edv-buchversand.de

SAP Business Data Cloud

Das umfassende Handbuch

» Hier geht's direkt zum Buch

DIE LESEPROBE

Diese Leseprobe wird Ihnen von www.edv-buchversand.de zur Verfügung gestellt.

Kapitel 4

SAP Databricks: Datenaustausch und KI

In der Wirtschaft bringen Daten und ein effizienter Umgang mit ihnen einen entscheidenden Wettbewerbsvorteil. Das Ziel vieler Unternehmen besteht daher nicht nur in der Persistierung dieser Daten, sondern vor allem in der Gewinnung wertvoller Erkenntnisse.

Während SAP-Systeme seit Jahren die Grundlage für diverse Unternehmensprozesse bilden, eröffnet die Plattform *Databricks* eine neue Dimension für Künstliche Intelligenz (KI) und Machine Learning (ML). Durch die im Mai 2025 durchgeführte Zusammenführung beider Welten in *SAP Databricks* wird ein neuer Maßstab für prozessorientiertes Handeln und innovative Datenintelligenz gesetzt.

In diesem Kapitel erfahren Sie, welche neuen Möglichkeiten sich mit SAP Databricks für die Datenverarbeitung und Analytics eröffnen. Wir zeigen Ihnen, wie sich SAP-Daten effizient in die Lakehouse-Architektur integrieren lassen und welche Performance-Vorteile Pipelines bieten. Praxisnah wird vermittelt, wie SAP-Daten für KI-Modelle, BI-Anwendungen oder weiterführende Analytics-Anwendungen nutzbar gemacht werden können. Nach der Lektüre dieses Kapitels werden Sie verstehen, wie Sie die Stärken strukturierter SAP-Prozesse mit der Flexibilität und Skalierbarkeit von SAP Databricks vereinen können, um Ihre Unternehmensdaten auf ein neues Level zu bringen.

In Abschnitt 4.1 lernen Sie die Kerneigenschaften von SAP Databricks sowie dessen Architektur kennen. Außerdem ordnen wir SAP Databricks in die SAP-Landschaft ein. In Abschnitt 4.2 thematisieren wir den Delta Share und besprechen die Verwendung von Datenprodukten in SAP Databricks. Abschnitt 4.3 beschreibt, wie Sie die SAP BDC mit möglichen Partnerplattformen verbinden können. In Abschnitt 4.4 stellen wir Ihnen den Unity Catalog vor und beschäftigen uns mit der Governance von SAP Databricks. Ab Abschnitt 4.5 wird es mit der Einführung von Notebooks in SAP Databricks deutlich praxisorientierter. Den Einsatz von SQL in Bezug auf SAP Databricks beleuchten wir in Abschnitt 4.6. Zum Abschluss geht es in Abschnitt 4.7 um Data Science und Mosaic AI sowie deren Umsetzung in SAP Databricks.

4.1 Kerneigenschaften von SAP Databricks

Einheitlicher Rahmen

Mit der Einführung der SAP BDC wurde SAP Databricks nativ verfügbar. Es ist nun standardmäßig in der SAP BDC enthalten. *SAP Databricks* schafft einen einheitlichen Rahmen für Datenmanagement und -analysen. Bislang mussten KI-Modelle in externen Systemen entwickelt werden. Nun besteht die Option, kuratierte SAP-Datenprodukte durch eine schnelle Datenextraktion und kurze Aufbereitungszeit in SAP Databricks zu nutzen, ohne die Daten kopieren zu müssen. SAP Databricks unterstützt die Entwicklung von KI-Modellen und -Anwendungen über die gesamte Prozesskette hinweg. Außerdem können SAP-Geschäftsprozesse mittels KI automatisiert werden. Dabei ist ein Zugriff auf SAP- und Nicht-SAP-Datenquellen möglich. Die Plattform stellt GenAI-Fähigkeiten bereit, sodass Unternehmen einen Wettbewerbsvorteil aus ihren Daten erzielen können. Die Umgebung ist kollaborativ und bietet erfahrenen Entwicklerinnen und Entwicklern viele Optionen, da bspw. Spark-Pipelines geschrieben oder SQL-Analysen ausgeführt werden können. Insbesondere die innovativen SQL-Analysen bieten die Möglichkeit, serverlos Einblicke in Echtzeit zu gewinnen.

