IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Modernizing Data Integration: SAP Datasphere's Integration with Data Visualization Tools Versus Traditional Data Warehouse Architectures

[1]Govinda Rao Banothu
SAP Analytics Cloud Technical Specialist
NJ, USA.

[2]Fabio De Castro Fialho
SAP Analytics Solutions Architect
RJ, Brazil.

Abstract: In today's rapidly evolving digital landscape, the demand for agile, real-time analytics has driven organizations to reassess traditional data warehouse architectures. While these systems have been long essential for providing structured, historical data for business intelligence, they often struggle to meet the needs of modern, decentralized, and real-time decision-making. SAP Datasphere, a cloud-native data integration platform, addresses these challenges by offering flexible, scalable solutions that align with the demands of contemporary businesses. This paper presents a comprehensive comparison between SAP Datasphere and traditional data warehouse architectures, with a focus on real-time data access, scalability, and seamless integration with leading data visualization tools such as SAP Analytics Cloud (SAC), Tableau, and Microsoft Power BI. Through a detailed analysis, this paper explores the technical advantages, performance improvements, and strategic benefits of SAP Datasphere, particularly in enhancing data accessibility, fostering collaboration, and optimizing data governance. Real-world use cases are provided to demonstrate how SAP Datasphere empowers organizations to meet the growing need for real-time insights, transforming data management and visualization for more informed decision-making.

Index Terms - SAP Datasphere, Data Warehouse, Real-Time Analytics, Data Integration, SAP Analytics Cloud, Business Intelligence, Cloud-Native Architecture**.**

I. INTRODUCTION

In today's hyper-competitive market, businesses must leverage data for more than just historical reporting—they need real-time insights to respond to fast-changing environments. Data has become a core asset that influences strategic decisions, product development, marketing initiatives, and operational improvements. Traditional data warehouse systems have served enterprises for decades, offering a centralized repository for structured data consolidated from various operational systems.

However, the explosion of data sources—including IoT devices, social media, customer touch points, and external APIs—along with the need for real-time data analysis has pushed traditional data warehouse architectures to their limits. These systems, while reliable for batch processing and historical analysis, often cannot meet the agility, scalability, and speed demands of modern enterprises.

To address these challenges, cloud-native platforms like SAP Datasphere have emerged, offering decentralized data integration and seamless connectivity to visualization tools, thus enhancing business agility and enabling faster, more collaborative decision-making processes.

II. TRADITIONAL DATA WAREHOUSE ARCHITECTURE

A. Overview of Legacy Data Warehousing: -

Traditional data warehouses are built on centralized, structured architectures designed to store transactional data from various operational systems, such as ERP, CRM, and manufacturing applications. This data undergoes an extract, transform, and load (ETL) process, during which it is cleansed, aggregated, and organized into relational database structures like star or snowflake or Galaxy/Fact constellation schemas for the purpose of analysis and reporting.

Two influential methodologies for data warehouse design were introduced by W.H. Inmon and Ralph Kimball, widely regarded as the pioneers in the field of data warehousing. Their respective approaches have had a profound impact on both the theoretical and practical development of data warehouse architecture.

Top-Down Approach (W.H. Inmon): Inmon's methodology defines a data warehouse as a subject-oriented, integrated, time-variant, and non-volatile collection of data designed to support decision-making processes. The top-down approach emphasizes the creation of a centralized, enterprise-wide data warehouse, which serves as the foundation for subsequent data marts.

Bottom-Up Approach (Ralph Kimball): In contrast, Kimball advocates for a more incremental approach, where businesses build smaller, specialized data marts focused on specific business processes. These data marts are then integrated to form a comprehensive data warehouse. This methodology focuses on delivering business value early, with flexibility to expand the data warehouse over time as the organization's needs evolve.

Both approaches have shaped the evolution of data warehousing, providing businesses with different strategies for organizing and utilizing data to enhance decision-making.

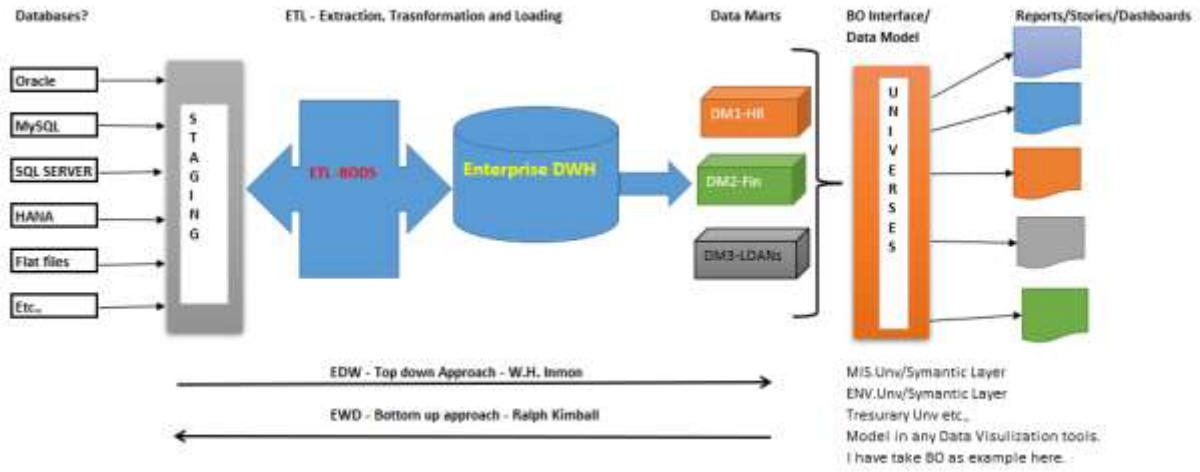


Fig 1: Traditional Data Warehouse Architecture

Legacy data warehousing refers to traditional, on-premises systems used to store and manage large volumes of structured data for business intelligence and reporting. These systems typically feature centralized architectures that aggregate data from multiple operational sources like ERP, CRM, and transactional applications. The main goal of legacy data warehouses is to provide a consolidated repository of historical data for analysis and decision-making.

Key characteristics of legacy data warehousing include:

Centralized Data Storage: Data is stored in relational databases (RDBMS) and organized using complex schema designs, such as star or snowflake or Galaxy/Fact constellation schemas, to optimize querying.

ETL Processes: Data from various systems is extracted, transformed, and loaded (ETL) into the warehouse, ensuring consistency but causing delays due to batch processing, typically scheduled daily or weekly.

Rigid Schema: Data warehouses are built on predefined schemas, requiring significant effort to modify when adding new data sources or adapting to changing business needs.

Limited Real-Time Capabilities: Traditional data warehouses rely on batch processing, resulting in delays between data creation and its availability for analysis.

High Operational Costs: These systems require significant investments in hardware, storage, and IT resources to maintain, particularly as data volumes grow.

Despite their foundational role in business intelligence, legacy data warehouses struggle to meet the demands of real-time analytics, flexibility, and scalability. As a result, many organizations are turning to modern, cloud-based data solutions to overcome these limitations.

