

SAP Business Data Cloud: A Unified Approach to Enterprise Analytics and AI with Databricks

Nareshbabu Sigamani
Independent Researcher, USA

Abstract

The current data challenges of enterprises revolve predominantly around fragmentation; critical business data across heterogeneous systems creates significant hurdles for organizations that want to derive timely and meaningful insights. SAP Business Data Cloud intends to remedy those challenges with a fully managed Software as a Service for organizations that want to unify and govern SAP and non-SAP data in one trusted environment. The combined platform will serve as the data foundation for advanced analytics and enterprise-scale Business AI by leveraging a landmark partnership with Databricks that builds industry-leading data intelligence capabilities directly into the SAP ecosystem. This integrated platform utilizes zero-copy data architecture, enabled by open-source Delta Sharing, that allows data to flow bi-directionally, with no need for physical replication, complemented by end-to-end governance through Unity Catalog. For existing SAP customers, BDC provides a clear cloud-centric modernization path with extended maintenance for legacy systems, while consolidating previously separate data and analytics products into a unified offering. The platform's curated data foundation accelerates the development and deployment of intelligent applications, including next-generation AI agents that require trusted, high-quality data to deliver reliable business outcomes across diverse enterprise contexts.

Keywords: SAP Business Data Cloud, Enterprise Data Management, Zero-Copy Architecture, Semantic Data Products, AI-Driven Analytics

1. The SAP Business Data Cloud: A New Era of Enterprise Data Management

1.1. The Strategic Rationale for a Unified Platform

The disconnectedness of enterprise data has always been a significant barrier to effective digital transformation and converting data into value. Organizations typically manage data across many disconnected systems, especially in large enterprises that might manage hundreds of disconnected data repositories across their operating environment. The data needed for comprehensive analysis can be spread across several different transactional systems, multiple data lakes, and data warehouses, leading to multiple data silos that limit effective decision-making. This scattered landscape leads to significant governance and lineage challenges, as data is duplicated across multiple environments. Research indicates that data engineers dedicate a substantial portion of their working hours to data preparation and integration tasks rather than value-generating analytics activities, with some studies suggesting this proportion reaches well over half of total effort. This duplication and manual intervention slow down the development of advanced analytics, machine learning, and artificial intelligence applications, with typical AI project timelines extending considerably, of which data preparation accounts for a dominant share of the total effort [1].

SAP Business Data Cloud is engineered as a definitive response to these challenges. As a fully managed SaaS solution, BDC is designed to break down data silos and unify all structured and unstructured data within a single trusted environment. The platform architecture is built to handle enterprise-scale data volumes while maintaining optimal query response times for standard analytical operations on large-scale datasets. Early adoption patterns demonstrate that BDC can substantially reduce data pipeline development time compared to traditional custom-built integration architectures, while simultaneously

10.48047/jocaaa.2025.34.11.58

decreasing storage costs through the elimination of redundant data copies. The unified approach enables organizations to establish a single source of truth that preserves critical business context and semantics inherent in SAP applications, eliminating the need for extensive reverse-engineering of complex ERP data structures [1].

The market discussion surrounding BDC has centered on whether it represents a fundamental strategic shift or merely a rebranding of existing products. Analysis of product announcements and strategic documentation reveals a deeper transformation beyond simple renaming. BDC represents a new go-to-market offering that bundles previously separate products and natively integrates partner technology at its core. The bundling and significant level of integration create an important, cohesive platform aimed at solving more comprehensive and complex problems, compared to predecessors, combined with the out-of-the-box unified architecture, together with the streamlined access to semantically rich SAP data, which generally represents a significant friction point for organizations. Internal analyses suggest that organizations invest substantial development hours annually in maintaining custom SAP data extraction and transformation pipelines, representing significant labor expenses for mid-sized and large enterprises [2].

A crucial architectural clarification involves the relationship between SAP Business Data Cloud and SAP Business Technology Platform. While BDC and BTP are separate solutions, BDC operates as a fully managed cloud solution built on and running atop the BTP platform. BTP serves as an innovation and integration platform supporting the entire SAP portfolio, with thousands of active enterprise deployments, whereas BDC is designed specifically to create a unified data foundation. This architectural relationship means BDC leverages BTP's underlying capabilities to provide managed services, including AI model training infrastructure that supports workflows for models ranging from modest to extremely large parameter counts [2].

1.2. Timeline and Controlled Availability

SAP Business Data Cloud was formally revealed during the second half of February 2025, complementary to the formal announcement of a strategic partnership with Databricks. The first release was done through a model of controlled availability, first made available to a select pilot customer cohort in manufacturing, retail, financial services, and health care, across several published regions globally. These pilot organizations collectively represent substantial annual revenue and manage significant volumes of SAP-related business data. The controlled availability phase was strategically designed to validate platform performance under production workloads, with pilot participants reporting substantial query performance improvements compared to previous data architectures and significant reductions in data preparation time [1].

The staged rollout plan encompasses multiple deployment phases across major cloud infrastructure providers. The native Databricks integration was scheduled to achieve full availability by mid-spring 2025, commencing with deployment in additional AWS data centers across key global regions, including North America, Europe, and Asia-Pacific. The initial AWS deployment phase targets serving hundreds of enterprise customers during the first quarter of general availability. Subsequent availability on Microsoft Azure and Google Cloud Platform is planned for the second half of 2025, with Azure deployments targeting mid-summer across multiple regional zones and Google Cloud Platform deployment scheduled for autumn across strategic global regions. This multi-cloud strategy provides customers with deployment flexibility and ensures proximity to existing workloads, with projections indicating that the vast majority of the global customer base will have access to a regional BDC deployment in close geographic proximity to their primary data center by year-end 2025 [2].

10.48047/jocaaa.2025.34.11.58

The phased approach prioritizes platform stability and seamless user experience as the deeply integrated solution rolls out globally. The controlled rollout strategy incorporates rigorous service level agreements guaranteeing high platform availability, minimal data latency for transactional queries, and rapid data synchronization intervals for real-time analytics workloads. Infrastructure resources have been committed to scale support for thousands of concurrent enterprise deployments, with each deployment engineered to process substantial volumes of active analytical data and support high query concurrency levels per tenant environment [1, 2].

2. The Paradigm-Shift Collaboration: The Synergy of SAP and Databricks

2.1. The Strategic Collaboration

The collaboration between SAP and Databricks represents a cornerstone collaboration in the enterprise data management space, which combines business-critical data management with advanced analytics. The collaboration focuses on reimagining enterprise data management and enabling the acceleration of AI adoption for organizations managing mission-critical business operations that are mission-critical. The alliance emphasizes the strategic pairing of business-critical transactional data with leading data platform capabilities, addressing a fundamental challenge where SAP systems manage a substantial portion of global transaction revenue and process trillions of financial transactions annually across hundreds of thousands of customer organizations worldwide [3].

The core objective focuses on turbocharging Business AI by establishing a reliable data foundation that fuses deep expertise in mission-critical business processes with world-class data engineering, data warehousing, and AI capabilities. To support this strategic alliance and ensure comprehensive customer success, a substantial financial investment has been committed to the partnership ecosystem. An investment has been earmarked specifically to assist customers and system integrator partners with deployment, migration activities, and unlocking business value from SAP data through the integrated offering. This multi-year investment program is structured to support thousands of customer implementations across diverse industries, including manufacturing, financial services, retail, healthcare, telecommunications, and public sector organizations [3].

The investment allocation encompasses several strategic initiatives distributed across multiple focus areas. A significant portion is dedicated to direct customer migration and deployment services, providing technical consulting and implementation support for complex enterprise environments. Additional funding supports partner enablement programs targeting hundreds of system integrator organizations worldwide to build specialized expertise in the integrated platform. Further investment supports the development of industry-specific solution templates and accelerators designed to reduce time-to-value for common use cases. The remaining allocation is reserved for ongoing innovation initiatives and platform enhancement efforts. This comprehensive financial commitment underscores the long-term strategic nature of the partnership and highlights the shared objective of addressing persistent data challenges affecting large enterprises that collectively manage exabytes of business-critical data in SAP environments globally [3], [4].

2.2. The Technical Blueprint: Zero-Copy and End-to-End Governance

Organizations seeking to leverage SAP data for advanced analytics and AI have historically faced significant obstacles involving costly, time-consuming, and complex data extraction processes. Traditional data integration approaches typically require organizations to invest extended timeframes in establishing production-grade data pipelines, with total implementation costs varying substantially based on enterprise scale and complexity. The traditional approach often led to the creation of multiple data

copies across various data lakes and warehouses, with enterprises commonly maintaining several replicated versions of their core SAP datasets across different analytical and reporting environments. This proliferation results in data duplication consuming substantial portions of total data infrastructure budgets, generating numerous data quality incidents annually in typical large enterprises, and creating security risks where each additional copy increases potential exposure surface area [3].

The SAP Business Data Cloud technical architecture specifically addresses this challenge through innovative architectural patterns. The core mechanism employs a zero-copy data architecture enabled by an open-source data sharing protocol that has gained widespread industry adoption. This protocol allows secure, bi-directional data sharing between the SAP-managed BDC environment and other platform deployments without physically moving or duplicating data. Data remains governed and secured in its original location within SAP-managed cloud storage infrastructure while being read and processed at high speed by analytics engines. Performance benchmark testing demonstrates that query operations on shared datasets achieve throughput levels approaching those observed on locally stored data, with latency overhead typically remaining minimal depending on data volume and network conditions [4].

This capability dramatically increases data team productivity, with early adopter organizations reporting substantial reductions in time spent maintaining data pipeline infrastructure and reducing operational overhead associated with complex data pipelines compared to traditional ETL-based architectures. Storage cost analysis from pilot implementations indicates significant savings on infrastructure expenses previously allocated to redundant data storage. The zero-copy architecture is complemented by end-to-end governance through a unified catalog system. The catalog provides a centralized metadata layer tracking not only tables but also notebooks containing analytical code, ML models across their entire lifecycle, and complete lineage information tracing data from source systems through all transformation steps to final consumption points [3].