Zusätzlich wird eine Vielzahl an Funktionen bereitgestellt. Eine dieser Funktionen ist *Delta Sharing*. Delta Sharing verzichtet auf Datenduplizierung, indem ein bidirektionaler Austausch zwischen SAP-Datenprodukten und dem bereits bestehenden Databricks-Lakehouse hergestellt wird. Die *Data-Lakehouse-Architektur* ermöglicht die Integration von strukturierten und unstrukturierten Daten. Die Verwaltung der Funktionen wird mittels eines *Unity Catalogs* realisiert. Dieser schafft die Grundlage für eine optimale Compliance, da die Datenlandschaft so an einem Ort gesteuert und überwacht werden kann.

In diesem Abschnitt lernen Sie zunächst die Architektur von SAP Databricks kennen. Anschließend erläutern wir die Entstehung von SAP Databricks, stellen Ihnen die Bausteine von SAP Databricks vor und zeigen Ihnen ein mögliches Ende-zu-Ende-Integrationsszenario.

4.1.1 Architektur und Verortung von SAP Databricks in der SAP BDC

Damit Unternehmen mit Daten und KI erfolgreich sein können, ist eine Architektur erforderlich, die es ermöglicht, Daten anzureichern. Die Architektur von SAP Databricks erfüllt diesen Anspruch und fügt sich nahtlos in die Architektur der SAP BDC ein. SAP Databricks wird von SAP verwaltet und bietet somit die Möglichkeit, geschäftsrelevante Daten im Rahmen von Data Science und Data Engineering zu erforschen.

ML-/KI-Modelle

SAP Databricks unterstützt die Bereitstellung von Daten an andere SAP-Anwendungen sowie die Implementierung aussagekräftiger ML-/KI-Modelle. Das vorherige Verfahren mit ausgelagerter Implementierung von KI-Modellen führte zu zusätzlichem Arbeitsaufwand und Qualitätsproblemen, da die Daten teilweise

nicht live waren. Durch den Einsatz von SAP Databricks wird die bestehende SAP-Landschaft gezielt optimiert, indem die teilweise vorherrschende mangelhafte Datenqualität bekämpft und deutlich verbessert wird.

Datenprodukte

Die sich in der untersten Architekturschicht befindlichen *Datenprodukte* können mit SAP Databricks zunächst einmal geteilt werden. SAP verwaltet Datenprodukte und sorgt somit dafür, dass Kosten vermieden werden, die ansonsten bspw. im Zuge von Datenextraktionen anfallen würden.

Datenprodukte

SAP-Anwendungen sorgen durch vorkonfigurierte Datenprodukte mit von SAP bereitgestellten Inhalten für eine noch schnellere Wertschöpfung.

Datenmodelle

Im nächsten Schritt werden die Datenprodukte mittels KI erweitert und in die SAP BDC zurückgeführt (siehe Abschnitt 4.2.4). Dort können sie verwendet werden, um bspw. in Datenmodelle innerhalb von SAP Datasphere integriert und in SAP Analytics Cloud zu Reporting-Zwecken erweitert zu werden. Anwenderinnen und Anwender profitieren von dieser Möglichkeit, da so passende KI-Modelle direkt in den Geschäftsprozess integriert und zusätzlich um Drittanbieter-Daten erweitert werden können.

Python und SQL

In SAP Databricks können Daten interaktiv über Notebooks oder SQL-Editoren abgefragt werden. Dafür stehen Python und SQL als Programmiersprachen zur Auswahl. Die Daten werden in verschiedenen Formaten unterstützt.

Business-Data-Fabric

Wie Sie in Abbildung 4.1 sehen, lässt sich SAP Databricks im Business-Data-Fabric-Teil der SAP-BDC-Architektur verorten. Dabei spielen insbesondere Zero Copy und benutzerdefinierte KI/ML eine besondere Rolle. SAP Databricks wurde in den Business-Data-Fabric-Teil integriert, um bei der Prozessoptimierung zu unterstützen und eine Integration der KI-Funktionen direkt in bestehende SAP-Systeme zu ermöglichen.