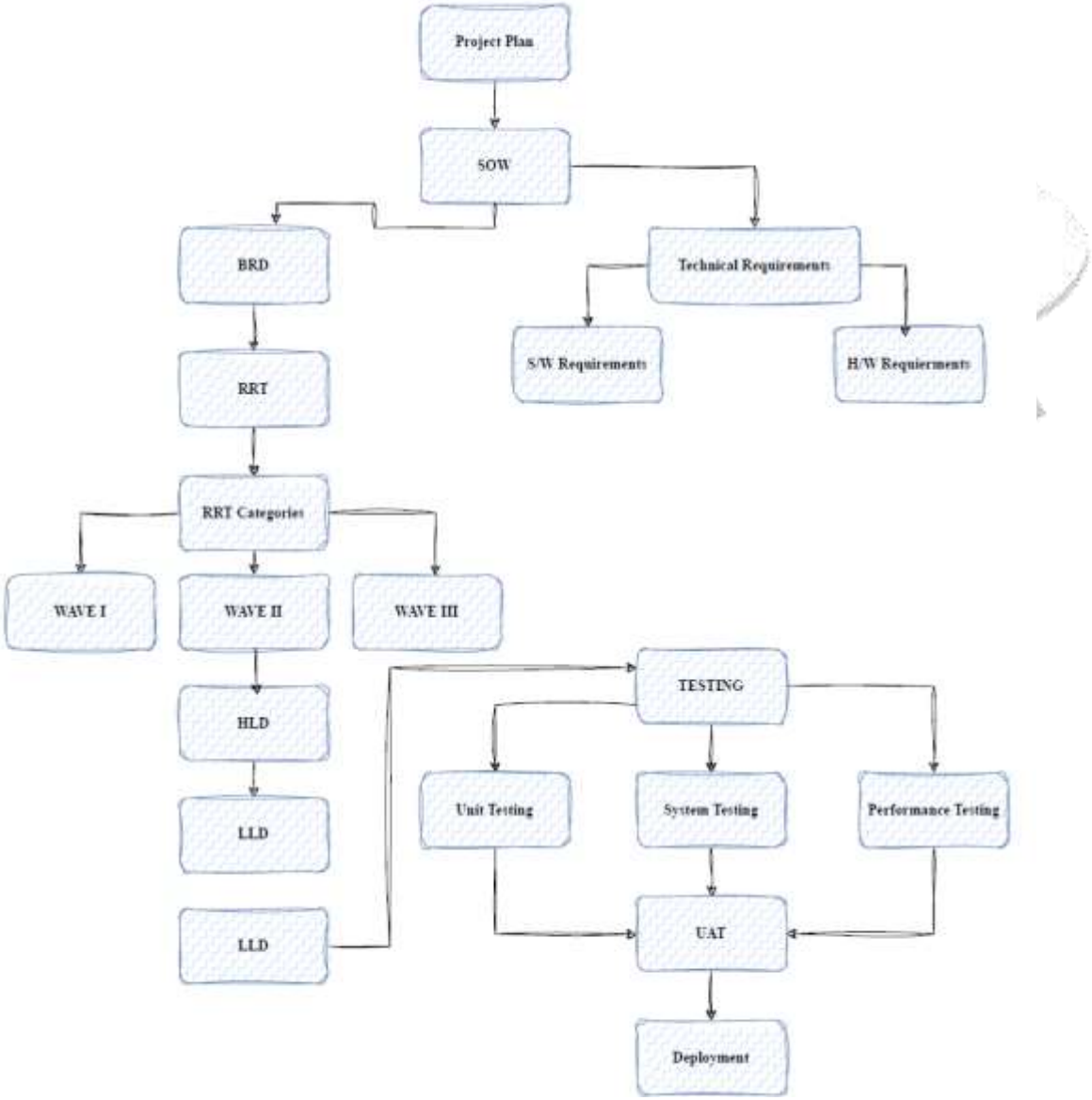


Fig 2: General Life Cycle of Data Visualization Project

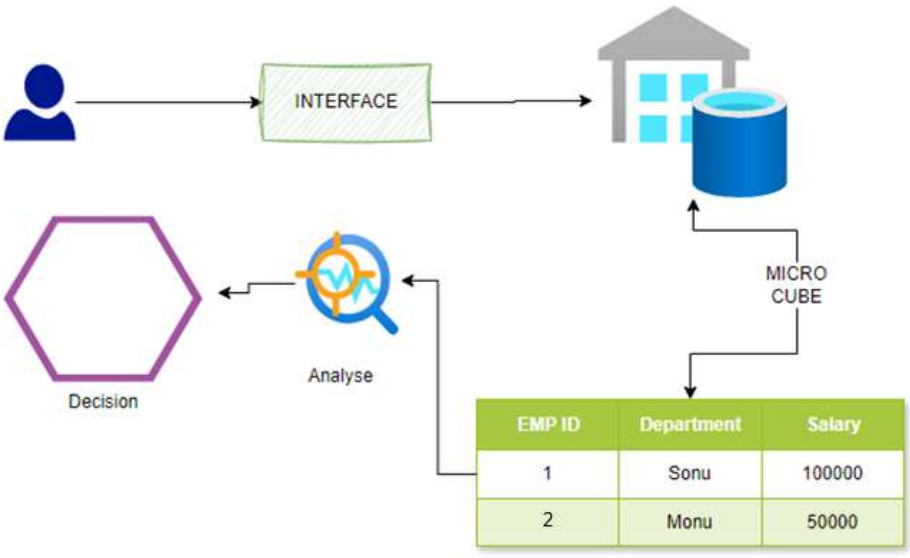


Fig 3: How the data extract from Data Warehouse for reporting / Analytics

Interface / Semantic Layer – A well-designed interface ensures efficient SQL execution and accurate data retrieval from the underlying database. It includes logical pointers, or metadata, that reference the actual data stored in the data warehouse (DWH). A semantic layer is built on top of the source data, such as a data warehouse, data lake, or data mart. Rather than storing the data itself, the semantic layer acts as an abstraction and metadata layer, helping users interact with and analyze data more effectively.

In the context of data visualization tools like SAP Business Objects (BO), SAP Analytics Cloud (SAC), Power BI, and Tableau, the semantic layer plays a crucial role by connecting the underlying data sources to the analytics tools. It simplifies data access, enhances user understanding, and ensures consistency across data sources.

The semantic layer serves several important functions:

- **Simplifies Data:** It transforms complex, raw data into a more user-friendly format, organizing it into tables, columns, measures, and relationships that are easier to navigate and understand.
- **Connects Data Sources:** The semantic layer integrates multiple data sources, including databases, cloud services, and on-premises systems, into a unified view for analysis.
- **Enhances Data Governance:** By standardizing data definitions and access policies, the semantic layer helps establish consistent data governance, ensuring that users work with reliable, accurate information.
- **Improves Data Analysis:** It democratizes data access, making it easier for both technical and non-technical users to explore and analyze data without needing in-depth knowledge of the underlying data structures.
- **Facilitates Reporting:** A single semantic model can serve as the foundation for generating multiple reports, ensuring consistency in how data is presented across the organization.
- **Reduces Administrative Overhead:** Through scheduled model refreshes and automated error resolution, the semantic layer minimizes the administrative burden associated with data management.

While this architecture has been widely adopted across industries, several challenges have emerged as data volumes and varieties continue to grow:

Challenges of Traditional Data Warehouse Architectures: -

Batch Processing Bottlenecks: Traditional ETL processes often run in scheduled batch intervals (e.g., daily or weekly), which results in delays between when data is generated and when it becomes available for analysis. This delay can be a significant disadvantage in today's data-driven environment, where real-time insights are critical for timely decision-making.

Rigid Schema Definitions: Data warehouses typically rely on predefined schemas for storing and querying data. Adding new data sources or adjusting to changing business requirements often requires substantial modifications to the schema. This can be time-consuming and costly, especially as the data landscape evolves.

