manufacturing company

A large retail corporation with a global presence decided to migrate from their traditional SAP BW system to SAP BW/4HANA to leverage the benefits of real-time analytics and improved performance. The migration involved a complete overhaul of their existing data models and the introduction of new LSA++ architecture. The company employed SAP ABAP CDS for change data capture to achieve automated handling of incremental data, providing near real-time data feeds to SAP BW/4HANA. Post-migration, the corporation reported a significant improvement in report generation speed and data load times, leading to more timely and accurate business decisions.

6.2 Multinational Medical Technology Company

A multinational manufacturing corporation adopted SAP BW/4HANA to enhance their production planning and inventory control. They leveraged the real-time data ingestion features of SAP SLT to directly input production and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 5.731
Volume 9, Issue 1, September 2021

sales data into SAP HANA. This enabled them to produce real-time reports on their production efficiency and stock levels, resulting in improved production planning and decreased inventory expenses.

6.3 Global Biotechnology Company

A global biotech organization implemented SAP BW/4HANA to manage their extensive and intricate datasets across multiple SAP BW systems, aiming to leverage the benefits of Cloud Infrastructure as a Service and modernize their legacy platform. They used a variety of ingestion mechanisms to meet the batch, real-time, near real-time, and advanced data modeling capabilities of SAP BW/4HANA and Native HANA data warehousing to create adaptable and scalable data models. As a result, the organization reported enhanced data management, quicker report generation, faster time-to-market, and elimination of on-premises hardware limitations.

VII. CONCLUSIONS

This paper has explored the extensive data ingestion capabilities provided by SAP BW/4HANA to integrate diverse data sources into a unified enterprise data warehouse. The native SAP HANA-based Operational Data Provisioning framework offers an efficient mechanism to extract and load data into the system. ODP enables optimized change data capture and provides interfaces to numerous SAP and non-SAP sources.

In addition to ODP, SAP BW/4HANA supports ingestion from varied source types including SAP HANA, Big Data, and flat files based on specific integration requirements. It facilitates batch, real-time and hybrid patterns allowing flexibility in data loading strategies. Direct data provisioning into SAP HANA database using integration tools is also viable for enhanced agility.

The insights presented equip SAP BW/4HANA adopters with in-depth knowledge of the available data ingestion options. By assessing their analytics and reporting needs, organizations can design optimal ingestion architectures leveraging the right frameworks, sources, and patterns. Efficient data integration is key to realizing the full potential of SAP BW/4HANA for accelerated data-driven decision making.

As big data growth proliferates, future research could explore adopting emerging technologies like blockchain, IoT and artificial intelligence to make data ingestion smarter, more secure, and scalable. With sound ingestion strategies, businesses can gain a competitive edge with timely data-backed insights.

REFERENCES

[1] “SAP BW/4HANA | Data Warehouse and EDW Solutions,” SAP. Accessed: Jul. 06, 2021. [Online]. Available: http://www.sap.com/bw4hana

[2] “SAP_BW4HANA_en_latst.pdf.” Accessed: Jul. 06, 2021. [Online]. Available: https://help.sap.com/doc/24b2b055b00143c5bb552edff7cc57c4/1.0.12/en-US/SAP_BW4HANA_en.pdf

[3] “DataStore Object with Write Interface | SAP Help Portal.” Accessed: Jul. 25, 2021. [Online]. Available: https://help.sap.com/docs/SAP_BW4HANA/107a6e8a38b74ede94c833ca3b7b6f51/8127edbcf2fa488ba04dfc1751d5261d.html?version=2.0.0

[4] “SAP Operational Data Provisioning (ODP),” SAP. Accessed: Aug. 09, 2021. [Online]. Available: https://www.sap.com/documents/2017/06/66673acb-c37c-0010-82c7-eda71af511fa.html

[5] “Process Chains for Streaming | SAP Help Portal.” Accessed: Aug. 13, 2021. [Online]. Available: https://help.sap.com/docs/SAP_NETWEAVER_750/2e90b26cf7484203a523bf0f4b1bc137/ac029de05e164a12ac1ce08d16180f05.html?version=7.5.4

[6] “SAP HANA Data Provisioning | SAP Help Portal.” Accessed: Aug. 18, 2021. [Online]. Available: https://help.sap.com/docs/SAP_HANA_PLATFORM/eb3777d5495d46c5b2fa773206bbfb46/d4eb257ebb571014928bd908f34f3f0a.html?version=2.0.01&locale=en-us