Zero Copy

Zero Copy und benutzerdefinierte KI/ML sind zwei relevante Konzepte auf denen SAP Databricks basiert. *Zero Copy* zielt darauf ab, sämtliche Datenbewegungen und -replikationen zu vermeiden oder zumindest zu minimieren. Dies ist insbesondere im Rahmen der SAP BTP und der SAP BDC ein entscheidender Faktor, da oftmals sensible Daten in verschiedenen Anwendungen verarbeitet werden und diese somit in ihren Ursprungssystemen verbleiben. Zudem steigt die Effizienz, da stets auf Basis aktueller Daten operiert wird und somit keine Datenschiefstände zu befürchten sind.

Benutzerdefinierte KI-/ML-Modelle

Der Einsatz *benutzerdefinierter KI-/ML-Modelle* ermöglicht die Entwicklung maßgeschneiderter KI-Lösungen und -Modelle. Diese Modelle können von Unternehmen entwickelt bzw. auf entsprechende Geschäftsprozesse zugeschnitten werden. So kann der Einsatz generischer Modelle, die nicht gezielt an die Unternehmensprozesse angepasst wurden, vermieden werden.

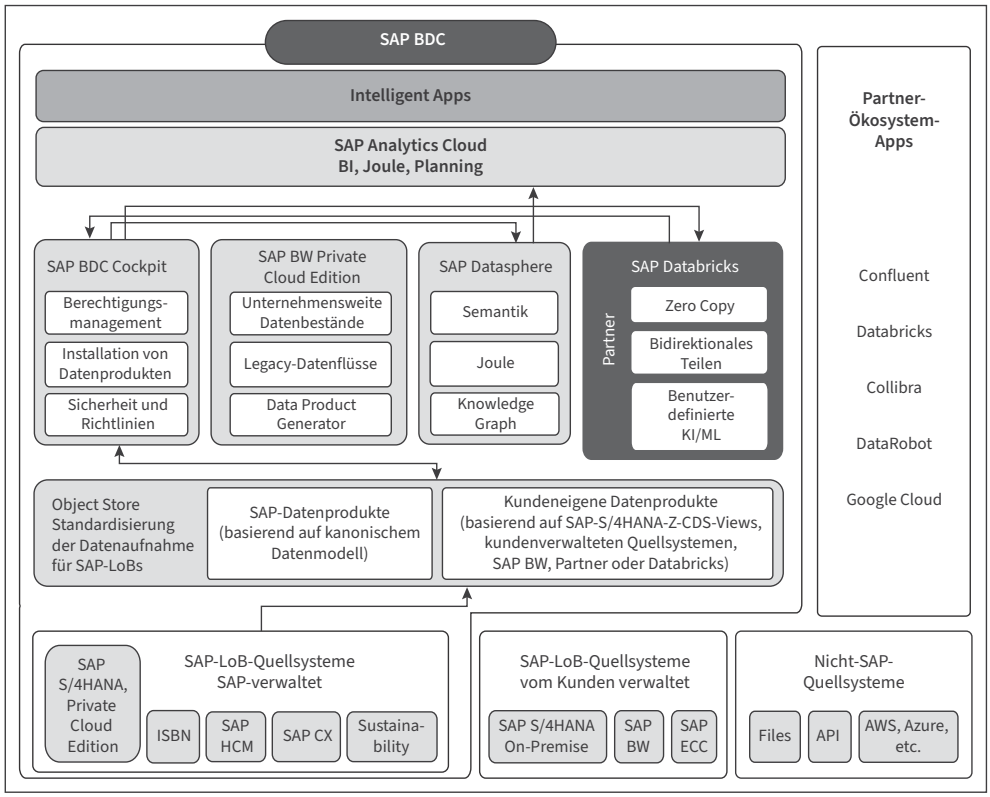


Abbildung 4.1: Architektur der SAP BDC mit SAP Databricks

Trainingsdaten

Im Rahmen der SAP BTP können diese benutzerdefinierten KI-/ML-Modelle in laufende Prozesse, wie bspw. in der Automatisierung, eingesetzt werden. Die Trainingsdaten der Modelle stammen direkt aus den SAP-Systemen und müssen aufgrund des Zero-Copy-Konzepts nicht vor der Verwendung exportiert werden. Außerdem sind diese Modelle in der Lage, Daten in Echtzeit zu konsumieren, sodass das Training bzw. die Modellanpassung in Echtzeit umgesetzt werden kann. Somit ermöglichen Zero Copy und benutzerdefinierte KI-/ML-Modelle eine effiziente KI-Integration für die Analyse und Verwendung von Echtzeitdaten.