Data Silos: While data warehouses consolidate data, they can also perpetuate data silos. Data is often replicated across multiple systems to meet various business needs, making it difficult to maintain consistency and integrity across the organization. This fragmentation can lead to issues with data quality and accuracy.

High Costs: Maintaining a traditional data warehouse requires significant investment in infrastructure, especially as data volumes increase. On-premises data warehouses incur costs for hardware, storage, and ongoing maintenance, as well as the need for specialized technical staff to manage complex ETL processes and optimize system performance.

Inability to Handle Unstructured Data: Traditional data warehouses are optimized for structured, relational data, but the growing volume of unstructured data (e.g., text, images, videos, social media feeds) presents new challenges. Managing and analyzing unstructured data often requires separate systems, creating additional complexity and inefficiencies in the data ecosystem.

As businesses increasingly demand real-time, flexible, and scalable data solutions, these challenges highlight the limitations of traditional data warehouse architectures and the need for more modern, cloud-native alternatives.

B. Limited Integration with Modern Analytics Tools: -

Another limitation of traditional data warehouses is their limited ability to integrate effectively with modern data visualization tools. Organizations typically rely on business intelligence (BI) platforms like SAP Business Objects/Analytics Cloud/Analysis for Office/Lumira, Tableau, or Power BI to query and generate reports from the data warehouse. However, because traditional data warehouses often operate on batch processing, the data available for visualization is frequently outdated. This delay makes it challenging to leverage these BI tools for real-time decision-making.

III. SAP DATASPHERE: A MODERN ANALYTICS TOOLS

A. Architectural Innovations in SAP Datasphere: -

SAP Datasphere, previously referred to as SAP Data Warehouse Cloud, addresses the challenges associated with traditional data warehousing solutions. As a comprehensive data service built on the SAP Business Technology Platform (SAP BTP), it serves as the backbone of a business data fabric, empowering organizations to provide meaningful data to all data consumers while maintaining the integrity of business context and logic.

The platform facilitates a range of functionalities, from data ingestion to self-service data access, encompassing both SAP and non-SAP data sources. It offers a multi-cloud, multisource business semantic service that supports enterprise analytics and planning.

As a cloud-native solution, SAP Datasphere enables flexible, real-time data integration, allowing organizations to connect, manage, and analyze data from diverse sources without the need for replication. This capability fosters a business data fabric architecture that harmonizes mission-critical data across the organization, empowering business experts to make informed and impactful decisions.

SAP Datasphere integrates previously siloed functionalities into a cohesive service encompassing data integration, cataloging, semantic modeling, data warehousing, and workload virtualization across both SAP and non-SAP data. It preserves the complete meaning and context of SAP data across various systems and cloud environments, while also integrating with other data vendors' platforms. This integration facilitates seamless and scalable access to a single authoritative source of an organization's most valuable enterprise data.

Moreover, SAP Datasphere capitalizes on existing data investments, eliminating the need to transfer data into additional storage solutions. This approach significantly simplifies the data landscape while ensuring robust governance throughout the data lifecycle.

SAP has established strategic partnerships with prominent data and Artificial Intelligence (AI) companies, including Collibra, Confluent, Databricks, and Data Robot, to enhance SAP Datasphere. These collaborations facilitate the development of a unified data architecture that securely integrates SAP software with third-party data.

Collibra will be integrated with SAP via a customized solution, enabling customers to implement an enterprise-wide data management strategy. This integration allows organizations to create a comprehensive data catalog that details the origins of data across their entire landscape, encompassing both SAP and third-party data. Collibra ensures that organizations can locate and leverage trustworthy data effectively.

Confluent plans to integrate its data streaming platform with SAP, empowering companies to access valuable business data and connect it to external applications in real time. Confluent's cloud offering serves as a central platform for continuous, real-time data streams sourced from various organizational inputs.

Databricks enables customers to integrate their data Lakehouse with SAP software, facilitating data exchange while preserving semantic integrity. This integration simplifies the overall data landscape for organizations.

Data Robot provides customers with automated machine learning capabilities for multimodal data within SAP Datasphere, allowing for direct integration into their data fabric for effective business data management, irrespective of the cloud platform in use.

SAP Datasphere's architecture is founded on the concept of data spaces—logical partitions that allow users to manage and visualize data in real time.

This architecture confers several significant advantages over traditional data warehouse systems:

Real-Time Data Access: SAP Datasphere eliminates the delays associated with batch processing by enabling users to access live data as it is generated. This capability is crucial for organizations that require immediate decision-making based on the most current information.

Data Virtualization: Instead of replicating data across multiple systems, SAP Datasphere allows organizations to query data directly from its original source in real time through data virtualization. This approach minimizes data duplication and storage costs while ensuring that data remains up to date.

Cloud Scalability: As a cloud-native platform, SAP Datasphere provides elastic scalability, allowing organizations to adjust their data storage and processing capabilities on demand. This flexibility removes the necessity for upfront infrastructure investments and enables organizations to scale resources according to their needs.

Unified Data Model: SAP Datasphere accommodates both structured and unstructured data, enabling organizations to create a unified data layer that integrates information from various systems, including both SAP and non-SAP sources. This integration provides a holistic view of enterprise data assets, facilitating deeper insights.

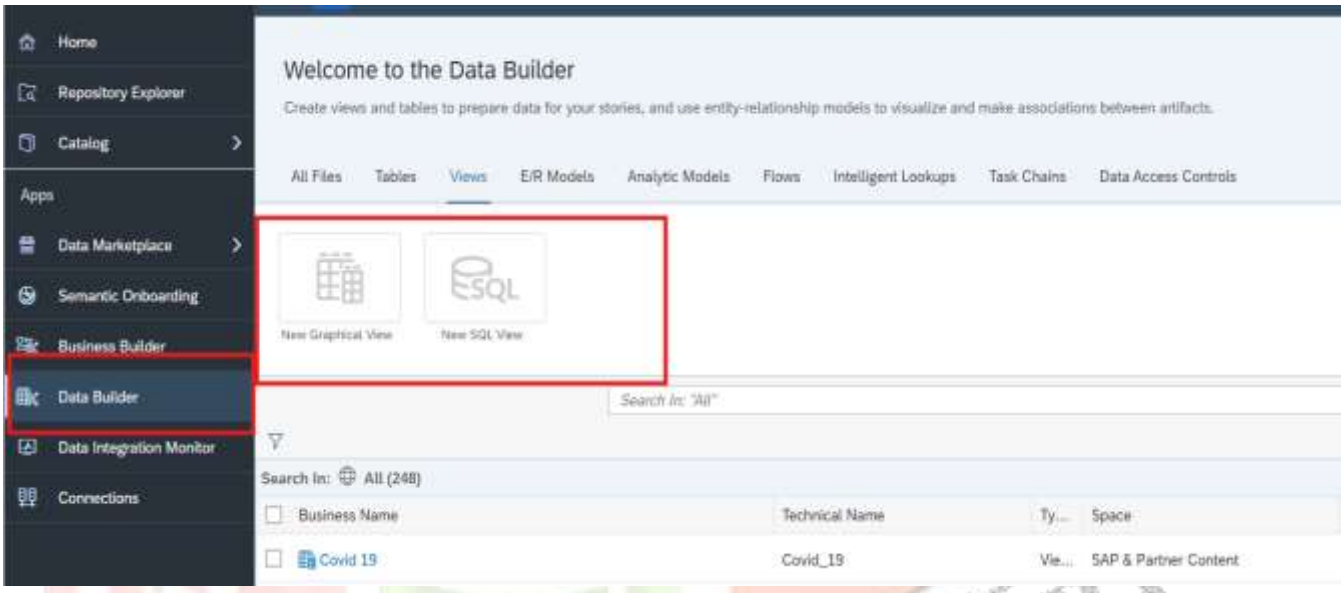
B. Integration with SAP Analytics Cloud and Third-Party Tools: -

