Deloitte. | databricks | SAP

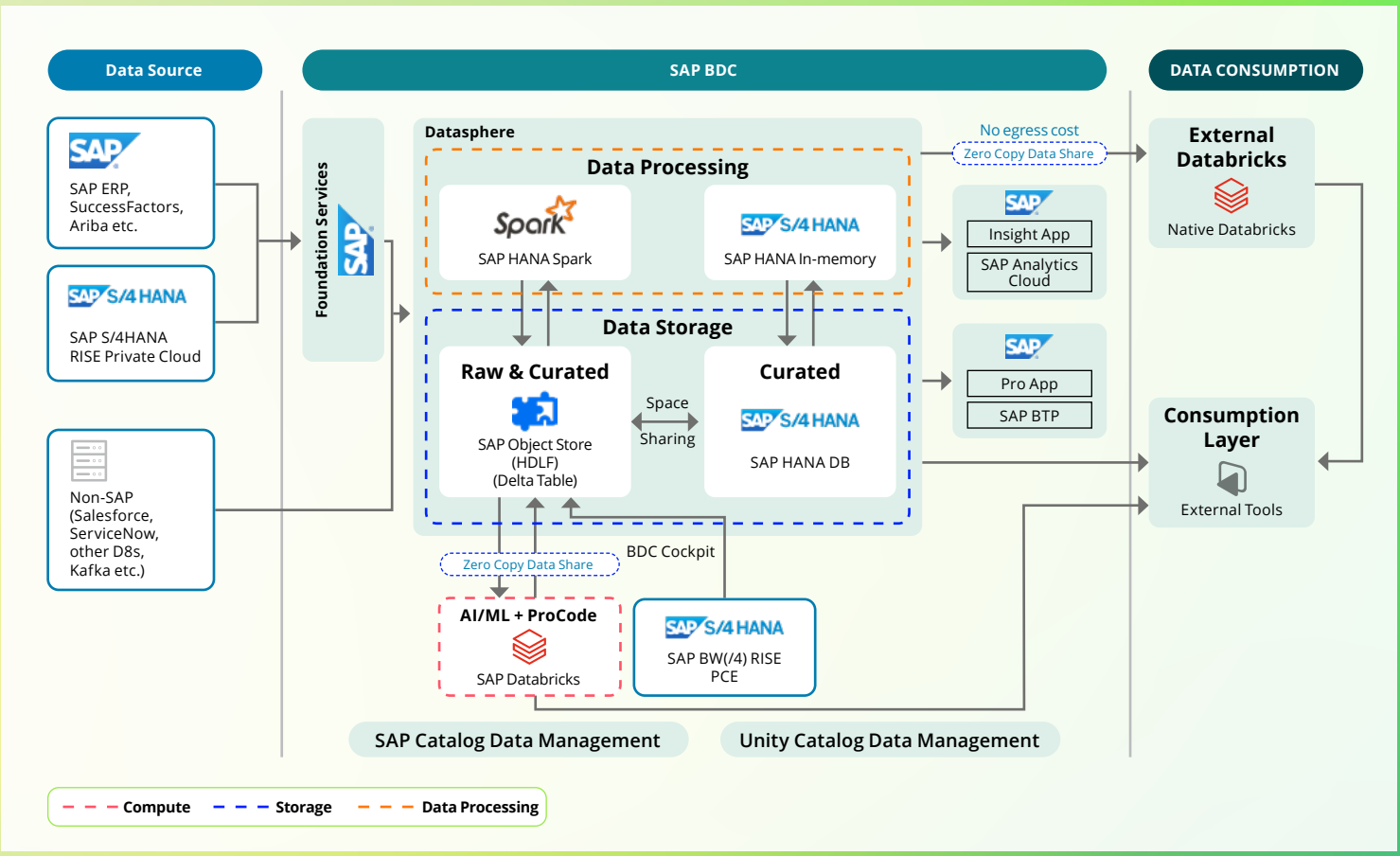
SAP BUSINESS DATA CLOUD AND DATABRICKS INTEGRATION:

TURNING SAP DATA AND ANALYTICS INTO ACTION

Have you heard the buzz surrounding the integration between SAP Business Data Cloud (BDC) and Databricks? The collaboration brings together leading data and analytics capabilities to minimize Total Cost of Ownership (TCO), maximize Return on Investment (ROI), and equip new and existing clients to use more data more powerfully.

Let's explore the significance of the SAP and Databricks integration across three key areas:

- *Overcoming data integration challenges*
- *Turning information into action*
- *Streamlining implementation for current Databricks clients (brownfield) and new clients (greenfield)*



OVERCOMING DATA INTEGRATION CHALLENGES

The SAP BDC integration empowers organizations to bring together and analyze previously siloed data. The integration architecture helps organizations extend their data and analytics capabilities across (1) SAP and non-SAP system data and (2) structured and unstructured data.

1 Integrating SAP and non-SAP data

Before SAP BDC, integrating SAP and non-SAP data typically required either (a) extracting data from SAP to a secondary platform or (b) transforming non-SAP data for integration.

