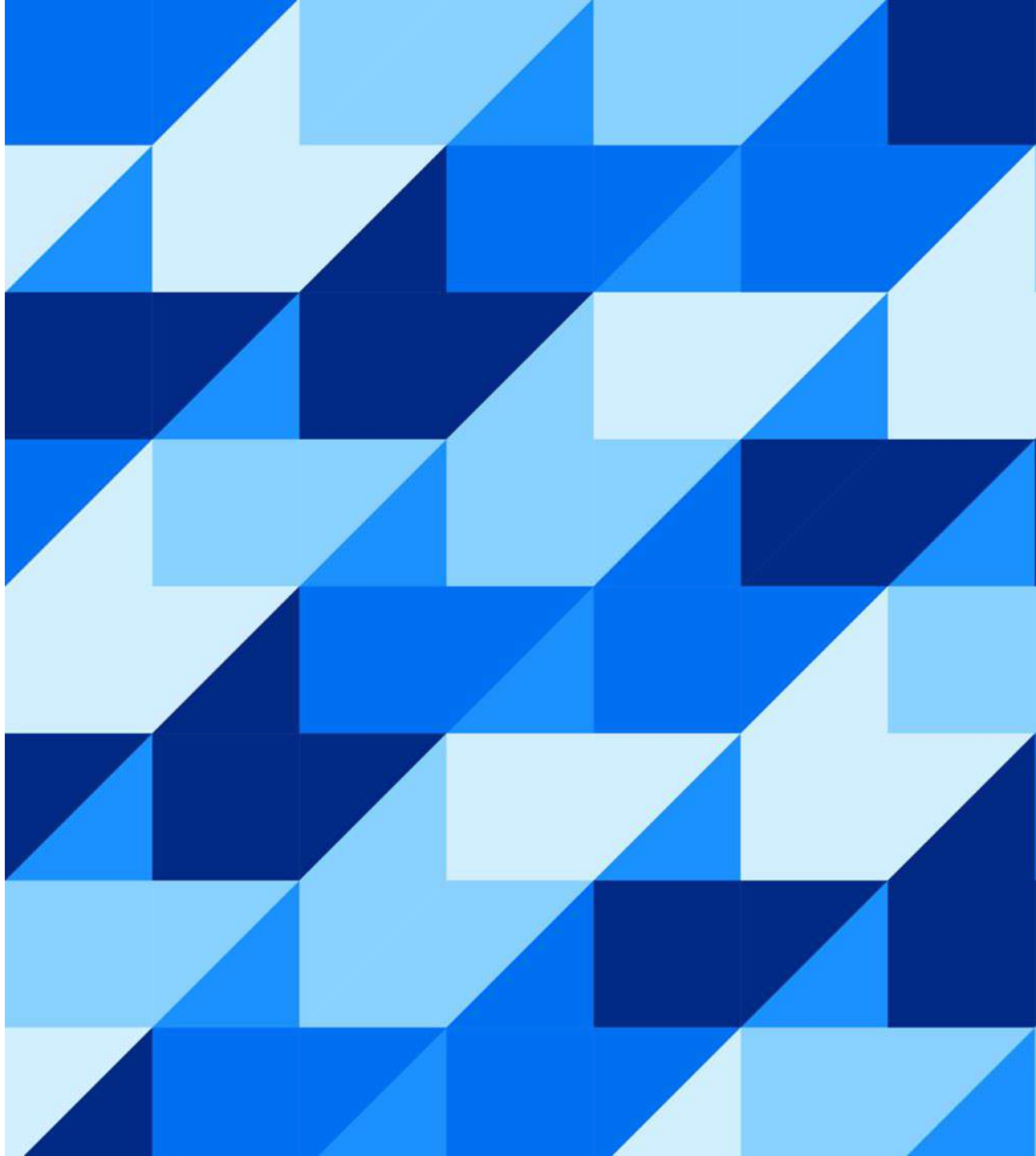


SAP Datasphere Unleash the power of business data

Speaker Name, SAP
Date

Public

Data needs context for authoritative decision-making

Innovation has happened in silos

Data, especially metadata, has been removed from its context

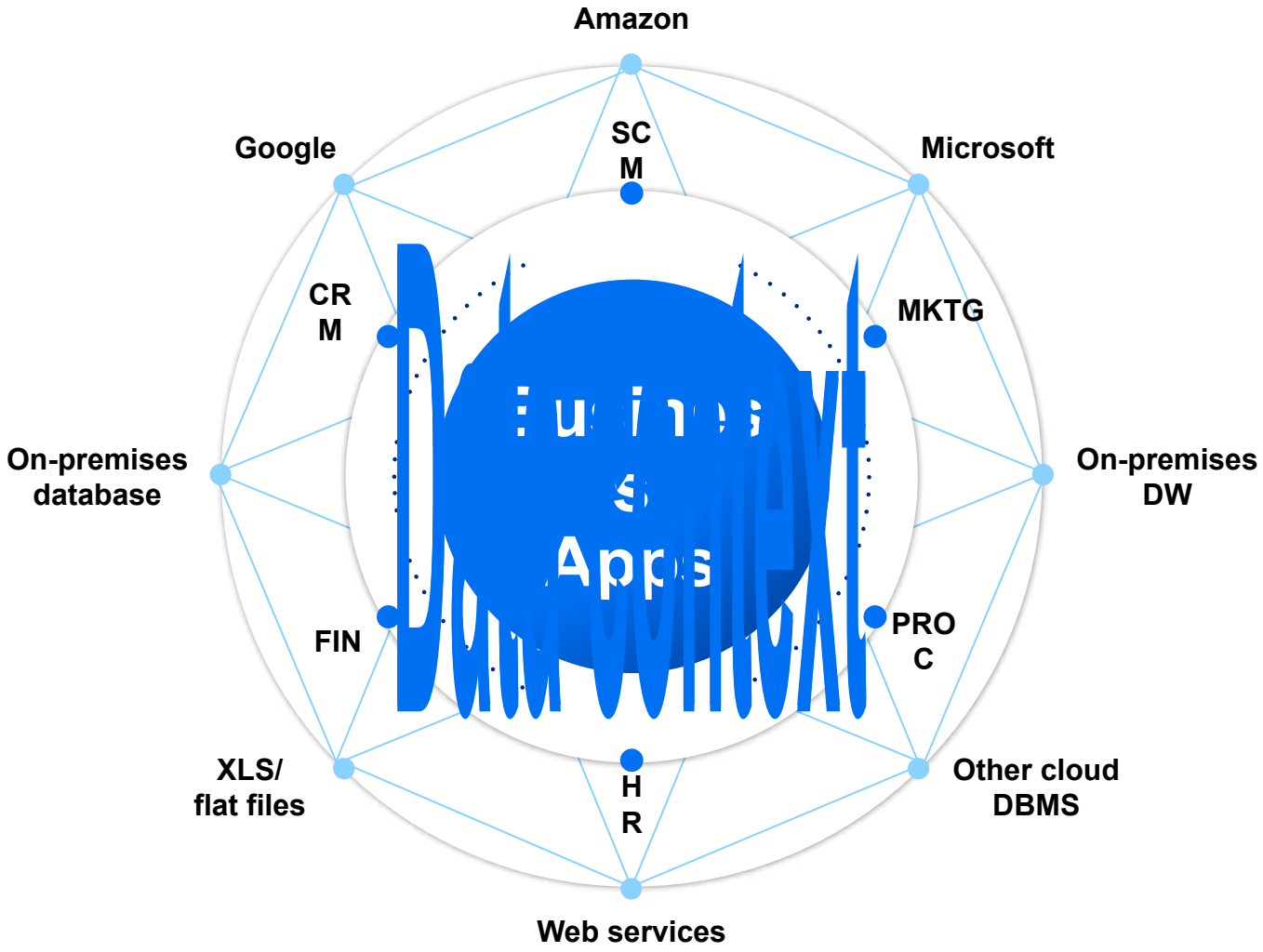
There's been a fundamental error—focusing solely on technology rather than the data

(Multi-) cloud brings new opportunities and challenges

New data sets spawn new projects

On-premises sources remain

Complexity has increased, shifting governance and integration work back to customers



Universal yet opposing data needs for business and technology stakeholders

Technology Needs

Centralized governance (data, financial, privacy and compliance)

Simplification of the ever-growing data landscape

Business Needs

Self-service access to trusted business data—delivered in business terms

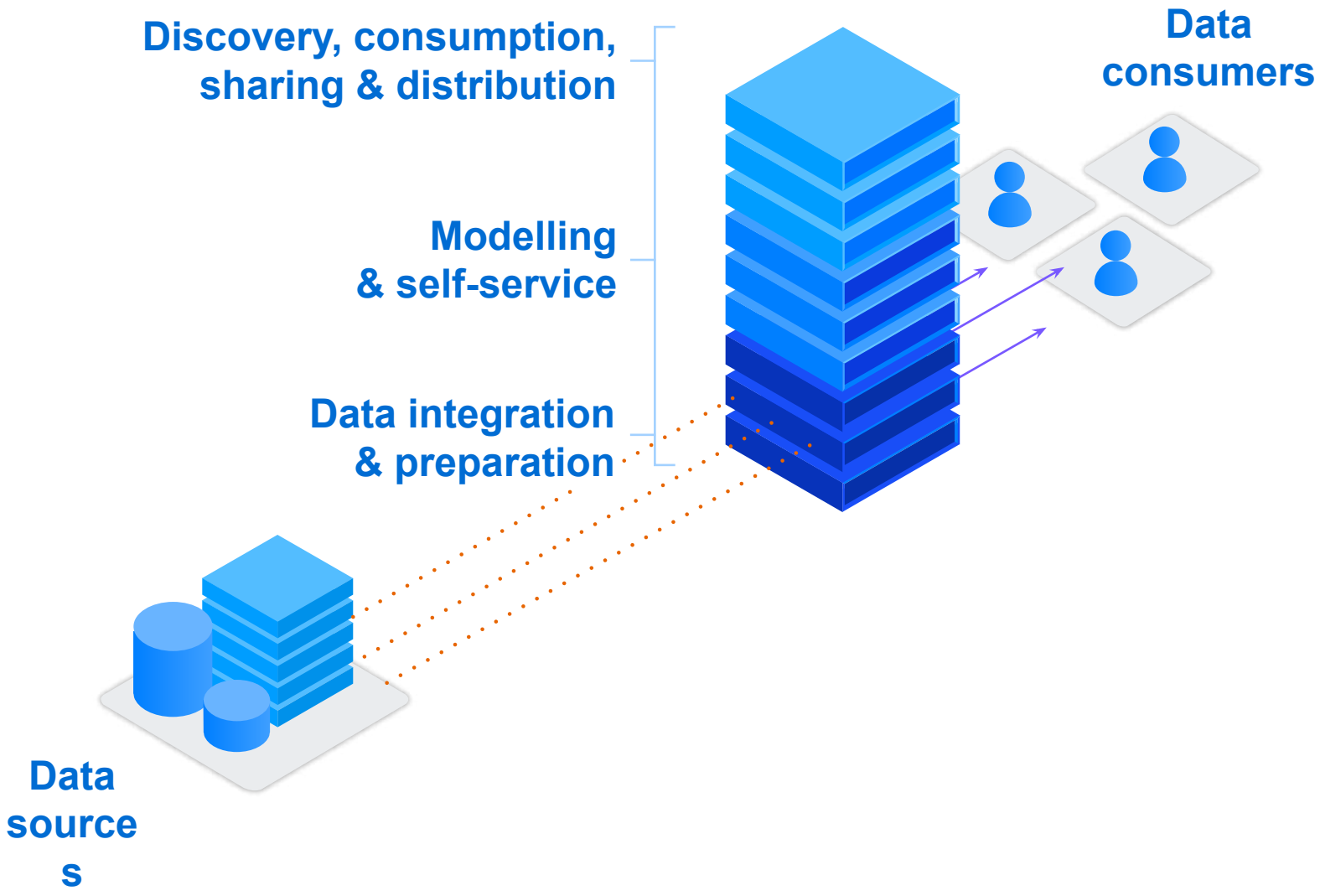
Real-time data to support business apps, insights, and analytics across hybrid and multi-cloud

Towards a business data fabric with SAP Datasphere

Data Products

Agile Governance

Business Semantics



SAP Datasphere is the foundation for a business data fabric architecture

Extended planning and analysis with SAP Analytics Cloud

Enterprise planning | Business simulations | Analytics and BI

Intelligent data applications with SAP HANA Cloud

Insights-driven transactions | Multi-model database | GenAI

Business data fabric enabled by SAP Datasphere

Security

Access control

Availability

Discovery, consumption, sharing & self-service
space, data product, marketplace

Governance
cataloging, metadata management, lineage, privacy, quality

Business semantics
semantic onboarding, entity import, business content, knowledge graph

Modelling
relational & rich analytic models, lifecycle management, authorizations

Data integration & preparation
real-time replication & federation, transformation, scheduling, monitoring

SAP and non-SAP data

Applications

On-premise data warehouses

Cloud data warehouses and lakehouses

relational databases

Unstructured and semi-structured data

Why a business data fabric

- **Self-service access** to trusted data breeds agility and accelerated, accurate decisions
- **Comprehensive data governance** assures every stakeholder that private data stays private
- **Real-time data**, infused throughout your data architecture helps business users, partners, and employees make in-the-moment improvements
- **A simplified data landscape** reduces costs and technical debt while maximizing current investments

SAP Datasphere provides capabilities for a complete business data fabric architecture

Analytics & planning consumption

Powerful end-to-end analytics & planning integration with SAP Analytics Cloud & Microsoft Office

Discovery, consumption, sharing & self-service

space, data product, marketplace

Governance

cataloging, metadata management, lineage, privacy

Business semantics

semantic onboarding, entity import, business content

Modeling

relational & rich analytic models, lifecycle management, authorizations

Data integration & preparation

real-time replication & federation, transformation, scheduling, monitoring

SAP BW move & hybrid

Leverage investment and extend on-prem scenarios with SAP BW Bridge and hybrid landscapes

Open data ecosystem partners

databricks

CONFLUENT

Collibra

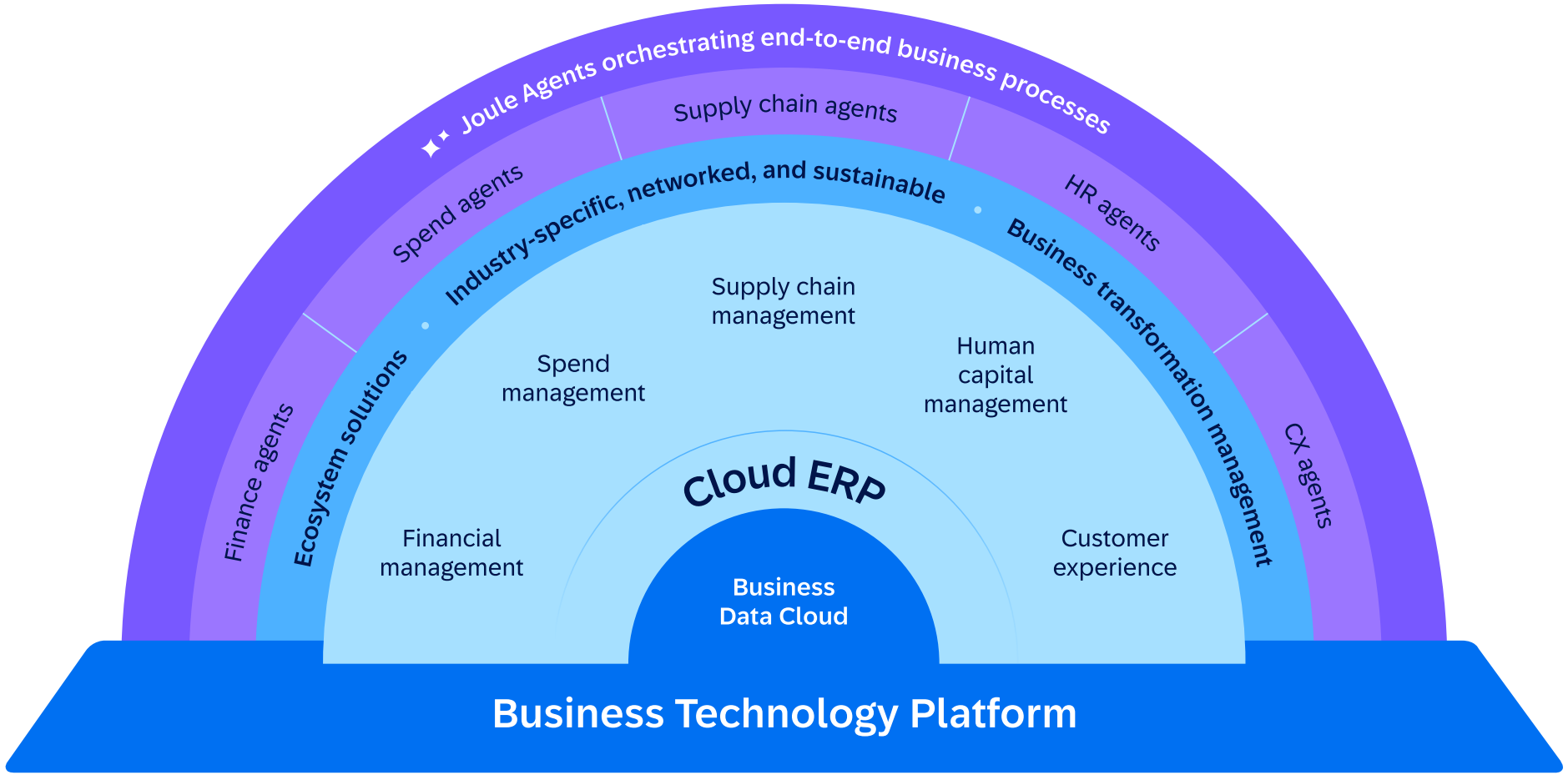
DataRobot

Google Cloud

Overview

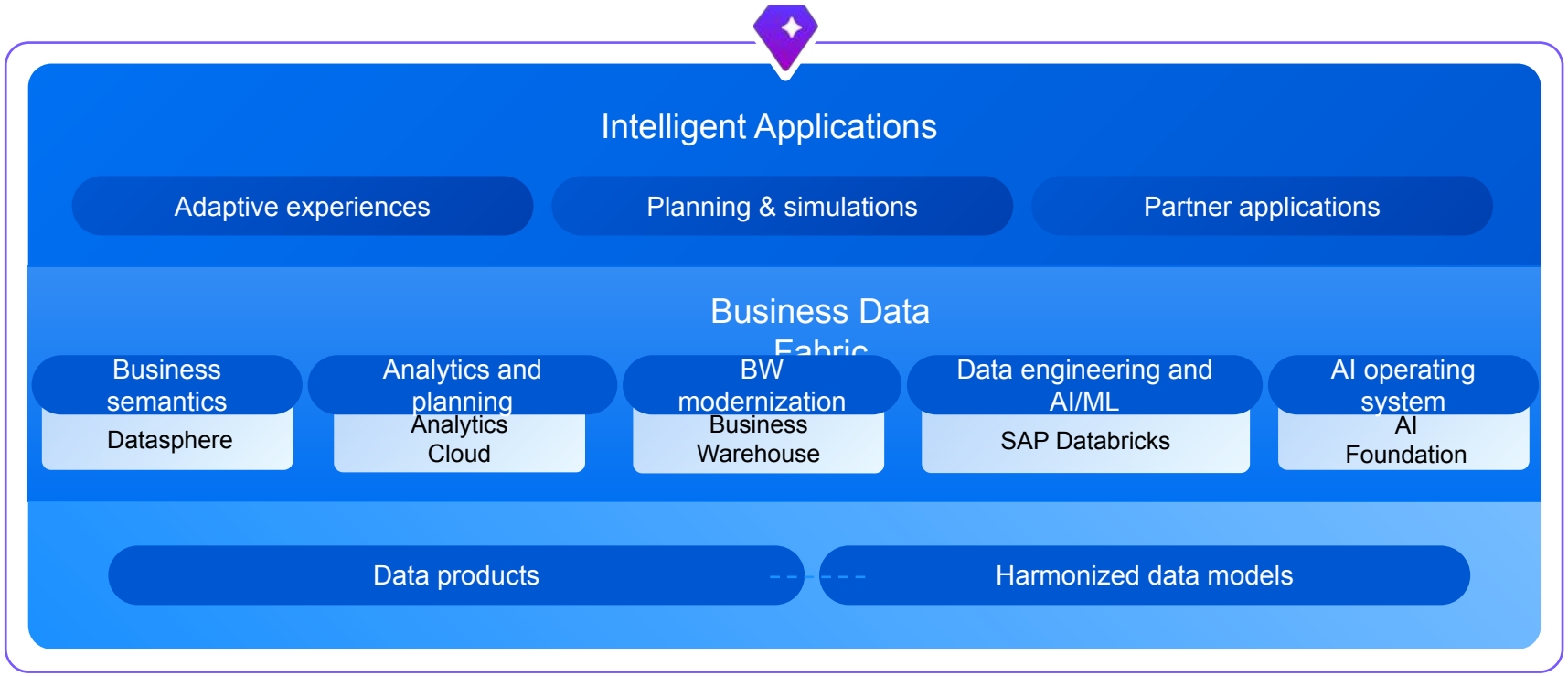
SAP Business Suite

Turns our vision into reality



SAP Business Data Cloud

Continue our path towards a Business Data Fabric



Evolution of the Business Data Fabric journey which started in 2023 with SAP Datasphere and SAP Analytics Cloud powered by SAP HANA Cloud enhanced by SAP Databricks and SAP BW

Next pivotal step in SAP's data strategy on delivering out-of-the-box always on data products following a harmonized data model to provide scalable access without duplication to reduce time to value

Adapting to the rapidly evolving needs by adding more services supporting use cases such as generative AI, data products with rich semantics, and analytical intelligent applications

SAP Apps

Hybrid • SaaS • Cloud-native • On-premises • Edge

Non-SAP Data

SAP data and analytics solutions

Maximize the value of all your data across SAP and third-party systems

Extended planning and analysis

Take enterprise planning to the next level with an optimized, vertically integrated consumption layer for planning and analytics workloads.

SAP Analytics Cloud

Business data fabric

Deliver an integrated, semantically rich data layer over underlying data landscapes to provide smooth and scalable access to data without duplication.

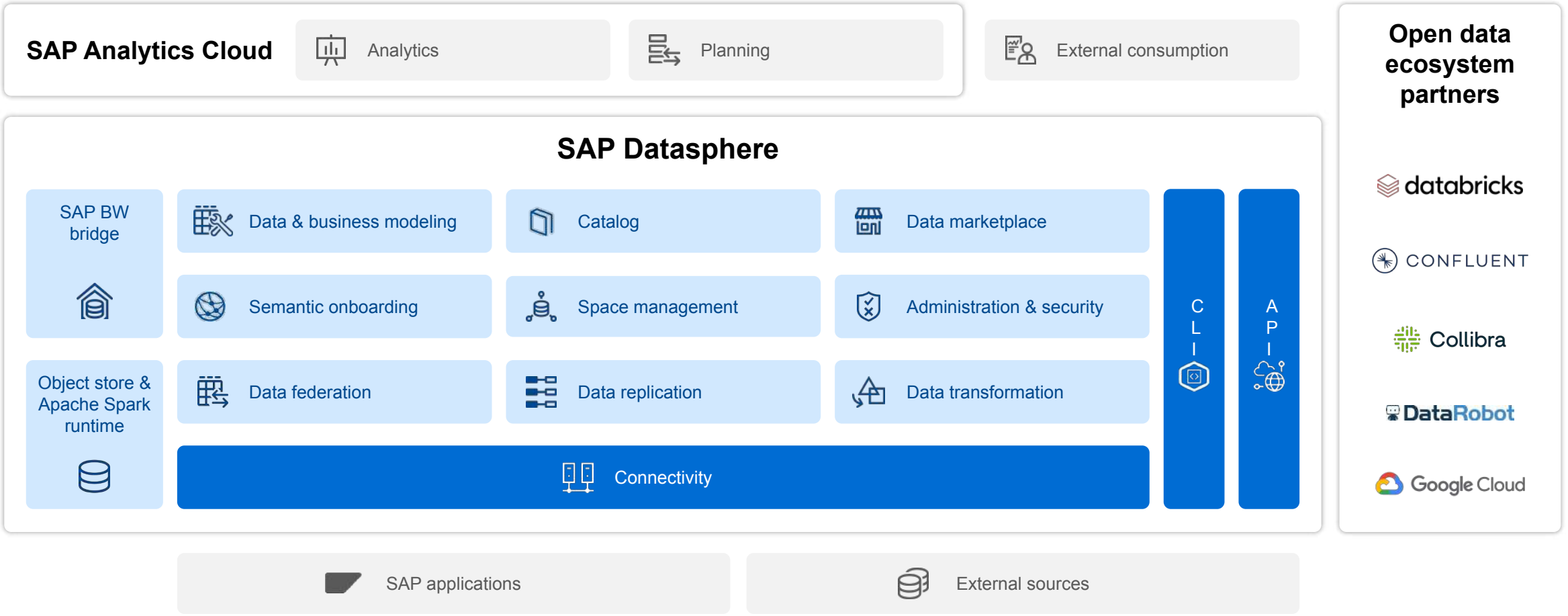
SAP Datasphere

Intelligent data applications

Build apps leveraging analytical capabilities (machine learning, graph, spatial, vector DB), combining historical and real-time data.

SAP HANA Cloud

SAP Datasphere overview



Data integration and preparation

Different ways to connect data to SAP Datasphere

- Built-in connectors
- Data marketplace
- Open connectors
- SAP BW bridge
- External tooling

Connection overview

SAP applications

- SAP S/4HANA
- SAP S/4HANA Cloud
- SAP ECC
- SAP ABAP
- SAP BW/4HANA Model Transfer
- SAP BW
- SAP HANA Data Lake
- SAP HANA, Data Lake Files
- SAP SuccessFactors
- SAP Fieldglass
- SAP Marketing Cloud

Databases & generic connections

- SAP HANA on-prem & Cloud
- Oracle
- Microsoft SQL Server
- Generic JDBC
- Generic OData
- Generic SFTP
- Local CSV File
- Generic Cloud Data Integration
- Confluent
- Adverity
- Precog

Hyperscaler & cloud apps

- Google Cloud Storage
- Google BigQuery
- Azure SQL Database
- Azure Blob Storage
- Azure Data Lake Store Gen1 (deprecated)
- Azure Data Lake Store Gen2
- Amazon S3
- Amazon Athena
- Amazon Redshift
- HDFS
- SAP Open Connectors
- SAP HANA, Data Lake Files

Outbound targets

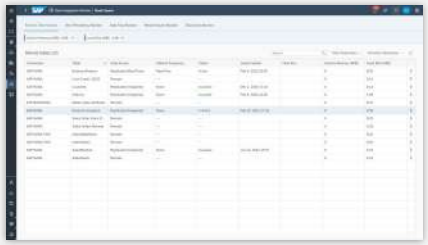
- SAP Datasphere
- SAP HANA on-prem & Cloud
- SAP HANA, Data Lake Files
- Amazon S3
- Azure Data Lake Store Gen2
- Google Cloud Storage
- Google BigQuery
- Apache Kafka
- Confluent
- Generic SFTP

SAP Help:

Federation and replication options

Remote tables

Virtual Access



Remote table federation

- Leave data in the source system and access remotely in real-time when needed
- No upfront data movement
- Federation is supported across various sources and hyperscalers

Replication

Remote table replication

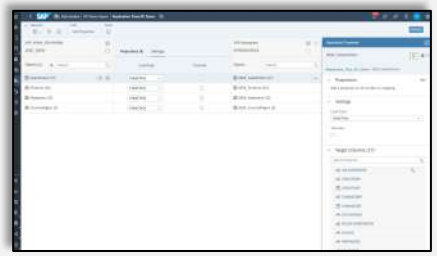
- Real-time replication or snapshots using single entities
- Schedule via task chains

View persistence

- Materialize view output results in a stable persistence
- Schedule via task chains

Cloud-native data integration

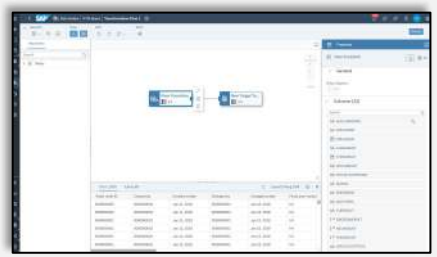
Replication flow



Mass data replication

- Data replication with multiple entities from a variety of sources
- Flexible targets to Datasphere and outbound target systems
- Support initial and delta replications from various sources

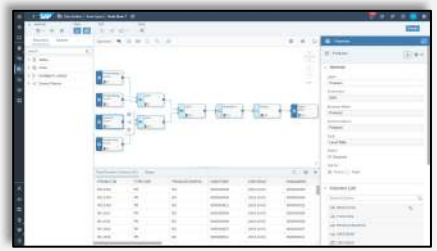
Transformation flow



Cloud-based data transformations

- Load data from one or more source tables and apply transformations (e.g. SQL, field-based)
- Store output results in a local target table
- Supports to process delta changes

Data flow



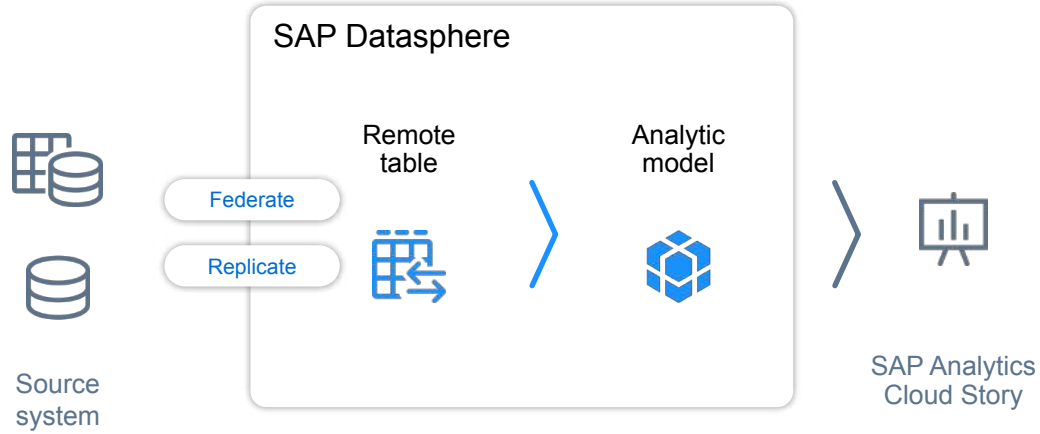
Batch loads & transformations

- Combine structured and semi-structured data while defining ETL processes
- Advanced transformation capabilities leveraging Python 3
- Schedule via task chains

Data integration using remote tables

Remote table federation and replication

- A remote table points to data in an external system without copying the data
- Handling remote tables is similar to local tables, however the data is transferred through the network each time a query is executed
- Remote data access can be restricted using central filters and selected columns only*
- Transparently switch from remote access to snapshots or real-time replication for change data capture (CDC) enabled tables without the need to change the data models
- Schedule snapshot loading for single remote tables, orchestrated via task chains (optional)
- Ability to start & stop, pause & resume, and cancel real-time replication
- Optional partitions to split larger data transfers and execute these single transactions in parallel

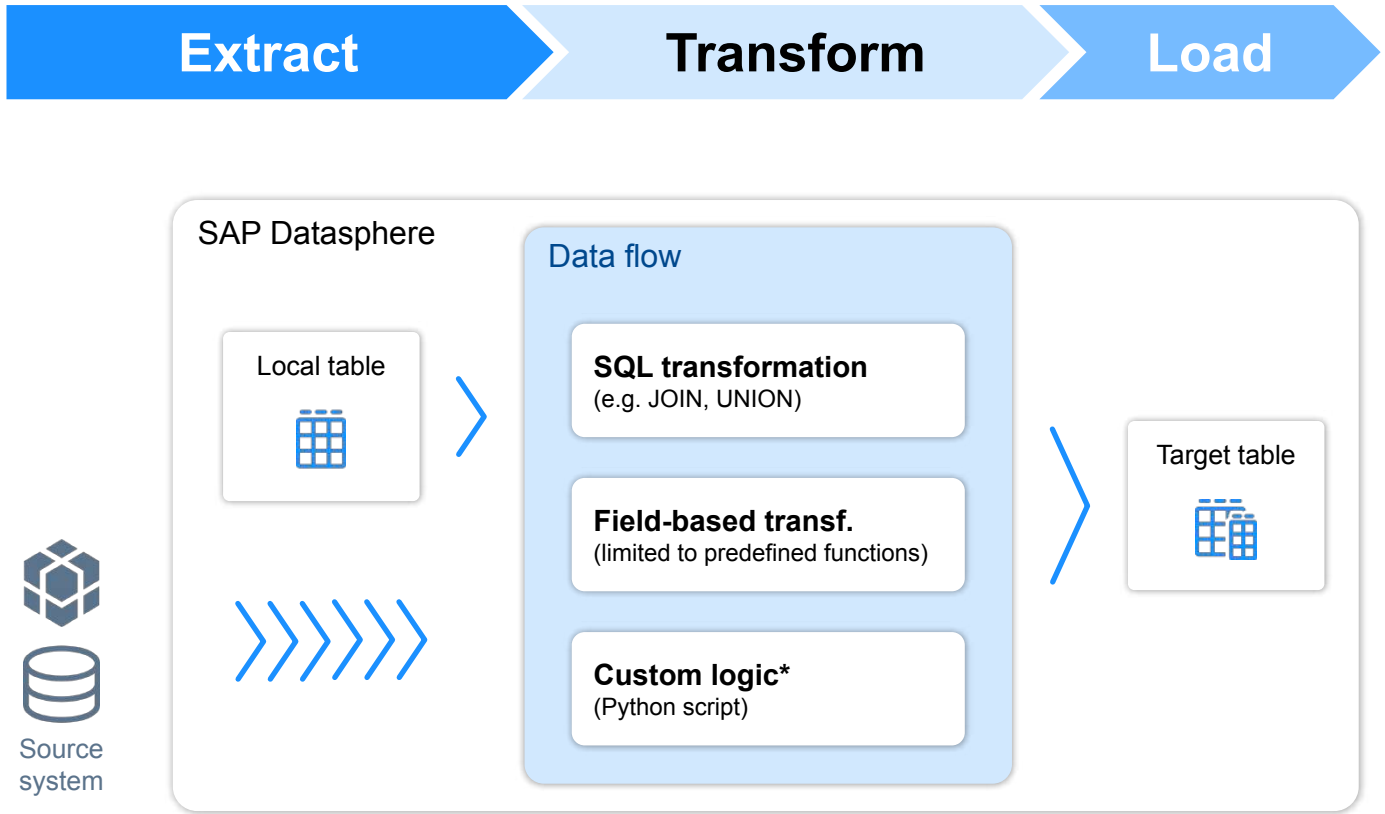


* Filtering depends on source connection and column data types. See more information in the online documentation.