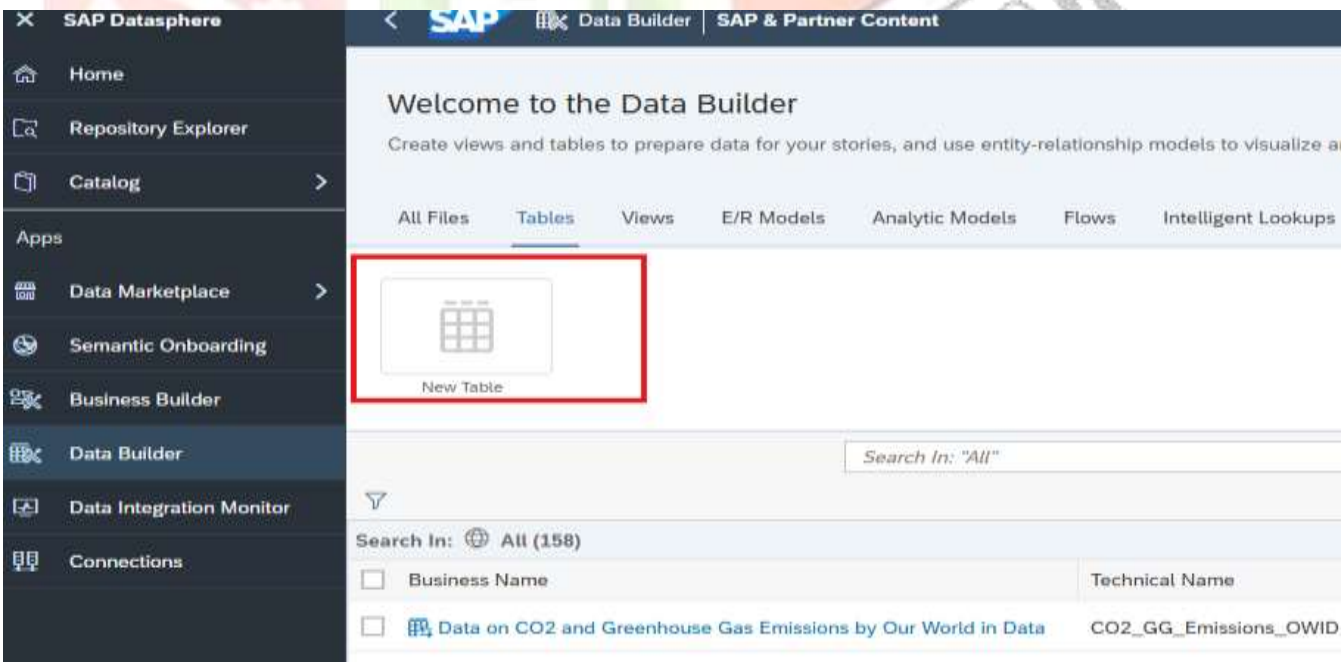
A key strength of SAP Datasphere is its seamless integration with modern data visualization tools. Platforms such as SAP Analytics Cloud (SAC), Tableau, and Power BI can connect directly to SAP Datasphere, enabling real-time data access and interactive visualizations.

SAP Analytics Cloud (SAC): As SAP's premier business intelligence and planning platform, SAC offers integrated analytics and planning capabilities. Its integration with SAP Datasphere allows users to generate real-time dashboards and reports using live data from across the enterprise. This capability eliminates the need for periodic data refreshes and empowers users to make informed, data-driven decisions based on the most current information available.

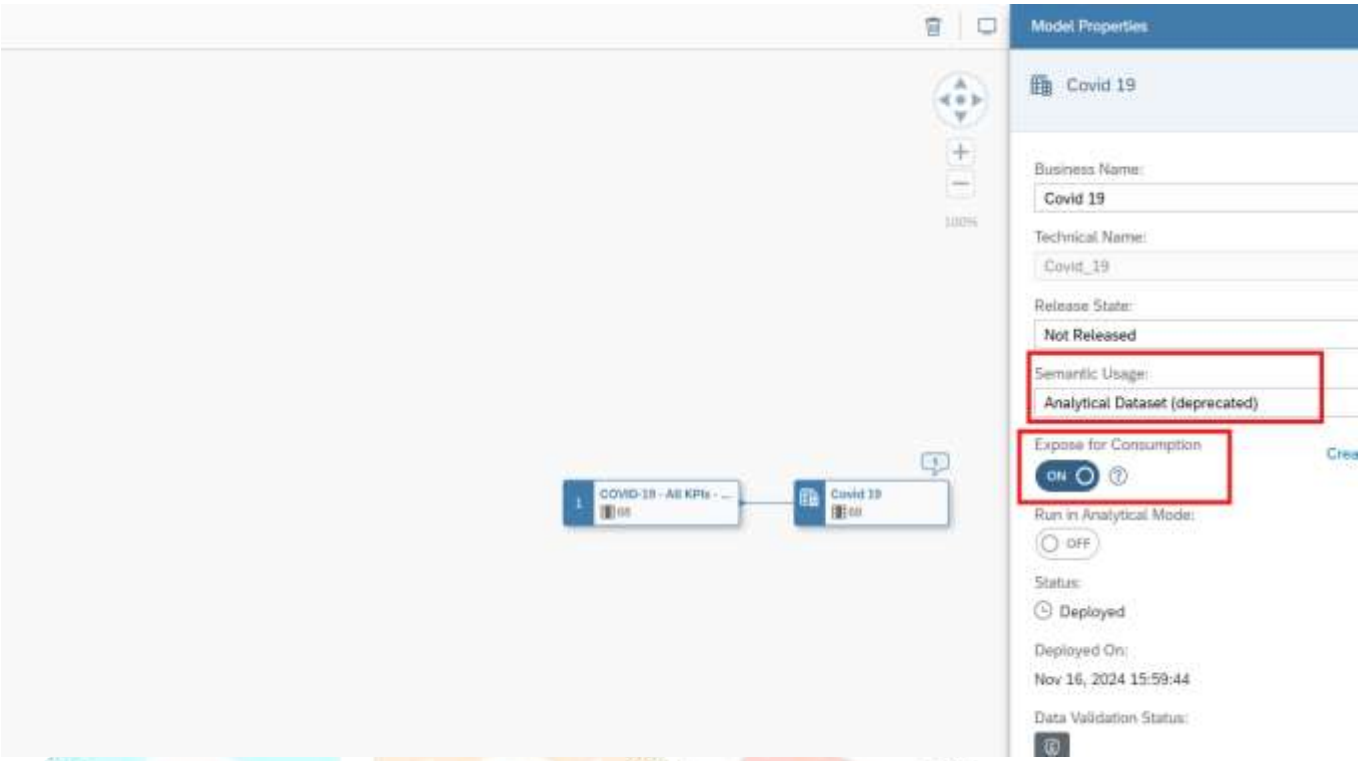
Steps to Integrate SAP Datasphere with SAC: -

1. Enable Business Contents in SAP Datasphere.
2. Bring data using either Live data connection projection in view or persist data in Remote Tables using Data Builder.