Large enterprise implementations manage metadata for environments containing extensive collections of distinct data assets, transformation logic definitions encoded in SQL and code-based pipelines, and ML model versions spanning development, staging, and production environments. This capability ensures comprehensive audit trails are available with retention periods configured to meet regulatory compliance requirements, and that data access controls are consistently enforced across all analytical workloads. The governance framework supports fine-grained access control policies, with typical enterprise implementations managing thousands of distinct permission rules governing access to sensitive data elements such as personally identifiable information, financial data, and proprietary business information. Access control enforcement operates in real-time with minimal decision latency to ensure a seamless user experience, utilizing distributed caching mechanisms that maintain frequently accessed permission sets in memory across compute clusters [4].

Consequently, even generative AI bots and agents can only access data they are explicitly permitted to use, with access decisions evaluated at query execution time based on the security context of the requesting user or service account. Security audit logs capture dozens of data points per access event, including user identity, timestamp, data accessed, query patterns, and access decision rationale, generating substantial log volumes daily for large enterprise deployments. This comprehensive governance fully aligns with stringent access policies required by regulations such as GDPR, HIPAA, SOX, and industry-specific compliance frameworks, with compliance reporting capabilities generating audit documentation rapidly rather than requiring the extended periods typical of traditional manual audit processes [3].

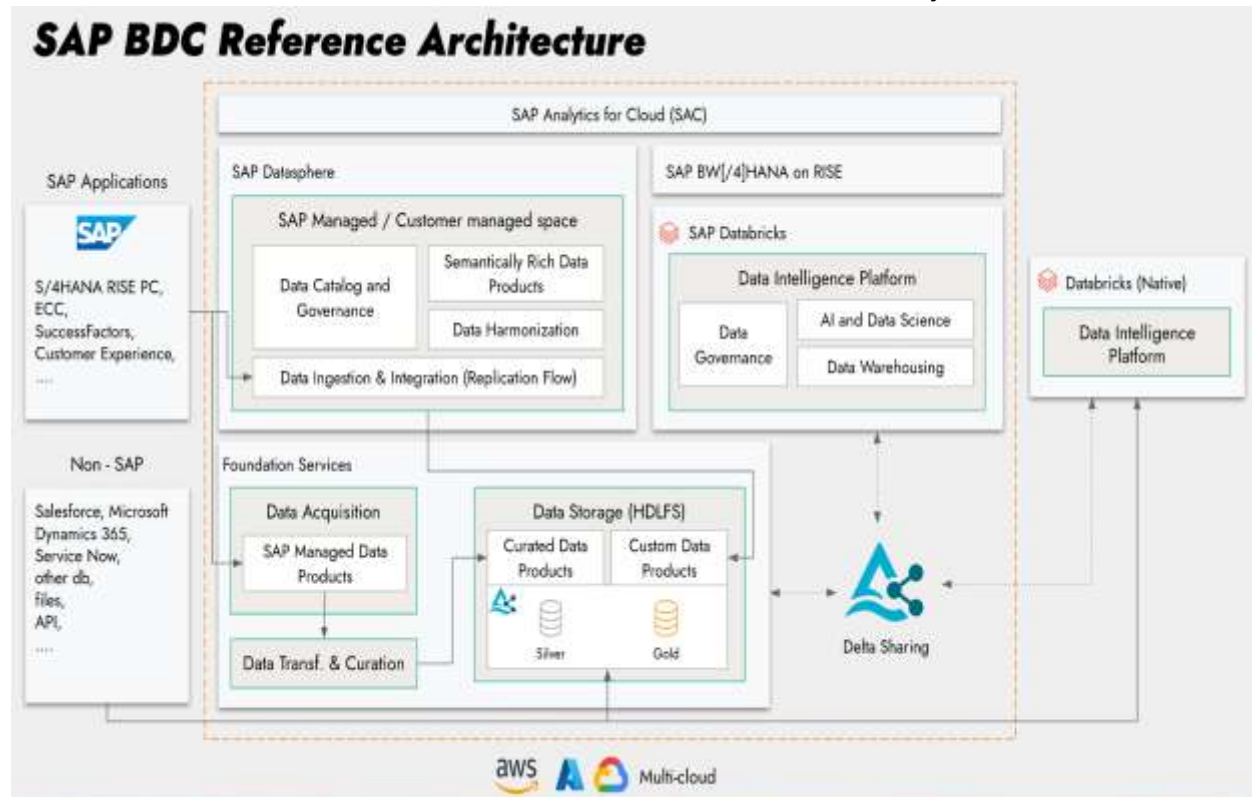


Fig. 1: SAP BDC Reference Architecture

2.3. Architectural and Functional Deep Dive

The SAP Business Data Cloud represents a comprehensive, end-to-end data and analytics platform designed to support enterprise-scale analytical workloads ranging from basic reporting to advanced AI model development. The solution comprises multiple core components working in concert, bundled to provide a unified user experience and cohesive technical architecture. These components include data integration and federation services capable of connecting to numerous different source systems including SAP applications, third-party enterprise software, cloud services, and on-premise databases; semantic modeling and business context layers providing business-friendly abstractions over technical data structures; analytical query engines optimized for complex aggregations and joins across datasets containing billions of records; visualization and reporting interfaces supporting both self-service analytics for business users and developer-oriented programmatic access; AI and ML development environments providing integrated workflows from data preparation through model deployment; and governance and security frameworks ensuring policy enforcement across all platform operations [3].

While all components provide essential capabilities, primary research focus and public discourse have centered on three key pillars representing the most significant differentiators: the semantic data modeling layer preserving SAP business context, the analytics and visualization suite enabling self-service insights, and the embedded advanced analytics offering providing professional-grade development environments. The integration of these components aims to create a streamlined, powerful environment for data management and AI-driven insights, with platform architecture supporting substantial concurrent user access per tenant environment and high query throughput for large enterprise deployments during peak business periods. Load testing performed during platform development validated stable performance

under sustained loads with response times remaining optimal for standard dashboard refresh operations [4].

The architecture supports seamless data flow from source systems through data modeling, advanced analytics, and ultimately to business intelligence and decision support. Data ingestion capabilities support both batch and real-time integration patterns, accommodating diverse organizational requirements. Batch processing windows are optimized for overnight refresh cycles, typically completing within defined timeframes for datasets containing hundreds of millions of records and processing substantial data volumes per batch cycle, depending on organizational scale. Change data capture mechanisms identify and process only modified records, substantially reducing processing volumes compared to full dataset reloads. Real-time streaming ingestion supports minimal latencies from source transaction commit to availability in analytical views, with streaming throughput capacity exceeding substantial event volumes per second for typical enterprise deployments [3].

The platform leverages distributed computing architectures capable of scaling compute resources elastically based on workload demands, implementing auto-scaling algorithms that continuously monitor query queue depth, CPU utilization across cluster nodes, memory consumption patterns, and I/O throughput metrics. Auto-scaling decisions typically occur within short timeframes of detecting sustained load increases, with compute clusters scaling from baseline configurations to peak configurations for the largest enterprise workloads. This elasticity maintains optimal response times for standard dashboard queries even during peak usage periods when concurrent user counts may increase substantially compared to average utilization levels. Cost optimization algorithms automatically scale down compute resources during off-peak periods, with organizations typically observing significant reductions in compute costs compared to static cluster sizing approaches. Storage architecture utilizes advanced columnar compression techniques, achieving substantial compression ratios for typical transactional datasets, significantly reducing storage costs while maintaining query performance through intelligent caching strategies that maintain frequently accessed data in high-speed storage tiers [4].

2.4. A Decision Framework for Implementation

Organizations evaluating whether to adopt advanced analytics capabilities within SAP Business Data Cloud require a structured decision-making process to ensure successful implementation, aligning with business objectives and technical constraints. This framework, based on key business and technical factors derived from analysis of numerous successful enterprise implementations, helps determine the optimal approach for integrating analytics platforms with SAP landscapes. Implementation success metrics indicate that organizations following structured assessment frameworks achieve production deployment substantially faster and report higher user satisfaction scores compared to organizations pursuing ad-hoc evaluation approaches that may overlook critical technical or organizational readiness factors. The structured approach typically involves several weeks of assessment activities, including current state analysis, future state design, gap identification, and roadmap development [3].

The fundamental question centers on whether the organization maintains SAP as a core enterprise system managing critical business processes. Statistical analysis across global enterprise technology landscapes indicates that a high percentage of Fortune companies and large enterprises with substantial annual revenues operate SAP systems as their primary ERP platform, with SAP installations managing a significant portion of global commerce by transaction value. Organizations in manufacturing sectors show the highest SAP adoption rates, followed by consumer products, financial services, and healthcare. If SAP is not part of the core technology landscape or represents a minor portion of enterprise data volume, then

10.48047/jocaaa.2025.34.11.58

SAP Business Data Cloud may not represent the optimal fit for the current data strategy, as platform strengths center specifically on SAP data contextualization and semantic enrichment [4].

For organizations confirmed as SAP customers where SAP systems manage substantial enterprise data, the next step involves understanding the deployment model that fundamentally determines integration patterns and available features. Readiness for advanced analytics in SAP BDC depends on whether implementations operate on-premise systems or have adopted cloud-based enterprise resource planning solutions. Current market analysis indicates that a portion of SAP customers continue operating primarily on-premise deployments, with others in hybrid configurations combining on-premise and cloud components, and the remainder operating fully in cloud environments. Migration patterns indicate accelerating cloud adoption, with customers initiating cloud migration projects annually. For on-premise SAP customers, organizations can leverage analytics capabilities in SAP BDC, but integration patterns are constrained to extractor-based replication flows utilizing standard SAP extractors covering hundreds of standard business content areas [3].