Data integration using data flow

Extract, transform and load data using data flow

- Use Dataflow to **extract**/read data from external sources or from a local table in SAP Datasphere
- Perform some **transformations** on the data using the existing operators as well as the limited set of built-in SQL functions
- Finally **load** the result into a target table in Datasphere without any delta capability
- Supports Python version switch from 3.9 to 3.11
  (migration recommended, as 3.9 support ends in Oct. 2025 to avoid disruption)



* Limited to NumPy and Pandas

Data integration using replication flow

Replication flows offer a cloud-based replication tool in which cloud-to-cloud based replication scenarios for on-premise components are not needed anymore (for example, data provisioning agent).

Mass data replication

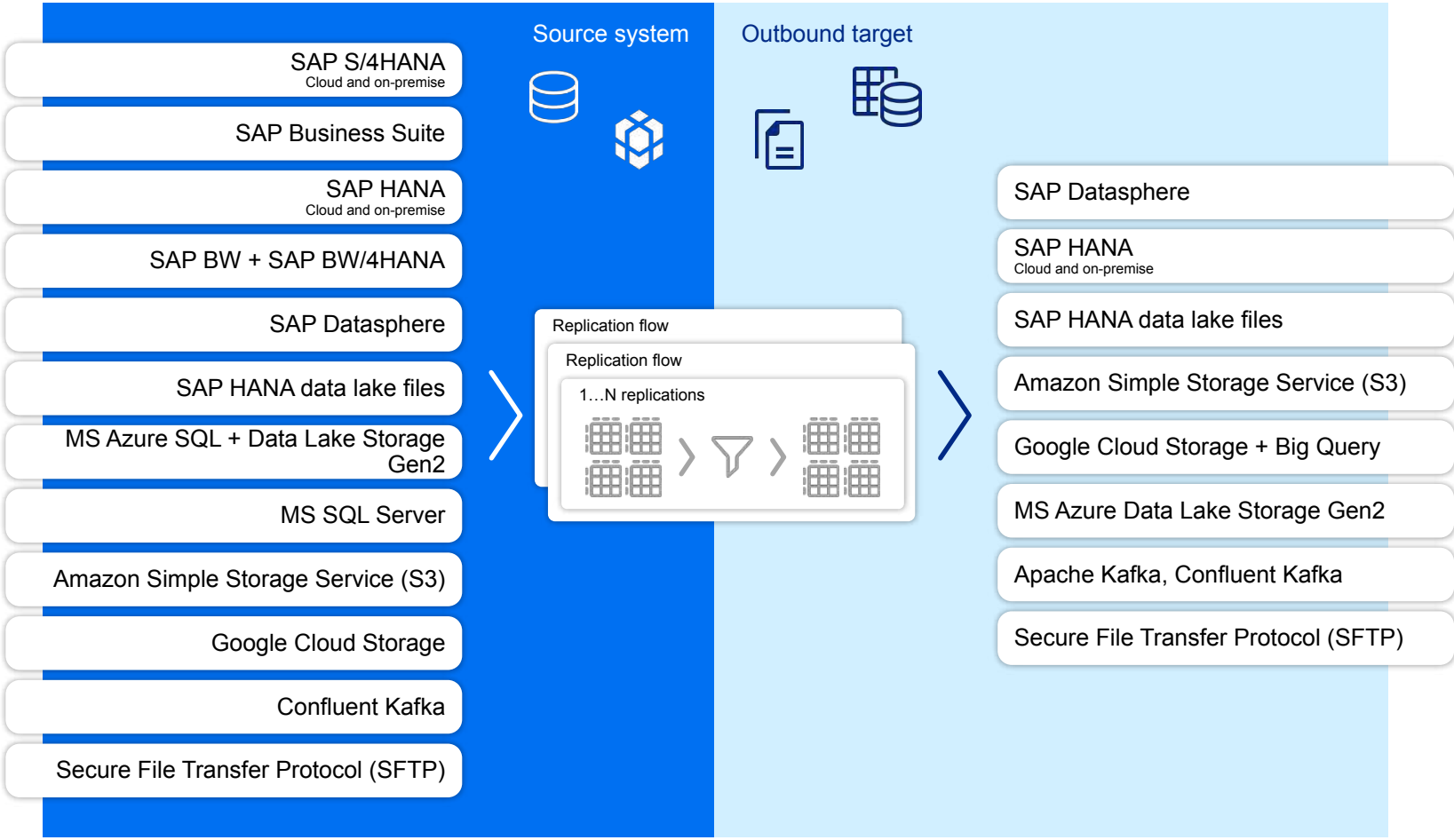
Possibility to replicate **parallelized** data from supported **source systems** into supported **target systems**

Integrated monitoring

Overview of monitoring features within **integrated** monitoring application

Capture changes

Support for **initial** data replications together with **delta*** replications

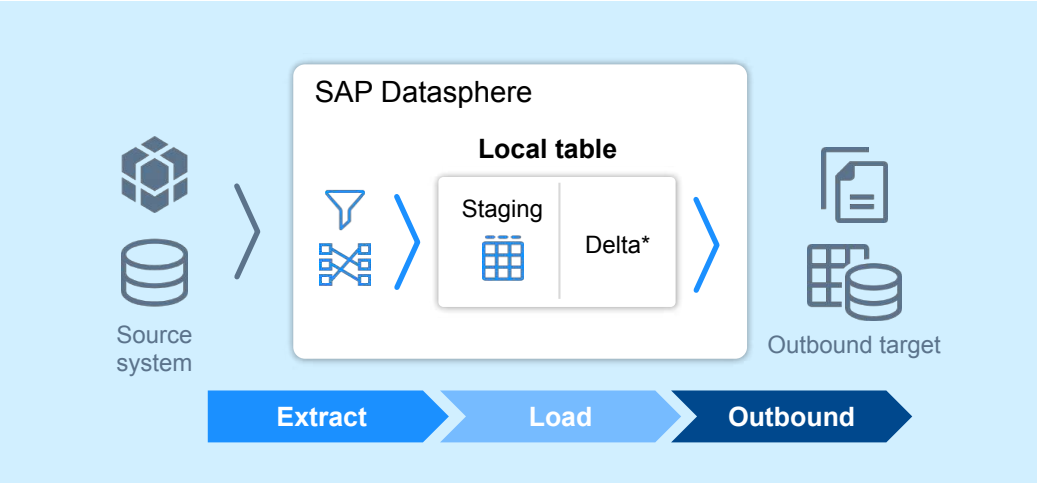


* Restrictions to delta capture apply: more information in the online documentation.

Data integration using replication flow

Simplify the realization of data replication use cases using replication flows

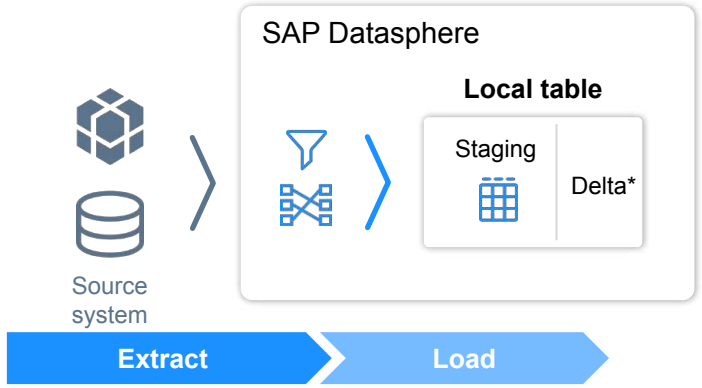
- **Model data replication** flows directly in the data builder
- Apply column projections and row-level filters in the replication flow
- Enable **cloud-to-cloud replication** without on-premise components
- Design **mass data replications** using a dedicated, embedded interface and allow the reuse of the same source objects in multiple flows
- Support **Initial Only, Delta Only** and **Initial & Delta** load types with **change data capture** mechanisms
- Accelerate initial loads through partition-based **parallelization**
- Ensure data **resiliency** with automated **error recovery** and **retry** logic
- Support HANA SQL and Script views as sources, enabling initial load extraction using semantic keys when needed across HANA Cloud, On-premise, and Datasphere



* Restrictions to delta capture apply: more information in the online documentation.

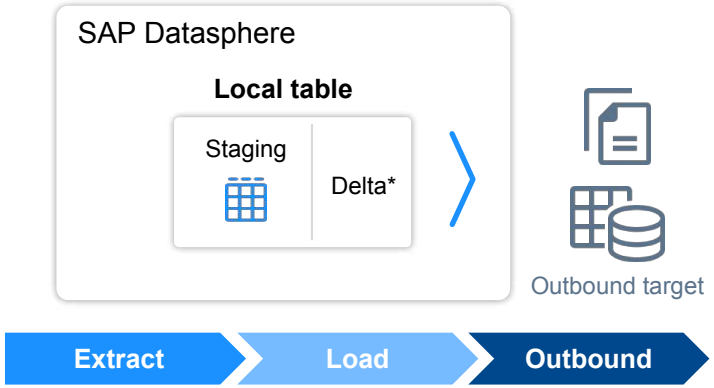
Replication flow use cases

1 Inbound data movement



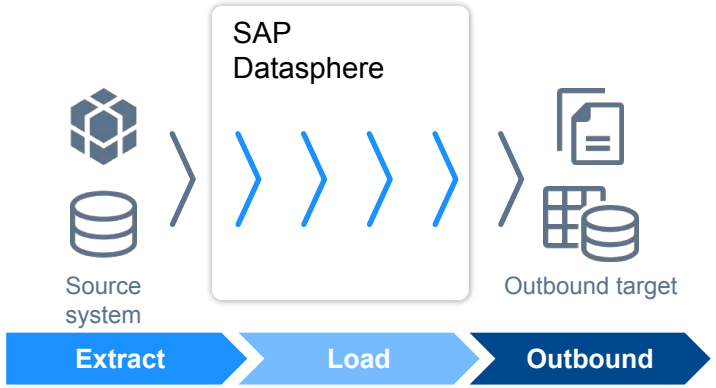
Replicate data from SAP or non-SAP sources into SAP Datasphere and capture data changes occurring near-real-time from the source to update the target table

2 Outbound data movement



Read data from SAP Datasphere and replicate it into supported SAP and non-SAP outbound targets while tracking data changes*

3 Pass-through option



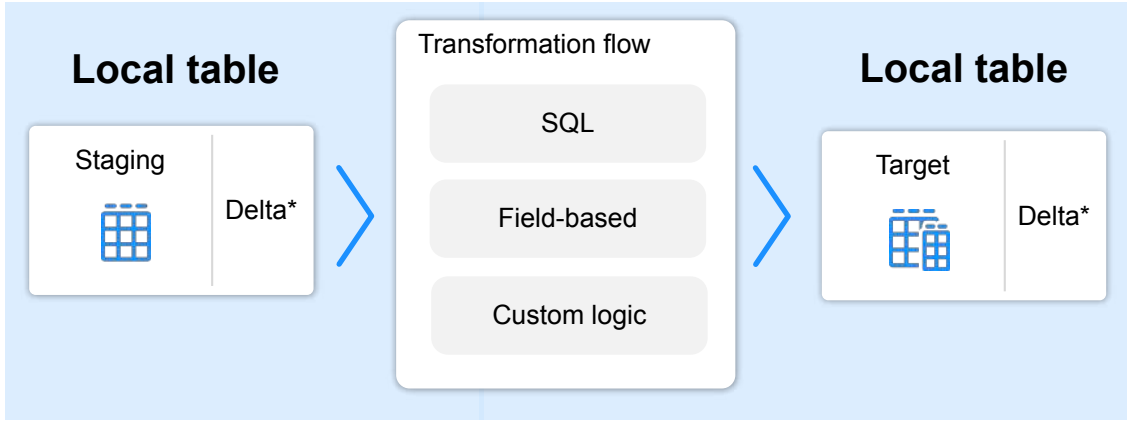
Use SAP Datasphere as a middleware and replicate data from supported sources into supported targets while tracking data changes* occurring near-real-time in that source

* Restrictions to delta capture apply: more information in the online documentation.

Data integration using transformation flow

Perform SQL-based transformations on delta and non-delta data

- **Delta staging*** in SAP Datasphere to enable SQL transformations, delta writing for multilevel staging, and storage of holistic truth after transformation
- **Support for initial and delta*** loading from delta tables
- **Schedulable** through task chains and integrated in data integration monitor
- **Support** various **SQL-based transformations** (for example, joins, aggregations, functions, calculated columns ...)
- Use **Python script operator** for transformation flow in file spaces
- Optional **Batch processing** when using SAP HANA Runtime
- Run mode setting for **performance and memory optimized consumption**
- **Simulate runs** and generate PlanViz files
- Generate **HANA ExplainPlan** for Transformation Flows using HANA runtime directly from the Data Integration Monitor to simplify performance analysis

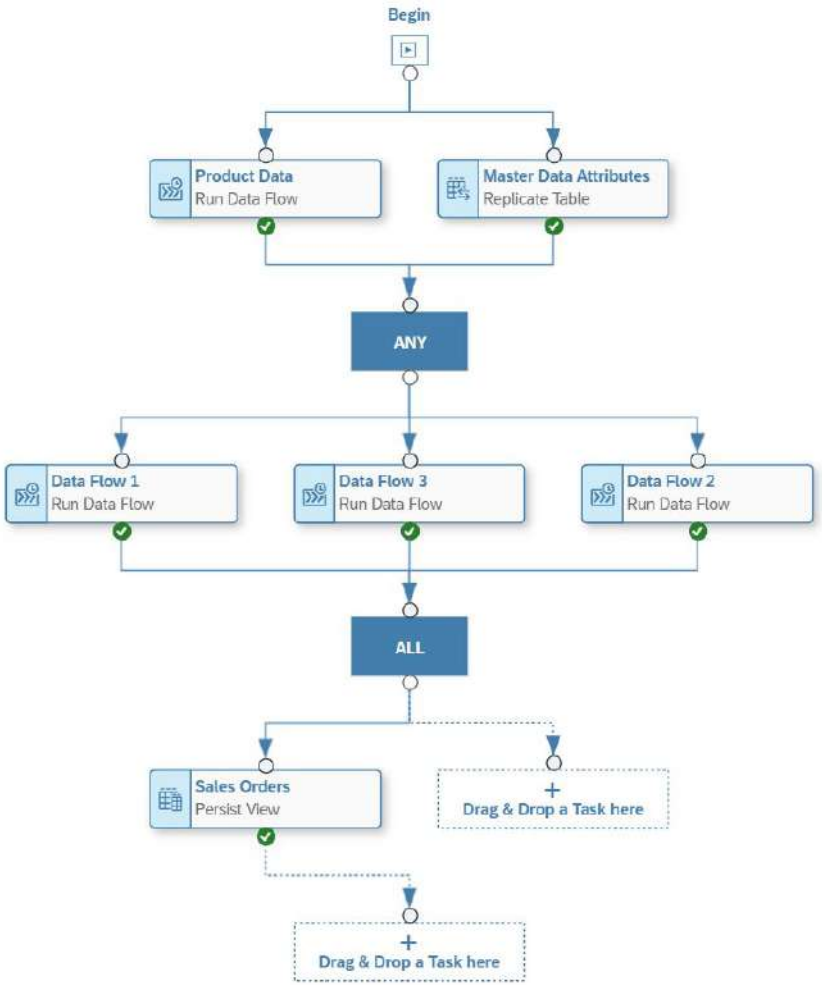


* Restrictions to delta capture apply: more information in the online documentation.

Task chains and scheduling

Add various tasks into a task chain and run them manually once, or periodically through a schedule

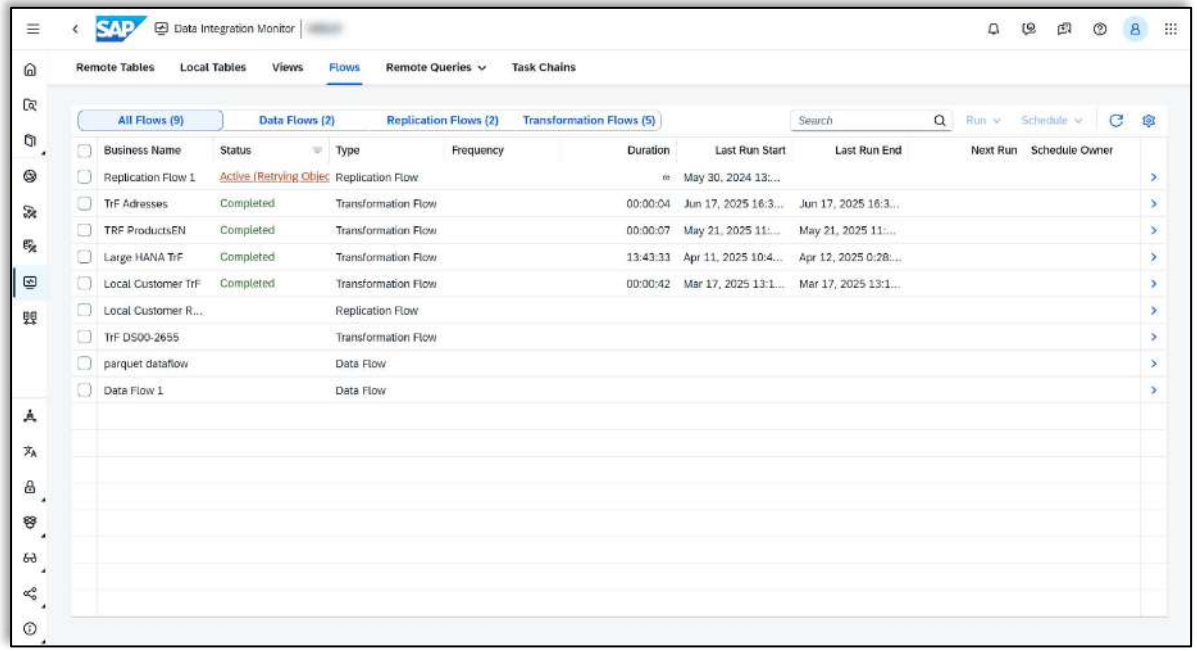
- Tasks chain scheduling may include execution of
  - remote table replication, view persistence, intelligent lookup,
  - other task chains (nested) and shared task chains,
  - call external (RESTful) APIs via HTTP(s) connections,
  - SQL procedures (open SQL schema), notification tasks,
  - remove replicated remote table data or persisted view data,
  - data flow, replication flow (initial only), and
  - transformation flow runs
- Serial & parallel processing supported with ALL or ANY operators and customizable email notification of success and/or failure of entire task chains
- Build global task chains using cross-space sharing
- Export and import of task chains CSN/JSON file definitions
- Create, manage, pause, and run schedules using simple time settings or advanced CRON expressions, with the ability to retry failed tasks within nested chains (deep retry)
- Scheduled tasks are executed asynchronously and recurrently in the background according to the defined schedule (frequency: minutes, hourly, daily, weekly, monthly). Tasks can be scheduled according to a preferred time zone with automatic daylight saving adjustments
- Browse and trigger BW Bridge process chains directly within Datasphere enabling seamless integration between task chains and BW Bridge processes
- More information in SAP note FAQ: 3018945



Data integration monitoring

Central monitoring and scheduling for remote tables, flows, views, remote queries & task chains

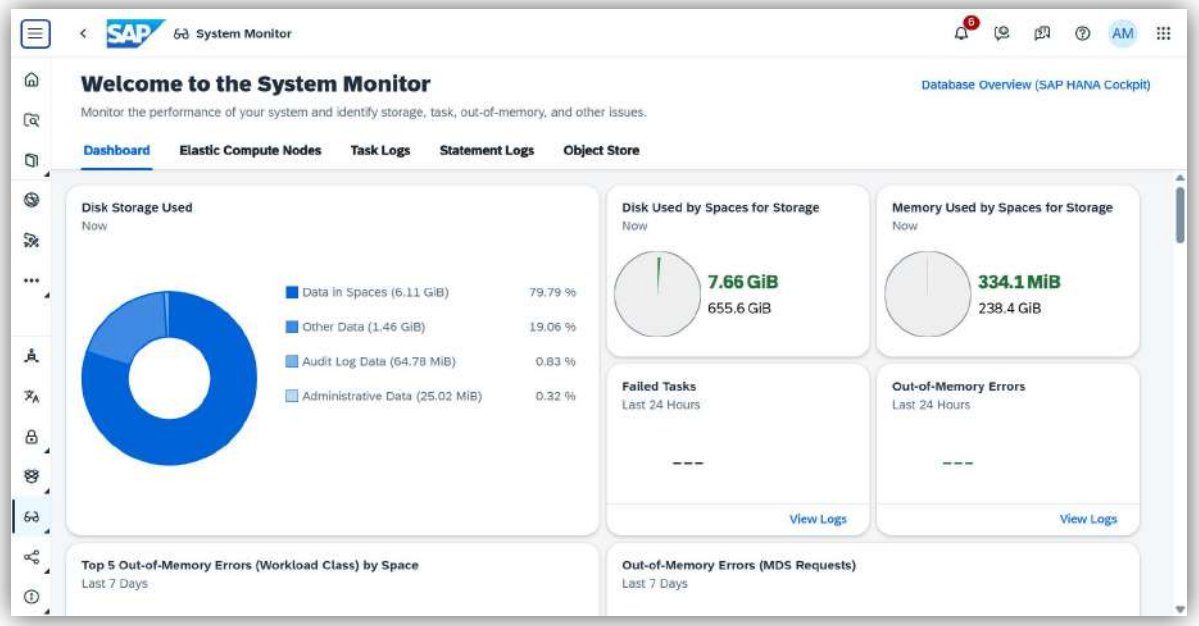
- **Remote tables**: Ability to start, stop, pause, resume and cancel real-time replication for change data capture (CDC) enabled tables for single and multiple objects and switch seamlessly between remote access, real-time replication and snapshots
- **Local tables**: Monitor local tables and check their metrics, delete records on-demand or scheduled using filter conditions
- **Views**: Add, monitor and schedule persisted views
- **Flows**: View and monitor the execution details of data flows, replication flows or transformation flows
- **Remote queries**: Check and analyze remote queries to the connected remote source systems and create statistics for your remote tables to improve federated query execution
- **Task chains**: Monitor the status and progress of running and past task chains
- **Spark Applications**: Default Spark application settings in HDLF spaces are now displayed in a dedicated table, laying the groundwork for future customization



System Monitor

Allows to monitor the system performance, under-stand storage breakdown, and identify task logs, out-of-memory, and other issues across all spaces

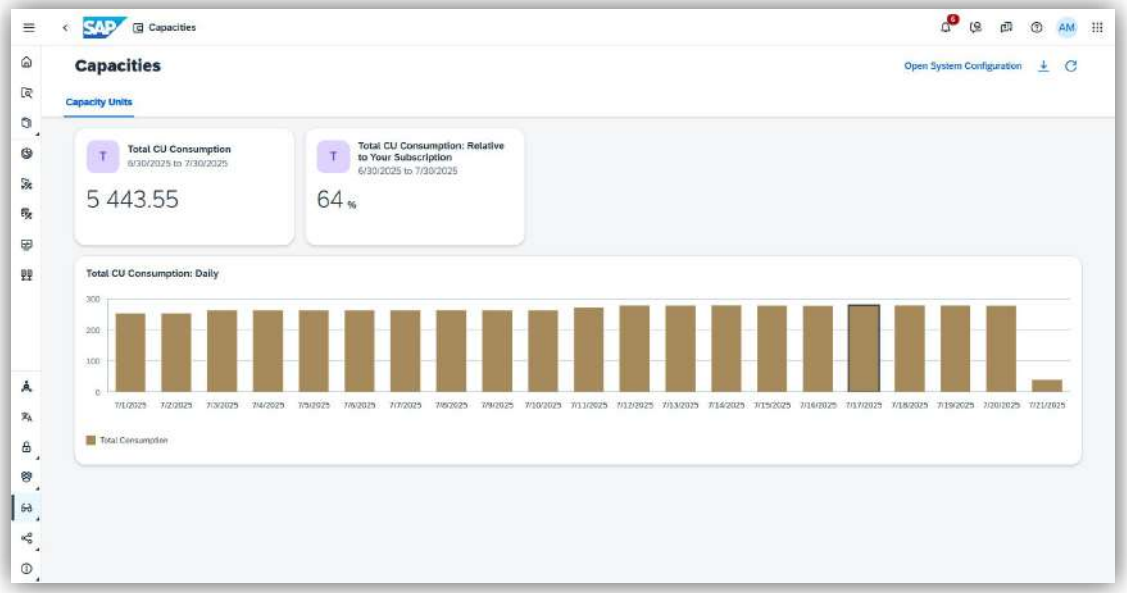
- Dashboard for a system wide overview
- Monitor the key figures for **Elastic Compute Nodes**, such as the start and end time of the last run or the amount of memory used for data replication
- Cancel long-running **Tasks** like Data Flows and Views directly from the System Monitor for improved execution control
- Enable expensive statement tracing by default in all SAP Datasphere tenants to capture key memory and performance diagnostics the **Statement Logs** tab
- Monitor HDLF storage and Spark usage in a new System Monitor tab for better visibility into **Object Store** resource consumption
- View monthly capacity unit consumption in the new **Capacities Dashboard**, accessible via System Monitor > Capacities



Consumed Capacity Unit Dashboard

Track daily and monthly capacity unit consumption to optimize resource allocation and manage subscriptions efficiently

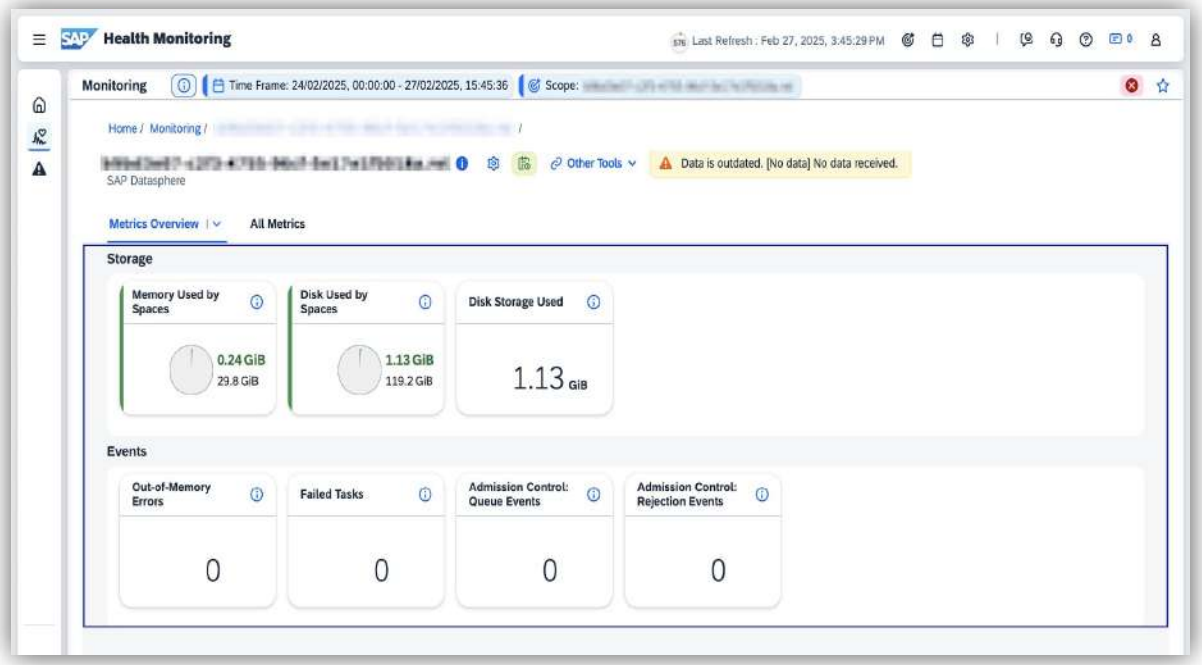
- Monitor total capacity usage for the current month to stay informed on overall consumption
- Track subscription utilization by viewing the percentage of consumed capacity relative to your plan
- Analyze daily trends with visual bar charts to identify usage patterns and peaks and identify Space and Object information when consumption is associated with a specific object
- Download hourly consumption data in CSV format for detailed, custom analysis and reporting



Health Monitoring

Monitor the health of connected cloud services and systems in SAP Cloud ALM with real-time metrics and historical insights.