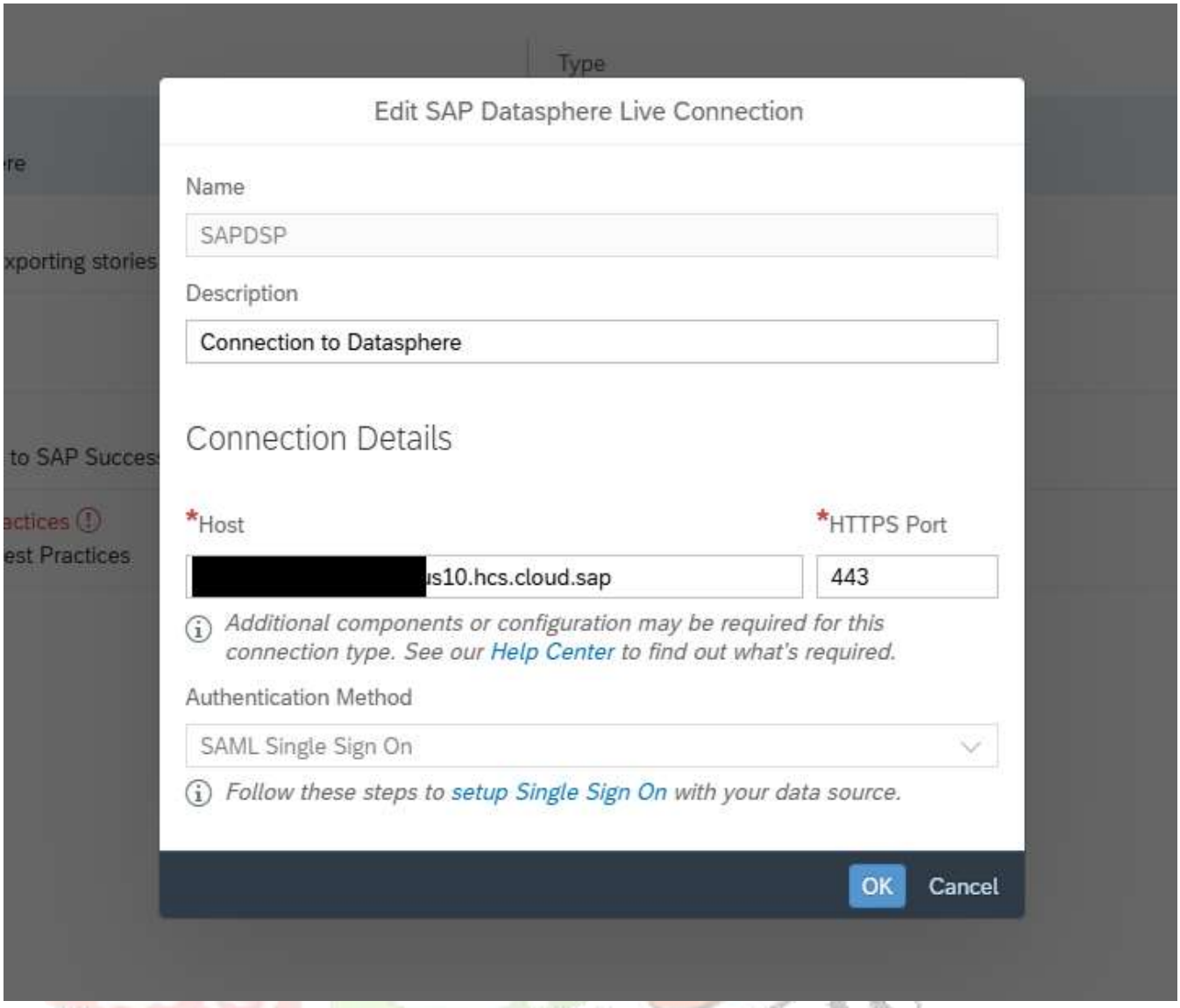




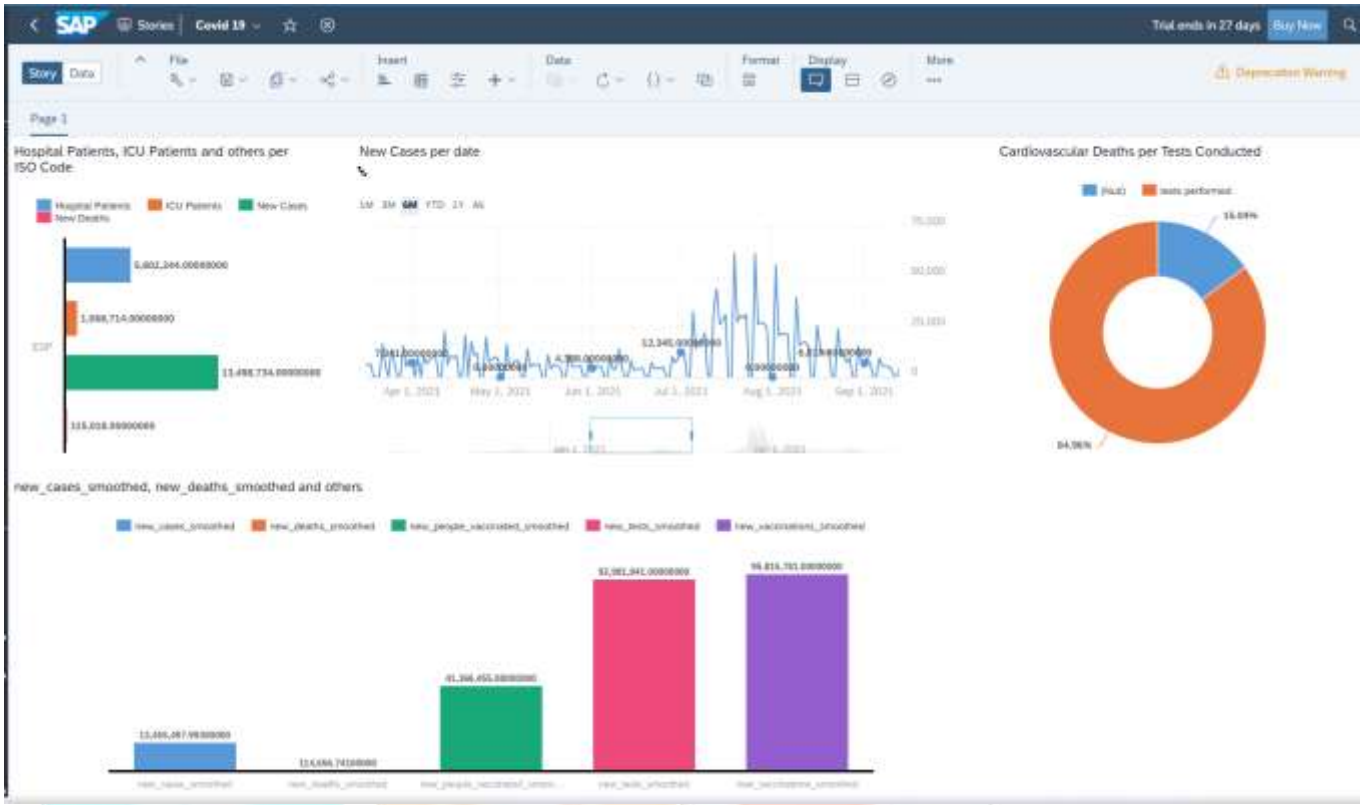
3. Create a view with Semantic Usage as Analytical Dataset or Analytic model and mark the view as exposed for consumption.