These on-premise customers cannot leverage pre-built semantic data products, requiring direct cloud-to-cloud integration APIs available only in cloud deployment models. Data replication latency for extractor-based patterns typically varies depending on dataset size, network bandwidth between source and cloud environments, and the complexity of data transformations applied during extraction. Organizations with on-premise deployments processing substantial daily transactional data may experience extended replication windows, potentially limiting near-real-time analytics use cases. For organizations already operating cloud-based SAP solutions, full platform readiness is achieved, enabling access to pre-built semantic data products covering hundreds of standard business processes across finance, supply chain, manufacturing, sales, and human resources domains. Cloud-based integrations leverage native APIs, delivering rapid data synchronization for most business entities [4].

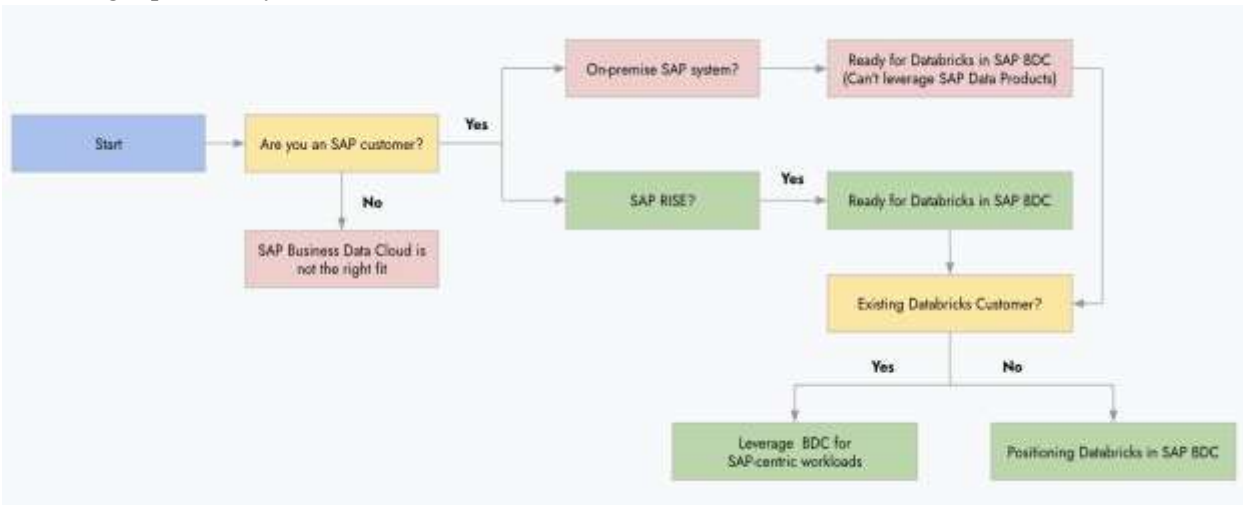


Fig. 2: Decision Framework for BDC Implementation

2.5. Datasphere vs. SAP Databricks: Roles and Responsibilities

Understanding the distinct yet complementary roles of different components within the BDC architecture represents a central point of clarity for organizations architecting their data and analytics ecosystem. These two components serve different user personas with divergent skill sets and address specific, interdependent stages of the data and analytics lifecycle that collectively enable comprehensive analytical capabilities. The semantic modeling component is designed primarily for business data users, including

10.48047/jocaaa.2025.34.11.58

data modelers and data analysts who typically possess SQL proficiency ranging from intermediate to advanced levels but may not have extensive programming expertise in languages such as Python or Scala. User demographic analysis from platform telemetry data indicates that a majority of semantic layer users hold business analyst, data analyst, or business intelligence developer titles, with additional users serving in data architecture or data governance roles, and the remainder representing power users from finance, supply chain, sales, and other business functions [3].

The core function of the semantic modeling component centers on connecting, integrating, federating, preparing, and modeling enterprise data while preserving crucial business context and semantics that provide meaning to raw data structures. It acts as the semantic layer, providing a business-friendly model of data that is comprehensible and governable in business terms rather than technical database schemas. Typical implementations manage hundreds to thousands of business entities representing key organizational concepts such as customers with associated demographic and behavioral attributes, products with hierarchical categorization and pricing structures, orders with complex lifecycle states and fulfillment information, financial accounts organized by chart of accounts structures, and organizational units reflecting reporting hierarchies. Large enterprise implementations with complex business models may manage thousands of distinct business entities spanning multiple lines of business, geographic regions, and legal entities [4].

The semantic modeling environment serves as the primary tool for creating and managing analytical models, with platform capabilities supporting model complexity ranging from simple star schemas optimized for specific departmental reporting needs to complex multi-fact models supporting enterprise-wide analytics. User productivity metrics gathered from platform usage telemetry indicate that business analysts using the semantic modeling interface can develop and deploy new analytical models substantially faster than traditional data warehouse modeling approaches, with model development timelines significantly reduced for moderately complex business domains compared to conventional dimensional modeling techniques requiring extensive technical specification documentation and hand-off between business and technical teams. Productivity gains derive primarily from the visual modeling interface, reducing specification and documentation overhead, pre-built SAP business content providing starting templates that eliminate substantial custom modeling effort, and integrated testing capabilities enabling rapid iteration cycles [3].

In contrast, the advanced analytics component represents the domain of technical data users, such as data engineers, data scientists, and AI/ML developers who possess programming expertise in languages such as Python, Scala, R, and SQL, along with an understanding of distributed computing concepts, machine learning algorithms, and software engineering best practices. Demographic analysis of user populations indicates that substantial portions hold data engineer or analytics engineer titles focused primarily on data pipeline development and optimization, serve as data scientists or ML engineers focused on predictive modeling and AI application development, or operate in analytics engineering or advanced analytics roles that bridge business and technical domains. Educational backgrounds show that most hold technical degrees in computer science, engineering, mathematics, or statistics, while others come from quantitative business, economics, or science backgrounds [4].

The advanced analytics environment operates as a pro-code development environment built for deep technical work, with development interfaces based on notebook paradigms that integrate executable code cells, visualization outputs, markdown documentation, and collaborative commenting in unified workspaces. Primary functions include building complex Extract, Transform, and Load or Extract, Load, and Transform pipelines processing data volumes ranging from gigabytes to petabytes, developing

10.48047/jocaaa.2025.34.11.58

custom AI/ML solutions including supervised learning models, unsupervised clustering algorithms, deep learning neural networks, and natural language processing applications, and performing ad-hoc SQL analytics and data engineering workloads. The strong point is the capability to communicate between all data classes—structured, semi-structured, and unstructured—with powerful and scalable computing capabilities using distributed processing frameworks [3].

The two components fundamentally have a symbiotic relationship to create a virtuous cycle of business insight and technical innovation. The semantic modeling layer is where the business context is curated and modeled, providing trusted definitions of critical business metrics and hierarchies, and relationships between entities. This business context creates important framing for analytical work, ensuring that any technical analyses relate to business definitions and can be understood as such by business stakeholders. Workflow pattern analysis indicates that a majority of advanced analytics projects begin with semantic data products as their primary source, ensuring that AI models are trained on business-contextualized data rather than raw transactional records. Conversely, sophisticated analytics and AI models developed in the advanced analytics environment generate insights and predictions that enhance business understanding, with workflow patterns enabling model scoring results to be materialized as new entities in the semantic layer accessible through standard reporting interfaces [4].

| Metric | SAP Datasphere | SAP Databricks |
|--------------------------|---|---|
| Target Persona | Business users, data modelers, data analysts | Data engineers, data scientists, AI/ML developers |
| Primary Function | Data modeling, data federation, semantic layer, data governance, business catalog | Pro-code development, ETL/ELT, AI/ML model training, custom application development |
| Key Features | Spaces, Insight Apps, Data Fabric, Data Mesh | Databricks SQL, Python Notebooks, MLOps, Delta Sharing, Unity Catalog |
| Data Types | Primarily structured data, with connectors to various sources | Structured, semi-structured, and unstructured data |
| Primary Interface | Business-friendly graphical interface, with SQL access for advanced users | Jupyter-style notebooks, Databricks SQL editor, API |

Table 1: SAP Datasphere vs. SAP Databricks - A Functional Comparison

2.6. The Data Product Economy: Delivering Context at Scale

A key strategic shift with BDC introduction centers on embracing the data product concept as a core architectural and operational principle. Instead of requiring customers to spend significant time and effort reverse-engineering the logic of complex SAP data structures involving hundreds of interrelated tables and thousands of business rules encoded in application logic, BDC delivers pre-curated, ready-to-consume data products encapsulating both data and essential business context. These are not simply raw tables extracted from transactional systems but are semantically rich data sets for core business processes spanning diverse functional areas, including General Ledger accounting, working capital management, supply chain inventory optimization, order-to-cash flows, procure-to-pay cycles, and human capital analytics [3].

The initial platform release includes over a hundred standard data products covering the most common business processes implemented across major industries, including manufacturing, retail, consumer products, financial services, healthcare, and professional services. Each data product incorporates business logic, which otherwise would take hundreds of hours of analysis and development time to create through custom data modeling. Development roadmap commitments target significantly expanding the library by the end of the first year of general availability, and the anticipated coverage will extend to industry processes such as loan origination for banking, claims management for insurance, patient journey analytics for healthcare, and merchandising analytics for retail. Data product development prioritization is informed by customer feedback and usage analytics [4].