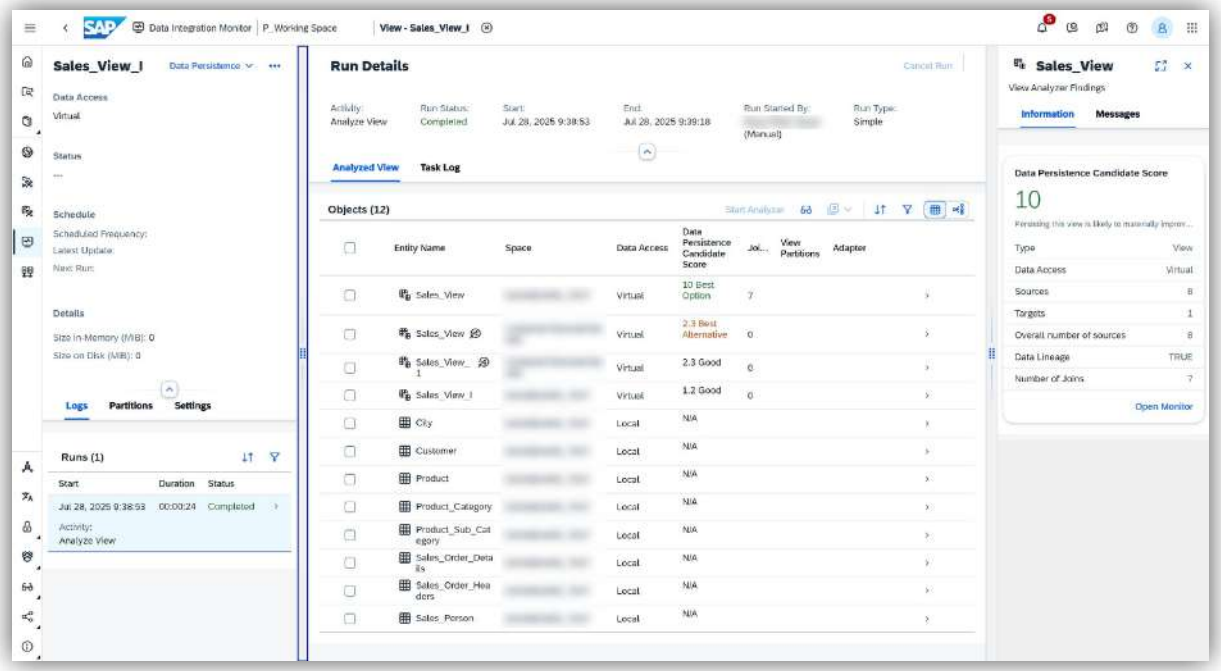
- Integrate SAP Datasphere with SAP Cloud ALM to centrally monitor system health, including memory, disk usage, task failures, and out-of-memory events
- View overall system health by service type, with visual indicators for warning and critical statuses
- Drill down into metric details for each service or system, including component-level tiles and current values
- Analyze historical data to identify trends and investigate past resource issues
- Access-controlled visibility ensures users see only the systems they're authorized to monitor



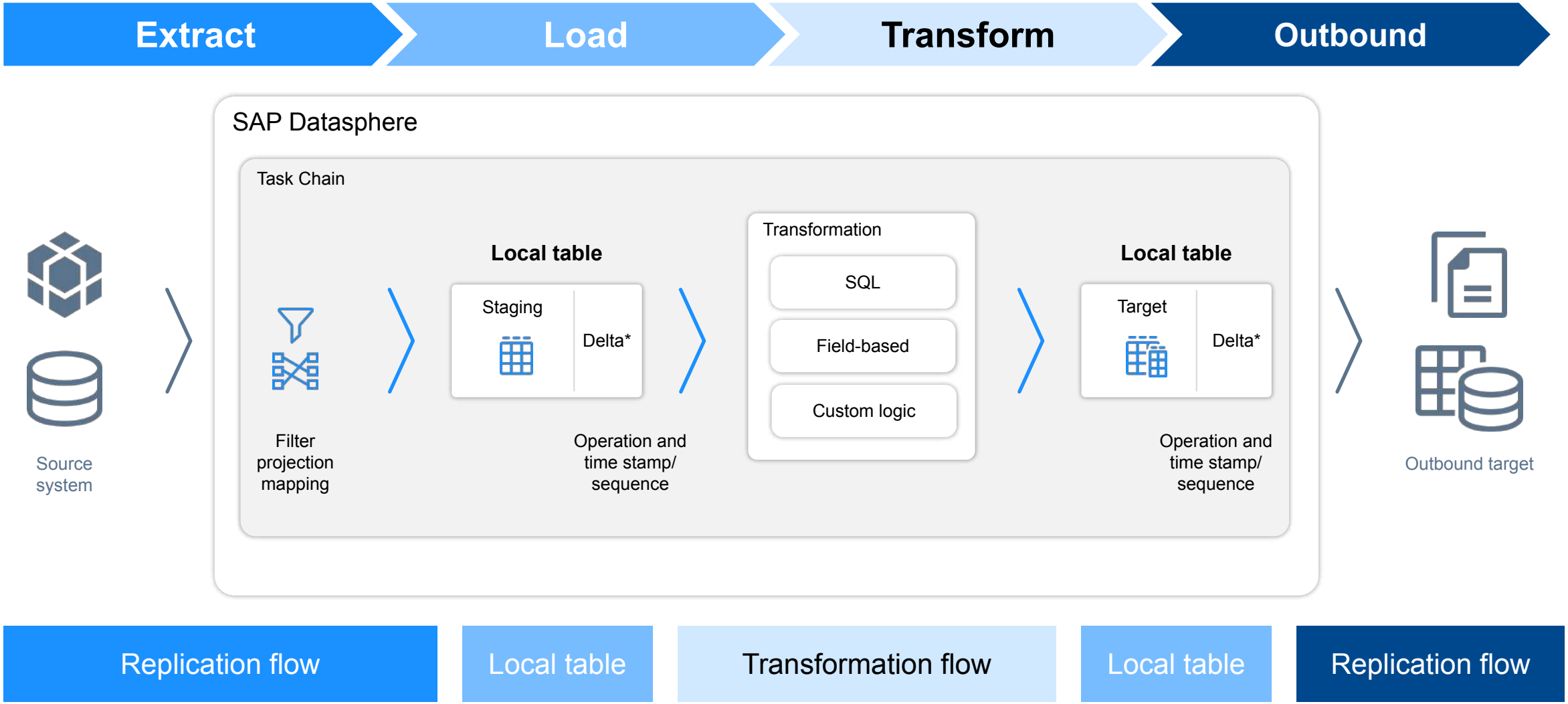
View Analyzer

The View Analyzer provides you with detailed information on each entity as part of your view

- Analyzes each view definition, the consumed views, the local tables and the data sources used by remote tables that are part of your data model. Considers the performance properties of local tables residing in object store in its view persistency guidance
- Use the analyzer results to **optimize your data model** and decide which view to persist
- Generate an **explain plan** to get a quick overview of your model and analyze the performance of a view
- The **data persistence guidance** provides recommendation on what views to persist, as the entire lineage graph is reviewed and relative metrics are analyzed to provide guidance for the optimal views to persist to improve performance, runtime, and memory consumption
- The **data persistence candidate score** provides a recommendation for the best views to persist to increase performance
- Supports **cross-space analysis,** including models with local tables (files), enabling improved partitioning recommendations and faster data preview across Datasphere spaces



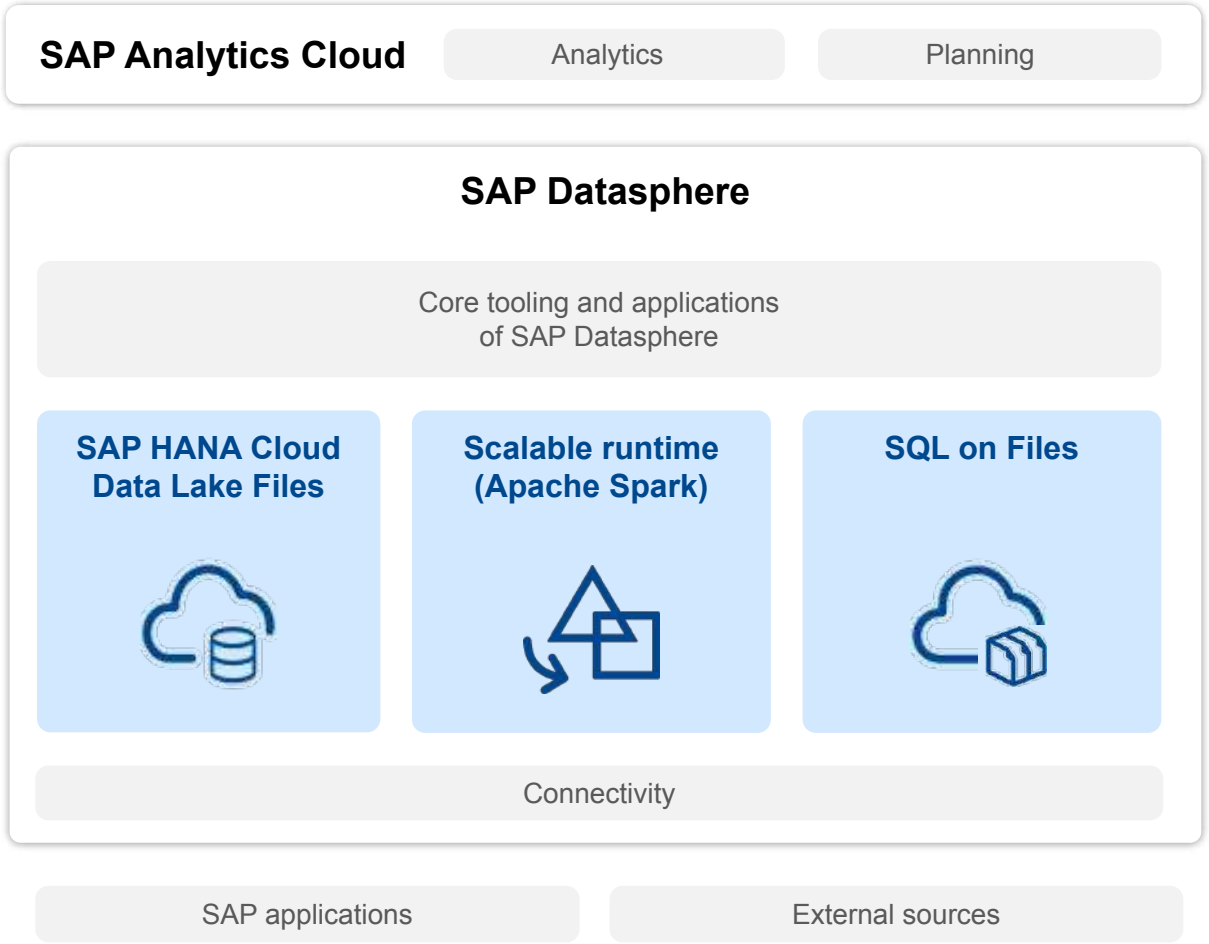
Loading and transforming data in SAP Datasphere



* Restrictions to delta capture apply: more information in the online documentation.

SAP Datasphere – embedded object store capabilities

Simplified interoperability by reusing existing SAP Datasphere applications

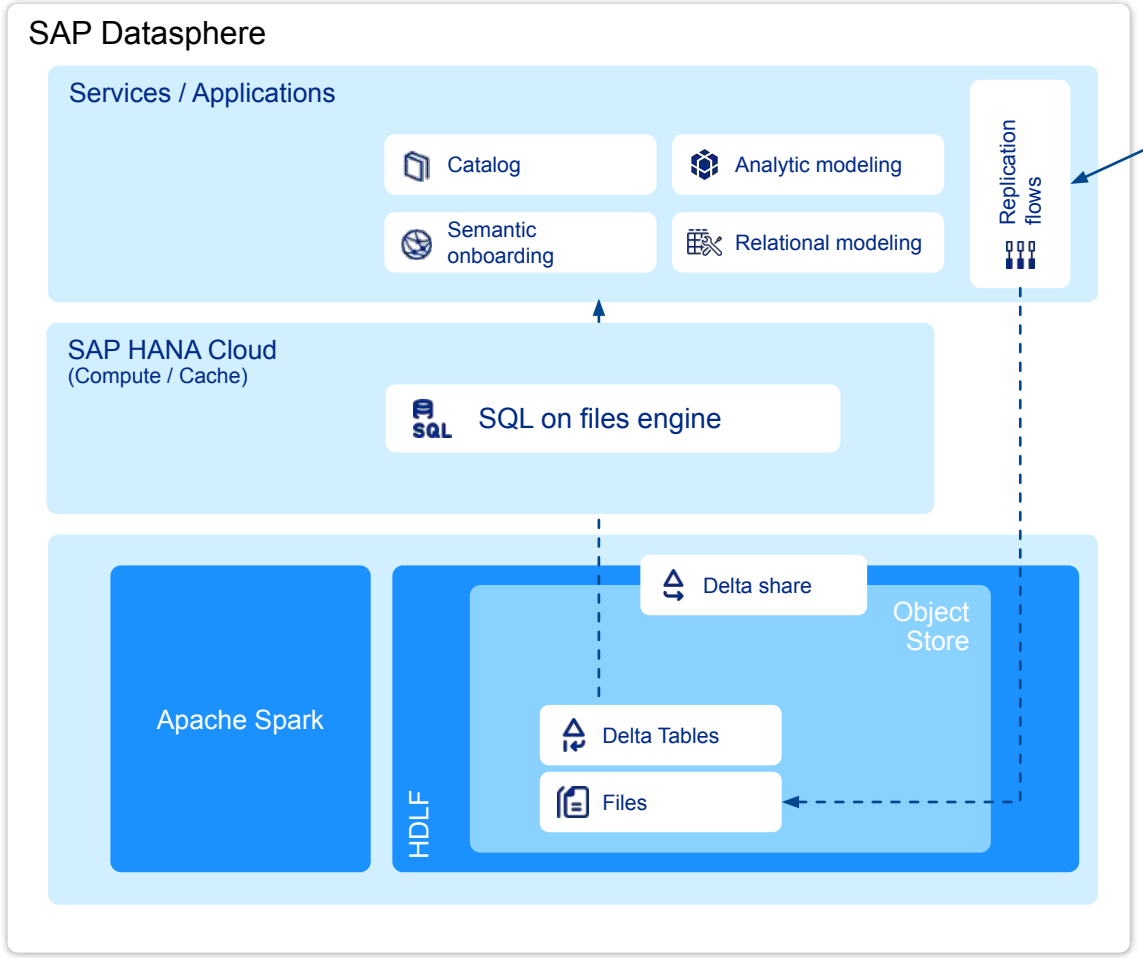


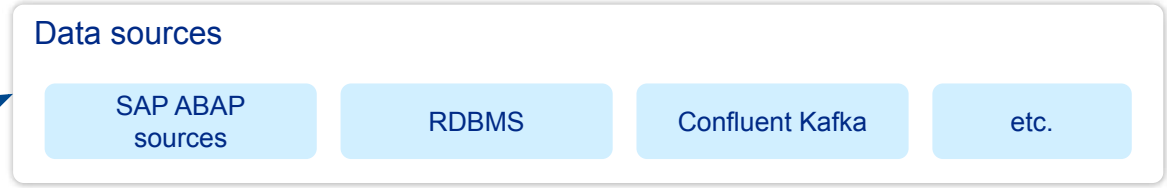
Benefits of Object Store capabilities

- Extended & cost-efficient storage of data
- Scalable runtime for data transformations
- Utilize existing applications incl. SAC integration
- Use SAP HANA Cloud as cache for optimal performance
- Reuse existing services for replication & transformation
- Tight integration into existing monitoring applications

Embedding Object Store in SAP Datasphere

Handle large volume scenarios

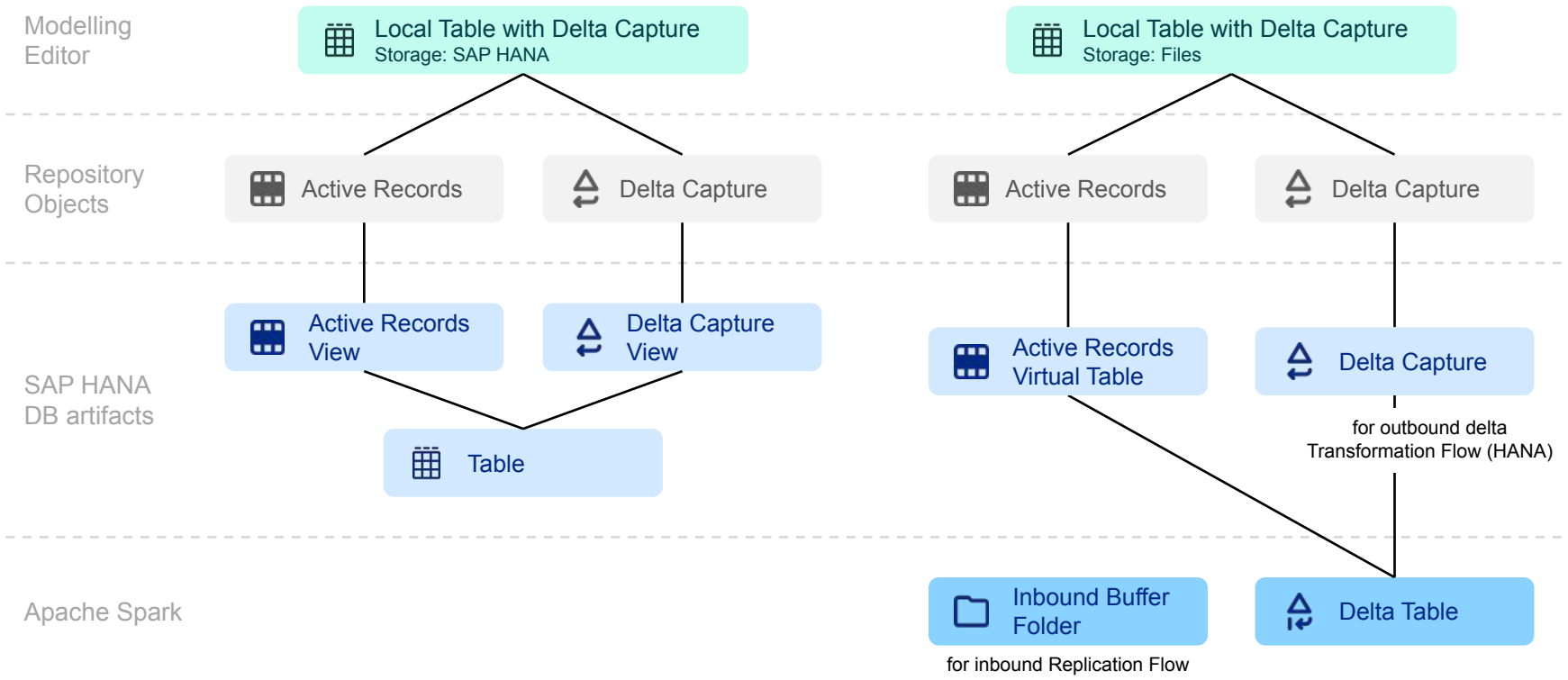




- The **embedded objects store** can be added via the Flexible Tenant Configuration (FTC), which includes SAP HANA Data Lake Files (HDLF), SQL on Files engine, and Apache Spark Service
- Create a **file space** and allocate compute resources to it
- New artefact with **local table (files)** as delta tables with primary persistency in HDLF
- Data ingestion via **Replication Flow** with direct write to HDLF object store
- Data transformations on object store using **Apache Spark runtime** and Transformation Flows
- Virtual read access via **SQL on Files** engine where SAP HANA Cloud acts as compute layer with cache and persistency
- **Export** objects from and **import objects** to file spaces via CSN files or standard transport capabilities

Local Table Objects

Overview of local table objects when using embedded object store in SAP Datasphere

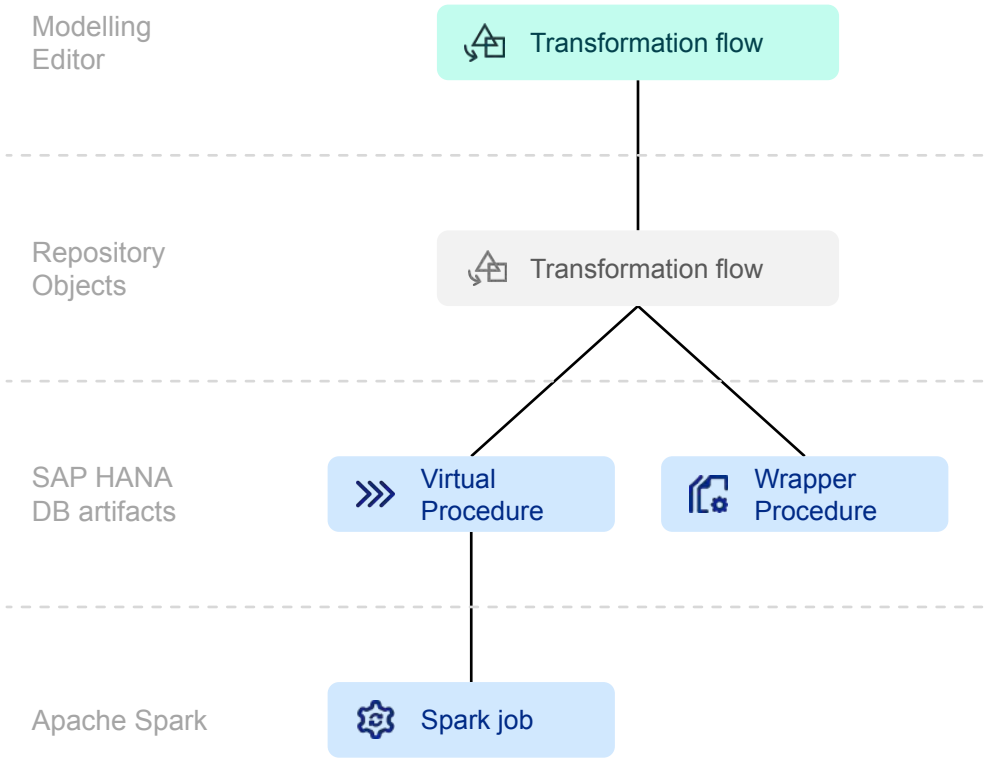


Key Features for Local Table on File (LTF)

- Delta Capture (always on)
- Partitioning (on column level, no ranges)
- Data Preview
- Delete All (Logical)
- Cross-Space Sharing
- Monitor details screens to optimize and merge logs in Data Integration Monitor
- Replication triggers automatic merge
- Track record and storage metrics in Monitor
- Execute vacuum on local table (file) to reduce storage and retain delta loads

Fundamentals of Transformation Flow

Overview of transformation flow when using embedded object store in SAP Datasphere



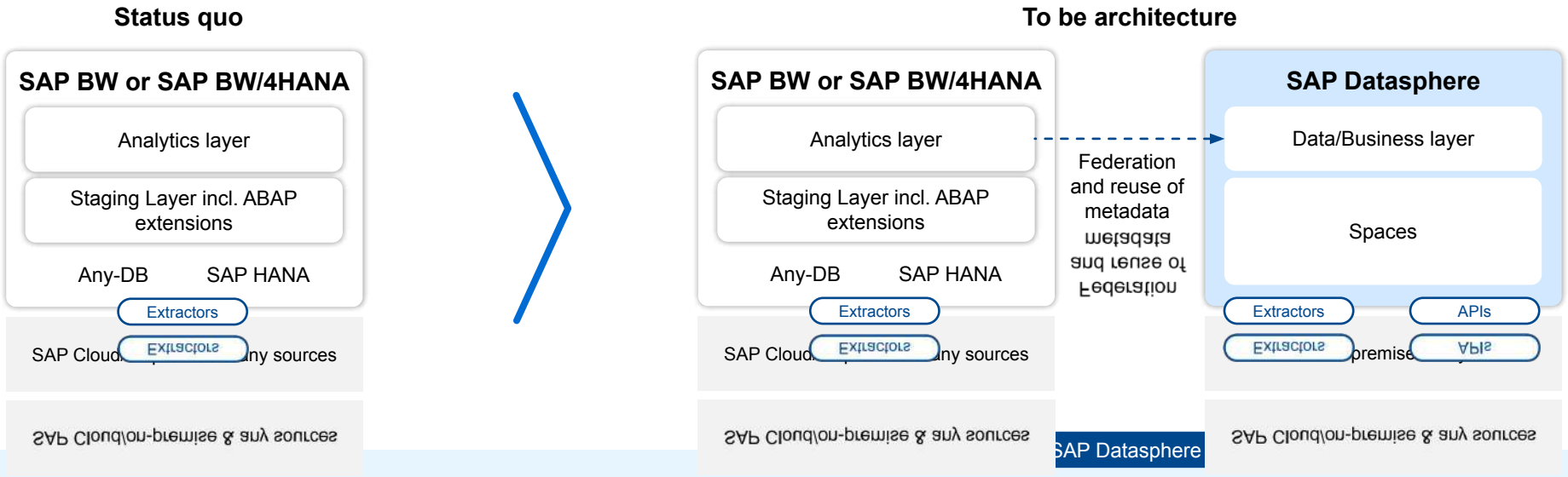
1. In object store spaces, Transformation Flow runtime is set to Apache Spark. UI checks space type and flags the transformation flow accordingly and with this it restricts features to those supported by Apache Spark
2. During deployment, SQL is generated and saved as part of the deployed metadata. Generated SQL is validated for Spark SQL compliance and Python scripting is supported for advanced logic beyond SQL
3. A virtual procedure is created at deployment and executed at runtime
4. A wrapper SQL procedure invokes it asynchronously and Spark Transformation Flows support mid-execution cancelation with rollback to the last stable state to maintain data consistency
5. Transformation flow will read data from LTF delta tables via Apache Spark or HDLF catalog and writes directly to LTF delta tables, bypassing the inbound buffer
6. For delta loads, watermark is retrieved via stored procedures and used to filter input data