Here's how the BDC and Databricks integration addresses those integration challenges:

a) **Extracting data from SAP to a secondary platform:** Moving data out of SAP to a secondary platform can cause data duplication, security concerns, and higher processing costs. The need for data governance and lineage also adds overhead.

With BDC, data stays within SAP. SAP Datasphere is used for data pipelines and SAP Databricks is used for advanced coding (Pro-Code*) and AI/ML tasks. This approach harnesses the high-performance capabilities of SAP Databricks for faster, more cost-effective, superior outcomes.

b) **Transforming other enterprise data for integration into SAP:** Transforming and loading data into HANA-based SAP data platforms for analysis can be costly as HANA 2.0 frequently requires complex, manual intervention to scale up. This scenario has no option to scale down to accommodate workload fluctuations—even for short periods—which increases inefficiencies and expenses.

The BDC and Databricks integration retains only the most critical data in HANA. The remaining data and associated processing tasks are transitioned to SAP data pipelines and/or Databricks pro-code development. This can enhance value and efficiency.

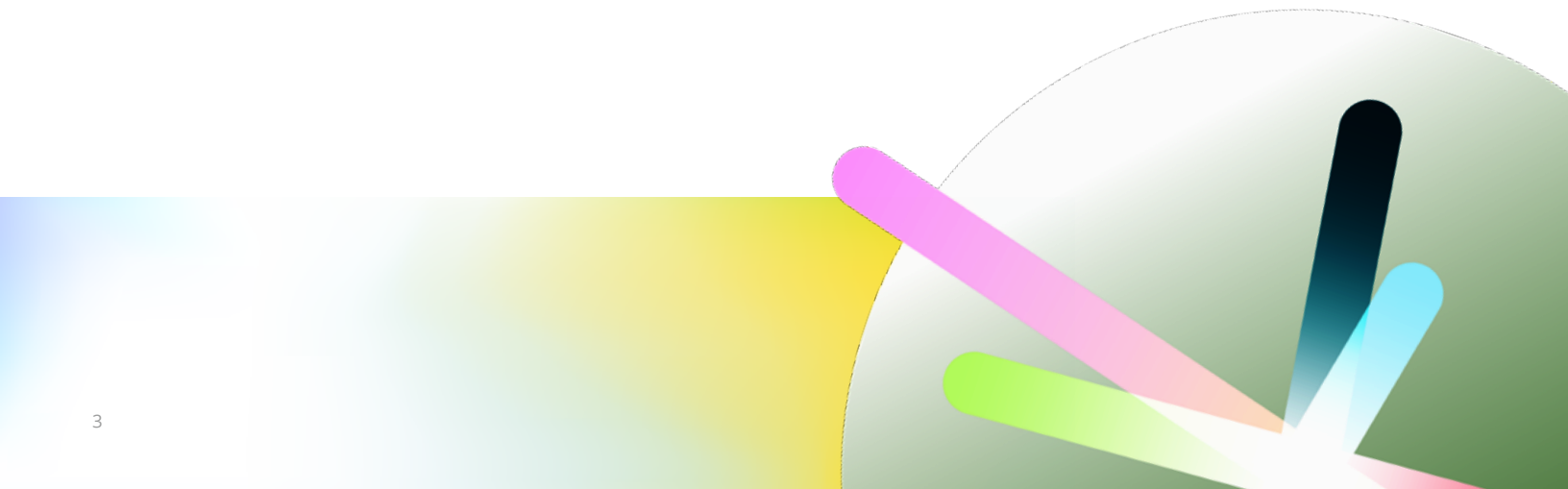
These integrated processes help reduce TCO while improving ROI, governance, real-time data access, and security across all data without duplication.

**Pro-Code: When using pro-code development, clients benefit from a multi-language notebook experience to solve complex data engineering problems.*

2 Using structured and unstructured data

SAP Business Warehouse (BW) and SAP HANA users often struggle to combine unstructured data with structured data due to each application's primary focus on structured data. This can lead to scalability and performance issues, limited integration with big data frameworks, and higher costs.

However, SAP BDC can process structured and unstructured data with unlimited scalability. SAP Datasphere pipelines and Databricks both excel in processing unstructured data, supporting diverse data formats, and real-time streaming. SAP BDC is also backed by low-cost, virtually unlimited lakehouse capabilities supported by HANA Data Lake Files (HDLF) and Delta tables.



TURNING INFORMATION INTO ACTION

Once data from different systems and formats comes together, SAP BDC and Databricks work together to analyze, interpret, and use data in powerful ways. This includes:

- **Developing semantically aligned data products**
- **Intelligently and safely using AI and deeper insights**
- **Enhancing governance capabilities**
- **Powering the data warehouse**
- **Sharing data in real time**

Developing semantically aligned data products

SAP BDC understands data from SAP sources, including applications like S/4HANA, SuccessFactors, and Ariba. SAP BDC also helps create data products that align with client needs using pre-built pipelines and semantic layers. This helps avoid extensive data cleansing or transformation, accelerate insights, and enables real-time decision-making by eliminating the need for laborious extract, transform, load (ETL) processes.