These data products come with business context already built in, including standardized naming conventions using business terminology rather than technical field names, pre-defined hierarchies enabling drill-down analysis from summary to detail levels across dimensions such as time periods, organizational units, and product categories, calculated measures implementing standard business formulas, and business rules reflecting industry best practices and standard SAP process configurations. Time-to-value analysis conducted across early adopter organizations indicates that projects leveraging

10.48047/jocaaa.2025.34.11.58

pre-built data products achieve production deployment substantially faster compared to projects requiring custom data modeling from raw transactional sources, representing significant reductions in time-to-value. Development effort analysis shows that projects using data products require substantially fewer developer hours compared to custom development, translating to significant cost savings per analytics initiative [3].

The availability of these governed data products also simplifies the consumption of enterprise data across organizational boundaries, addressing a persistent challenge in large enterprises where different departments often develop conflicting definitions of seemingly simple concepts. By providing standardized definitions and consistent business logic, data products ensure that analytics used for decision-making are consistent, trusted, and correctly understood across the organization, eliminating the common challenge of multiple versions of truth. Enterprise governance assessments conducted post-implementation indicate that organizations leveraging standardized data products experience substantially fewer data quality incidents, fewer metric definition conflicts, and reduced time spent on data validation activities. Data products also include embedded governance metadata such as data classification levels, retention policies, access control recommendations, and compliance tags, enabling consistent policy enforcement across analytical workloads and simplifying compliance with regulatory requirements [4].

2.7. The Role of Insight Apps

Insight Apps represent a new capability within BDC designed to provide pre-built, targeted solutions for specific business needs, operating at a higher abstraction level than data products by combining data, analytics logic, and user interface elements into complete applications. They offer a streamlined deployment process with one-click installation mechanisms, automating environment provisioning, user access configuration, and security policy implementation, and automated data integration and staging workflows, eliminating manual configuration steps typically requiring substantial technical effort per application. Installation automation includes dependency resolution, data product activation, semantic model deployment, and user interface configuration, reducing deployment complexity from multi-week projects to rapid implementations for standard scenarios [3].

These applications cover areas like core enterprise analytics, including financial performance monitoring, supply chain visibility, sales pipeline analysis, and human resources workforce analytics, built on the trusted data foundation and data products within BDC. The current portfolio includes dozens of Insight Apps available at platform launch, spanning multiple functional areas across finance, supply chain, sales, human resources, procurement, manufacturing, customer service, and corporate performance management. Roadmap commitments target substantial expansion covering specialized industry scenarios, with planned additions including retail merchandising analytics, healthcare patient journey analysis, banking loan portfolio management, insurance claims analytics, and telecommunications network performance monitoring [4].

As fully managed applications, Insight Apps simplify the delivery of advanced analytics and planning capabilities, with deployment timelines substantially reduced compared to custom-built analytical applications requiring requirements gathering, design, development, testing, and training phases. User adoption metrics from pilot implementations indicate that Insight Apps achieve high active user adoption within the first month of deployment compared to lower adoption rates for custom applications over similar timeframes. This accelerated adoption is attributed to intuitive, pre-configured user interfaces designed based on user experience research, immediate availability of relevant business insights, and pre-built navigation patterns aligned with common analytical workflows. However, Insight Apps are designed primarily for immediate use and therefore offer limited customization capabilities, with configuration

options typically restricted to parameter adjustments, user access controls, and visual theme customization [3].

In cases where organizations have specific needs or may want to make specialized changes beyond the standard configuration scope, recommend copying the Insight App template from the Semantic Model environment and developing modifications within the semantic modeling environment. This development pattern allows the organization to use pre-built logic as a starting point, and then add custom business rules, additional data sources, or specialized analytical calculations to the organization's needs. The time to develop custom Insight Apps varies based on complexity, but it is still a time savings over time spent developing the same functionalities from scratch. This structured approach ensures that core, pre-built logic remains governed and benefits from ongoing platform updates delivered through regular release cycles, while allowing necessary flexibility to meet specific business demands. A substantial portion of Insight App implementations involves some level of customization beyond standard configuration, indicating a balance between out-of-the-box usability and extensibility requirements [4].

3. A Detailed Example: Optimizing Supply Chain with AI

To demonstrate the power of this integrated ecosystem, consider a global manufacturing organization using SAP BDC to optimize supply chain operations across multiple continents. The organization operates numerous manufacturing facilities, manages thousands of suppliers, and processes substantial volumes of sales orders monthly with considerable inventory value. Historical challenges have included revenue impacts from stockouts and expedited shipping costs that necessitate more sophisticated forecasting and disruption management capabilities [5].

In establishing the data foundation within BDC, historical sales orders spanning multiple years and containing millions of transaction records, inventory levels tracked across hundreds of warehouses and distribution centers, and production schedules from manufacturing sites are unified from SAP S/4HANA into SAP BDC. The platform consolidates terabytes of operational data, with daily incremental updates adding substantial volumes of new transactional information. BDC also ingests external data from third-party sources, including weather patterns from meteorological services covering numerous geographic markets, market trends from industry research providers tracking multiple product categories, and economic indicators encompassing GDP growth rates, currency exchange fluctuations, and commodity prices for key raw materials. Social media sentiment data from millions of customer interactions is processed monthly to capture real-time consumer perception shifts. All this data is modeled and governed within SAP Datasphere, creating a unified semantic layer with hundreds of business entities, thousands of calculated measures, and dozens of pre-defined hierarchies enabling multi-dimensional analysis [6].

For advanced analytics in the embedded environment, the curated, business-contextualized data products from BDC are shared via Delta Sharing, eliminating the need to physically copy terabytes of historical data and ongoing daily increments. Data scientists train machine learning models to predict future demand with enhanced accuracy. The model leverages historical sales data from SAP and correlates it with external factors like weather forecasts affecting seasonal product lines, consumer sentiment scores derived from social media analytics, promotional calendar events spanning marketing campaigns annually, and competitive pricing intelligence from primary competitors. The gradient boosting regression model is trained on millions of historical data points, achieving substantial improvements in mean absolute percentage error during validation testing compared to previous statistical forecasting approaches. A separate classification model is trained to predict supply chain disruptions by analyzing data on supplier performance metrics, including on-time delivery rates, shipping delays tracked across

10.48047/jocaaa.2025.34.11.58

major logistics carriers, port congestion indicators from critical shipping hubs, and geopolitical risk scores for sourcing countries. This disruption prediction model achieves high accuracy rates in identifying high-risk scenarios days before occurrence [5].

The demand forecasts and disruption alerts generated by the models are sent back to BDC, where they are materialized as new semantic entities. These visualizations are developed into a custom dashboard in SAP Analytics Cloud, delivering supply chain managers a real-time snapshot of potential risks and opportunities. The dashboard shows demand forecasts by product line and their associated confidence intervals, takes note of SKUs at risk of stock-out based on specified time horizons, identifies suppliers with risk scores above some level of concern, and recommends proactive mitigation actions based on projected revenue impact, ranked from highest to lowest. The BDC platform can be configured to automatically trigger actions within the SAP ERP system through pre-defined business rules. For example, if a high-risk disruption is predicted for a specific product with a disruption probability exceeding defined thresholds and a potential revenue impact surpassing specified amounts, the system automatically generates purchase requisitions to increase safety stock levels or initiates communication with alternative suppliers from pre-qualified vendor pools. For critical components where lead times are extended, the system automatically triggers early ordering protocols when demand forecasts indicate significant consumption increases [6].

In operational deployment, the integrated system reduced stockout incidents substantially, decreased expedited freight costs significantly, improved forecast accuracy considerably, and reduced excess inventory holding costs through more precise demand alignment. This example illustrates how SAP BDC, with advanced analytical capabilities, transforms traditional supply chain management into a proactive, intelligent, and automated system that continuously learns from new data and adapts recommendations based on evolving business conditions [5], [6].

| Component | Description | Role in Supply Chain Optimization Example |
|----------------------------|--|--|
| SAP ERP (S/4HANA) | The source of truth for all business transactions. | Provides historical sales orders, inventory, and production data. |
| SAP BDC (incl. Datasphere) | Unified data management and governance platform. | Unifies internal and external data, creating a trusted data foundation for AI. |
| Databricks | Advanced analytics, data engineering, and ML platform. | Trains and runs AI models for demand forecasting and disruption prediction. |
| SAP Analytics Cloud (SAC) | Business intelligence and planning front-end. | Visualizes AI-driven insights on a dashboard for managers to take action. |
| Joule (SAP AI Copilot) | Generative AI assistant for business tasks. | Could recommend or automate actions in SAP ERP based on the AI insights from Databricks. |

Table 2: Key Components and Their Roles in the Supply Chain Optimization Solution

4. Broader Strategic Implications

4.1. The Business Data Fabric

SAP BDC is the technological underpinning of what is referred to as the business data fabric; this is a fundamental architectural paradigm for enterprise data management. This kind of architecture is focused on developing a semantically rich data layer that creates connections between disparate data sources while maintaining the critical context and meaning of the data relative to the business. A business data fabric model differs from the concept of a data lake, which is simply a repository of raw data, where information is held in its native format without any contextual enrichment. A business data fabric is designed to enable either business user or business analyst to make sense of all of the data coming from the ERP system or third-party application without he or she translating it into a usable format [7].

The architectural approach implements a logical data layer that virtualizes access to data across multiple physical storage locations, supporting federation patterns where data remains in source systems while being accessible through unified interfaces. Enterprise implementations typically federate numerous different source systems, creating unified views that span on-premise databases, cloud applications, data warehouses, and data lakes. This comprehensive approach eliminates data silos that have historically fragmented enterprise information assets and reduces the time spent on data wrangling activities. Industry research indicates that data preparation and wrangling consume the majority of data professionals' time in traditional architectures, whereas business data fabric implementations substantially reduce this proportion by providing pre-contextualized data products [7].