SAP BW Bridge & hybrid

Simplify the data landscape

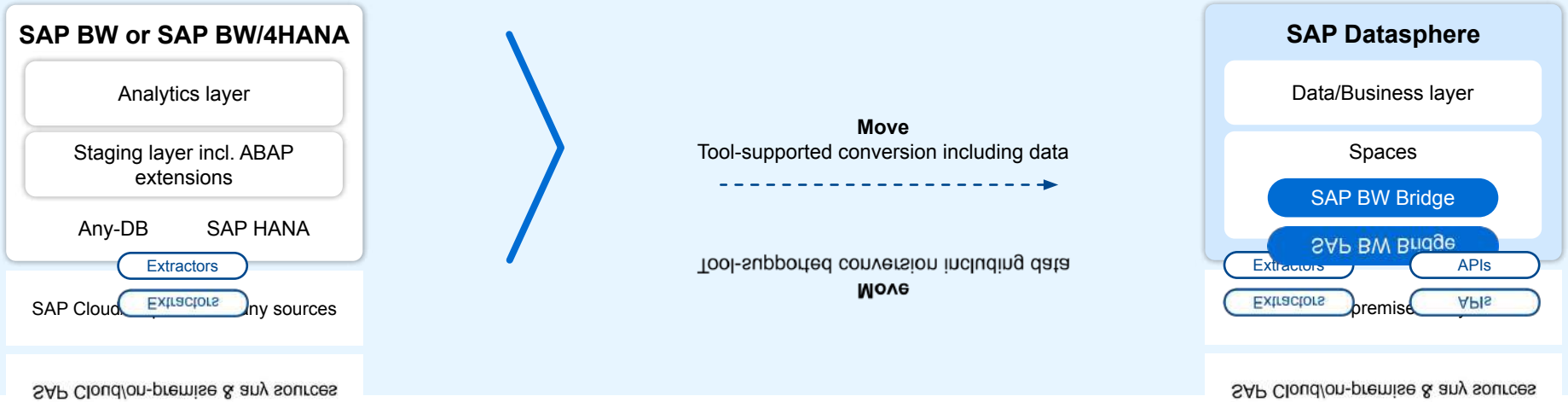
Hybrid

Combining SAP BW or SAP BW/4HANA with SAP Datasphere



SAP BW bridge

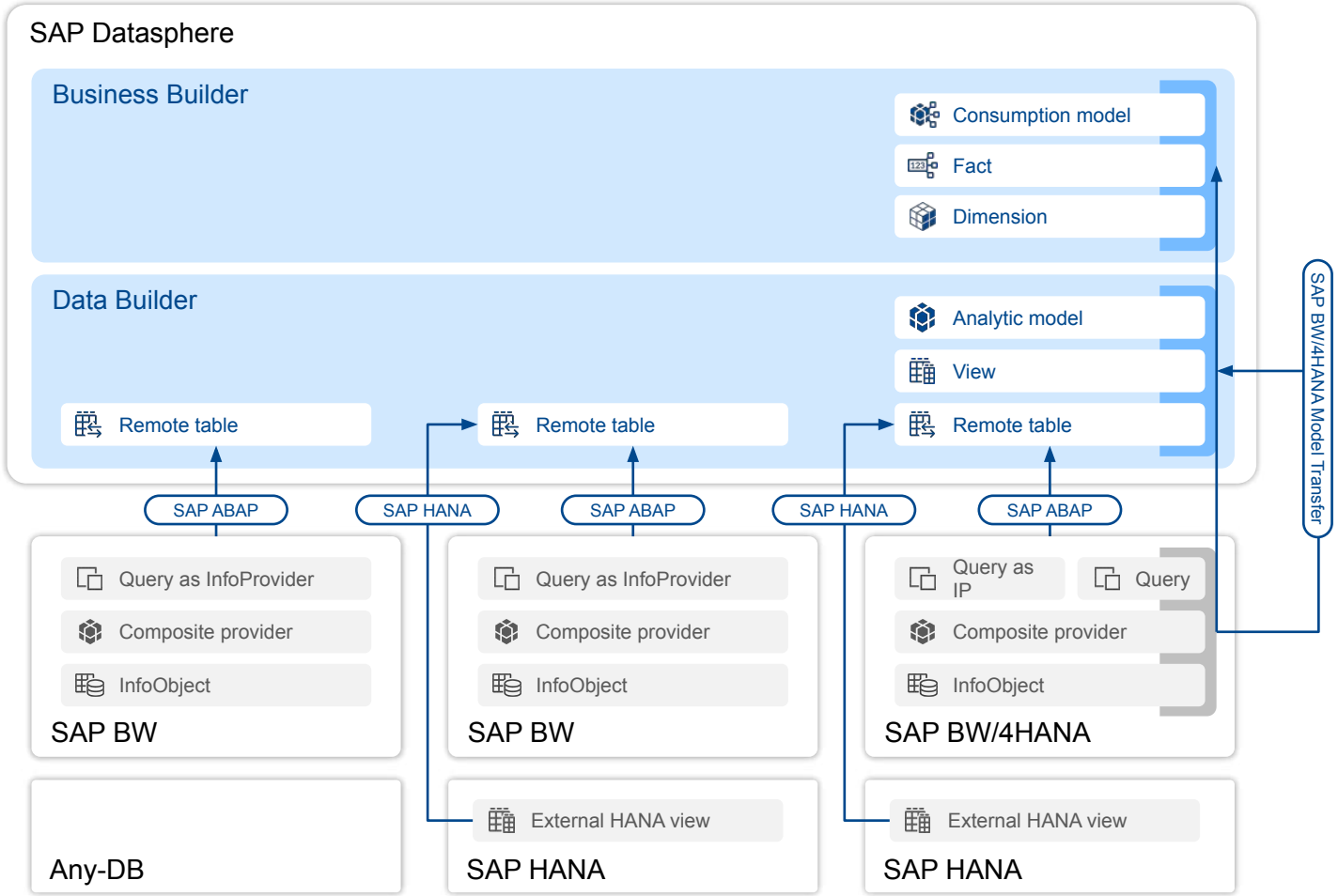
Replace parts of SAP BW or SAP BW/4HANA with SAP Datasphere



SAP BW and SAP BW/4HANA hybrid integration

Different ways to integrate SAP BW and SAP BW/4HANA systems

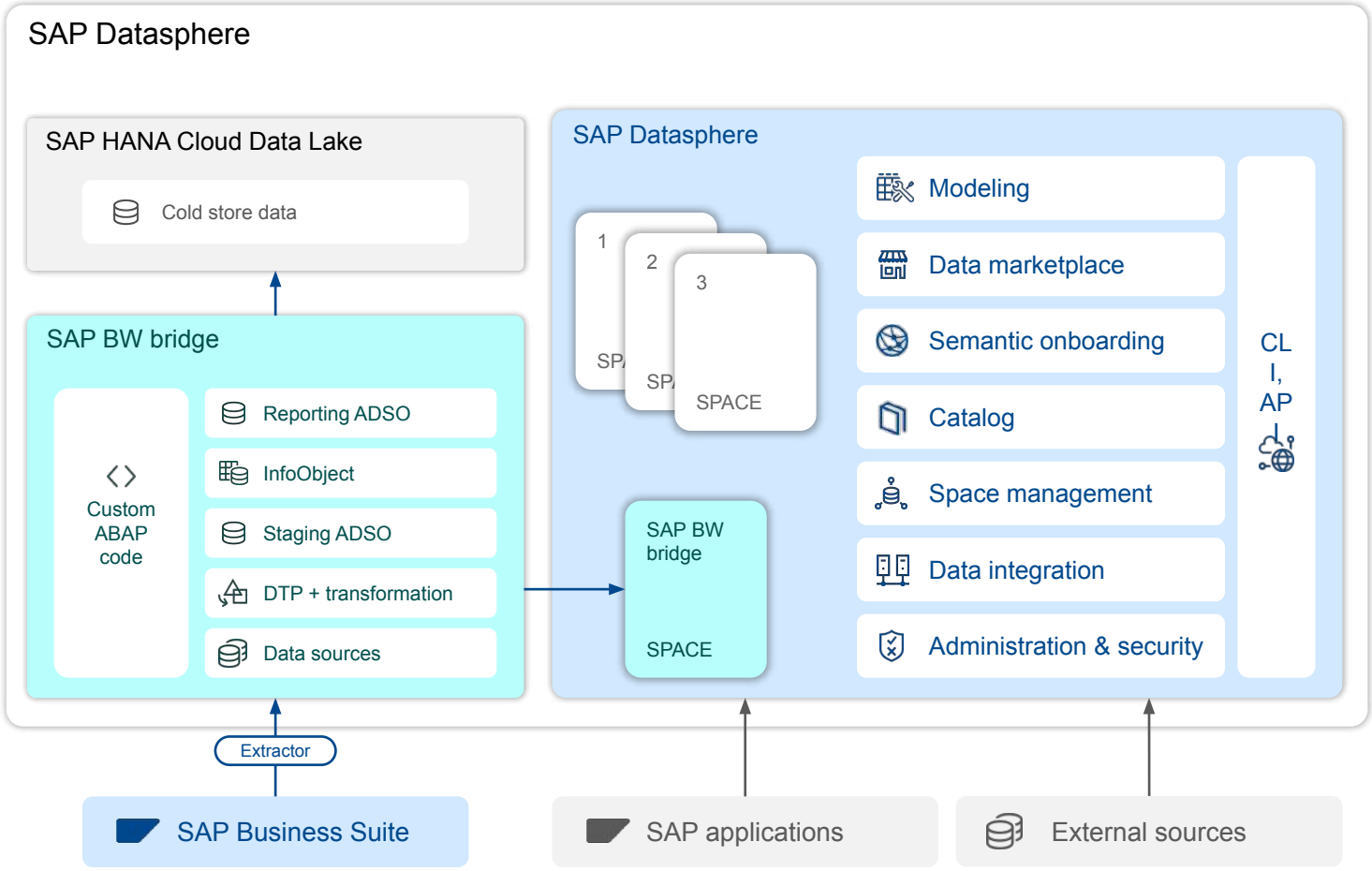
- Remote or replicated data consumption scenarios supported with SAP BW & SAP BW/4HANA via
  - Operational Data Provisioning framework (SAP ABAP connection)
  - External SAP HANA views (SAP HANA connection)
- Business semantics onboarding support for SAP BW/4HANA objects only via SAP BW/4HANA Model Transfer connection more information



SAP Datasphere, SAP BW bridge

SAP BW bridge provides a path to the public cloud for SAP BW NetWeaver and SAP BW/4HANA customers

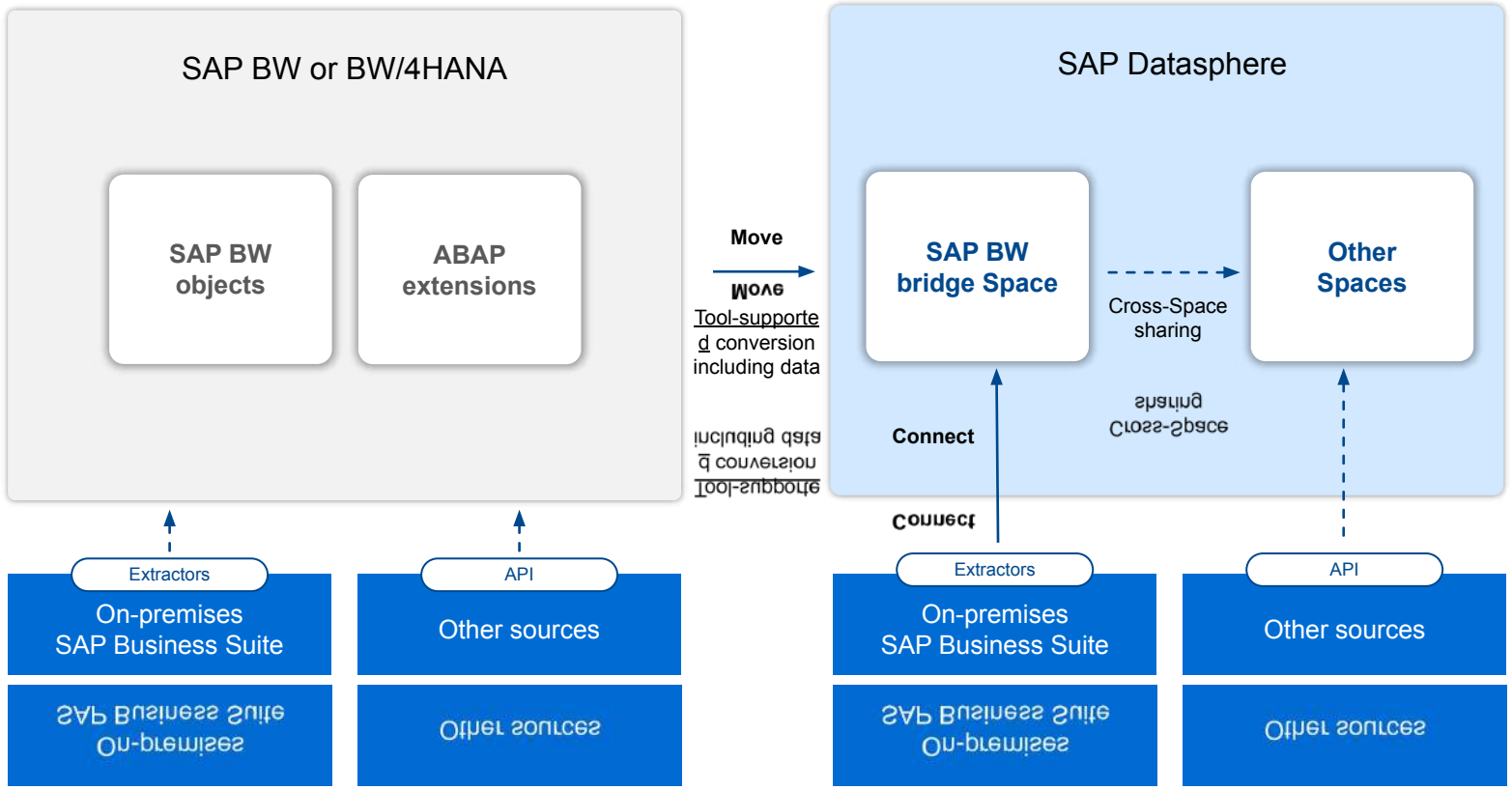
- It offers SAP BW capabilities directly in SAP Datasphere:
  - Connectivity & business content providing proven SAP BW extractors from SAP ECC and S/4HANA
  - Staging layers of SAP BW for managing data loading with partitioning, monitoring, error handling
- Benefits of SAP BW bridge
  - Leverage SAP BW data structures, transformations, customizations, and skills – quickly extending your SAP BW investments to the public cloud
  - Integrate with on-prem SAP ECC data with familiar connectivity & semantic flexibility – retaining instant access while expanding your analytics depth
  - Empower your business to rapidly innovate on SAP BW data with an open, unified data & analytics cloud service



Cloud journey for SAP BW customers

Leverage existing SAP BW models and transformations to innovate in the cloud

- Reuse SAP BW content with SAP Datasphere, BW bridge
- Move existing SAP BW artefacts with transfer tools
- Transition to the cloud at your own pace

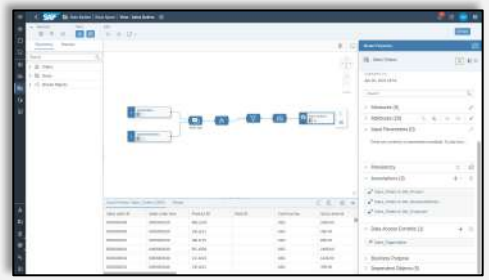


Modeling

Data modeling using integrated editors

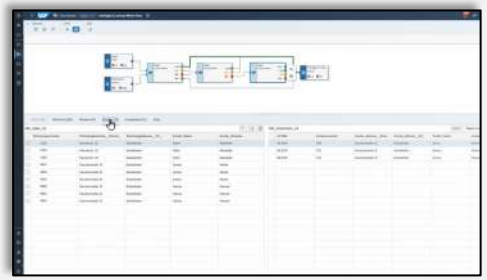
Built-in editors
covering no-code to pro-code

Graphical & scripted view



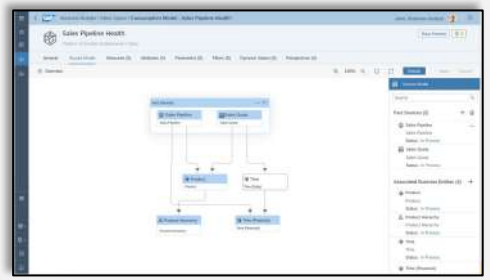
- Collection of no-code/low-code editors to support graphical modeling
- Model E/R-models, tables, views, SQL & SQL-script views
- Focuses on most-used modeling operators

Intelligent lookup



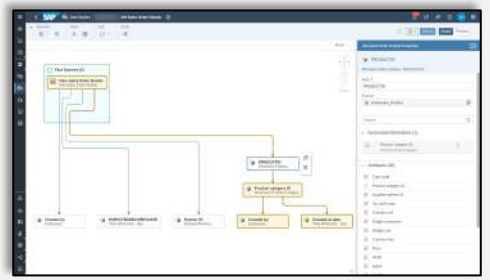
- Interactive data harmonization environment
- Iteratively join entities by defining rules to match records
- Mappings created once and can be automated through rules
- Use output view in other models

Business model



- Allow for a greater degree of self-service
- Non-disruptive environment for business scenarios – independent from the data integration layer

Analytic model

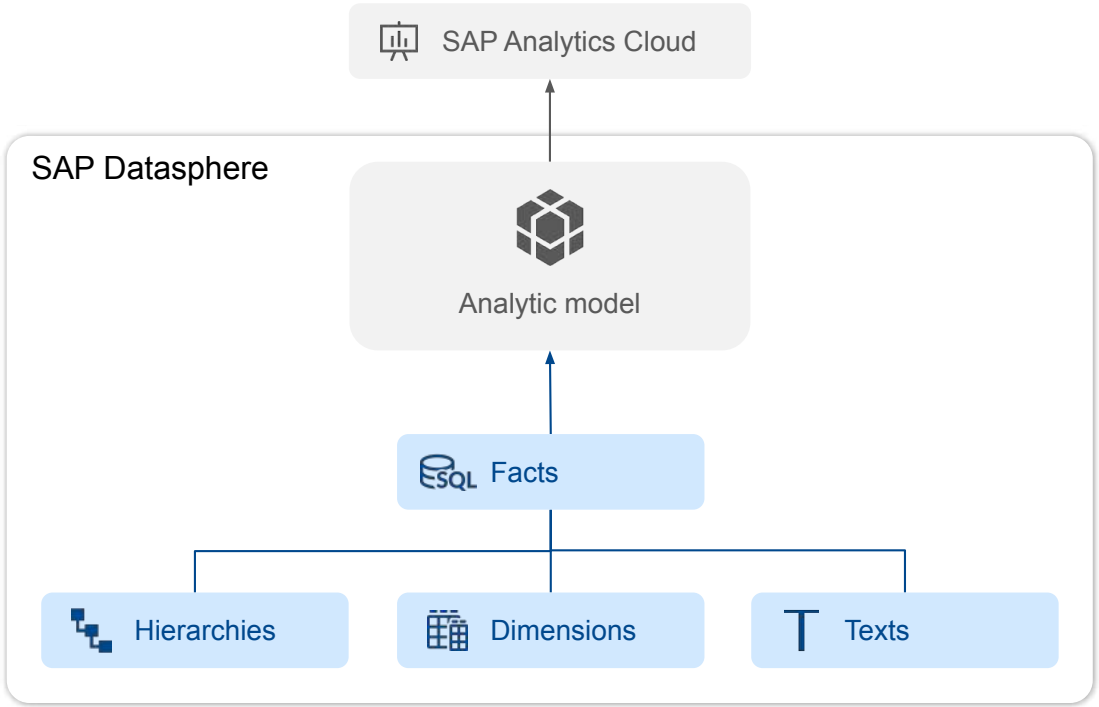


- Rich multi-dimensional & analytical modelling reusing pre-defined measures, hierarchies, filters, parameters & associations
- Built-in analytical preview incl. filtering, pivoting, hierarchies, etc.
- Seamlessly integrates with SAP Analytics Cloud and Office 365 to analyze data

Graphical & scripted view

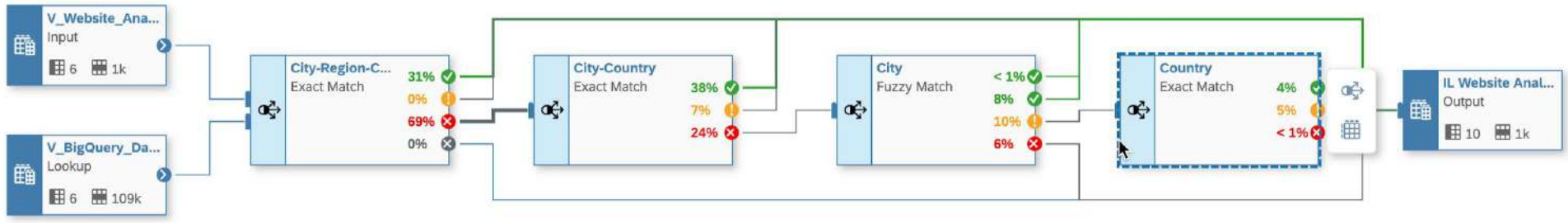
Building the foundation of your models

- Use graphical or SQL views to build the foundation of your data modeling on top of remote or local tables which can be consumed in the analytic model
- Choose the semantic usage type fitting your requirement:
  - **Fact** to indicate that your entity contains one or more measures that can be analyzed
  - **Dimensions** indicate that your entity contains attributes that are used to analyze and categorize measures defined in other entities
  - **Hierarchy** to indicate that your entity contains parent-child relationships for members in a dimension
  - **Text** to indicate that your entity contains strings with language identifiers to translate text attributes
  - **Relational dataset** contains columns with basically no immediate analytical purpose, typically used as interim entities
- Validation framework to check the consistency of view data with modelling metadata and analyze view performance directly in the View Builder with runtime metrics and memory usage insights
- Leverage release states to ensure released objects stay stable for data consumers
- Configure OData requests directly from the graphical and SQL view editors and expose it through the public OData API



Intelligent lookup

The **Intelligent lookup** is a business centric, interactive data harmonization environment, which lets you iteratively join entities by defining rules to match records, review the results, and use the output in other models.



Interactive environment for mappings

- Subject matter experts efficiently create mappings and mapping rules in an intuitive no-code user interface.
- Results are shown immediately.
  Data is taken as it is.

Create mappings only once

- All automated & manual mappings are persisted in a “pairing table”.
- No further user intervention is required ever again on rows where these mappings are applicable.

Mapping automation through rules

- Rules create mappings automatically.
- Users can add or rework mappings where required - down to record level.
- Run them manually or periodically via task chains.

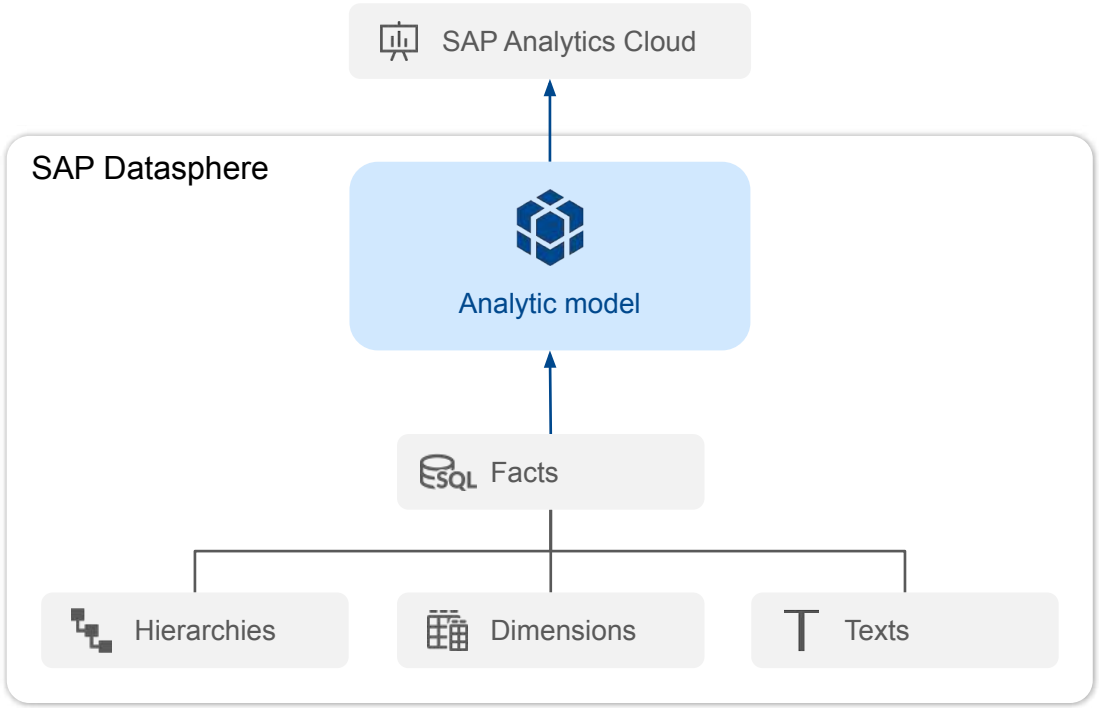
Full integration with other data models

- Harmonization result is a standard entity within the data builder that all modelling entities and consuming tools can leverage.

Analytic Model – Core Capabilities

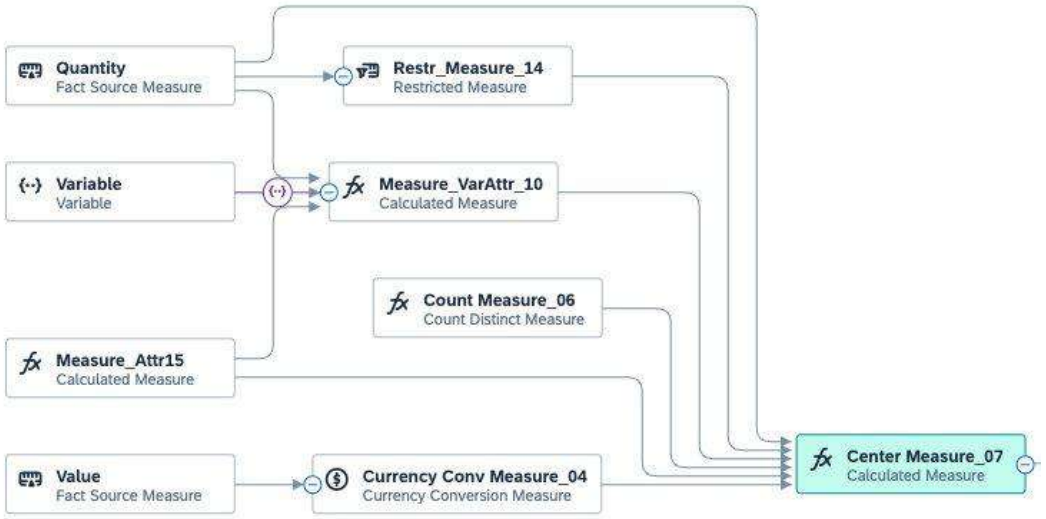
Building the analytical foundation to make data ready for consumption.

- Enables multi-dimensional and rich analytical modelling with less effort to answer business questions easier, faster, and more efficiently.
- Provides data for analytical purposes by reusing predefined measures, hierarchies, filters, parameters, and associations to answer different business questions.
- Supports stacking of analytic models to any level with inherited measures, dimensions, variables, and filters, enabling centralized governance and flexible reuse.
- Allows configuration of OData requests directly from the analytic model editor and exposes them through the public OData API.



Analytic Model – Advanced features for Modeling

- Offers many advanced features such as restricted and calculated measures, exception aggregation incl. unbooked values, non-cumulative measures, pruning, variable support, time-dependent dimensions and texts, currency and unit conversion, analytical preview (including download), multi-lingual metadata, and row-level security (DAC).
- Supports graphical visualization of measure, attribute, and variable dependencies with interactive icons and context-based navigation, enhancing model transparency.
- Enables default values for multiple range variables and derived variables in Analytic Models for more flexible and precise filtering.
- Allows replacement of fact and dimension sources in Analytic Models without losing existing logic, using name-based mapping for seamless updates.
- Defines standard measure formatting in Analytic Models (scale and decimals) with inheritance in stacked models and optional overrides in SAC.

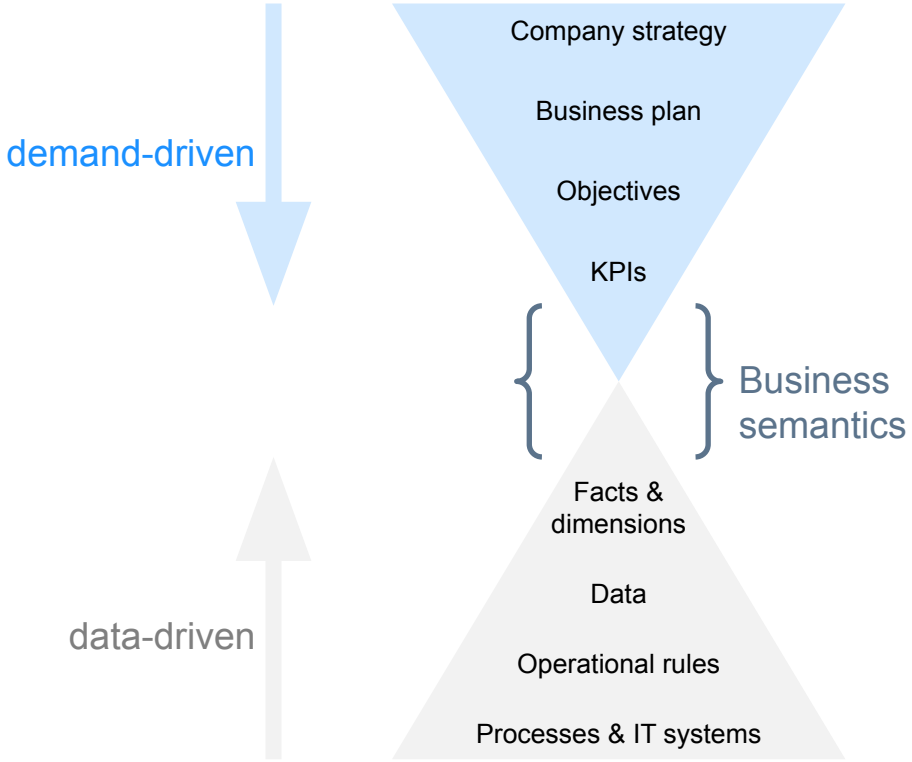


Impact and Lineage Analysis for Measures

Business modeling

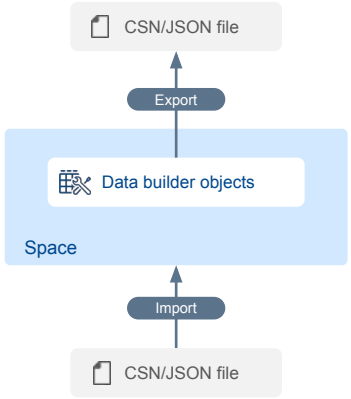
Empowers BI analysts to model using business terminology and supports demand-driven modeling.