4. Now switch to SAP Analytics Cloud and create a connection like below using the host details of Datasphere.



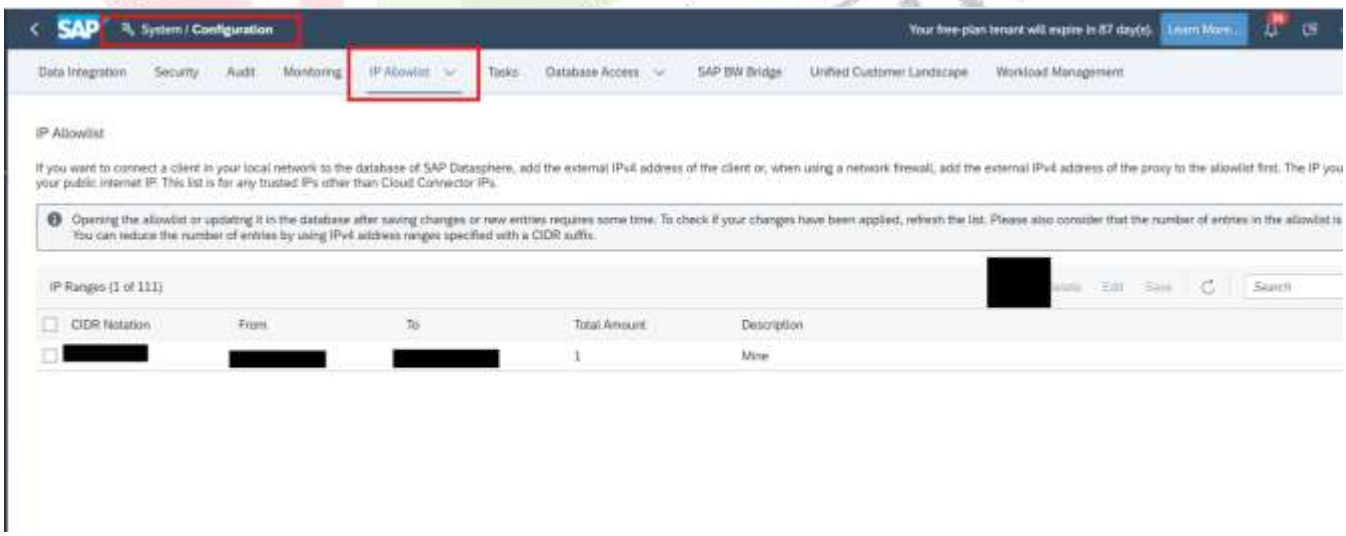
5. Using the same connection created as above Start creating Stories.



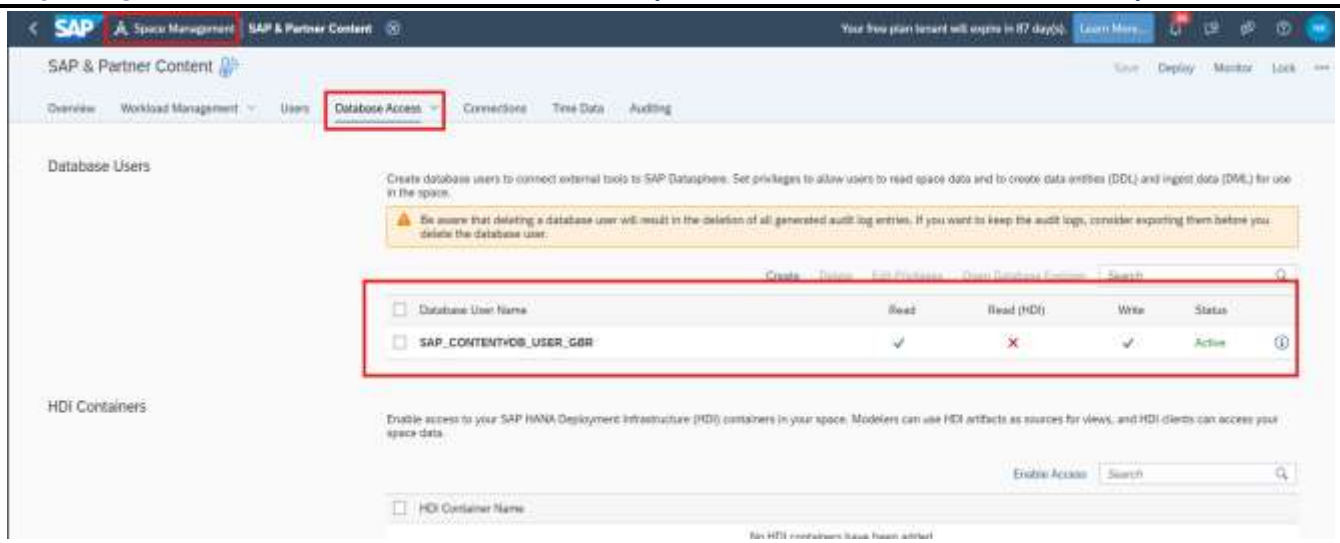
Third-Party Tools: SAP Datasphere also integrates with various third-party visualization platforms, including but not limited to Tableau, Microsoft Power BI, and Qlik. This integration enables organizations to continue utilizing their preferred analytics tools while benefiting from SAP Datasphere's real-time data access and decentralized data management capabilities. With standardized connectors and APIs, organizations can easily incorporate SAP Datasphere into their existing analytics ecosystems.

Steps to Integrate SAP Datasphere with Power BI: -

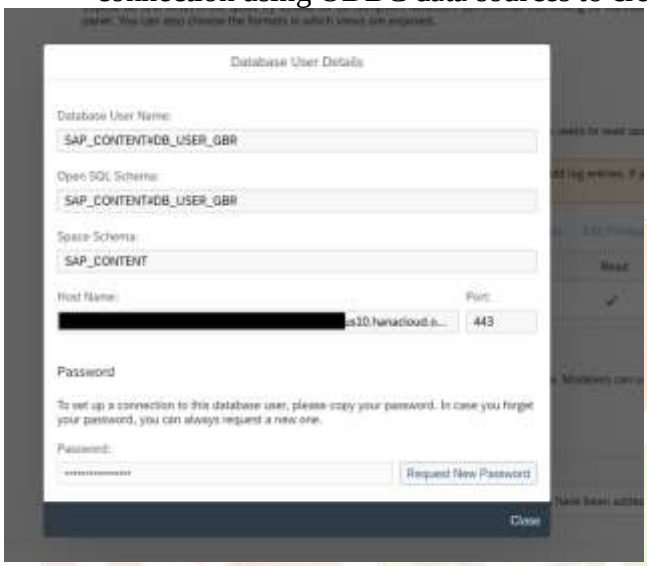
1. Download the HANA Client from SAP Development Tools.
2. Add your external IP address into the IP whitelist of SAP Datasphere.