Organizations implementing business data fabric architectures report substantial reductions in time-to-insight for new analytical initiatives, with project timelines decreasing significantly for comparable scope. The business data fabric concept represents a fundamental shift from data management to data value creation, where the focus transitions from technical activities such as data movement, transformation, and storage optimization to business-centric activities, including insight generation, decision support, and operational automation. This reorientation enables organizations to redirect technical resources from

10.48047/jocaaa.2025.34.11.58

infrastructure maintenance to value-generating activities, with typical implementations reallocating substantial portions of data engineering capacity from pipeline maintenance to new capability development [8].

The semantic enrichment layer within the business data fabric architecture maintains extensive standard business process definitions spanning financial accounting, supply chain operations, human resources management, customer relationship activities, and manufacturing operations. Each business process definition includes standardized terminology with comprehensive glossaries, relationship mappings defining connections between business entities per domain, business rules encoding validation and calculation logic statements, and data quality metrics establishing acceptable thresholds for completeness, accuracy, consistency, and timeliness. This semantic foundation ensures that data consumers across the organization interpret information consistently, eliminating the costly reconciliation efforts that arise when different departments maintain divergent definitions of seemingly straightforward concepts [7].

4.2. The Rise of "Insight Apps"

One of the most compelling features of BDC is the concept of Insight Apps, which represent a strategic evolution in enterprise software delivery models. These are pre-built, AI-powered applications that solve specific business problems, addressing well-defined use cases such as cash flow analysis, providing liquidity forecasts, supply chain resilience monitoring, identifying disruption risks across supplier networks, or talent acquisition analytics, optimizing recruitment funnel performance. These applications leverage the underlying BDC data foundation and advanced analytical capabilities to deliver out-of-the-box value with minimal configuration requirements. The catalog encompasses multiple Insight Apps at platform launch, with each application targeting specific functional domains and incorporating best-practice analytical frameworks developed through analysis of numerous customer implementations [7].

This approach significantly reduces the difficulty of moving to a cloud analytics model and enables companies to deploy sophisticated systems in drastically reduced timelines compared with building the same functionality from scratch. Development of custom analytics applications usually takes a long time to design and complete the phases of development, testing, and deployment, and consumes many developer hours and costs, depending on geographic location and complexity. In contrast, Insight Apps deployment requires abbreviated timeframes for installation, configuration, and user training, consuming far fewer implementation hours and substantially reduced costs for typical deployments. This represents dramatic reductions in both time-to-value and total implementation cost [8].

Organizations deploying Insight Apps report achieving positive return on investment within shortened periods compared to extended timeframes for custom-built solutions. The Insight Apps represent a paradigm shift where enterprise software transitions from being merely a tool requiring substantial configuration and customization to becoming a complete solution that delivers immediate business outcomes. Each Insight App includes pre-configured data models consuming relevant data products, embedded analytical logic implementing industry best practices, interactive visualizations designed based on user experience research with numerous business users, and automated workflows that trigger alerts and recommendations based on configurable business rules [7].

User adoption metrics demonstrate the effectiveness of this approach, with Insight Apps achieving significantly higher active user adoption rates within the first month of deployment compared to substantially lower rates for custom applications over similar timeframes. Higher adoption rates translate directly to faster value realization and stronger return on investment. The economic impact of Insight Apps extends beyond implementation cost savings to include operational benefits realized through improved decision-making. Organizations utilizing cash flow forecasting applications report reducing

10.48047/jocaaa.2025.34.11.58

working capital requirements through more accurate liquidity planning, while supply chain resilience applications have enabled substantial reductions in disruption-related costs through earlier risk identification and proactive mitigation. Talent acquisition applications have reduced time-to-fill for critical positions and improved candidate quality scores through data-driven recruitment optimization [8].

4.3. A Path for Modernization

For SAP customers with legacy on-premise systems, BDC provides a clear and managed modernization path that addresses one of the most significant challenges facing enterprise IT organizations. It allows organizations to leverage their existing investments in on-premise infrastructure, custom-developed applications, and accumulated business process knowledge while gradually transitioning to a cloud-native, AI-driven architecture. Current market analysis indicates that a substantial portion of SAP customers continue to operate primarily on-premise deployments, with many organizations facing considerable technical debt accumulated over extended periods of ERP operation. These legacy environments often include numerous custom-developed programs, hundreds of customized business processes, and deep integration with dozens of peripheral systems [7].

The BDC offering for SAP BW customers provides extended maintenance guarantees, ensuring support availability for multiple years ahead, providing a substantial planning horizon for modernization initiatives. This agnostic support window allows organizations to plan and execute a migration at a suitable pace and aligned with business priorities, rather than reactively migrating because of support expiration deadlines. The transition from on-premise to a fully managed cloud solution is intended to be carried out in phases. Similarly, modernization generally takes extended periods to complete, depending on legacy application complexity. Initial phases typically involve establishing the cloud data foundation and migrating portions of reporting workloads over defined timeframes. Subsequent phases expand coverage to the majority of analytical use cases over additional periods. Final phases complete the migration and decommission legacy infrastructure over the remaining duration [8].

By allowing for incremental change, customers can furthermore still benefit from their deep investment in SAP, which for large enterprises can be many millions cumulatively in licensing, implementation, customization, and operational costs over decades, while adopting the latest emerging technologies such as AI, machine learning, real-time analytics, and cloud-native scalability. Disruption and business interruption are not acceptable features of a legacy transition, and thus the modernization path is intended to be non-disruptive and gradual enough in phases for organizations to minimally disrupt their data infrastructure at their own pace with business continuity. Migration planning frameworks incorporate parallel operation periods where legacy and modern systems operate concurrently for extended durations per phase, enabling validation of new platform capabilities before decommissioning legacy components [7].

Organizations following the structured modernization path report maintaining business operations without service interruptions while achieving substantial benefits from cloud migration. Total cost of ownership analysis demonstrates that cloud-based architectures reduce infrastructure costs significantly through the elimination of data center expenses, hardware refresh cycles, and dedicated infrastructure management personnel. Operational benefits include dramatic reductions in time spent on system maintenance and upgrades, enabling IT organizations to redirect resources from infrastructure management to innovation initiatives. Performance improvements are substantial, with query response times improving significantly for standard analytical workloads and data refresh latency decreasing from overnight batch windows to near-real-time synchronization with minimal delays. These combined benefits enable organizations to modernize their data infrastructure while realizing compelling return on investment, with typical payback

periods falling within reasonable timeframes considering both cost savings and value generated through enhanced analytical capabilities [8].

| Strategic Benefit | Description |
|-------------------------------|--|
| Enhanced AI Reliability | BDC provides a trusted, business-contextualized data foundation, which is crucial for training reliable and bias-free AI models. |
| Reduced Data Silos | The business data fabric architecture unifies data from all sources, eliminating fragmentation and creating a single source of truth. |
| Accelerated Time-to-Value | Pre-built "Insight Apps" and the zero-copy integration with Databricks accelerate the deployment of advanced analytics and AI solutions. |
| Future-Proofing Data Strategy | BDC provides a scalable and flexible platform that can easily integrate new technologies and data sources as they emerge. |

Table 3: Key Strategic Benefits of SAP BDC

5. Unpacking the Economics: A Closer Examination of Licensing and Costs

5.1. A New Commercial Model: From Users to Consumption

SAP Business Data Cloud is a revolutionary change in commercial approach, shifting away from legacy user-based licenses to a more responsive, consumption-based model that more closely ties costs to platform use. The platform is provided on a subscription basis, and its utilization is quantified mainly in terms of Capacity Units and compute units, a departure from traditional named-user licensing that has long been the norm in enterprise software. Market analysis shows that consumption pricing patterns are being increasingly adopted in cloud data platform segments, where the patterns in workloads are characterized by high volatility and organizations require higher cost alignment to business value [9].

A major convenience for customers is the single licensing model, where independent licensing agreements are not needed because advanced analytics functionality is embedded as part of the core elements of BDC licensing. While processing units are consumed by analytics workloads, that consumption is incorporated within the overall Capacity Unit framework, making billing and contracting easier through the avoidance of having organizations deal with multiple vendors and reconcile independent invoices. This integration usually decreases procurement administrative costs by simplifying contract negotiations, invoice payment, and vendor management processes. The combined model also extends to analytics and visualization services, which are moving from user-based designs to units of capacity, tying costs to true use instead of fixed quantities of users. Companies with cyclic business models or fluctuating analysis needs are especially advantaged by such flexibility, where expenses grow in proportion to use instead of having fixed capacity allowances [10].

The consumption model solves a chronic problem wherein traditional user-based licensing tends to result in either over-provision with large amounts of spare capacity to meet peak loads and projected expansion, or under-provision resulting in license deficiencies and user access limitations during peak times for the business. Early adopter financial analysis suggests that consumption models can lower aggregate licensing expenses for organizations with fluctuating workload patterns, with enhanced flexibility to absorb demand variations without procurement lags. The model also democratizes access to analytics power among organizations because departments can provision to larger sets of users without incurring

significant incremental fixed costs, resulting in higher adoption rates and extended value realization throughout the enterprise [9].

5.2. Capacity Units (CUs) and Databricks Units (DBUs)

Capacity Units are a unit of measurement for platform services, with several infrastructure components packaged into one consumption measure. A Capacity Unit contains query processing and data transformation compute resources, active analytical data, and historical data archival storage capacity, data integration services for linking source systems and managing data flows, metadata cataloging and governance, and analytics capabilities for executing queries and rendering visualizations. This model makes budgeting easier by allowing a single, combined metric for a group of services in place of asking companies to forecast and procure compute instances, storage volumes, data transfer bandwidth, and software licenses separately [9].