- Self-service modeling for lines of business
- Enables business users to model their business scenario without knowing the underlying data models
- Hide data complexity through using business language terminology
- Flexibly change the underlying data layer when needed and without disrupting your consumers
- Run advanced data preview with filtering and sorting capabilities to validate the modeled business logic immediately



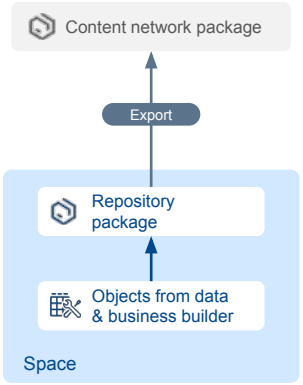
Lifecycle management

Export & import object definition



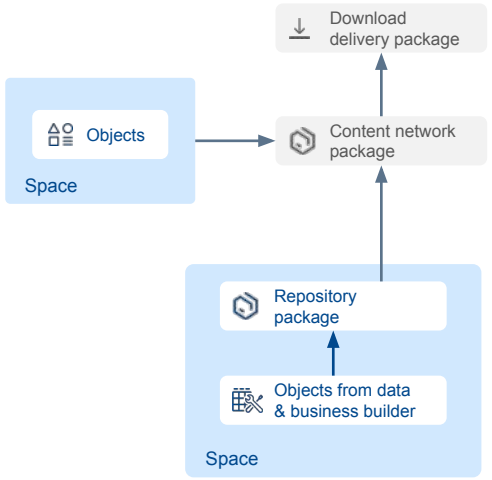
- Definitions for data builder objects like tables, views, flows, intelligent lookups can be exported as CSN/JSON files
- Exported object definitions can be imported into the data builder to create new objects or replace existing ones

Repository packages



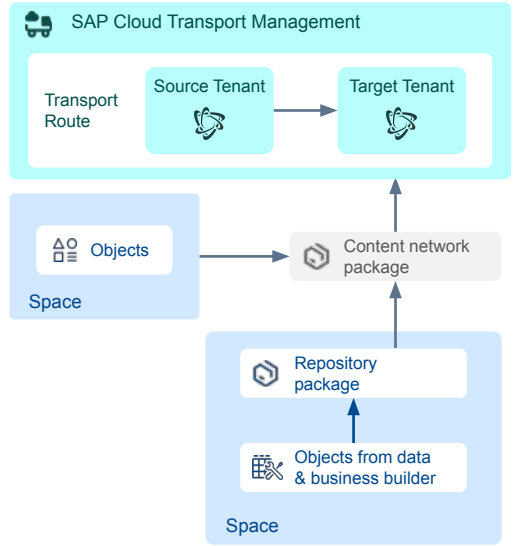
- Repository packages as a means to organize objects from a single space
- Modelers can assign objects to existing repository packages
- Repository package objects can be exported to a content network delivery package with a single click

Content network package



- Content network package can contain objects from one or multiple spaces along with their dependencies
- One or more repository packages can be included in a single content network package
- Content network packages can be downloaded from or uploaded into the content network
- Change space assignment on import to copy objects in various spaces of the same tenant

SAP BTP Cloud Transport Management service

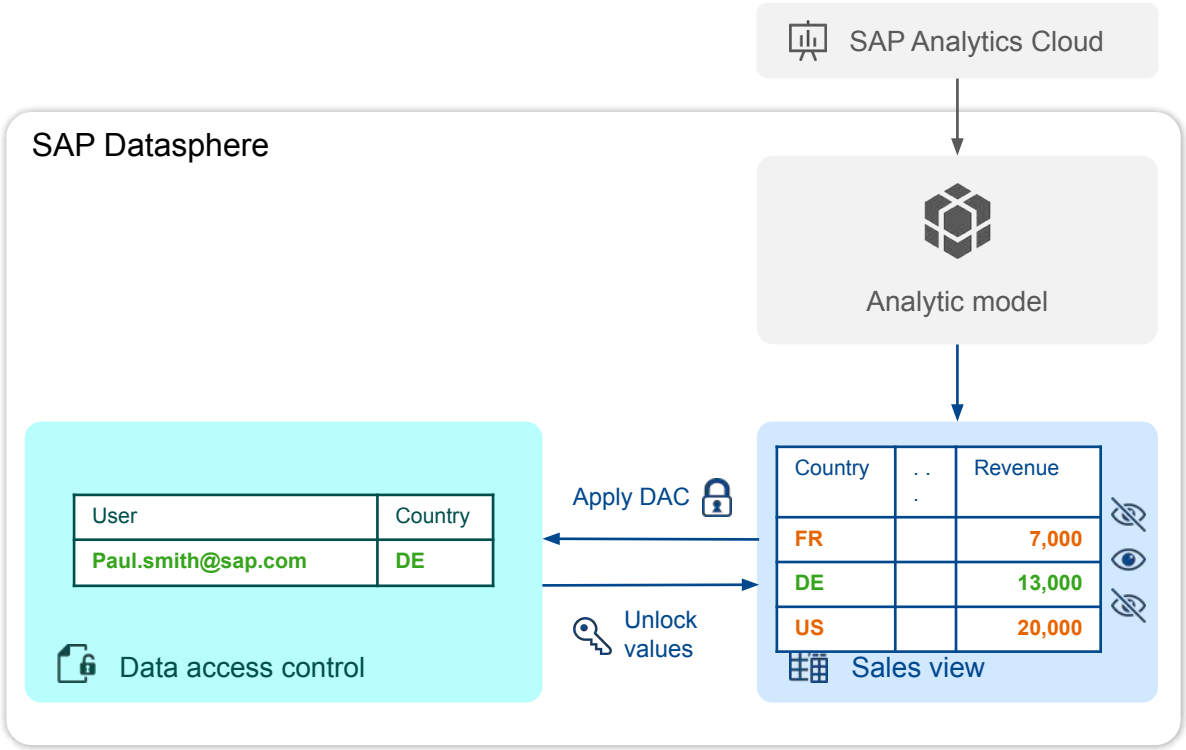


- SAP BTP Cloud Transport Management service (CTMS) integration with SAP Datasphere to transport content packages across different landscapes (blog)
- Configure your SAP CTMS system and connect it to your SAP Datasphere tenants
- Define your transport nodes and routes, then export your content network package to CMTS and trigger the transport

Authorizations using data access controls

Use data access controls to apply row-level security to your view or analytic model

- When a data access control is applied to a data layer view, analytic model, or a business layer object, any user viewing its data will only see the rows for which they are authorized, based on the specified criteria
- The following types of criteria are supported:
  - **Single values:** each user can only see the records that match any of the single values they are authorized in the permissions entity
  - **Operator and values:** each user can only see the records that fulfill the operator-value pairs they are authorized for in the permissions entity
  - **Hierarchy:** Users can view records matching their assigned hierarchy values and descendants, as defined in the permissions entity. Supported types are parent- child and hierarchies with directory
- Data access controls are defined once and can be applied to multiple artifacts in data layer, and cannot be overruled
- Use the “View as User” tool to review the effects of the data access controls you apply by checking the records that another user will be allowed to see

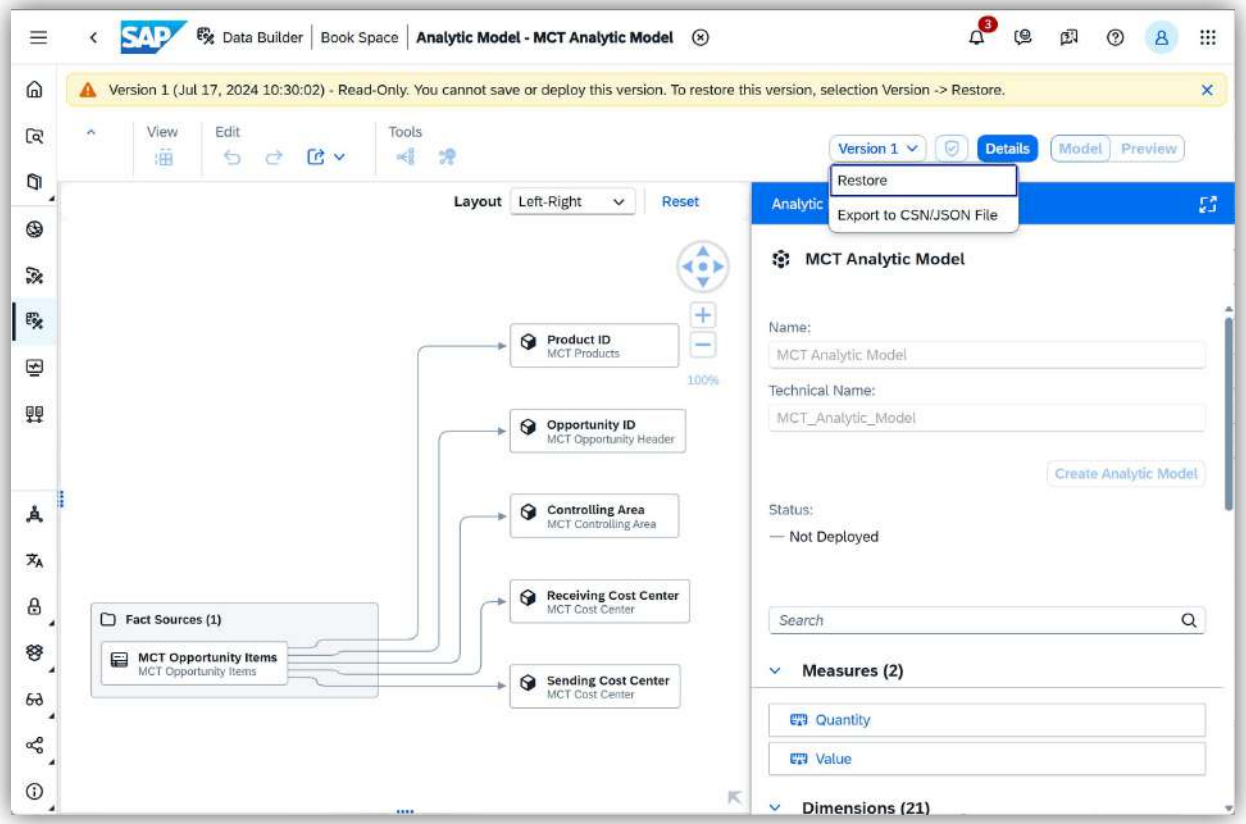



```
SELECT * FROM SALES_VIEW WHERE 'Country' == 'DE'
```

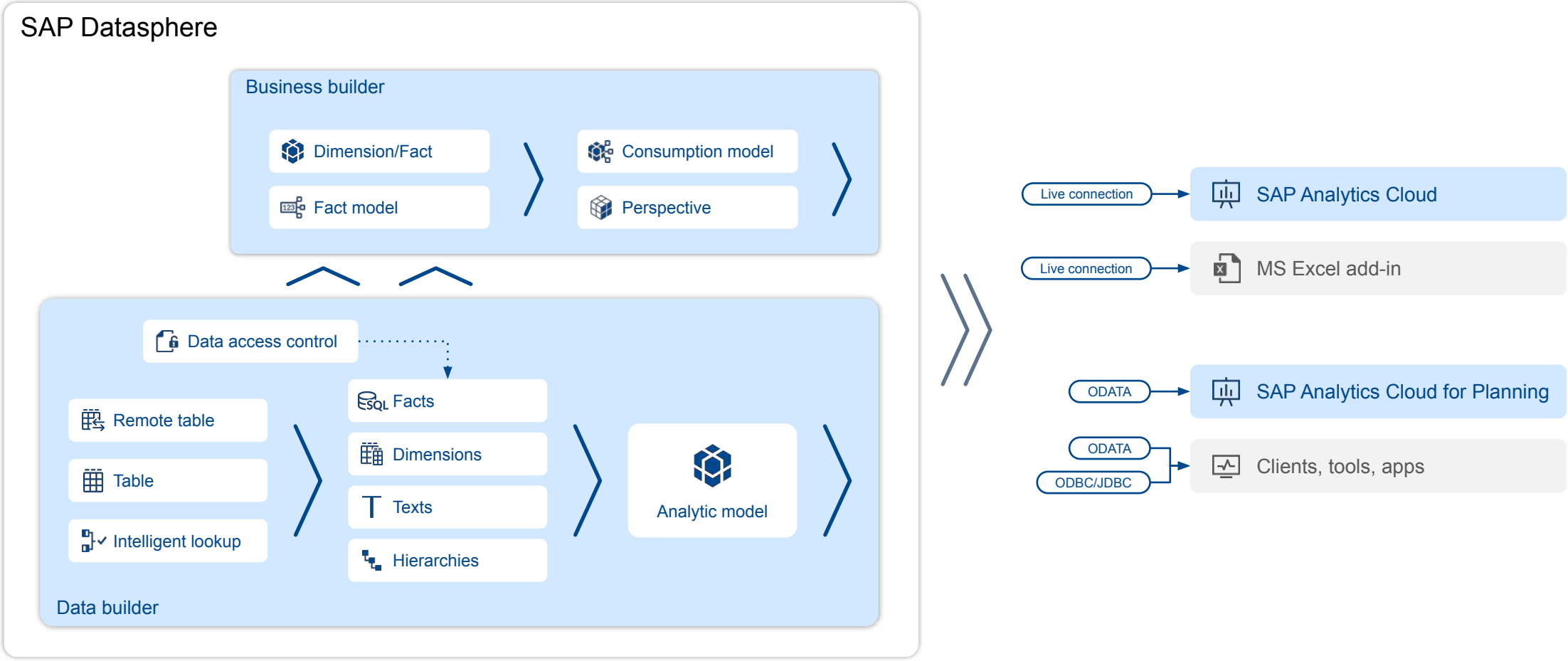
Object Versioning

Object versioning allows you to review the history of changes to your objects

- Automatic version creation on each deployment
- List and view all past versions of supported objects and restore their metadata
- Export and download CSN definitions or open old versions in a new tab for visual comparison for graphical and SQL views, analytic models, local and remote tables, E/R models, transformation and replication flows, task chains, data access controls, and packages
- Restore previous versions with full validation for graphical and SQL views, analytic models, local tables, data access controls; only error-free versions can overwrite the current object



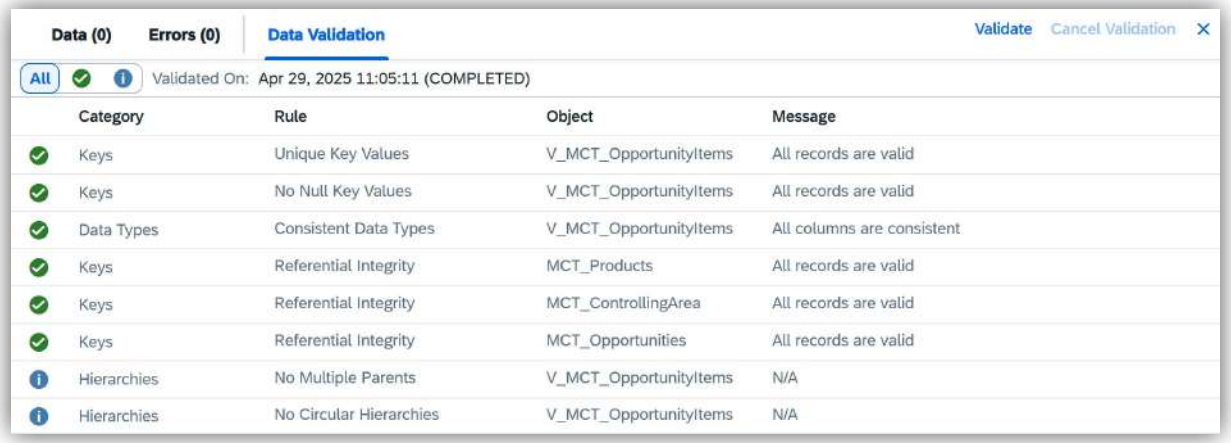
Data modeling and consuming with SAP Datasphere



Validating View Data

The validation framework provides initial validation rules to check the consistency of views

- Automatically validates objects against a set of predefined rules and highlights violations with detailed feedback:
  - key validation rules (unique key, nulls, referential integrity)
  - hierarchy validation rules (missing nodes, circular hierarchies, finding & resolving unassigned nodes in hierarchies w/ directory)
  - data type consistency
- Shows the last validation status and the last validation run (date and time)
- Validates data models for key issues like data type mismatches, null or duplicate keys, and hierarchy errors to prevent runtime failures
- Checks foreign key referential integrity between facts and dimensions, improving data quality and troubleshooting



Business Semantics

Reuse SAP business semantics with a single entry point

Semantic onboarding

Bringing together all features & processes that allow SAP customers to make semantics available at ease & scale

Import entities

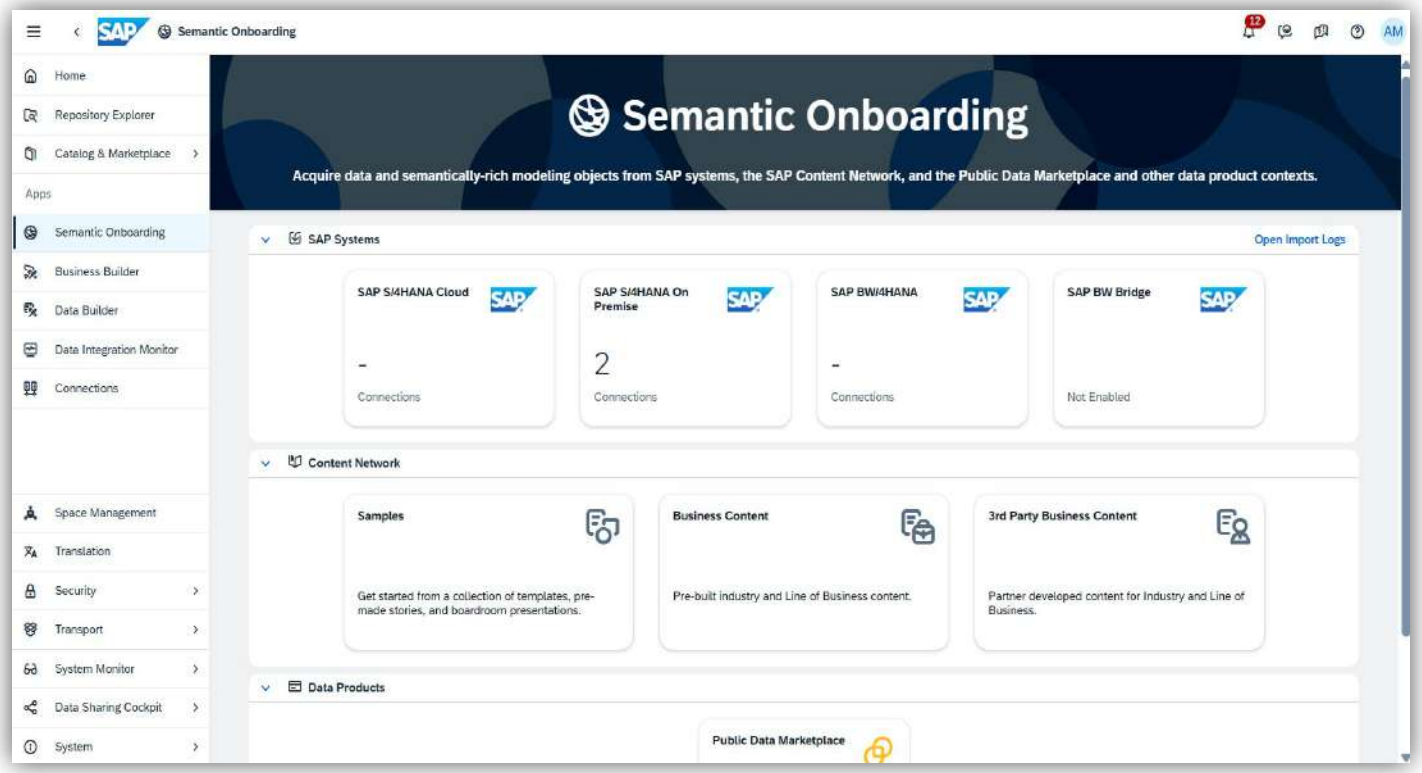
Harmonize existing metadata import options from SAP S/4HANA Cloud, SAP S/4HANA, SAP BW, SAP BW/4HANA

Content network

Bring in prebuilt SAP and 3rd party business content as accelerator

Data products

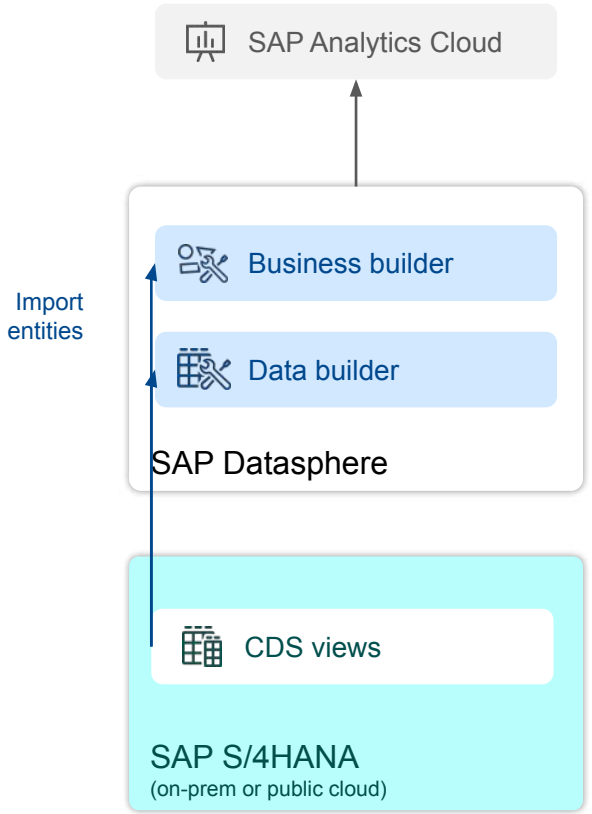
Access customer and partner managed data products to democratize SAP data within SAP Datasphere & beyond via API



Entity import for SAP S/4HANA on-premises & SAP S/4HANA Cloud Edition

Load metadata from your SAP S/4HANA Cloud and SAP S/4HANA on-premise connections

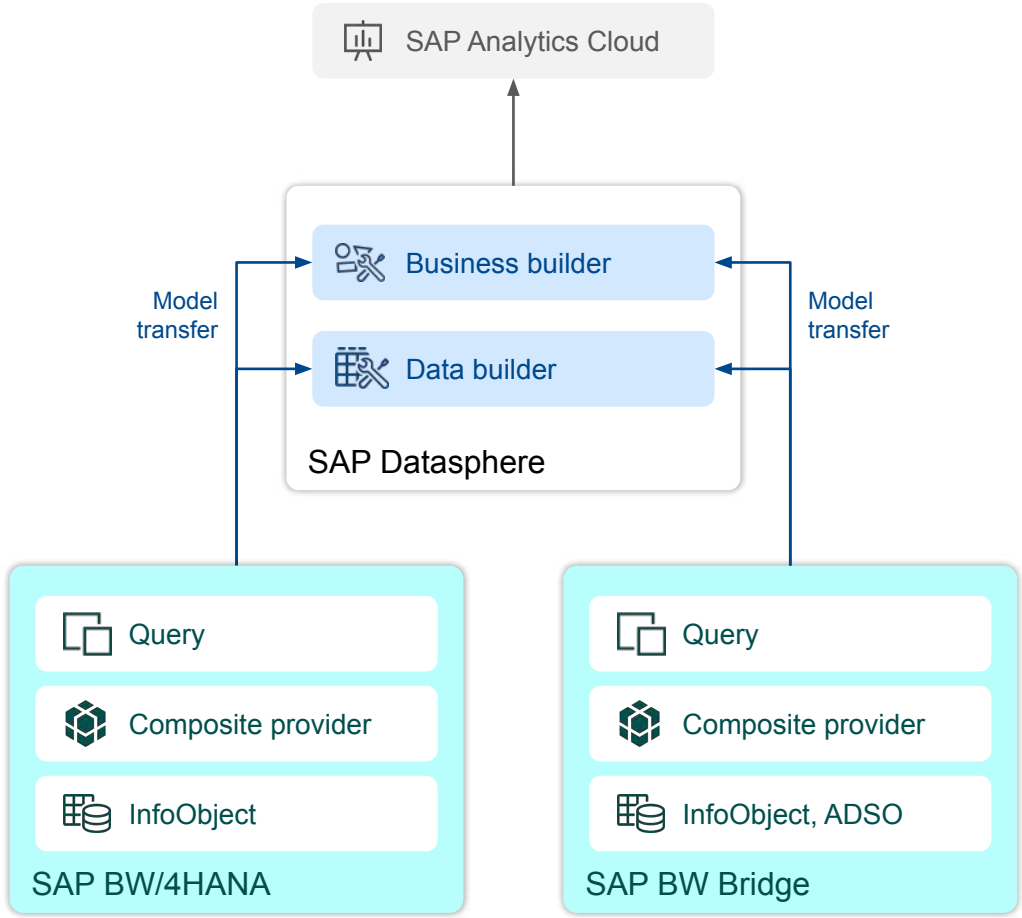
- Convert the SAP S/4HANA semantics to SAP Datasphere entities
- The wizard creates entities in data and business builder
- Ease the customer's journey from operational reporting to cross-analytics reporting in the public cloud
- Unleashing metadata & domain knowledge from SAP source systems
- Relevant SAP notes: 3081998, 3283282, 3463326



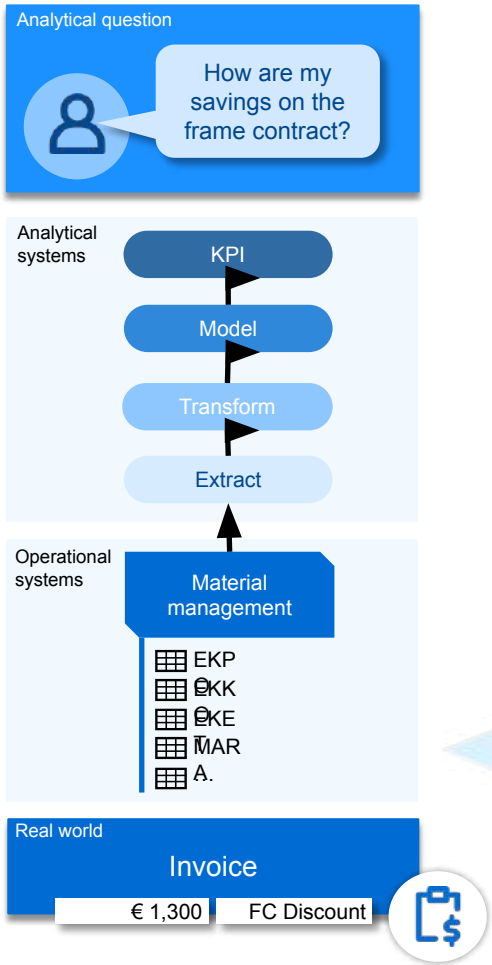
SAP BW/4HANA model transfer

Load metadata from your SAP BW/4HANA and SAP BW bridge connection

- Provide SAP BW/4HANA and SAP BW bridge semantics to SAP Datasphere
- Build new models on top of them or enhance them
- Transfer BW query as native entity (KPI, analytic model)
- Supports the import of BW analysis authorizations via data access controls
- More information about supported features in SAP note 2932647



SAP business content

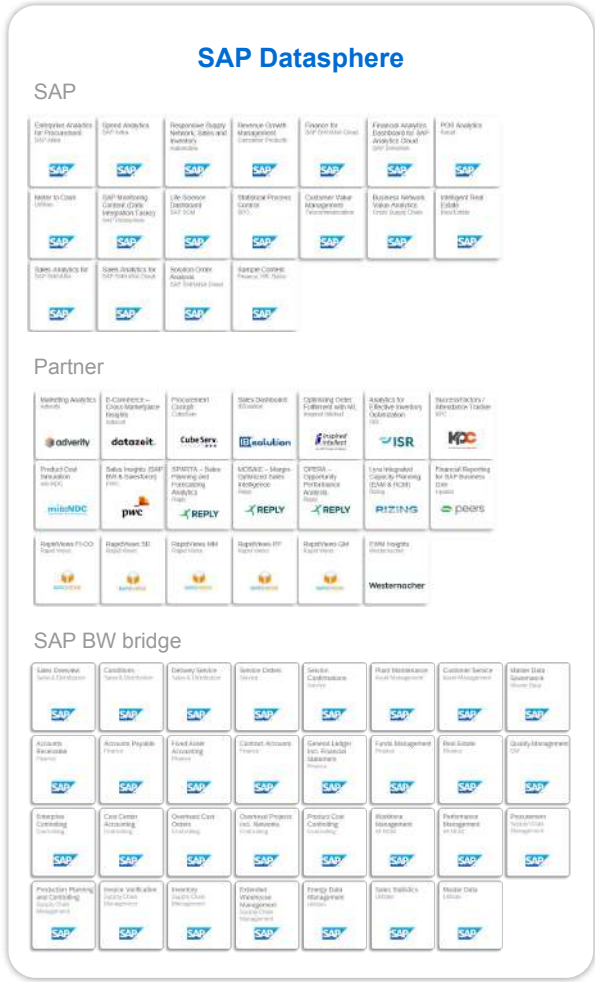


Business content collection

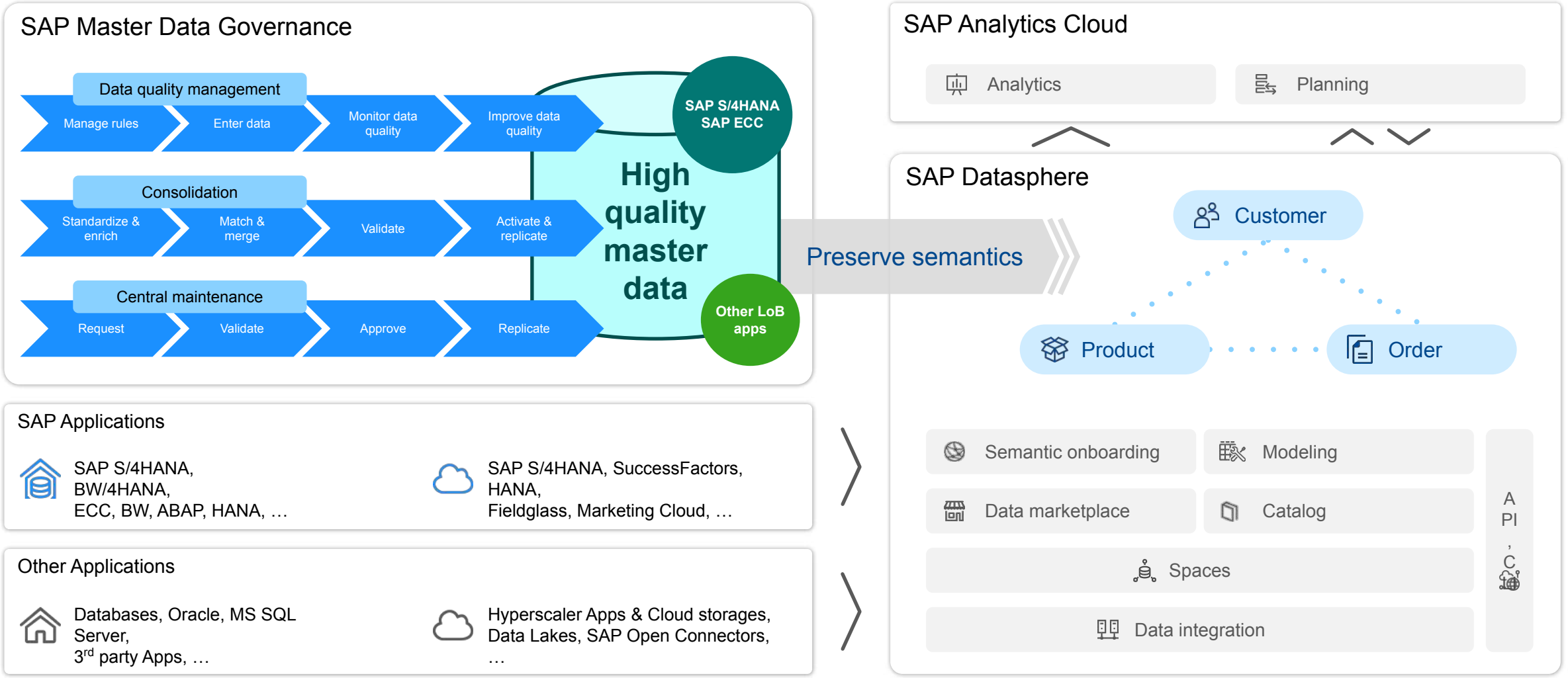
Leverage SAP's business expertise and knowledge from our ecosystem of partners with prebuilt data models, semantic views of SAP application data, and transformation logic.