4.1.2 Entstehung von SAP Databricks

Produktportfolio

SAP-Daten gelten in der Regel als strukturiert, geschäftskritisch und reglementiert. Databricks fokussiert sich hingegen auf andere Datenmengen, bspw. heterogene Daten, sowie auf verteilte Verarbeitung und agiert datengetrieben. Das Produktportfolio von Databricks wurde in Bezug auf den Daten- und KI-Bereich kontinuierlich erweitert. Ein zentraler Bestandteil war dabei die Weiterentwicklung der Lakehouse-Plattform war.

Die Anfänge von Databricks

Databricks wurde im Jahr 2013 gegründet und sollte zunächst als Plattform für Big Data und Machine Learning dienen. Das Kernelement war *Apache Spark* – ein Framework zur verteilten Datenverarbeitung. Da die Analytics-Plattform viele Funktionen vereinte, gewann Databricks schnell an Beliebtheit. So fand eine rasante Entwicklung von einem Tool für Big Data hin zu einer umfassenden Data-Science- und KI-/ML-Plattform statt. Im Jahr 2017 entwickelte sich Databricks weiter in Richtung Lakehouse. Weitere Informationen zu SAP Databricks Lakehouse finden Sie in Abschnitt 4.1.5. Kurz zusammengefasst vereint ein Lakehouse die Vorteile von Data Lakes, die zur Speicherung von Rohdaten dienen, mit denen eines Data Warehouses.

Der Delta Lake, auf den Databricks zurückgreift, verwaltet Daten mit ACID-Transaktionen. So haben bis 2020 immer mehr Unternehmen Databricks eingesetzt, um ihre großen Datenmengen zu verarbeiten und KI-/ML-Modelle auf ihnen zu trainieren. Seit 2021 gilt Databricks als eine der am schnellsten wachsenden Technologieplattformen, die stetig in die Einführung weiterer Funktionalitäten für Data Science und Machine Learning investiert. In vielen Unternehmensarchitekturen wird Databricks mittlerweile eingesetzt.

Weiterentwicklung

Während der Weiterentwicklung von Databricks standen folgende Themen im Fokus:

- Datenverarbeitung
- Data Warehousing
- Governance
- Business Intelligence
- KI

Kollaboration von SAP und Databricks

Insbesondere Kunden, die neben Databricks klassische SAP-Produkte wie SAP ERP einsetzen, zeigten Interesse an einer Kollaboration zwischen SAP und Databricks, um vertiefende und gewinnbringende Insights zu generieren. Dies liegt unter anderem daran, dass SAP-Daten für Unternehmensprozesse essenziell sind. Databricks verfügt über optimale Datenverarbeitungs- und KI-Funktionen. Allerdings stellte es lange eine Hürde dar, die Daten aus den SAP-Produkten auf eine externe Datenplattform zu übertragen. Dies ist darauf zurückzuführen, dass bislang umständliche ETL-Prozesse verwendet werden mussten, um die Daten für weitere Analysezwecke bereitstellen zu können. Mit der Planung und Entwicklung der SAP BDC wurde eine Möglichkeit geschaffen, die gewünschte Kollaboration zu realisieren.

Vorteile

Die zur SAP BDC zugehörige Databricks-Version wird als *SAP Databricks* bezeichnet. Databricks, von SAP losgelöst, ist unter *Databricks* bekannt. SAP Databricks bringt nun einige Vorteile mit sich. Dazu zählen die native Verarbeitung in Echtzeit und die zukünftige Vermeidung von aufwendigen Workarounds, die aus

Datenreplikation resultieren würden. Dadurch können die Daten dort genutzt werden, wo sie tatsächlich entstehen. Gleichzeitig können andere Speicheroptionen neben SAP HANA eingesetzt werden. Der größte Vorteil ist, dass der Geschäftskontext der Daten erhalten beleibt. Metadaten gehen bei einem Transfer nicht verloren. Zudem kann mithilfe des Unity Catalogs sichergestellt werden, dass die im Unternehmen definierten Geschäftsregeln für SAP-Daten eingehalten werden.