Organizations buy capacity pools and draw down from them according to usage patterns, which may ramp up for high-use periods like month-end financial closes, quarterly business reviews, or yearly planning cycles, and ramp down during less busy periods. Standard enterprise deployments provision baseline capacity to manage mean workload requirements with auto-scale, meaning adding incremental capacity during high-use periods. Capacity planning analysis determines that an organization's size baseline capacity to support average levels of utilization, with the peaks being addressed through burst capacity. This provides the elasticity that avoids the over-provisioning associated with earlier fixed-capacity methods, where organizations had fixed infrastructure of peak sizes [10].

Processing units for advanced analytics represent a normalized unit of processing capability on the platform. Usage depends on the size and nature of the compute instance as well as the nature of the workload being run. Workloads like data engineering with intense transformations on large datasets, data warehousing with frequent concurrency patterns, or AI model training requiring special acceleration are charged at varying rates corresponding to their computation load and resource intensiveness. Usage is charged on a per-second basis, which provides fine-grained and equitable billing for computationally intensive operations. Second-level billing granularity avoids organizations' paying for unutilized capacity that is inherent to finer billing increments. Organizations running many short-lived queries stand to benefit most from per-second billing over hourly or daily billing schemes that would charge across full time increments irrespective of utilization time duration [9].

5.3. Total Cost of Ownership (TCO) and Pricing Tiers

Pricing tiers are flexible, allowing for different customer usage patterns and needs. The option of a pay-as-you-go model is present, charging no upfront costs and billing for compute capacity per second on an as-used basis, allowing the greatest level of flexibility for organizations with sporadic workloads or proof-of-concept testing. This model is especially appealing for first-time deployments where usage patterns are still unknown and companies want to reduce financial risk through assessment periods. With organizations that have stable workloads and long-term utilization, committed use contracts and prepaid capacity units provide significant discounts over pay-as-you-go pricing. Discounting levels generally rise with commitment term and size, with larger commitment terms and larger sizes having more attractive pricing arrangements that can have a significant effect on effective per-unit costs [10].

In considering the total cost of ownership of BDC, both direct and indirect costs have to be considered in a general picture. Simply concentrating on the direct licensing costs is not enough, as this is only one of the components of total economic impact. Organizations should also take into account tremendous savings in terms of less manual effort since data pipeline development and maintenance can be automated, effectively removing huge amounts of data engineering work that could take tremendous

10.48047/jocaaa.2025.34.11.58

annual efforts and signify major labor expenditures. The removal of data redundancy and cumbersome pipeline administration also creates savings because storage infrastructure expenses fall when there are fewer copies of datasets necessary, and pipeline maintenance overhead is not needed. Enterprises with multiple replicated data copies across analytics environments can achieve significant storage cost savings through zero-copy architectures [9].

Furthermore, quicker time-to-value from new analytics and AI initiatives drives economic return through the faster realization of revenue or cost savings from analytical initiatives. Projects that once took protracted development lifetimes can now be rolled out in significantly shorter timescales, enabling organizations to achieve business value several months sooner than traditional methods. This ramp-up accumulates over a series of projects, allowing organizations to deliver more initiatives over annual planning horizons and elevating the overall volume of the analytics function. The new model of licensing also supports improved financial control by providing visibility to track and allocate the cost of data and analytics spent on particular use cases or business domains using detailed cost allocation methods. This ability establishes a one-to-one association between investment in technology and business results, allowing better prioritization decision-making and more transparent return on investment demonstration to business stakeholders and executive management [10].

| Service | Core Metric | Pricing Model | Billing Provider |
|-----------------------------------|-----------------------|---|-----------------------------------|
| SAP Datasphere | Capacity Unit (CU) | Consumption-based (pay-as-you-go and committed) | SAP (as part of BDC contract) |
| SAP Databricks (Data Engineering) | Databricks Unit (DBU) | Consumption-based, with pay-as-you-go and committed contracts | SAP (as part of BDC contract) |
| SAP Databricks (Data Warehousing) | Databricks Unit (DBU) | Consumption-based, with pay-as-you-go and committed contracts | SAP (as part of BDC contract) |
| SAP Databricks (AI/ML) | Databricks Unit (DBU) | Consumption-based, with pay-as-you-go and committed contracts | SAP (as part of BDC contract) |
| Insight Apps | Capacity Unit (CU) | Subscription-based (requires a minimum license commitment) | SAP (as part of the BDC contract) |

Table 4: BDC Licensing and Pricing Breakdown

6. Competitive Landscape: BDC versus the New Data Stack

6.1. BDC versus Snowflake

Snowflake leads the market in data warehousing and has a much greater market share in the cloud data warehouse space than SAP Data Warehouse Cloud, as indicated by enterprise customer count. Snowflake's customer base crosses thousands of organizations worldwide, with annual revenue showing significant market adoption and enterprise purchase. The main strength of Snowflake is its highly scalable, cloud-native architecture, which is a multitenant, general-purpose environment well-optimized for various types of analytical workloads and varied data sources. The solution accommodates concurrent query execution for large user bases, with published implementations supporting large query volumes per day and data warehouses with enormous data volumes. Performance benchmarks exhibit maximum query response times for typical business intelligence loads in data sets with large record volumes [11].

The competitive strength of SAP BDC is not in sheer market share but in its native, in-depth integration in SAP applications and its data modeling in the context of the business. Whereas Snowflake is a great generalist data analysis platform for analyzing data from any source with native connectivity to hundreds of sources through its partner network, BDC is specifically designed for SAP-heavy enterprises. Industry research shows that organizations running SAP systems controlling the vast majority of their core transactional data realize much more value from BDC's pre-integrated solution. Its capacity to offer trusted, semantically enriched data products out of the box greatly lessens the effort involved in getting mission-critical ERP data ready for analytics and AI. Relative implementation studies show that organizations using BDC's pre-built SAP data products realize significantly quicker initial analytics deployments compared to similar Snowflake implementations with custom SAP data modeling, which is a meaningful time-to-value reduction [12].

The effort gap is largely due to the necessity of reverse-engineering SAP business logic for generalist platform usage. SAP S/4HANA projects usually involve thousands of intricate database tables with relationships, for which hundreds of domain SMEs are needed to properly represent business entities in a model. Companies developing custom SAP semantic layers on Snowflake indicate spending significant

developer hours per major business domain, and overall semantic layer development takes lengthy time periods for extensive enterprise coverage. This renders BDC the better option for organizations where the main business logic and core transactional information are in the SAP universe, especially those where SAP data is the majority of the analytical needs. Nevertheless, for organizations with heterogeneous data worlds where SAP is a minority of analytical data requirements, Snowflake's wider ecosystem and generalist appeal might provide greater flexibility [11].

6.2. BDC vs. Microsoft Fabric

Microsoft Fabric and SAP BDC have a common strategic vision to offer a single, end-to-end platform for data, analytics, and AI. Microsoft Fabric, introduced as an overarching data platform that brings together several erstwhile standalone services, is supposed to be one pane of glass that blends well with cloud services and offers robust orchestration for hybrid environments. The platform integrates capabilities across data engineering, data warehousing, data science, real-time analytics, and business intelligence into one SaaS solution. Both platforms compete head-to-head in the data integration and cloud database management spaces, with total addressable markets estimated to be substantial annually across enterprise data management, analytics, and AI spaces [11].

Microsoft provides a broad set of built-in connectors, with rich native data source connectors provided by Power BI and Azure Data Factory, and a partner ecosystem to enable data pull from SAP systems. Microsoft SAP connectivity includes qualified connectors for SAP ERP, S/4HANA, BW, and HANA databases, enabling batch as well as real-time data patterns of extraction patterns. But these connectors offer technical data access independent of inherent business context and necessitate custom semantic models by organizations. BDC's differentiator, though, is pre-packaged, semantically enriched data products provided on platform launch. Whereas a Microsoft Fabric customer would have to construct their own data pipelines and semantic models to access SAP data, normally demanding many development hours per business area and lengthy periods for first-time implementation, a BDC customer merely has to tap into these pre-curated data sets ready to be consumed by BI tools or AI models within reduced platform provisioning timeframes [12].

This offers significantly quicker time-to-value to organizations that need to gain thorough insight into their SAP core business processes. Comparative analysis shows that BDC implementations attain initial user productivity significantly quicker, while Microsoft Fabric implementations to SAP data usually take longer to achieve parity in productivity levels. The effort gap is especially evident in sophisticated SAP business processes like order-to-cash, procure-to-pay, and financial close, where pre-configured data products reduce most custom development effort. For enterprises deeply committed to the Microsoft environment with large non-SAP data sets, Microsoft Fabric brings value through native integration with Azure services, Microsoft 365, Dynamics 365, and Power Platform. But for SAP-centric businesses whose SAP data constitutes most of their analytical needs, BDC's purpose-built method provides better time-to-value and lowered total cost of ownership [11].

6.3. BDC versus Amazon Redshift

Amazon Redshift is a reliable cloud data warehouse product from Amazon Web Services that is popular due to its excellent performance, scalability, and affordability. Redshift has been used by thousands of enterprise customers worldwide since its release, serving data warehouses from small scales to huge petabyte-scale deployments. It provides fast data processing functionality, with columnar storage and massively parallel processing architecture, providing query performance from subsecond responses for basic aggregations up to reasonable timescales for advanced analytical queries over billions of records. Seamless integration with other AWS services, such as S3 storage, Lambda serverless computing,

10.48047/jocaaa.2025.34.11.58

SageMaker machine learning, and Kinesis streaming, makes it a good contender for companies operating intensively within the AWS environment [11].

Redshift's cost is competitive, with on-demand pricing and reserved instance pricing providing significant discounts for longer commitments. Its pay-as-you-go approach presents a compelling return on investment to cost-conscious organizations, with enterprise deployments commonly serving analytical workloads for large-scale user groups at modest annual expenses. SAP Business Data Cloud, on the other hand, is designed specifically for enterprises that have a mission-critical SAP presence. While Redshift excels at general-purpose data warehousing and offers powerful tools for data integration and scalability, integrating SAP's complex and semantically rich data typically requires additional third-party tools such as SAP Data Services, Informatica, Talend, or custom-developed ETL processes. These integration projects typically consume extended timeframes for initial implementation and substantial developer hours, with ongoing maintenance requiring significant portions of initial development effort annually [12].