Continuously growing number of content packages

- 250+ content packages built by SAP & partners for SAP Analytics Cloud and SAP Datasphere
- 31 content packages with SAP BW bridge (11,000+ reusable objects)
- SAP S/4HANA content generation via semantic onboarding (1,000+ Fiori apps ready for import)



SAP Master Data Governance providing high quality master data



SAP Master Data Governance and SAP Datasphere

SAP Master Data Governance		SAP Datasphere
▪ Focus on master data from SAP S/4HANA and operational SAP applications ▪ Process oriented ▪ Supported by workflows ▪ Most common master data domains (e.g. business partner, product, finance master data)	FOCUS	▪ Focus beyond SAP S/4HANA, could be data from all kinds of systems from SAP and 3rd party vendors ▪ Analytic context ▪ Supported by rich analytical modeling capabilities ▪ Master data and transactional data
▪ High master data quality for SAP S/4HANA and operational SAP processes	GOAL	▪ Good (enough) data quality for analytic and planning purposes

COMPLEMENT

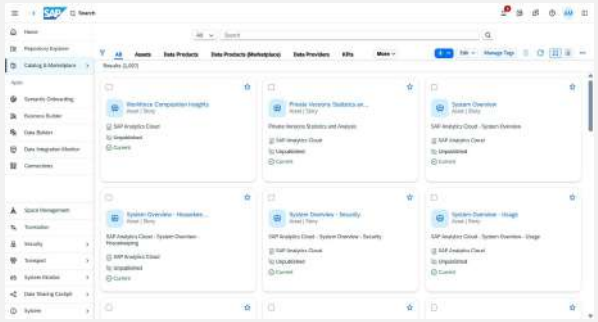
- One of the key concepts of a business data fabric architecture, is to preserve the business context of the data. SAP Master Data Governance maps and consolidates the master data objects in the operational layer across different applications, while SAP Datasphere provides the analytical layer to combine data of any type even virtually from the different applications while preserving the semantics.
- Data products managed by SAP Master Data Governance can directly be consumed by SAP Datasphere in dedicated spaces to create further data products by joining with data from other data sources.
- SAP Master Data Governance provides key and value mappings to improve joining of datasets from different sources within SAP Datasphere.
- Top quality master data is key foundation for operational, analytical and planning purposes as well as for AI use cases.

Governance

Governance in SAP Datasphere

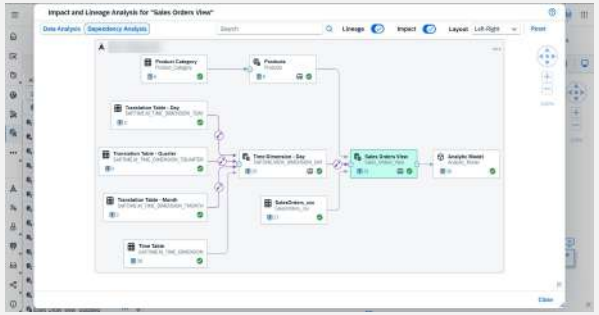
Governance across all areas

Catalog



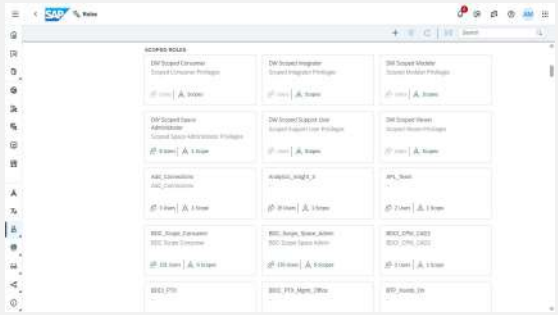
- The central place to discover, classify, profile, understand, and prepare all the data assets in your enterprise beyond SAP Datasphere only, incl. SAP BW Bridge, SAP BW/4HANA and SAP Analytics Cloud
- Map out a connected view of your data sources, build & apply data governance processes, and offer trusted, self-service data discovery to your users

Metadata & lineage



- Use the repository explorer to browse, create and import objects in SAP Datasphere
- Search and filter the list of existing objects or open them centrally
- Leverage GenAI to automate business description and auto-tagging

Data privacy



- SAP Datasphere is designed to fulfill the highest security standards which guarantee the safety of your data
- Security options range from cloud network communication and identity provider settings incl. SCIM API support, to user and access management as well as spaces, roles, scoped roles and data access controls

Catalog enabling governance

Discover

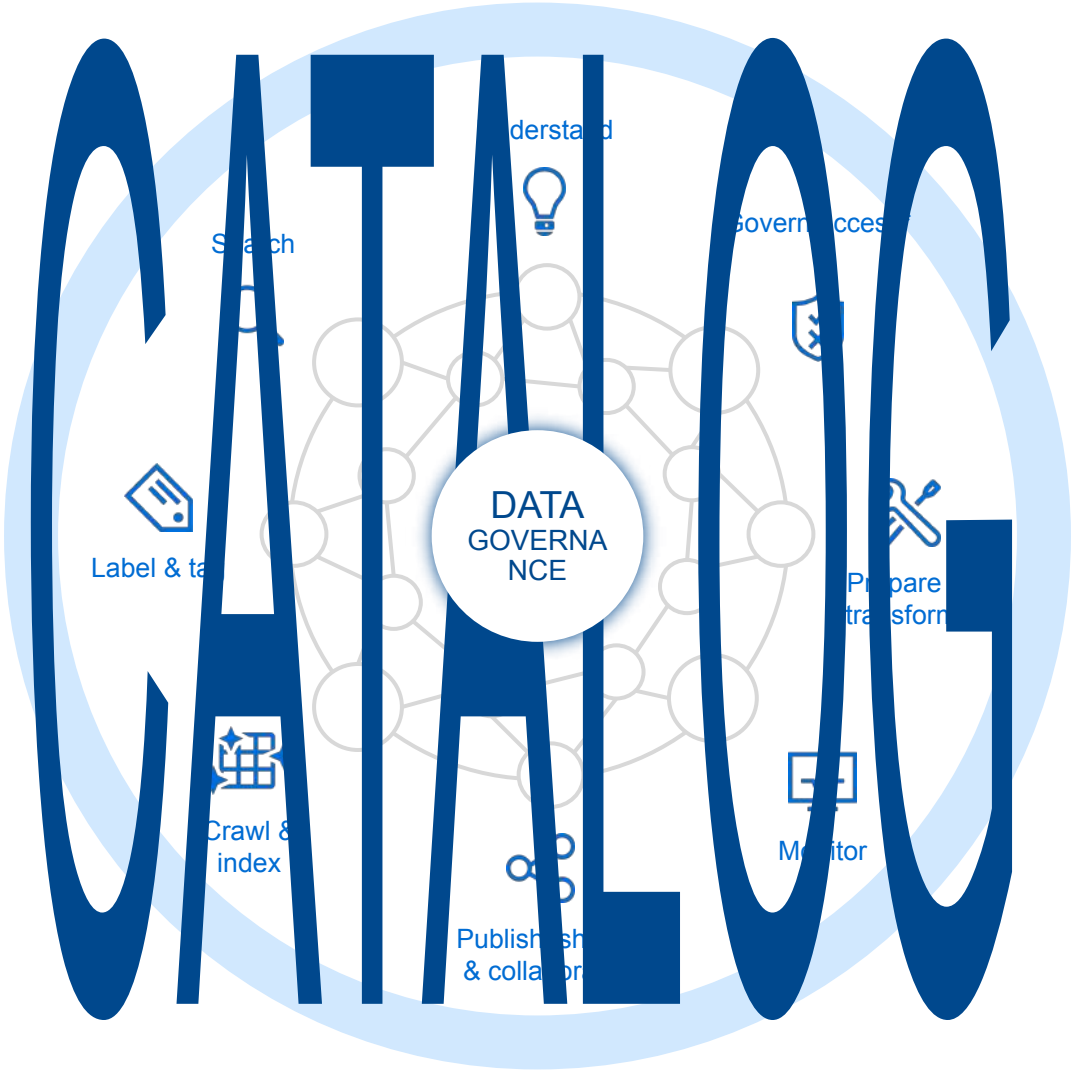
Discover data (*crawl*, *profile*, *organize*, *link*, and *enrich*) and make data **accessible** (*browse*, *search*)

Organize

Classify, organize, and **understand data** (*location*, *attributes*, *quality*, *lineage*, *sensitivity*)

Curate

Enforce **centralized authorization** & **security** for data orchestration, and control **data quality standards**



* not a current feature, roadmap items.

Catalog use cases

Data discovery for self-service & data democratization

- Metadata extraction from SAP sources including SAP Datasphere, SAP BW Bridge, SAP BW/4HANA, SAP Analytics Cloud, SAP business applications and others
- Empower users to be data-driven
- Increase productivity with search and filtering capabilities

Data governance for driving collaboration *

- Ensures the proper usage of data throughout the organization
- Exchange metadata with 3rd party catalog providers
- Monitor activity, track workflows, and enforce organizational policies

Comprehending data with metadata & lineage

- Lineage and Impact analysis for data provenance incl. lineage from SAC planning models to data in Datasphere and generated tables or views
- Minimize impact of changes by understanding relationships between assets
- Track transformation of data from source to point of use

Data literacy through glossary & KPIs

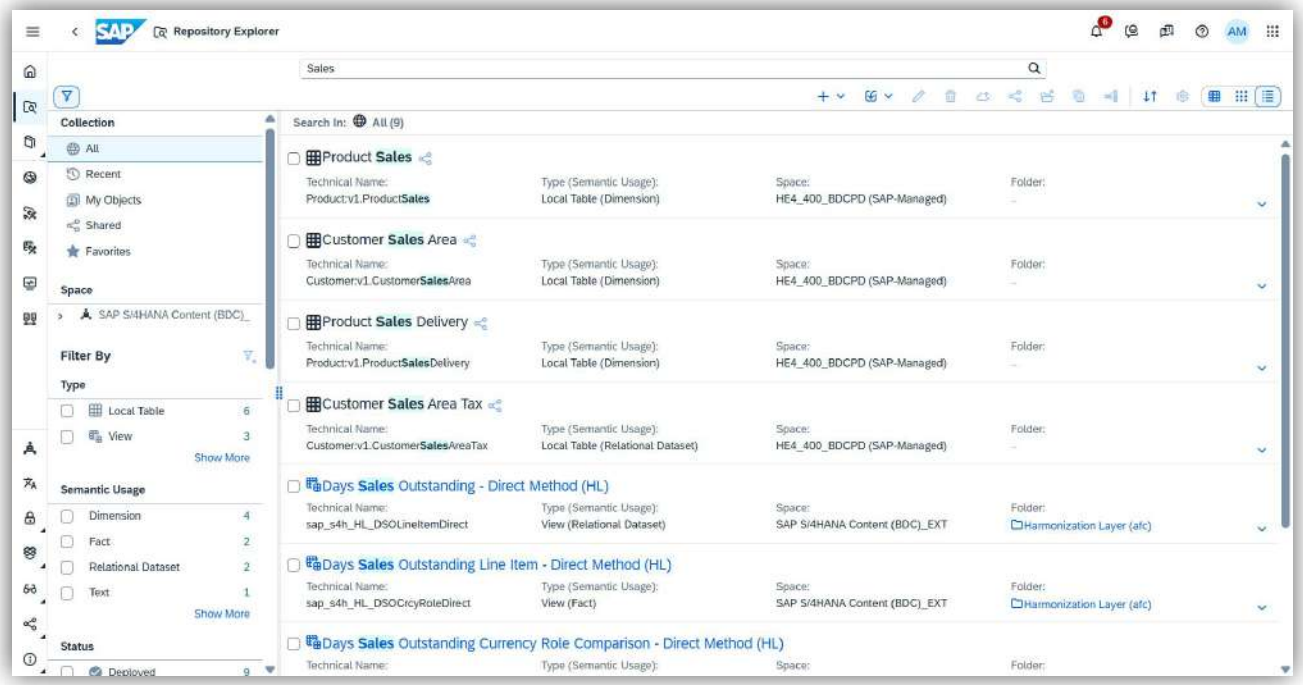
- Establish common understanding of data across enterprise
- Define KPIs for quantifiable, outcome-based strategic goals
- Import glossary terms and KPIs
- Save time and focus attention on what matters most

* not a current feature, roadmap items.

Repository explorer

The repository explorer gives you access to create, find, organize, and share objects in your SAP Datasphere tenant

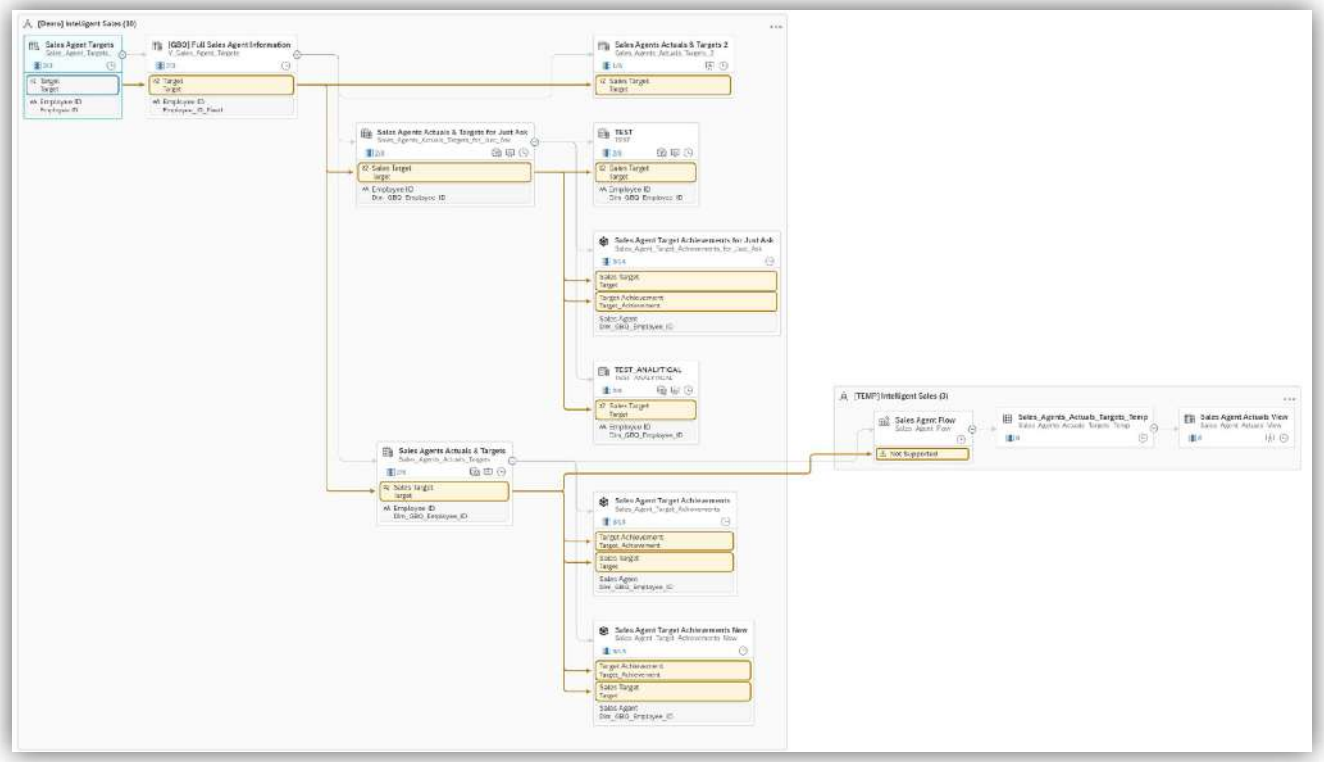
- Start all modeling activities from one place
- Get an instant overview on all available objects
- Create folders to organize objects in spaces
- Mass deployment of objects
- Intuitive fuzzy search function
- Sorting, filtering, grouping
- Favorites list



Impact and lineage

Helps you to understand the lineage of a selected object, along with its impacts

- Impact and data lineage diagrams are available from repository explorer, data builder and various editors
- Diagrams contain information about the source of the data, transformations, associations, data products, data providers, data access controls, affected objects, incl. shared and unauthorized objects:
  - **Lineage** (displayed to the left of the object) shows objects that are used as sources and allows to trace errors back to the root cause
  - **Impact** (displayed to the right of the object) shows objects that use the object as a source and allows to understand the impact of changes on dependent objects
  - The **data analysis** diagram focusses on data movements and transformations from source to destination across spaces
  - The **dependency analysis** diagram additionally displays objects connected through associations and data access controls
  - Review the lineage and impacts of **one or more columns** in your analyzed object



Different levels of authorizations in SAP Datasphere

Before introducing scoped roles

User level

Providing access to a **SAP Datasphere tenant** is the most fundamental way to control access of users.

Role level

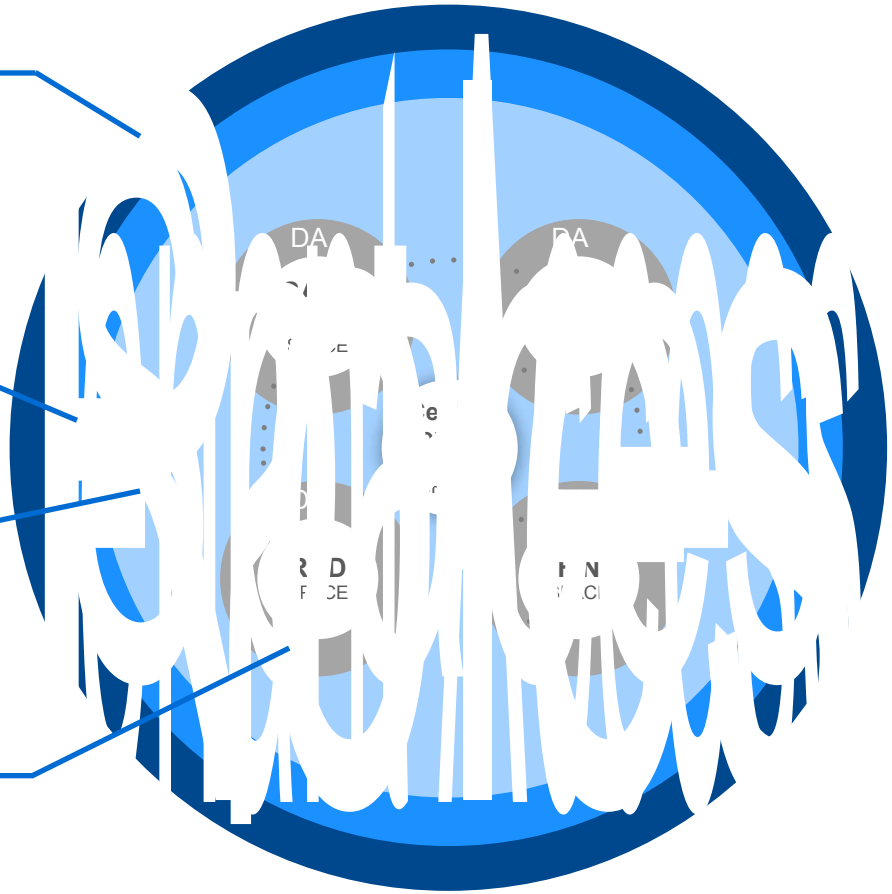
Assigning **roles** with dedicated privileges gives users access to functionalities according to their tasks in all spaces the user is assigned to.

Space level

Users can be assigned to **spaces** by a DW space administrator inheriting the functionalities from the roles.

Row-level security

Data access controls establish authorizations on the data itself based on row-level access for each user.



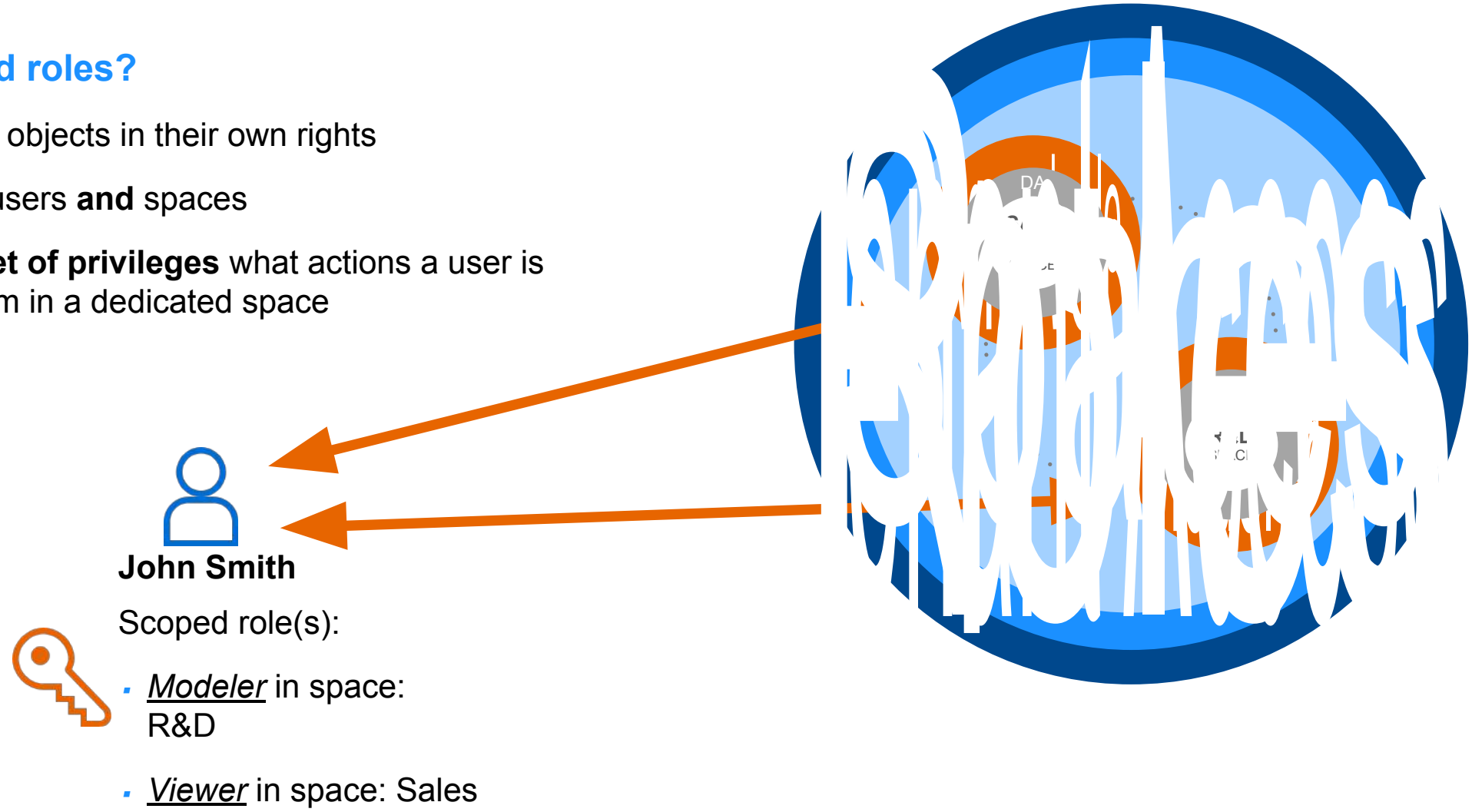
Challenge
The roles of users are the same for all spaces they are assigned to, and cannot be different from space to space.

Different levels of authorizations in SAP Datasphere

Introducing scoped roles

What are scoped roles?

- Scoped roles are objects in their own rights
- Are assigned to users **and** spaces
- Define by their **set of privileges** what actions a user is allowed to perform in a dedicated space



John Smith

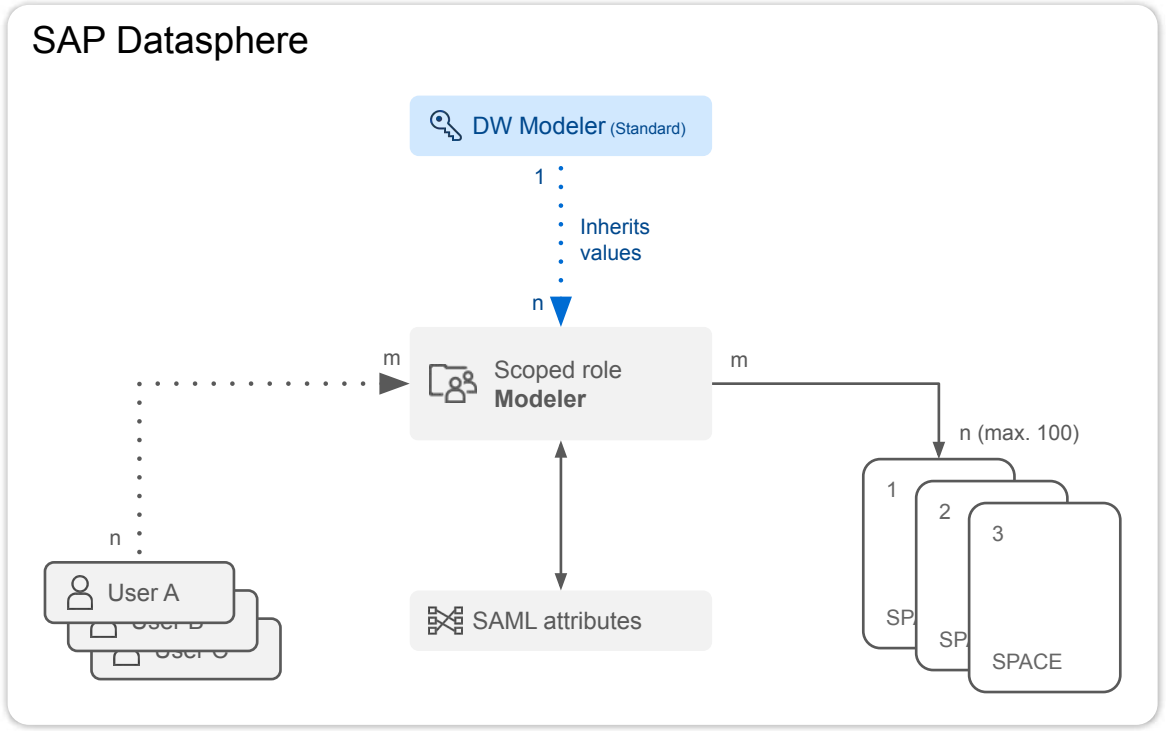
Scoped role(s):

- *Modeler* in space: R&D
- *Viewer* in space: Sales

Global standard and custom roles

Global roles are available as standard roles and custom roles

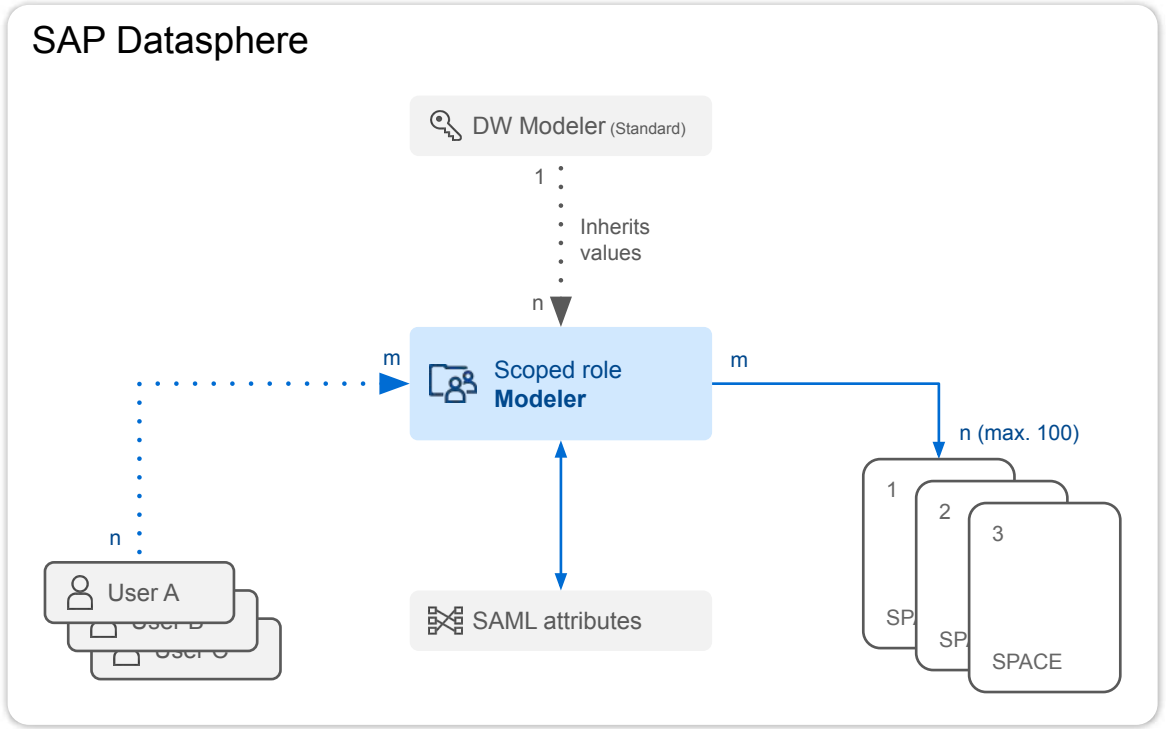
- Predefined standard application roles
  - System Owner, DW Administrator, DW Space Administrator
  - DW Consumer, DW Viewer
  - DW Integrator, DW Modeler
  - Catalog Administrator, Catalog User
- Global standard & custom roles
  - Carry tenant-wide privileges should be assigned to dedicated users by the tenant administrator
  - Serve as templates for scoped roles which inherit the according Space dependent privileges and assign them to users and spaces
  - Can still carry some privileges which are valid tenant-wide more information: privileges and permissions (help.sap.com)