Leveraging SAP's business data model and integrated products enhances the creation of unified data solutions, enriching the data ecosystem. SAP-curated data products are ready for analytics and can be blended with non-SAP data. Databricks is the exclusive partner for SAP-curated line of business data products made available via Delta Sharing.

Intelligently and safely using AI and deeper insights

With the integration of SAP and SAP Databricks into BDC, clients can leverage SAP Databricks AI/ML to build AI applications that understand both their SAP and external data. This integration helps manage the entire ML lifecycle with enterprise-level reliability, security, and scalability.

The SAP Databricks integration helps clients maximize AI and advanced analytics by offering AI/ML training, development, and deployment capabilities. These include collaborative features for data science teams, providing flexibility, boosting productivity and cost-efficiency using MLflow and large language model operations (LLMOps). It also tracks and secures ML model training and projects in production, with advanced GenAI and LLM support.

This helps clients to get more accurate answers, aiming to ensure AI applications are correct and safe. It also allows for the development of domain-specific AI agent systems using SAP Databricks' powerful tools.

Additionally, Databricks' pro-code environment can combine SAP and non-SAP data from finance, human resources, supply chain and more. Powered by a high-performance Spark engine and dynamically scaling clusters, this integration delivers faster and cheaper real-time insights with the SAP Databricks engine and HDLF.

Enhancing governance capabilities

Databricks' Unity Catalog provides enterprise governance capabilities such as discovery, lineage, data access, model management, monitoring, and automated intelligence. SAP BDC's Datasphere Catalog and SAP Databricks' Unity Catalog bi-directionally exchange metadata about SAP and Databricks curated data products via Delta Sharing.

Powering the data warehouse

In addition to HANA SQL on files, SAP Databricks offers robust data warehousing capabilities that let BI tools like Power BI and Tableau analyze Databricks curated data products as well as Delta Shared BDC data products through Databricks SQL (DBSQL).

DBSQL combines Databricks' scalability with SQL's familiarity, so clients can run complex queries on large datasets and get real-time insights. DBSQL also supports high concurrency—performing effectively even with many users and queries running at once—to help clients make informed decisions.

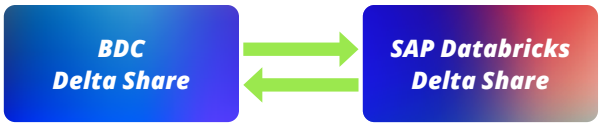
Sharing data in real time

Whether using SAP BDC as a new client (greenfield) or integrating with an existing Databricks instance (brownfield clients), there is no need to physically copy data.

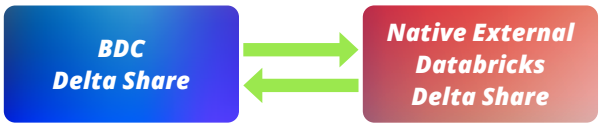
Data is shared in real-time with zero copying using Delta Sharing, which helps enhance data security, governance, and control. This can significantly enhance the end-user experience by providing access to high-quality data that was previously unavailable or required complex integrations.

Both new SAP Databricks clients and existing Databricks users can receive SAP BDC data products through "zero-copy" Delta Sharing in near real-time and also share derived data products, e.g., AI derived insights, back to BDC.

DELTA SHARING WITH BDC



DELTA SHARING WITH NATIVE DATABRICKS



STREAMLINING IMPLEMENTATION FOR GREENFIELD AND BROWNFIELD CLIENTS

Greenfield clients enjoy many benefits by not moving data outside of SAP, such as leveraging Datasphere pipelines, SAP Databricks Pro code, AI/ML capabilities, and HDLF.

Brownfield clients using native Databricks enjoy easy SAP data access without complex ingestions and data duplication. This helps users access SAP data in real-time and minimizes the need for intermediate native hyperscaler storage. In cases where native Databricks requires data to land in hyperscaler storage, access remains straightforward through Delta Sharing, eliminating the need for complex extraction processes.

Zero-copy Delta Sharing maintains data security during SAP and Databricks data sharing. External data and analytics platforms can use SAP replication flows for data not included in delta sharing

WHERE TO NEXT?

The SAP BDC and Databricks integration offers potential benefits for both greenfield and brownfield implementations, including:

- **Accelerating** advanced analytics and AI/ML and GenAI development for transformative outcomes
- **Unifying** SAP data with enterprise systems for comprehensive insights from exploratory analysis, built-in data intelligence, auto-optimization, and cost-effective SAP Databricks SQL
- **Streamlining** governance for data and AI assets with Databricks' Unity Catalog for fine-grained access controls, auditing, and data lineage. By simplifying data management across various environments, Unity Catalog ensures consistent security and compliance.

As integration capabilities continue to evolve, benefits will also expand. For organizations seeking to minimize TCO while maximizing ROI, embracing this integration now may deliver data and analytics advantages for years to come.

Mani Kandasamy
Databricks Alliance CTO
Technology Fellow
Deloitte Consulting LLP
mkandasamy@deloitte.com

Georgios Giannantonakis
SAP Analytics Strategies & Architecture
Senior Manager
Deloitte Consulting GmbH
ggiannantonakis@deloitte.de

Qi Su
SAP Solutions Architect
Databricks
qi.su@databricks.com