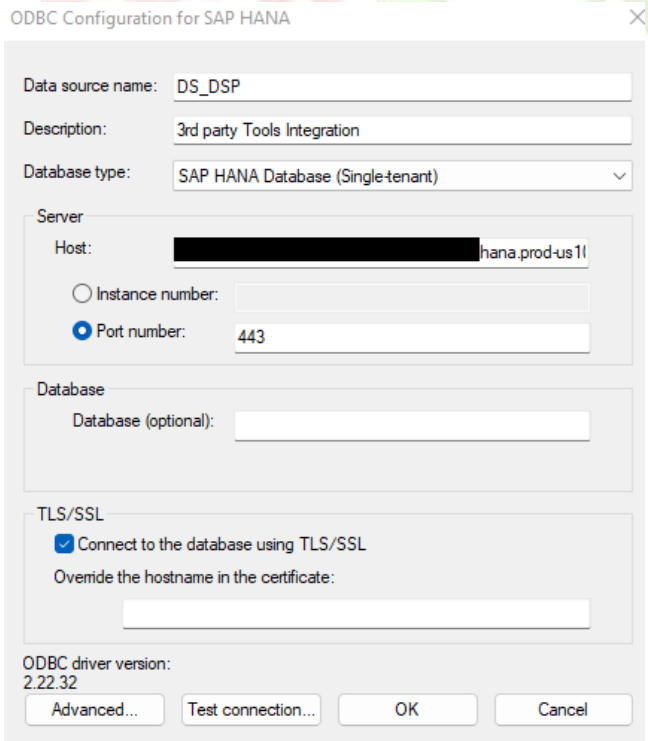


3. Create a database (DB) user for connecting to Datasphere from External tool via ODBC (Open Database Connectivity) connection.

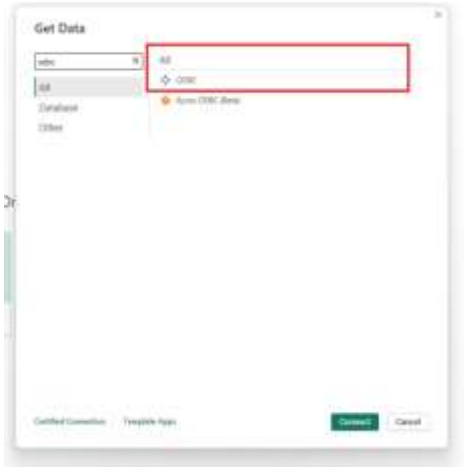


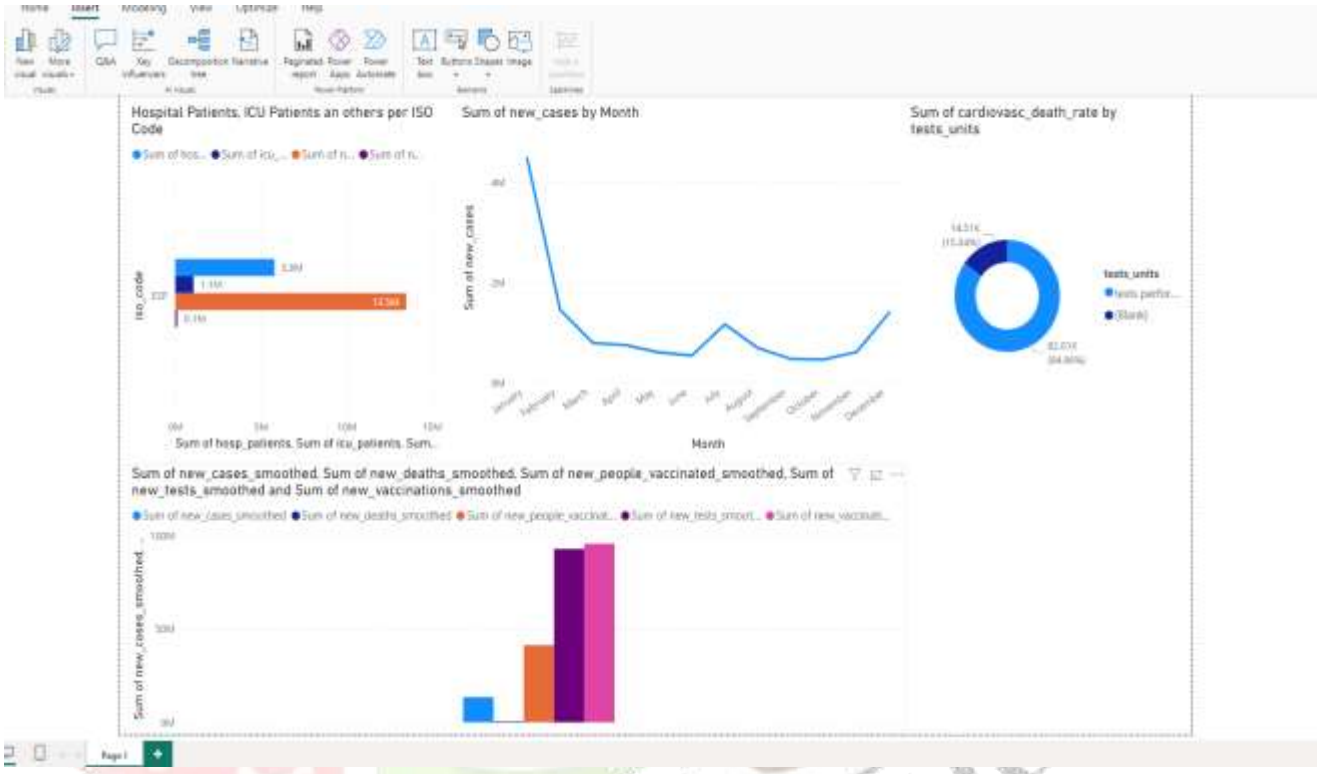
4. Use Hostname of Datasphere present in the Database user details dialog box and create the ODBC) connection using ODBC data sources to create an ODBC connection.





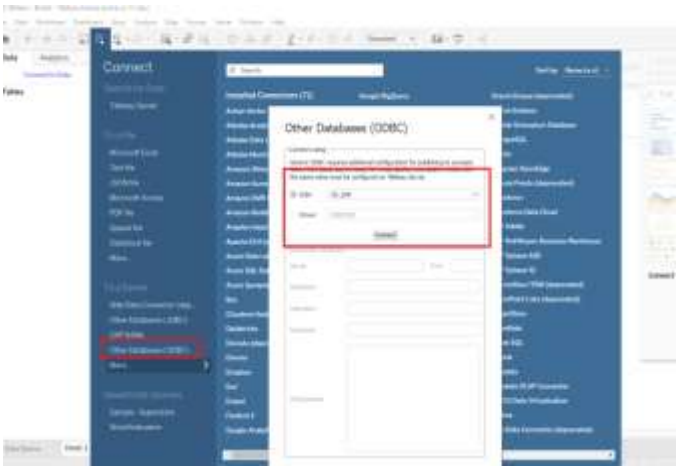
5. Open Power BI and select get data option and connect using ODBC connection just created to start consuming the Datasphere view in Power BI.