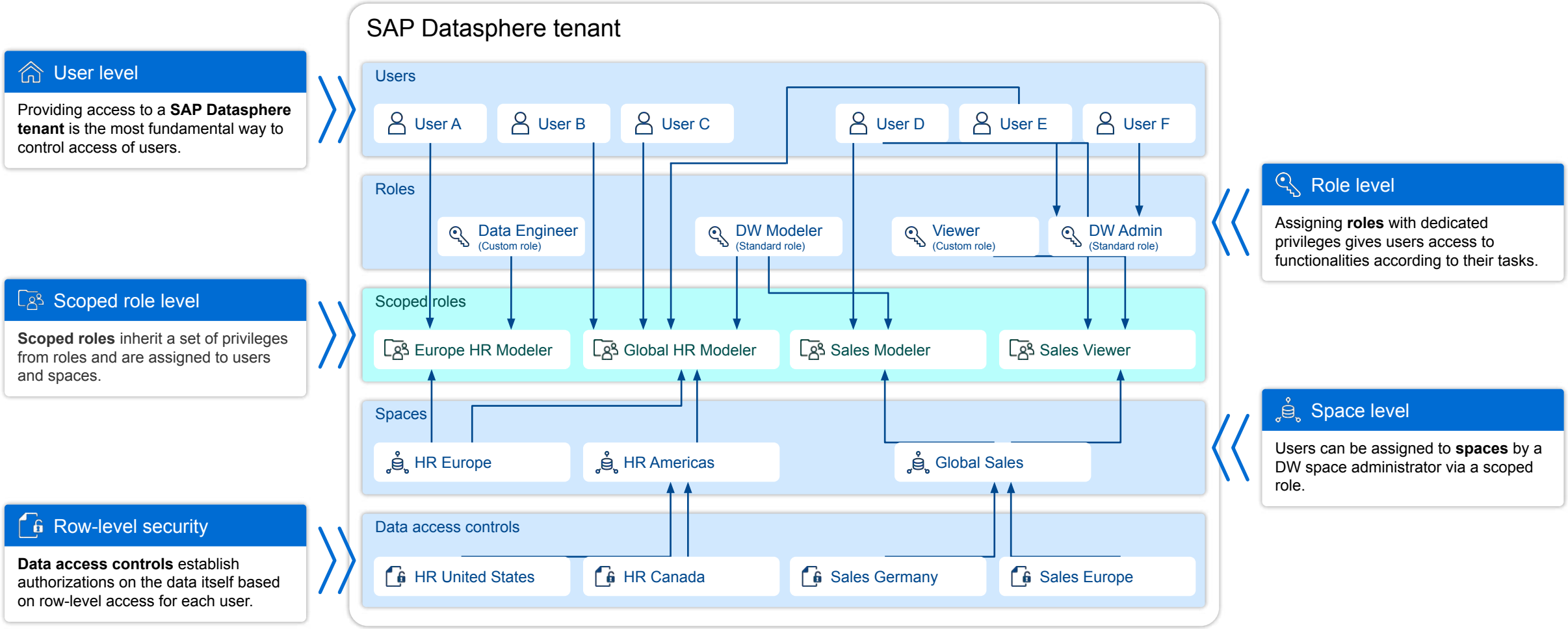
Scoped roles

Scoped roles inherit a set of privileges from roles and are assigned to users and spaces

- Scoped roles are objects in their own rights and allow users to have a different set of privileges dependent on the assigned space, e.g. be a modeler in space 1 and a viewer in space 2
- They inherit the space dependent privileges from a selected standard or custom role and can be assigned to spaces (scopes)
- Scoped roles are defined and assigned in central role management
- The assignment of users can be handled in addition in user management and decentral in space management



How the different levels of authorizations work together



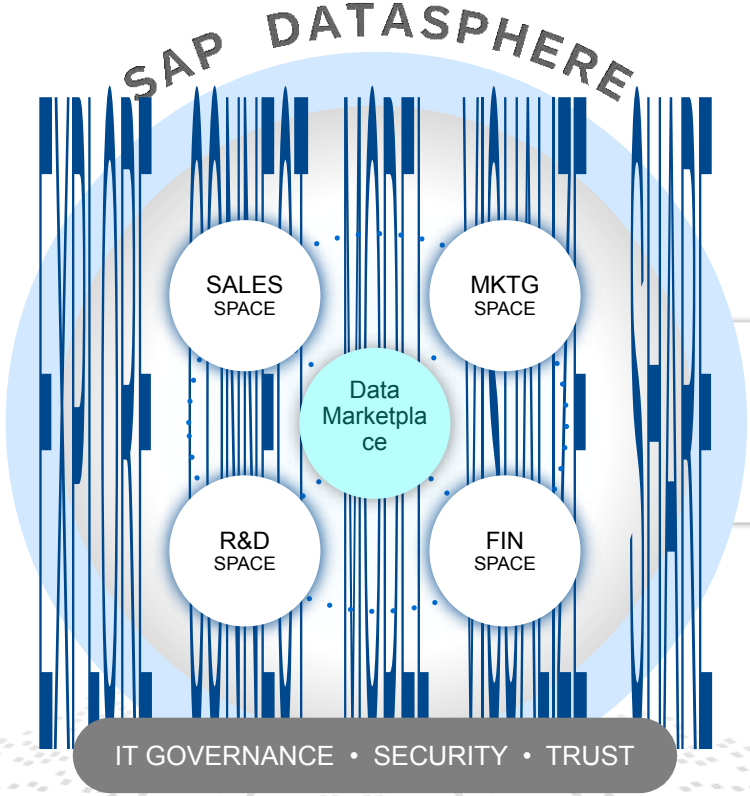
Discovery, sharing, self-service

Reusable data products can be accessed in clicks for better, faster & more scalable data collaboration across spaces

Independent virtual work environments
for departments, LOB's, data domains, project teams, and individual

Data sharing fosters collaboration
between IT, business units, and projects to connect global & local data

Get in control of data sprawl
by reusing shared data & avoid maintaining disconnected data sets



Public & private data exchange
with other enterprises to access external data, collaborate or monetize data

Open partner ecosystem

Data marketplace
3500+ data products

Customers with SAP Datasphere

SAP BW bridge | Native connectors | Open interfaces

SAP data | 3rd party application data

It all starts with a space in SAP Datasphere

What is a space?

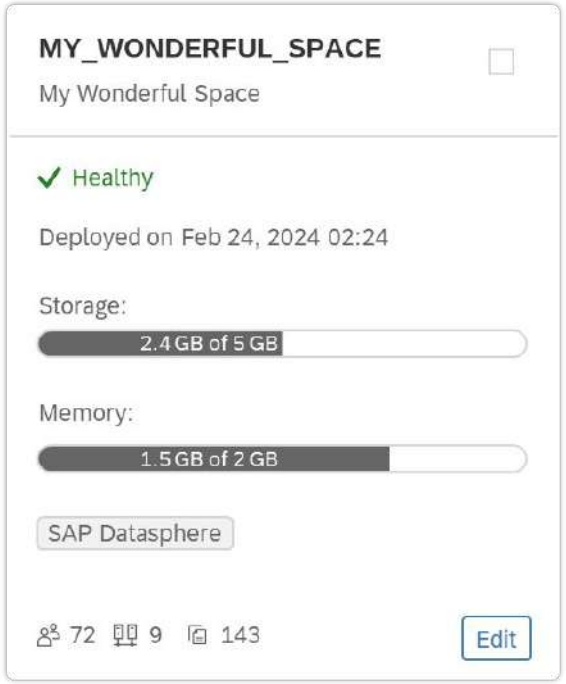
A space is a secure area created by an administrator, in which members can acquire, prepare, and model data.

Resources
Storage, memory, workloads

Storage type
SAP HANA, SAP HANA Data Lake Files

Compute runtime
Elastic compute nodes, SAP HANA, Apache Spark

Recycle bin
Logical space deletion



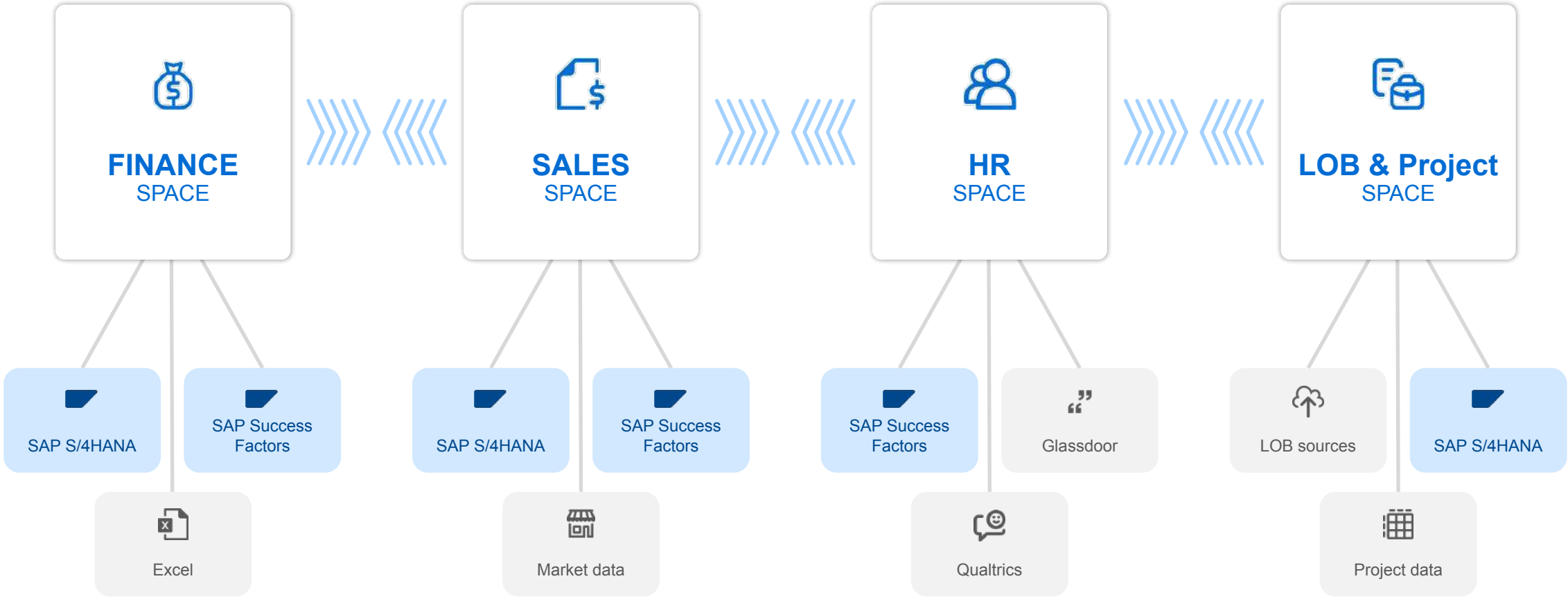
Users
Only members get access

Connections
SAP on-prem & cloud, non-SAP systems

Models
Tables, views, flows, translation, …

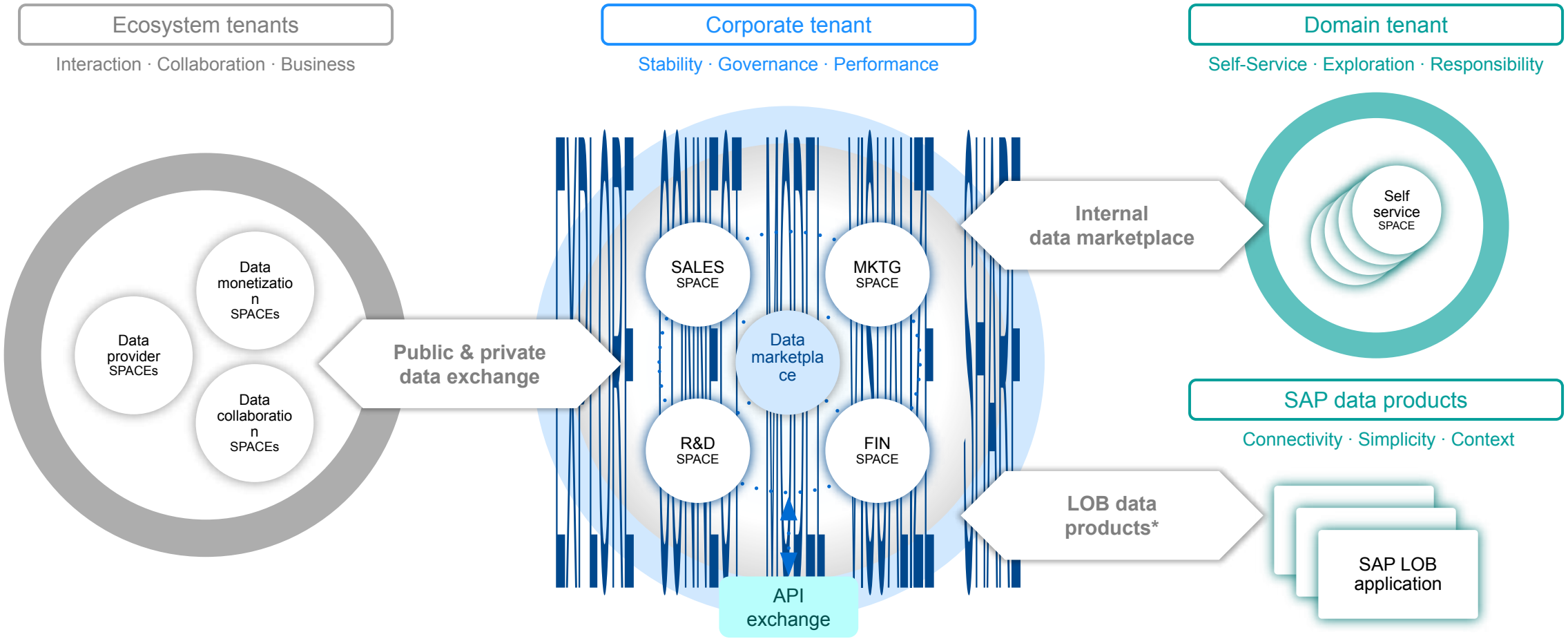
Sharing
Tables & views

Spaces provide your business process with the right data



The data marketplace for SAP Datasphere

Facilitates self-service data exchange within and between enterprises to be a driving force for data democratization



*see SAP Datasphere Roadmap

Key components of the data marketplace

Data Provider

- Create Data Products on top of existing data assets
- Democratize data within organizations

Context

- Define visibility of Data Products
- Clearly define group of Data Consumers per Data Product
- Invite users, tenants or 3rd party stakeholders
- Define who is able to ship and consume data with predefined context types

Data Product

- Contains Views on top of data
- Provides metadata information
- Different delivery methods available

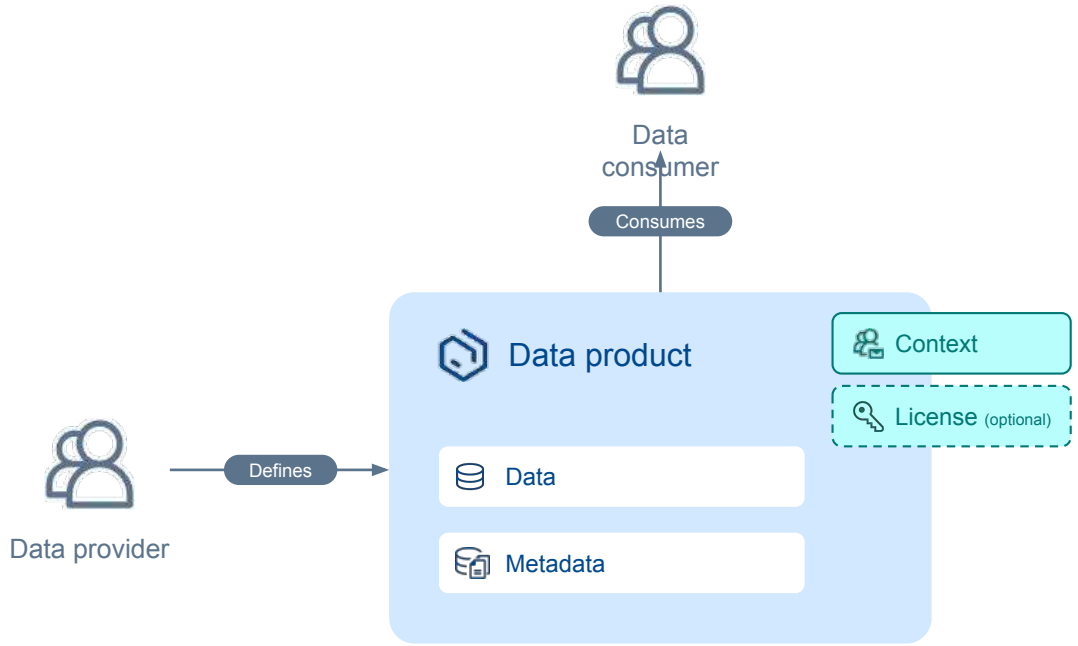
License

- Manage accessibility of Data Products
- Leverage licenses in commercial set ups and/or for sensitive data
- Allow license specific data selection

Data products in SAP Datasphere

Data products in SAP Datasphere inherit data & metadata from a variety of data models and sources

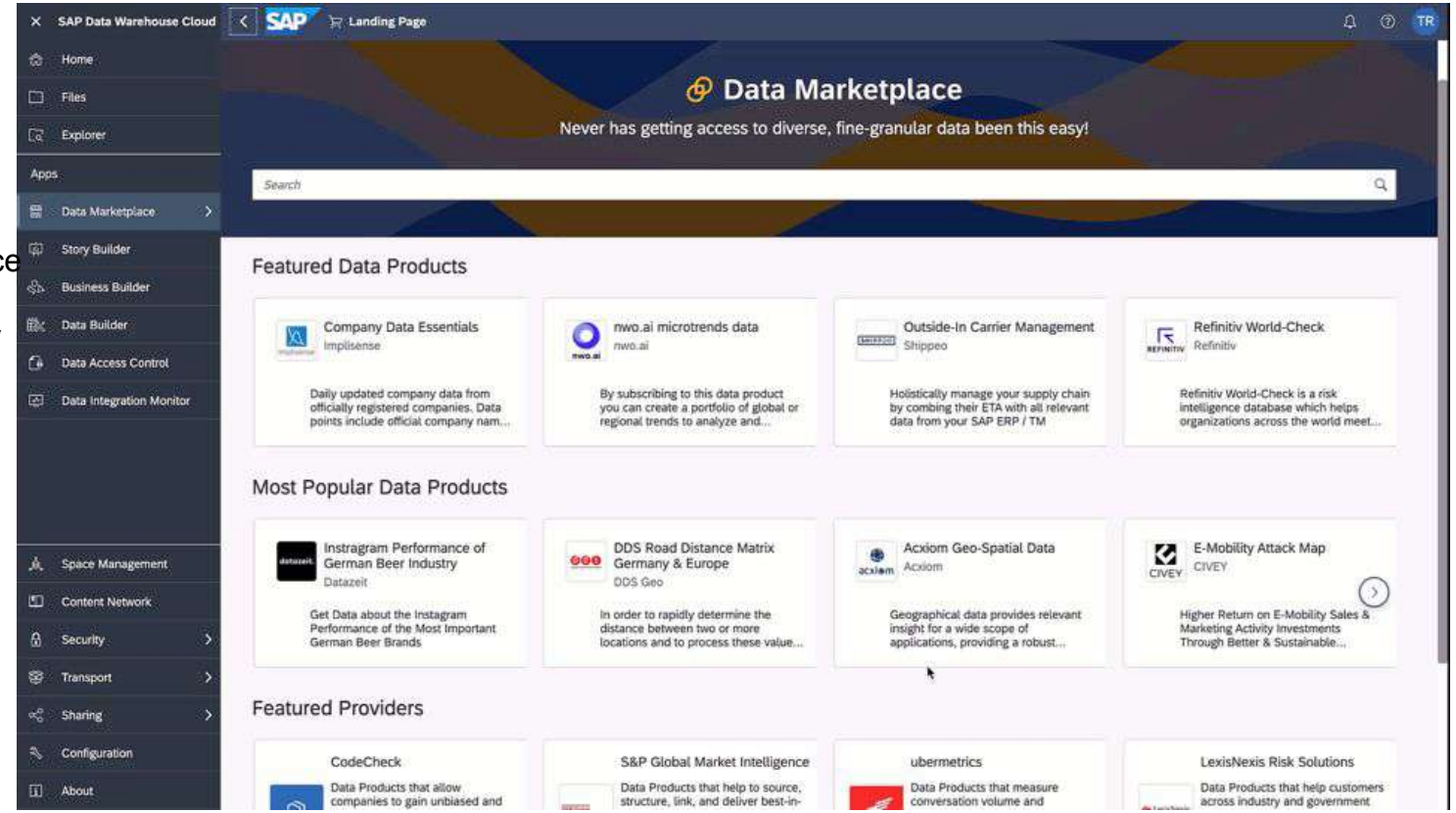
- Defined by data providers to be used by data consumers
- Contains data and describing metadata
- Controlled visibility via context management
- Controlled access via license (optional)
- Provides data directly to the selected space
- Can be provided within a single tenant and/or across tenants
- Easy to use onboarding processes for data consumers
- Light-weight processes for data product creation, duplication and distribution
- Allow license specific data selection



Data sharing & discovery using the data marketplace

1

In-app data discovery experience for transparent evaluation of internal and more 3000+ 3rd party data products through the cooperation with Datarade.ai. More can be added on request.



2

Embedded data sharing cockpit for self-service data sharing for LoBs and IT. Creates data product endpoints in a business user friendly way.

3

Powerful context management to steer visibility of data products in public, private and internal setups.

4

Access management through licenses handles access to data down to row- and column level where needed.

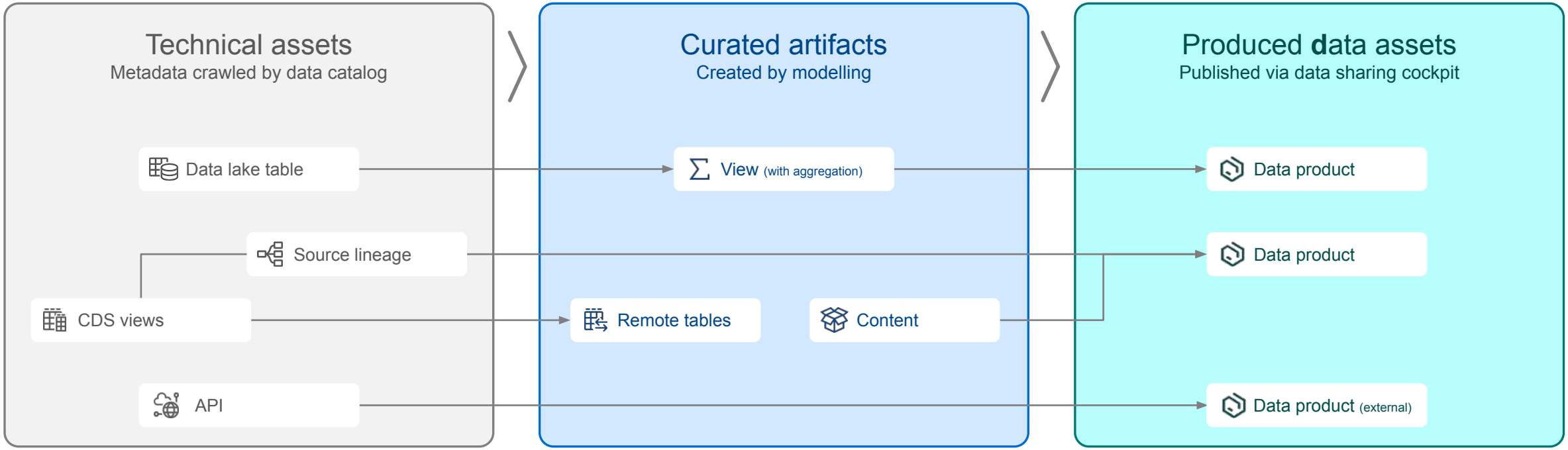
5

Managed data integration & update management for sustainable adoption of external data sources via several data delivery options.

Convergence of catalog & marketplace – one place to find all assets

Aspiration: converged user experience

One search. One process. One profile. One lineage.
One onboarding for E2E SAP & ecosystem managed content
(Catalog entities, Datasphere artifacts & data products)



Analytics & planning consumption

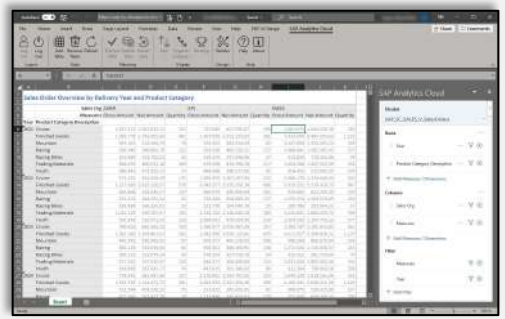
Consumption in SAP Analytics Cloud and add-in for MS Office

Tight integration with SAP Analytics Cloud for analytics and planning

SAP Analytics Cloud, analytics

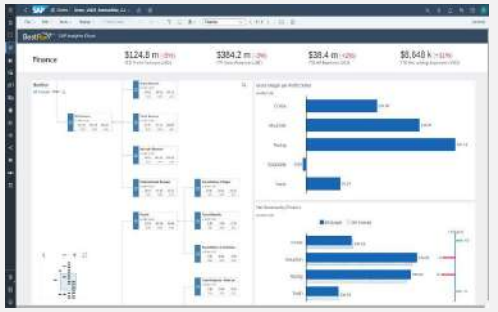
- Direct consumption of SAP Datasphere data models in SAP Analytics Cloud using a live connection
- Multiple SAP Datasphere tenants can be connected to any number of SAP Analytics Cloud systems

Microsoft Office



- SAP Analytics Cloud, add-in for Microsoft Office 365 with live connection for online or desktop version
- SAP Datasphere with SAP Analysis for Microsoft Office 2.8 SP14+ - limited[1]

SAP Analytics Cloud, planning



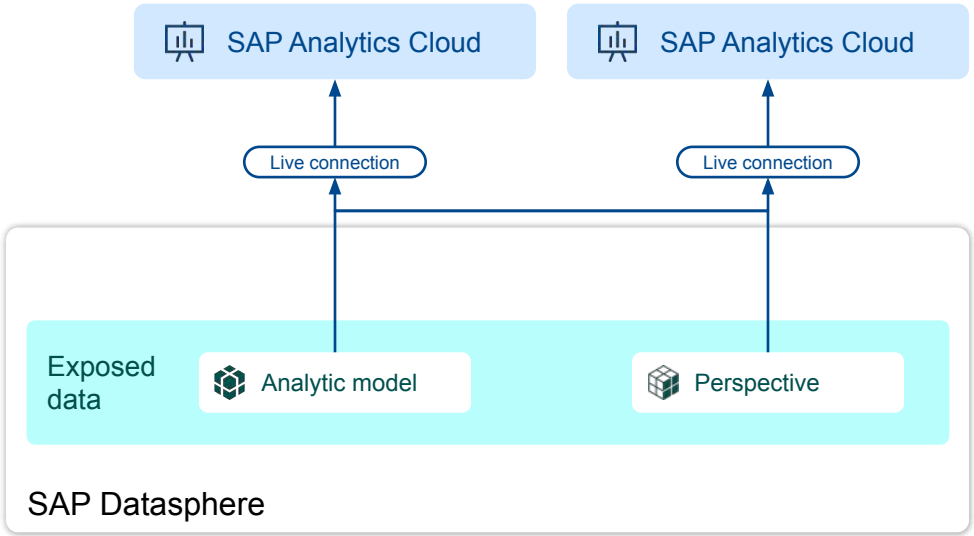
- Bi-directional integration of SAP Datasphere and SAP Analytics Cloud for planning
- Use OData API to pull data from SAP Datasphere into planning
- Use data export service API to pull data into SAP Datasphere

[1] No support of Analytic Models

SAP Analytics Cloud, analytics

Consume data in SAP Analytics Cloud via a live connection

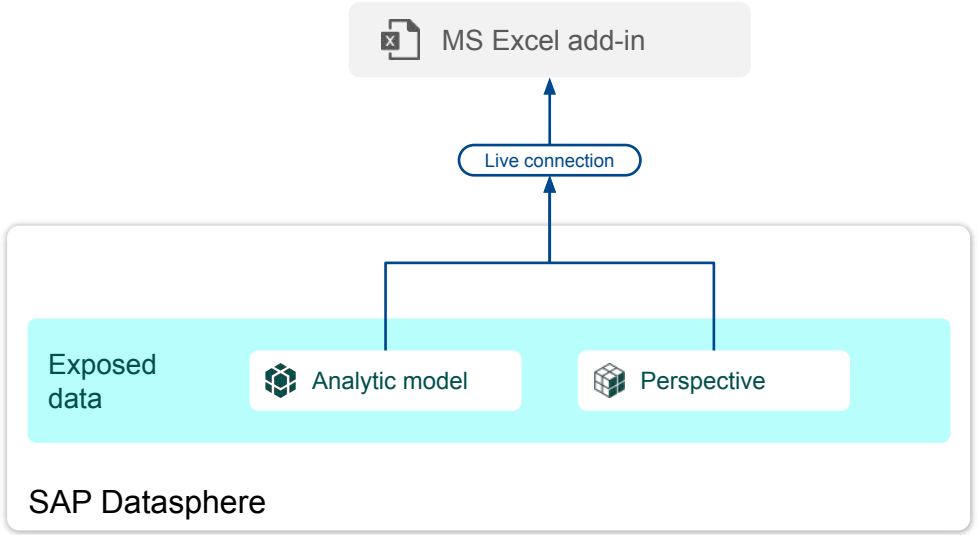
- Direct consumption of data models in SAP Analytics Cloud via live connection
- Ability to connect live to source data sets in SAP Datasphere, create stories SAP Analytics Cloud, and change the source model while preserving the content
- SAP Datasphere tenants can be connected to any number of SAP Analytics Cloud systems, and vice versa
- The live connectivity needs to be set up for each SAP Analytics Cloud system
- Metadata translation can be enabled for SAP Analytics Cloud stories
- SAP Note 2832606 - Limitations with live connections
- More information



SAP Analytics Cloud, add-in for Microsoft Office

Consume data in Microsoft Excel via the SAP Analytics Cloud add-in for Microsoft Office 365