The BDC platform, by contrast, offers this integration natively in the form of curated data products, without the burden of manual data modeling and replication. Pre-built SAP data products are available to organizations in abbreviated timeframes of platform provisioning instead of months of custom development. This differentiation renders Redshift an effective generalist tool for organizations with mixed data sources where SAP is a minority of analytical data, and BDC is a specialized, out-of-the-box solution for SAP-dominated workloads where SAP data forms the main analytical foundation. Total cost of ownership analysis shows that for SAP-heavy scenarios, BDC can achieve significantly lower TCO over several years based on integration development, maintenance costs, and reduced time-to-value, even with potentially higher per-unit platform prices, versus Redshift-based options. For organizations with little SAP data representation, the lower infrastructure costs and larger ecosystem of Redshift can offer better economics [11].

6.4. BDC vs. Google BigQuery

Google BigQuery is a serverless, fully managed, and highly scalable data warehouse service by Google Cloud Platform. Ever since its general availability launch, BigQuery has been embraced by many organizations around the world, with high query processing volumes on its customer base. Its key strengths are real-time analytics with high-throughput streaming ingestion, integrated machine learning with BigQuery ML supporting SQL-based model training and deployment, and multi-cloud support with BigQuery Omni, enabling users to analyze data from GCP, AWS, and Azure without data movement. The serverless architecture of the platform makes infrastructure management easy through automatic scaling of compute resources according to demand, and its SQL-based query makes it easily accessible to a variety of data professionals with knowledge about common ANSI SQL syntax [11].

Performance tests show how BigQuery can read and process huge sets of data efficiently, with costs for queries depending on the data volume processed. Organizations see the best query response times for typical analytical queries on datasets with billions of records, where sophisticated aggregations and joins take reasonable processing times based on data size and query complexity. The most significant difference between SAP Business Data Cloud and BigQuery is their fundamental philosophy. BigQuery is a robust, versatile product intended for use in general-purpose analytics, whereas BDC is a bespoke platform heavily integrated with the SAP environment. For an organization with a significant SAP footprint where SAP data constitutes most analytical needs, BigQuery can be an effective analytics engine, but SAP Datasphere or similar orchestration tools would be needed for it to be used as an orchestration layer so that data from S/4HANA could be properly transformed and loaded [12].

10.48047/jocaaa.2025.34.11.58

This pattern of integration usually calls for organizations to have data pipelines with solution tools like SAP Data Services, Google Cloud Data Fusion, or homegrown extraction logic, with deployment durations spanning over lengthy periods for the first deployment and considerable developer time. This implies that although BigQuery provides remarkable scalability with established implementations handling massive data quantities and sophisticated features such as geospatial analytics, time-series analysis, and federated querying across a multitude of sources, BDC offers a simplified, governed, and pre-integrated data foundation that enhances SAP-specific use case time-to-value by offering trusted data products with embedded business context. Comparative evaluations show that BDC deployments become production-ready much more quickly than corresponding BigQuery deployments targeting SAP data, with the first analytics functionality available within truncated timeframes. But for companies with extensive non-SAP data origins or those needing sophisticated geospatial or machine learning features not yet supported in BDC, BigQuery's wider set of features and established scalability at extreme data volumes may be a more suitable option [11].

6.5. The Platform vs. The Generalist Tool

The competitive dynamic of SAP Business Data Cloud is not merely a comparison of technical attributes like query performance metrics, storage limits, or support for concurrent users. It is a choice between a best-of-breed generalist tool and a strategically packaged platform. A CIO has a crucial choice: either adopt a generalist platform such as Snowflake, Microsoft Fabric, Amazon Redshift, or Google BigQuery and afterwards invest considerable time, expense, and effort into creating custom data pipelines and semantic layers to gain SAP data value, or spend money on the pre-integrated BDC solution. Economic comparative analysis shows that the custom integration route generally demands large initial investments in integrating SAP data across broad business processes, with longer implementation periods and continuing annual maintenance charges accounting for large shares of up-front development investment [11].

The strength of BDC resides in its potential for faster time-to-value and less complexity. Through the provision of an integrated managed service with pre-built data products for wide-ranging standard business processes and a single governance framework, BDC eliminates a significant source of friction for organizations where SAP data is the lifeblood of the business. Organizations whose SAP systems handle most financial transactions, control most supply chain operations, or hold most master data usually achieve best value from BDC's integrated solution. This strategic positioning makes BDC an attractive option, especially for companies that value out-of-the-box integration and less complex data architecture over pure platform agnosticism. Total cost of ownership modeling shows that for SAP-centric companies, BDC can provide significantly lower multi-year TCO than comparable custom-integrated generalist platforms, taking into consideration implementation costs, maintenance costs, opportunity cost of deferred insights, as well as lower technical debt [12].

Nevertheless, the open design of the SAP-Databricks collaboration makes possible a hybrid approach where customers can utilize the trusted SAP data of BDC within their current, native analytics ecosystem, allowing them to combine SAP and non-SAP data seamlessly. This hybrid model is especially useful for organizations with balanced data environments in which SAP data comprises a large but not dominant share of analytical needs, along with significant customer relationship management, marketing automation, e-commerce, IoT sensor, or external marketplace data. The connect service integration pattern allows organizations to leverage existing investments in analytics platforms while unlocking access to SAPHANA-curated data products, essentially offering the best of both worlds in the form of a single, unified data lakehouse combined with the ease of a managed platform. Adoption trends reveal that

most organizations with established analytics platforms opt for the hybrid model, leveraging existing capabilities while reducing SAP data integration timelines by a significant amount [11], [12].

| Metric | SAP Business Data Cloud | Snowflake | Microsoft Fabric | Amazon Redshift | Google BigQuery |
|---------------------------|---|--|--|---|--|
| Primary Strength | Native integration with SAP and semantically rich data products | Scalability, multi-cloud flexibility, and performance | Unified "single pane-of-glass" for data and AI workloads | High-speed data processing and cost-effectiveness | Serverless architecture and real-time analytics |
| Data Governance | End-to-end governance via Unity Catalog | Centralized governance via Snowflake Governance | Unified governance via Microsoft Purview | Strong data governance and security features | Advanced security features, fine-grained access control, and column-level encryption |
| Data Sharing | Zero-copy, bi-directional Delta Sharing with full semantics | Zero-copy, bi-directional Delta Sharing & Data Marketplace | Open Mirroring, OneLake for seamless data sharing | Data sharing and cluster snapshots for backups | Multi-cloud data analysis with BigQuery Omni |
| SAP Integration | Native integration via predefined Data Products | Strong partner ecosystem for data extraction and modeling | Robust connectors and open mirroring with certified partners | Requires third-party tools for SAP data extraction | Replication from SAP S/4HANA via SAP Datasphere as an orchestration layer |
| Target Audience | Enterprises with a large, critical SAP footprint | Multi-cloud enterprises with diverse data sources | Enterprises in the Microsoft ecosystem | Companies operating primarily on AWS | Organizations operating on GCP or in a multi-cloud environment |
| AI/ML Capabilities | Native Databricks integration for pro-code AI, Joule AI agents | Snowpark for data science, AI/ML workloads | Integrated AI capabilities within the Fabric platform | Built-in analytics and in-database ML functionality | Built-in machine learning (BigQuery ML) and AI/ML integration |
| Deployment Model | Fully managed SaaS | Fully managed SaaS | Fully managed SaaS | Fully managed, hosted data warehouse | Fully managed, serverless data warehouse |

Table 5: SAP BDC vs. The Modern Data Stack - A Competitive Analysis

7. Strategic Recommendations and Future Outlook

7.1. An Assessment of the Future State

10.48047/jocaaa.2025.34.11.58

SAP Business Data Cloud signifies a significant milestone in enterprise data strategy, showcasing a transformative architectural change in the manner in which organizations think about data management and analytics in the SAP ecosystem. The SaaS solution also reflects a trend in the wider industry toward consumption-based services and less on-premises infrastructure management, as research demonstrates that software spending based on consumption is gaining momentum, having a higher growth rate compared to more traditional software purchasing license models, which may be inhibiting growth to the single digits. Industry analysts project that within the next several years, the majority of enterprise data platform spending will shift to consumption-based models, representing a fundamental transformation in software economics and customer expectations around flexibility and cost alignment with business value [13].

The platform is positioned to become the central data hub for the SAP ecosystem, offering a simplified and governed way to leverage data for everything from BI reporting in SAP Analytics Cloud to the development of custom AI models within embedded analytics environments. Strategic positioning analysis indicates that the unified platform approach addresses critical market needs, with research showing that a significant majority of enterprises identify data fragmentation as a top barrier to AI adoption, and substantial portions cite lack of business context in data as a primary obstacle to analytics value realization. This strategic move is poised to strengthen market positioning as a leader in enterprise data management by providing a cohesive and powerful offering that directly addresses the persistent challenges of data silos and AI operationalization that have plagued enterprises for decades [14].

The platform's architecture, supporting real-time data synchronization, semantic enrichment of business data, and integrated AI/ML capabilities, positions it to capitalize on emerging market trends. Market sizing analysis projects substantial growth in the enterprise data management market over the coming years, with AI-augmented analytics representing the fastest-growing segment. Organizations implementing unified data platforms report substantially faster time-to-insight for new analytics initiatives and significant reductions in data infrastructure total cost of ownership compared to fragmented multi-platform approaches that require extensive custom integration work. The strategic integration of advanced analytics capabilities natively within the data platform eliminates traditional friction points where data scientists historically spend the majority of project time on data access and preparation rather than model development and insight generation [13].