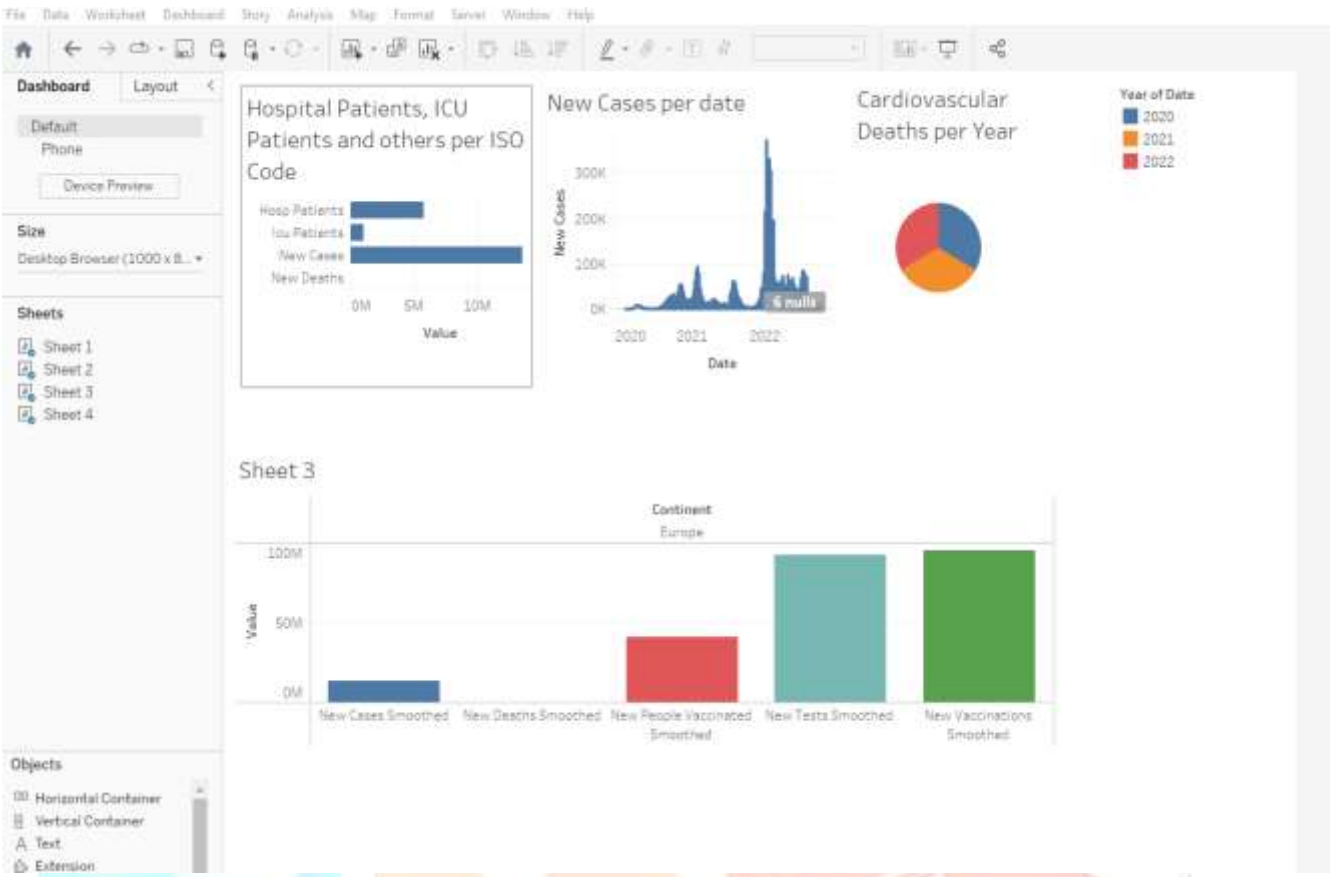


Steps to Integrate SAP Datasphere with Tableau: -

1. After opening Tableau use ODBC Connection like below to connect to the Datasphere Space (Can be used wherever there is a possibility to connect with an external tool).



2. Now start consuming the View from datasphere into Tableau.



IV. COMPARING SAP DATASPHERE AND TRADITIONAL DATA WAREHOUSING

To highlight the fundamental differences between SAP Datasphere and traditional data warehouse architectures,

Feature	Traditional Data Warehouse	SAP Datasphere
Architecture	Centralized, rigid schema	Cloud-native, flexible, and decentralized
Data Processing	Batch-based	Real-time, federated queries
Data Storage	On-premises or limited cloud	Scalable cloud storage with virtualization
Schema Management	Static, requires changes for new data sources	Flexible, schema-less for virtualized data
Integration	Limited integration, delayed reporting	Seamless integration with BI tools, real-time access
Maintenance Costs	High, requires dedicated infrastructure	Low, fully managed in the cloud
Scalability	Limited, costly for large-scale environments	Elastic, on-demand scalability

Table 1: Presents a comprehensive comparison across several critical dimensions.

Key Advantages of SAP Datasphere: -

Real-Time Insights: SAP Datasphere's capability to deliver real-time data access without the need for ETL processes provides significant advantages in industries where timely decision-making is crucial.

Cost Efficiency: By minimizing the necessity for data replication and centralization, SAP Datasphere significantly reduces the infrastructure and storage costs typically associated with traditional data warehousing.

Agility and Flexibility: The decentralized architecture of SAP Datasphere allows organizations to swiftly adapt to new data sources and evolving business requirements.

Enhanced Collaboration: With real-time data access and seamless integration with visualization tools, SAP Datasphere fosters collaboration across departments, enabling cross-functional teams to make more effective data-driven decisions.

V. REAL-WORLD USE CASES AND APPLICATION

A. Supply Chain Management: -

In the manufacturing and logistics sectors, real-time data is critical for optimizing supply chains. SAP Datasphere enables manufacturers to integrate data from IoT devices, Enterprise Resource Plannings (ERP), and external logistics providers into a single platform, providing real-time visibility into the entire supply chain. Using tools like SAC, stakeholders can visualize key performance indicators (KPIs), such as inventory levels, order fulfilment rates, and delivery times, enabling them to respond quickly to disruptions and optimize operations.

B. Financial Planning and Analysis: -

For financial institutions, SAP Datasphere offers the ability to integrate financial data from multiple systems and provide real-time access to financial planners. By leveraging SAC's planning and forecasting capabilities, organizations can create dynamic financial models that are continuously updated with the latest data, enabling more accurate forecasting and strategic planning.

C. Healthcare Data Integration: -

In healthcare, real-time access to patient data and operational metrics can be lifesaving. SAP Datasphere enables healthcare providers to integrate data from electronic health records (EHRs), medical devices, and external healthcare networks. Visualization tools like Power BI allow healthcare professionals to monitor patient outcomes and hospital performance in real time, improving decision-making and patient care.

VI. CONCLUSION

The swift advancement of digital transformation has introduced both challenges and opportunities for enterprises aiming to leverage the power of data. While traditional data warehouses have been instrumental in providing structured data for historical analysis, they increasingly fall short of meeting the needs of modern businesses that demand real-time insights, flexibility, and scalability. SAP Datasphere presents a contemporary, cloud-native solution that effectively addresses these challenges through real-time data access, decentralized integration, and seamless collaboration with data visualization tools.

By enabling organizations to dismantle data silos and democratize data access, SAP Datasphere empowers cross-functional teams to make real-time, data-driven decisions, enhancing business agility and outcomes. As organizations continue to adapt in the digital era, platforms like SAP Datasphere will be essential in reshaping how data is managed, visualized, and utilized for strategic advantage.

REFERENCES

[1] Schemas used in data warehouses: Star, Galaxy, and Snowflake.
Schemas used in data warehouses: Star, Galaxy, and Snowflake (softwareag.com)
[2] The Data Warehouse: From the Past to the Present.
The Data Warehouse: From the Past to the Present - DATAVERSITY
[3] A comparison of data warehouse design approaches: Kimball and Inmon.
A comparison of data warehouse design approaches: Kimball and Inmon | by Rahma Hassan | Medium
[4] SAP Innovation Awards 2024 - 4362 –Bright speed
SAP Innovation Awards 2024 - 4362 - Bright speed
[5] Enable Category Management with Big Query and SAP Datasphere
SAP Discovery Center (cloud.sap)