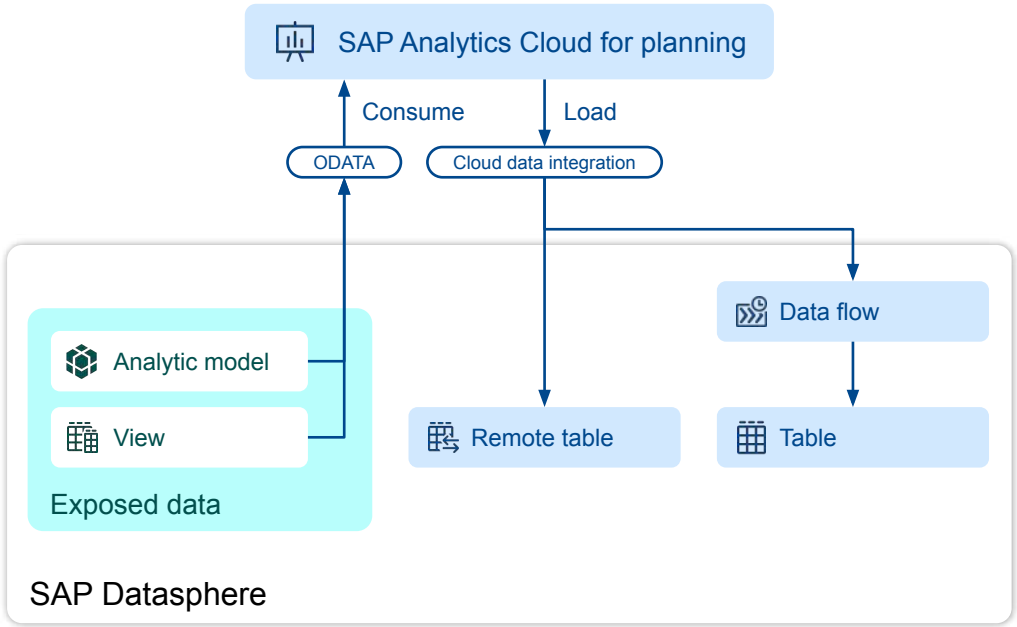
- Direct consumption of data models in SAP Analytics Cloud, add-in for Microsoft Office via live connection
- Ability to connect live to source data sets in SAP Datasphere and create MS Excel reports
- Use Microsoft Office 365 online or desktop version
- Add-in available via Microsoft store
- Find more information in the SAP community for the SAP Analytics Cloud, add-in for Microsoft Office and learn how to use it in this blog
- SAP Note 2832606 - Limitations with live connections



SAP Analytics Cloud for planning

Integrate with SAP Analytics Cloud for planning

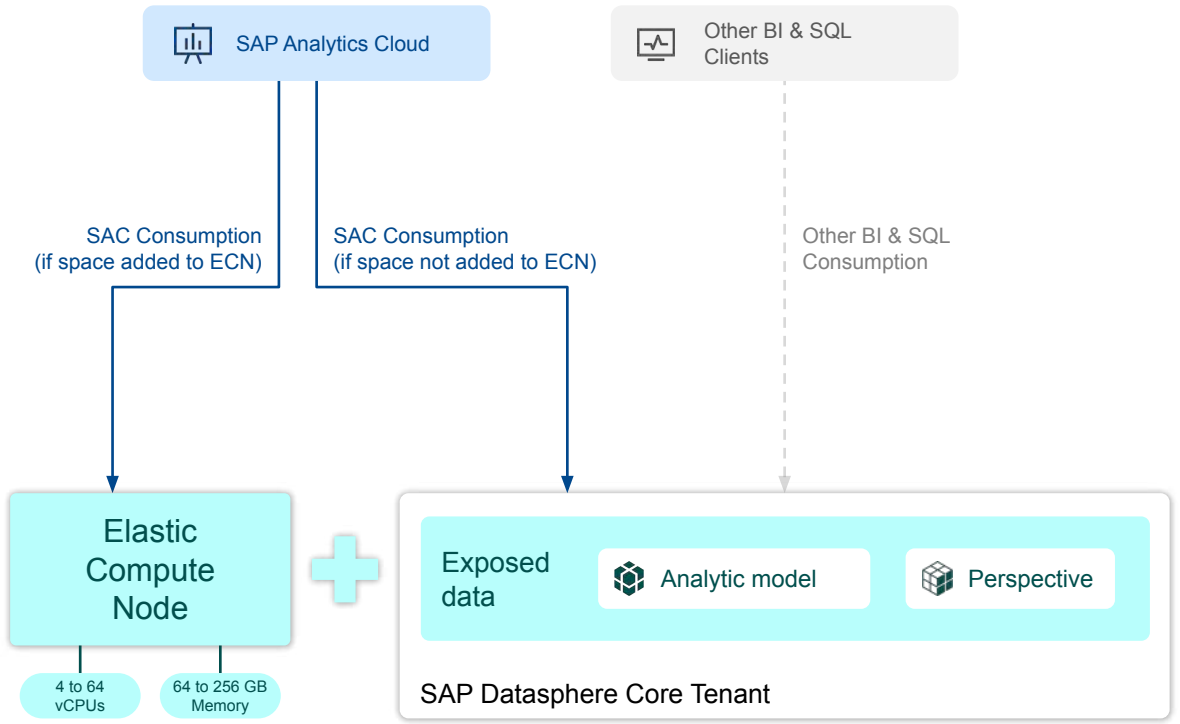
- Load actuals data from SAP Datasphere into SAP Analytics Cloud for use in your planning models via an OData Services Connection
- Load data from SAP Analytics Cloud into SAP Datasphere as a remote table or via a data flow and optionally combine it with actuals data or other sources
- Find more information in these blogs:
  - Introducing the bi-directional integration of SAP Datasphere and SAP Analytics Cloud for planning
  - Using the SAP Analytics Cloud data export API with SAP Datasphere and SAP Data Intelligence
- SAP Note 2832606 - Limitations with live connections



Elastic Compute Nodes (ECNs)

Provision additional resources for a specific time period to cover peak consumption loads

- Dedicated instances for serving SAP Analytics Cloud Consumption for selected space(s)
- SAP Analytics Cloud (SAC) consumption towards spaces not added to the ECNs, continue to be served by the main SAP Datasphere tenant
- Configure ECN instances with 4 different memory sizes – 64 GB, 128 GB, 196 GB and 256 GB
- Choice of a performance class for all the ECNs in the tenant – Memory, Compute and High Compute
- Start up to 4 ECN instances in parallel
- Start & stop an ECN instance manually or via a schedule



Openness

Open for data ecosystem partners

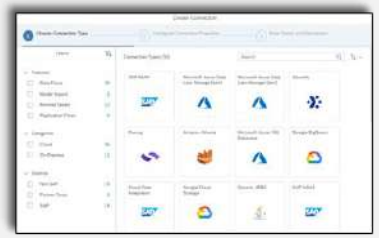
Together with our open data partner ecosystem

Open data ecosystem

Google Cloud · CONFLUENT · DataRobot · Collibra · databricks

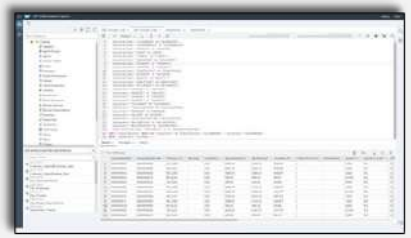
- Integrate with industry-leading data and AI platforms to capitalize on data investments
- Bring SAP and non-SAP data together with SAP Datasphere leveraging tailored integration with solutions from open data ecosystem partners

Data integration



- Allow external data integration tools like SAP Data Services, SAP Data Intelligence, SAP Open Connectors, SnapLogic, Precog, Adverity, etc. to bring data into the system using SQL interfaces and the open SQL schema

Data modeling

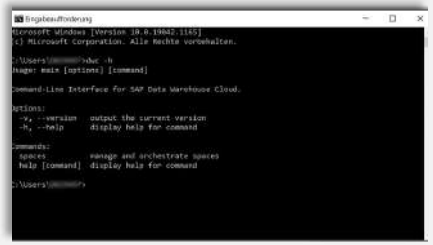


- Option to use SQL DDL & DML in the open SQL schema leveraging for instance SQL views, tables, procedures, etc.
- Leveraging existing SQL-tooling & skillset
- SAP Database Explorer

Data consumption

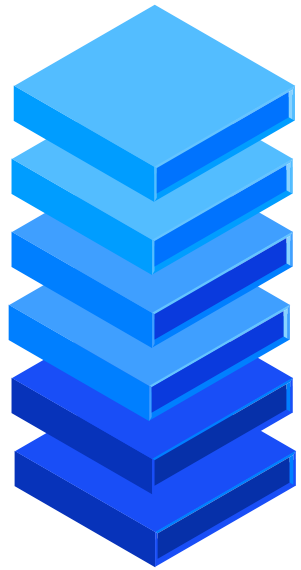
- External SQL & OData API for consumption
- Use consumption interface to connect any 3rd party front end tool to your exposed views
- Make your data models accessible for consumption tools & applications

Command line



- Support external automation tooling via Command line interface (CLI – API)
- The CLI allows you to outsource repeated and time-consuming tasks programmatically

Provide a holistic data solution together with our open data partners

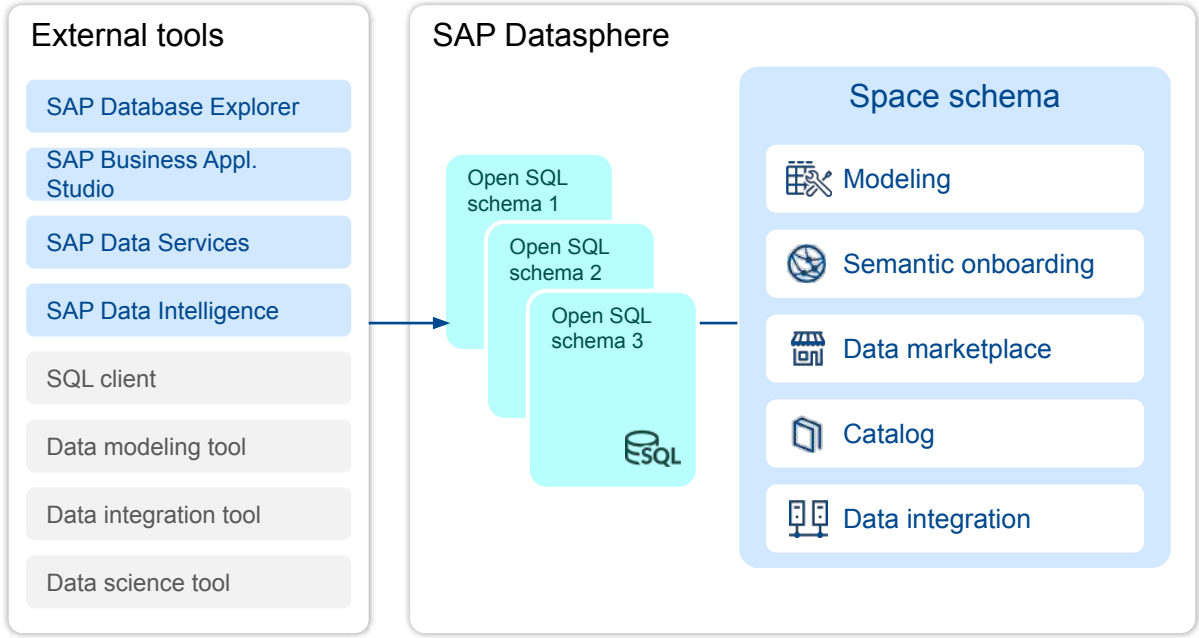


Collibra	Collibra delivers the best-of-breed cloud platform that tightly integrates data catalog, governance, lineage, quality, and privacy capabilities.
CONFLUENT	Confluent delivers the leading cloud service for connecting and processing real-time data within a fully-managed Apache Kafka environment.
DataRobot	DataRobot empowers organizations to leverage augmented intelligence by making the selection and training of ML models easy.
databricks	Databricks delivers an open and unified data lakehouse platform for data, analytics and AI uses cases, based on Apache Spark.
Google Cloud	Google extends SAP data models to Google Big Query or bring Google Big Query insights directly into SAP processes with the ability to gain new analytics and AI insights from unified data.

Open SQL schema for data integration & data modeling via SQL API

Enabling external tools to create artifacts and to write data to an open SQL schema

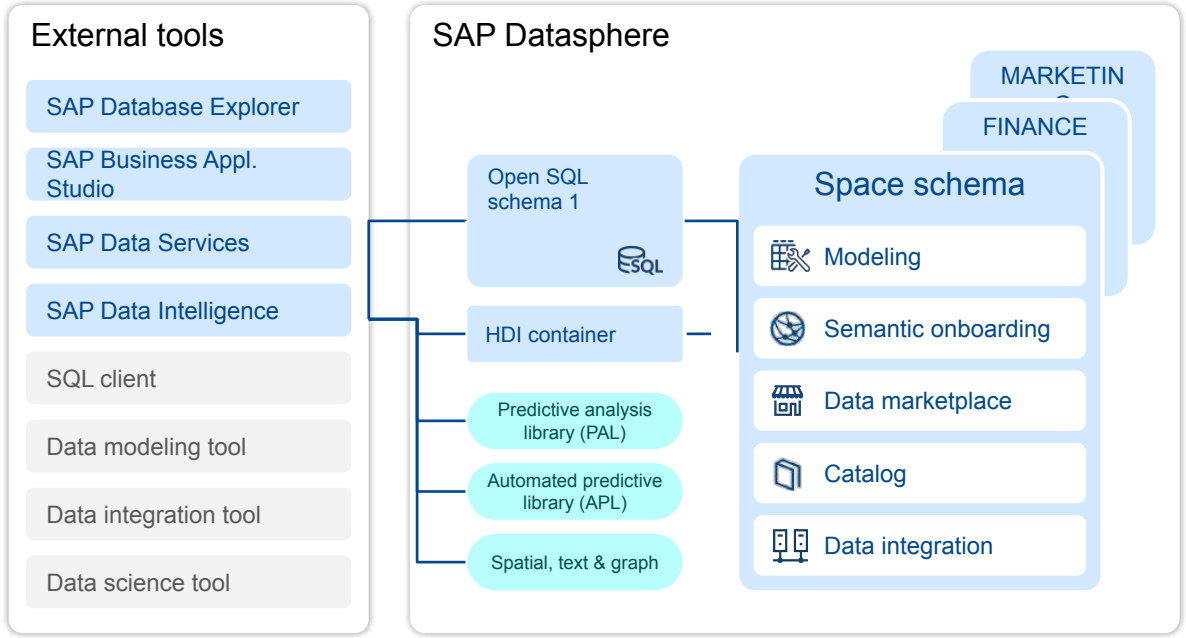
- Connect via an external data integration tool or SQL client to a SQL endpoint of an open SQL schema
- Enable write access (SQL, DDL & DML)
  - Create views, tables
  - Define & execute stored procedures
  - Leverage data anonymization and data masking
  - Automated Predictive Library (APL) and Predictive Analysis Library (PAL), if enabled in Space
  - Use data science tooling like Jupyter notebooks
- Consume data models from your space schema in the Open SQL schema



Embedded machine learning and advanced analytics

Data science enrichment in SAP Datasphere without data extraction

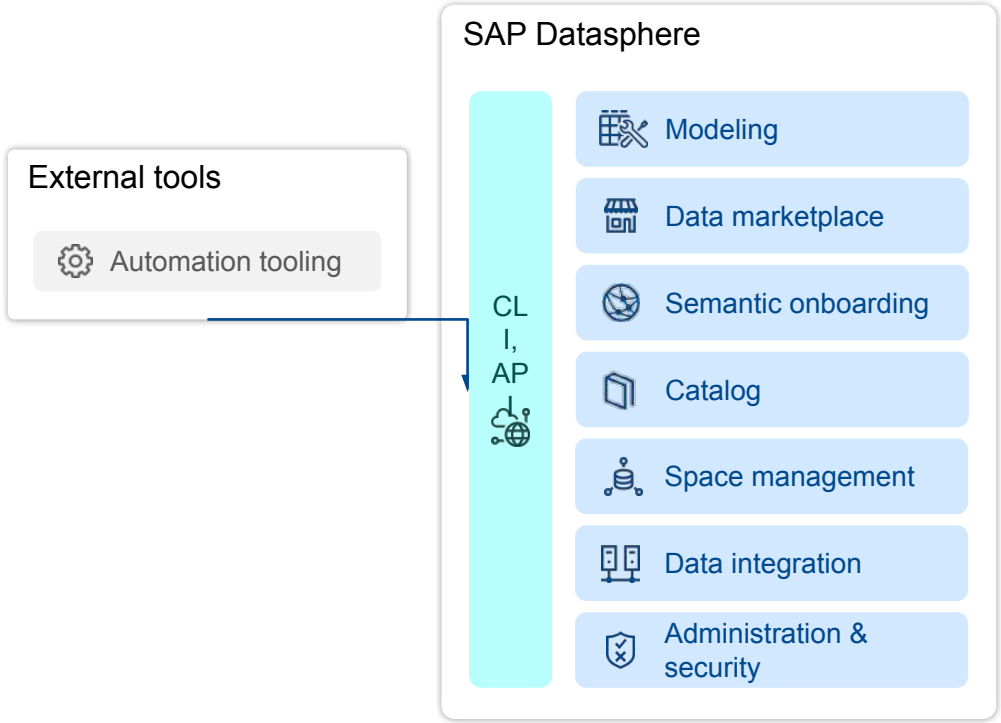
- Leveraging SAP HANA Cloud data science capabilities
  - Native Python machine learning client for SAP HANA Cloud
  - Native R machine learning client for SAP HANA Cloud
  - Exposing the data as SAP HANA data frame in Python
  - Remote use of SAP HANA Cloud machine learning, spatial, text and graph functions in Python or R
- Use your existing Python or R environment to trigger calculations in SAP Datasphere
- Train ML model based on persisted and managed data in HDI container or open SQL schema
- Store ML Prediction result in table in HDI container or open SQL schema
- Ensure that the SAP HANA Cloud script server is enabled on your tenant



Command line interface

External access via CLI-API to support automation tooling that allows you to outsource repeated and time-consuming tasks programmatically

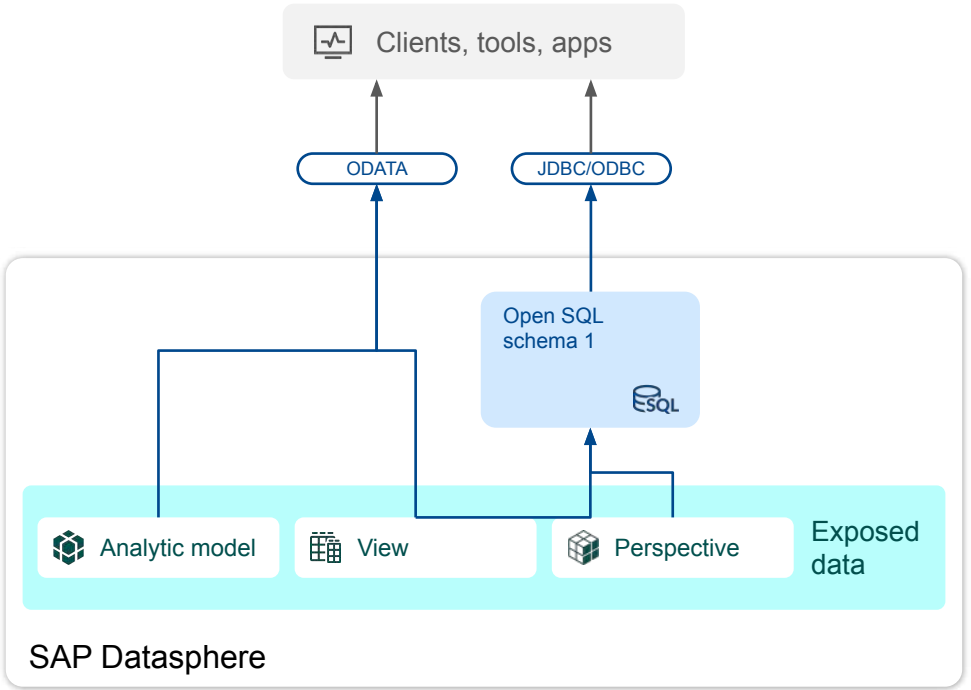
- Use the CLI to connect to SAP Datasphere and manage certain types of objects to be managed via CLI:
  - Manage users, global and scoped roles
  - Manage spaces (properties, members, database users, HDI containers, tables, views, and data access controls)
  - CRUD operations on modeling objects (create, read, update and delete)
  - Manage data marketplace data providers, data products, licenses, releases, and contexts
  - Manage tasks and task chain execution
  - Create, update and manage connections and certificates
- Supports OAuth client login and password reset
- Download the latest @sap/datasphere-cli from npmjs.com (requires the latest node.js version)
- More information



Outbound consumption

Consume exposed data in clients, tools, and apps via an OData service or JDBC/ODBC

- Data can be exposed as analytic models, views, and perspectives to external clients, tools, apps
- Connect to the OData API and consume data exposed as views or analytic models in clients, tools, and apps capable of accessing an OData service and authenticating via an OAuth client
- Connect via an Open SQL schema and JDBC/ODBC to consume data exposed as consume perspectives
- More information



Why SAP Datasphere

Foundation for business data fabric

Comprehensive data service that enables every data professional to deliver seamless and scalable access to mission-critical business data.

Connect to SAP data with business context

Out-of-the-box understanding of SAP data with semantic business context from SAP applications using semantic onboarding and prebuilt business content data models.

Connect and collect data from all sources

Data virtualization, replication, orchestration and distribution from and to SAP and external solutions across cloud and on-premises landscapes.

Agile self-service data modelling

User empowerment with self-service analytical data modeling, governance and IT control. Enable users to connect their own data, share it securely, and run analytics without affecting other users.

Space management & data sharing

Independent virtual work environments to foster collaboration and enable data sharing between IT, business units, and projects to connect global and local data.

Reuse & extend SAP BW investments

Leverage SAP BW data structures, transformations, customizations, and skills to quickly extend your SAP BW investments to the public cloud using the SAP BW bridge or hybrid scenarios.

Open interfaces and partner ecosystem

Hyperscaler choice with data centers in a variety of regions, open interfaces for data integration and consumption, and integration with industry-leading ecosystem of data and AI partners.

Governance and metadata management

Ensure governance, metadata management and data lineage throughout the data lifecycle and across systems: discover, manage, and govern all data using the catalog.

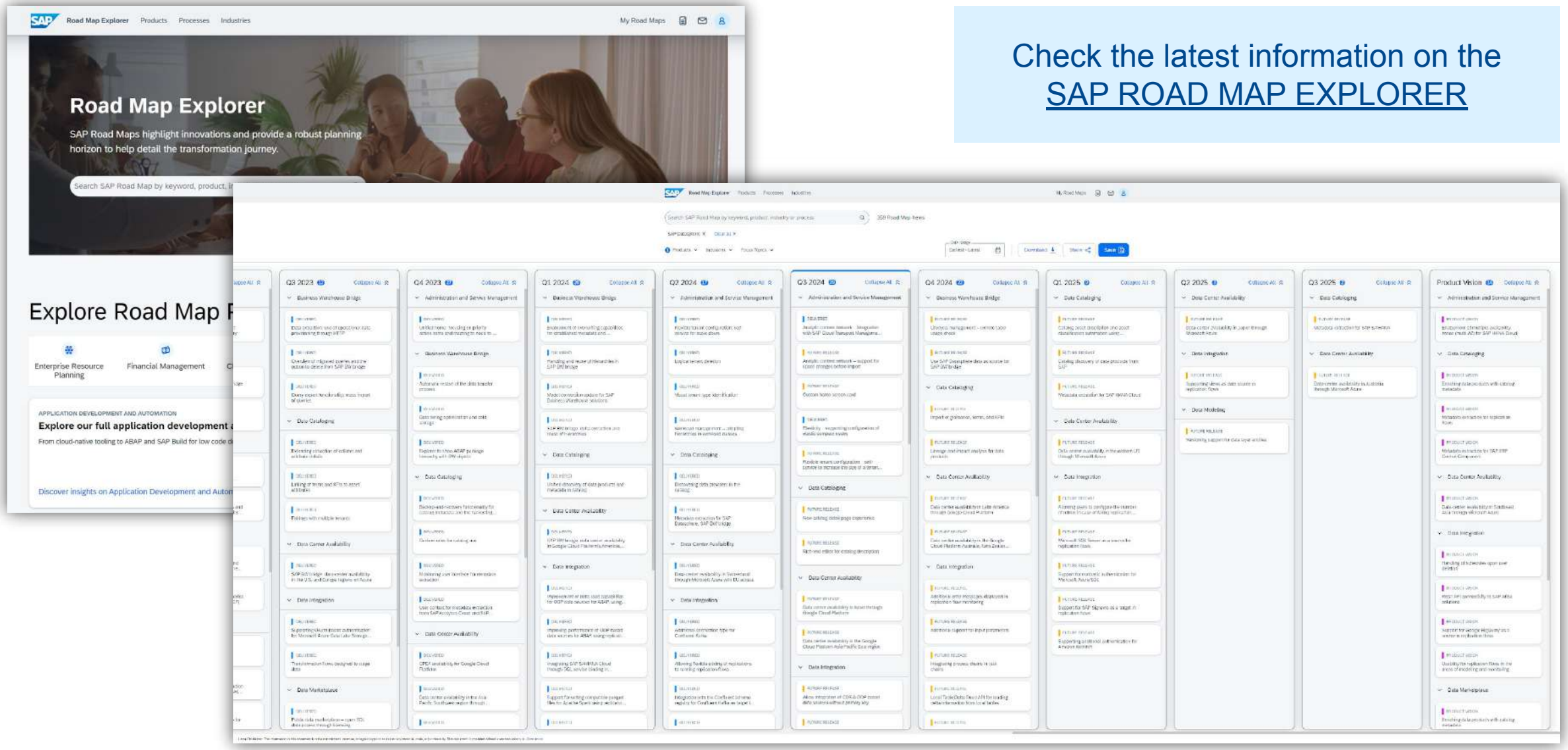
Integrated data marketplace

The data marketplace leverages trusted data partners and allows to securely share data across the organization with public, private and company-internal data products across tenants.

Foundation for planning

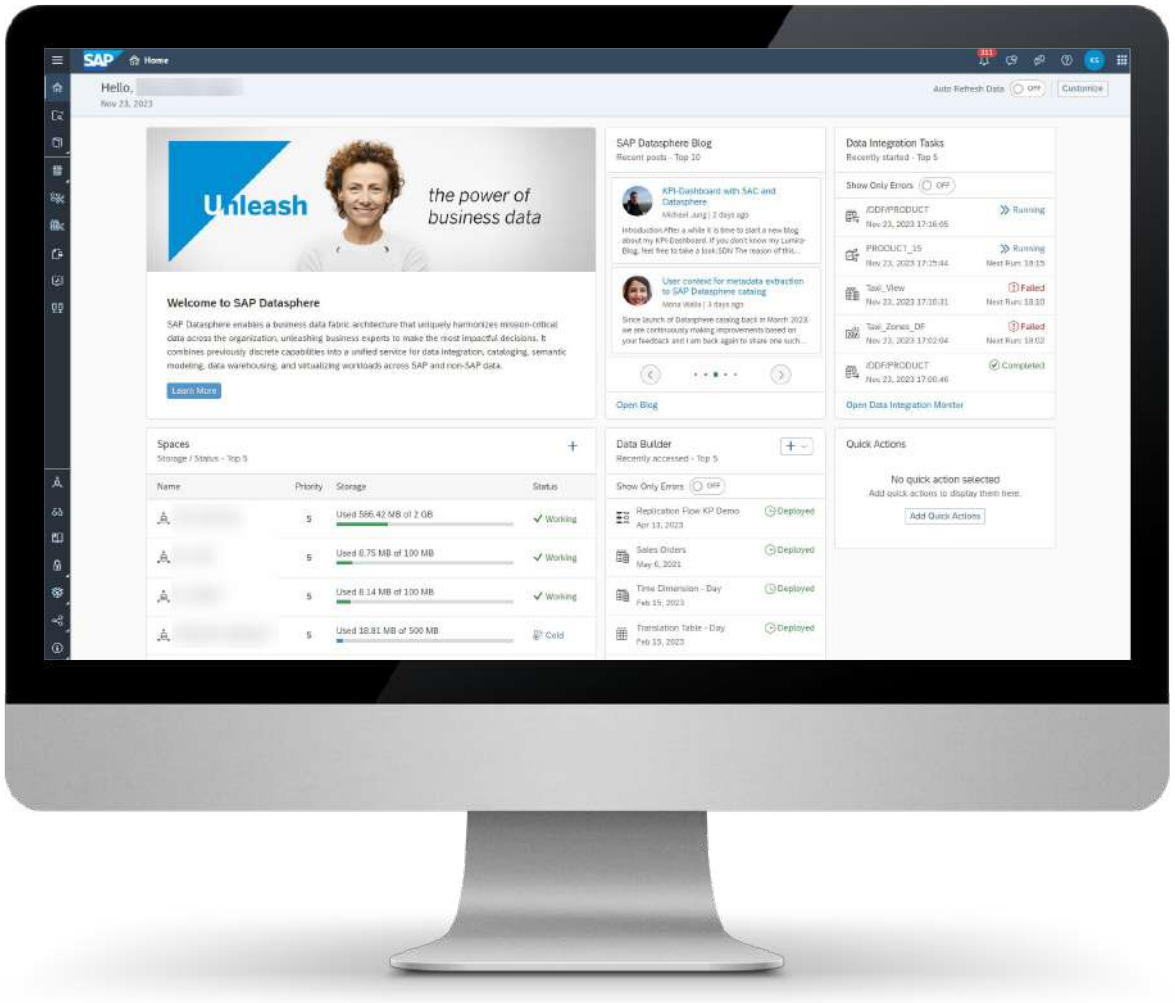
Streamlining planning and analysis with reduced modelling efforts, while operating on a single source of truth for analytics and planning across the organization.

SAP Road Map Explorer



Check the latest information on the SAP ROAD MAP EXPLORER

More information



- Check out the product page on sap.com and get access to a basic trial system
- Get started with SAP Datasphere components, concepts, and the video library
- Use the and our learning journey to learn more
- Stay updated using the release navigator
- Join the community and check out the best practices recommendations
- Add your ideas and requests for new features on the SAP Datasphere customer influence
- Online Documentation on SAP Help and support content for troubleshooting & analysis

Thank you.

Contact information:

Speaker Name