7.2. Potential Challenges and Risks

While the BDC offering presents significant advantages, a critical assessment of potential challenges is essential for organizations making long-term strategic technology investments with multi-year implications. Despite its use of open standards, including Delta Sharing for data sharing and Unity Catalog for governance, the bundling of services and the nature of the partnership contract create potential considerations around vendor relationships and platform flexibility. Customers must carefully weigh the benefits of a simplified, pre-integrated solution against the potential future costs and complexities of switching to a different platform should business requirements or vendor relationships change over time. Industry research on enterprise software switching costs indicates that platform migrations typically consume extended timeframes and require substantial investments depending on customization depth and data volumes managed within the platform [13].

Moreover, the platform is able to consume third-party data from multiple non-SAP sources by utilizing connectors; however, informed design decisions would note that the platform is specifically engineered to integrate with the SAP ecosystem. For organizations with a complex data landscape where SAP accounts for a minority of the analytical data, or large organizations with a majority of their data estate outside of

10.48047/jocaaa.2025.34.11.58

SAP (e.g., customer relationship management (CRM), marketing automation, e-commerce and IoT sensors with market data systems) the case for a more neutral data platform may provide some flexibility and possibly a more complete data solution. Economic analysis indicates that organizations where SAP data represents a smaller portion of total analytical requirements may experience a higher total cost of ownership with SAP-centric platforms compared to agnostic alternatives, as maintaining dual platform capabilities or undertaking extensive data replication becomes necessary [14].

These organizations must assess whether the out-of-the-box value of BDC's semantically rich data products, which can substantially reduce SAP data integration effort, outweighs the potential need for a more agile, platform-agnostic solution that provides equal native integration across diverse source systems. Risk assessment frameworks should evaluate several factors, including the current and projected SAP data percentage of total analytical requirements, anticipated growth in non-SAP data sources over multi-year planning horizons, organizational cloud strategy and multi-cloud requirements, existing analytics platform investments and user community size, and tolerance for vendor relationship concentration. Organizations with high scores across multiple risk dimensions may benefit from hybrid architectures leveraging the connect service pattern rather than fully managed deployments that consolidate all analytics workloads within a single vendor ecosystem [13].

7.3. Actionable Recommendations for a C-Level Audience

Based on this comprehensive analysis, strategic recommendations are provided for senior-level executives and IT decision-makers navigating complex technology investment decisions with significant budget implications. For existing SAP customers, conducting a BDC readiness assessment to evaluate the current data landscape, including existing SAP BW or S/4HANA environments, is crucial as a foundational first step. This assessment should quantify several key metrics, including percentage of analytical requirements sourced from SAP systems, current annual spending on SAP data integration and maintenance activities, time-to-delivery for new SAP-based analytics initiatives, and data quality incident rates for SAP-derived information. Organizations where SAP represents the majority of analytical data and current integration costs are substantial typically demonstrate the strongest business cases for platform consolidation [13].

For organizations with complex legacy footprints, including decades of accumulated custom development, extensive ABAP customizations, and deep integration with dozens of peripheral systems, a phased approach is strongly recommended to manage risk and validate value incrementally. Beginning with a greenfield project targeting a well-defined business domain, such as financial performance analytics or supply chain visibility, validates the value proposition and demonstrates positive ROI before embarking on full-scale migration of mission-critical systems. Pilot projects should be scoped to abbreviated timeframes for initial deployment, serving moderate user populations, and targeting several high-value use cases. Success metrics should be clearly defined, including time-to-insight reduction targets, user adoption rates exceeding specified thresholds, and measurable business outcomes such as improved forecast accuracy or reduced working capital requirements [14].

For new customers or those early in SAP cloud adoption journeys, evaluating SAP Business Data Cloud not as a collection of individual tools but as a unified platform is essential for an accurate total cost of ownership assessment. The business case should be centered on the total value provided by out-of-the-box access to semantically rich data, which promises faster time-to-value for high-impact analytics and AI projects. This approach should be rigorously compared against the time and resources required to build and maintain a similar capability using general-purpose data platforms. Comparative analysis should quantify implementation timelines, ongoing maintenance effort, and time-to-value for representative use

10.48047/jocaaa.2025.34.11.58

cases. Organizations typically find that BDC delivers substantially faster implementation and lower multi-year TCO for SAP-centric use cases compared to custom-integrated alternatives [13].

For stakeholders like chief financial officers, chief data officers, and line-of-business executives, this new consumption-based licensing model provides greater transparency and financial oversight than previous fixed licensing models could provide. As part of this analysis, consideration should be given to starting cross-functional conversations with IT and finance, as well as across different lines of business, to find common ground on the economics of the data platform. Being able to see data spent by specific business user cases through usage analytics and cost allocation creates a straight link between technology investment and demonstrated business value, which can then serve to align data strategy with enterprise goals. Organizations should establish governance frameworks for capacity planning, usage monitoring, and chargeback or showback models that assign costs to consuming departments based on actual utilization, creating accountability and encouraging efficient resource use [14].

Conclusion

SAP Business Data Cloud is a major shift in enterprise data strategy, representing a complete change in architecture that addresses historical data fragmentation, governance challenges, and AI operationalization barriers that have constrained organizational analytics capabilities. This managed SaaS model fits in well with major industry movements to consumption-based services and delegating infrastructure management, making it the primary data platform within the SAP ecosystem. The synergies BDC enables through its integration with Databricks, further combined with a focus on semantically-rich data products, provide significant value for organizations that have SAP systems as the primary source of essential business-critical transactional data. The zero-copy architecture, unified governance, and pre-built data products all help reduce time-to-value while reducing technical complexity and total cost of ownership. However, organizations need to consider their specific data landscape composition, which includes the SAP product percentage of overall analytical needs, investments in their current level of SAP solutions, and a capacity for vendor relationship risk concentration. For enterprises using SAP as a core platform, BDC can deliver favorable economic factors compared to custom integration solutions while providing a more timely path to implementation and adoption. The platform's support for hybrid architectures through connect services provides flexibility for organizations with balanced data landscapes to leverage curated SAP data while maintaining existing analytics infrastructure. As consumption-based licensing models gain prevalence and enterprises increasingly prioritize AI capabilities, BDC positions itself strategically at the convergence of data management, advanced analytics, and artificial intelligence, enabling organizations to transform data assets into competitive advantages through intelligent, automated business processes that continuously learn and adapt to evolving conditions.

References

1. SAP, "SAP Debuts Business Data Cloud with Databricks to Turbocharge Business AI," 2025. [Online]. Available: <https://news.sap.com/2025/02/sap-business-data-cloud-databricks-turbocharge-business-ai/>
2. Raja Gupta, "What Is SAP Business Data Cloud? Benefits, Architecture & Use Cases," SAP Community, 2025. [Online]. Available: <https://community.sap.com/t5/technology-blog-posts-by-sap/what-is-sap-business-data-cloud-benefits-architecture-amp-use-cases/ba-p/14033086>
3. Ali Ghodsi, et al., "Introducing SAP Databricks," Databricks, 2025. [Online]. Available: <https://www.databricks.com/blog/introducing-sap-databricks>
4. JansiMurugesan, "Session 6 Recap: SAP Business Data Cloud + Databricks – Why They’re Stronger Together," SAP Community, 2025. [Online]. Available: <https://community.sap.com/t5/enterprise-architecture-blog-posts/session-6-recap-sap-business-data-cloud-databricks-why-they-re-stronger/ba-p/14156762>
5. SAP Product Information, "SAP Databricks in Business Data Cloud." [Online]. Available: <https://www.sap.com/products/data-cloud/databricks.html>
6. Jing Wen, "SAP Databricks: Building an Intelligent Enterprise with AI Unleashed – Part 1," SAP Community, 2025. [Online]. Available: <https://community.sap.com/t5/technology-blog-posts-by-sap/sap-databricks-building-an-intelligent-enterprise-with-ai-unleashed-part-1/ba-p/14166813>
7. Gambit, "What is SAP Datasphere?" [Online]. Available: <https://www.gambit.de/en/wiki/sap-datasphere/>
8. Paresh Kolte, "AI in SAP: Why your data strategy will make or break your success," Tietoevry Tech Services, 2025. [Online]. Available: <https://www.tietoevrytechservices.com/en/blog/25/ai-in-sap-why-your-data-strategy-will-make-or-break-your-success/>
9. SAP, "What You Need to Know About SAP Datasphere Licensing and Its Role in Data-Driven Procurement," 2Data, 2025. [Online]. Available: <https://www.2-data.com/knowledge-hub/what-you-need-to-know-about-sap-datasphere-licensing-and-its-role-in-data-driven-procurement>
10. CloudChipr, "Understanding Databricks Pricing on AWS, Azure, and Beyond," 2025. [Online]. Available: <https://cloudchipr.com/blog/databricks-pricing>
11. 6sense, "Snowflake Vs SAP Data Warehouse Cloud: In-Depth Comparison." [Online]. Available: <https://6sense.com/tech/data-warehousing/snowflake-vs-sapdatawarehousecloud>
12. LeverX Team, "The Top DWH Platforms Compared: SAP Datasphere vs. Snowflake, BW, and Others," LeverX, 2025. [Online]. Available: <https://leverx.com/newsroom/the-best-dwh-platforms-comparison>
13. Carsten Bange, et al., "BARC Perspective: SAP Business Data Cloud (BDC) – Breaking Tradition and Embracing Data Products," BARC, 2025. [Online]. Available: <https://barc.com/sap-bdc-barc-perspective/>
14. Debraj Ray, "SAP Business Data Cloud: Rebrand or Game Changer," Capgemini, 2025. [Online]. Available: <https://www.capgemini.com/gb-en/insights/expert-perspectives/sap-business-data-cloud-rebrand-or-game-changer/>