4.1.3 Bausteine von SAP Databricks

Mit der SAP BDC werden SAP-Geschäftsprozesse, analytische Workloads und KI-Anwendungen miteinander vereint. Dabei spielt insbesondere SAP Databricks eine zentrale Rolle. In diesem Abschnitt stellen wir Ihnen die wichtigsten technologischen Bausteine vor und erklären, warum SAP Databricks ein essenzielles Fundament für datengetriebene Innovationen ist.

Delta Sharing

Zunächst betrachten wir das Prinzip des Delta Sharings, eine der zentralen Innovationen der SAP-BDC-Architektur, genauer. Delta Sharing ist ein cloudbasiertes Protokoll zum sicheren Austausch von Live-Daten. Es ist ein direkter Zugriff auf Delta-Tabellen möglich, um eine Echtzeit-Datensynchronisation durchzuführen, ohne dabei einen Kopieraufwand zu erzeugen. Delta-Tabellen im SAP-BDC-Rahmen erfassen und speichern ausschließlich die Datenänderungen (wie Hinzufügungen, Aktualisierungen oder Löschungen) innerhalb eines definierten Zeitraums, um so effiziente Datenverarbeitung und Aktualisierungen zu ermöglichen.

Mit Databricks wird es möglich, unternehmensübergreifende Datenräume mit fein abgestuften Berechtigungen zu schaffen und dabei den Anforderungen der Daten-Governance gerecht zu werden. Durch Delta Sharing haben Unternehmen den Vorteil, dass Daten nicht mehr bewegt werden müssen, sondern lediglich geteilt werden können. Dadurch werden Komplexität und Kosten reduziert. Typische Integrationshürden klassischer ETL-Prozesse werden somit vermieden.

Datenprodukte

In diesem Abschnitt stellen wir Ihnen Datenprodukte vor. Diese können als wertschöpfende Dateneinheit definiert werden. In Kapitel 5, »Datenprodukte und Object Store« steigen wir tiefer in die Thematik ein und erläutern weitere Details zu Datenprodukten in der SAP BDC. Datenprodukte zeichnen sich dadurch aus, dass sie fachlich zu verschiedenen Themengebieten passen und eine gute Dokumentation sowie Versionierung aufweisen. Die Konsumierung von Datenprodukten kann über standardisierte Schnittstellen erfolgen. SAP Databricks ermöglicht durch seine Architektur und die native Unterstützung von Delta Sharing den einfachen Einsatz von Datenprodukten. Dabei ist der Delta Lake als Speicherschicht, der Unity Catalog als Governance-Schicht und Delta Sharing als Distributionsschicht zu verstehen. Unternehmen haben mit Datenpro-

dukten die Möglichkeit, klar definierte, wiederverwendbare Datenbausteine zu erschaffen, die als Schnittstelle für Analytik, Reporting oder KI dienen können.

Entwicklungsfreiheiten

Mit den SQL-Möglichkeiten in SAP Databricks können Anwenderinnen und Anwender SQL-Abfragen ausführen, ohne sich um Cluster-Verwaltung, Skalierung oder Ressourcenallokation kümmern zu müssen. Damit wird eine Umgebung geboten, die den Komfort klassischer SAP-Systeme mit der Skalierbarkeit eines modernen Cloud-Backends vereint. Vorteile sind bspw. eine geringe Latenz beim Zugriff auf die Daten, ein automatisches Hoch- oder Runterskalieren nach Bedarf und ein nutzungsbasiertes Abrechnungsmodell.

SAP Databricks bietet Data Engineers umfassende Entwicklungsfreiheiten, darunter eine Pro-Code-Umgebung. Somit haben sie die Möglichkeit, Daten effizient zu bereinigen, zu transformieren und zu integrieren – unabhängig davon, wie komplex oder unstrukturiert die Quellen sind.

Feature-Engineering

Bislang setzte SAP auf vordefinierte ML-Bibliotheken wie Predictive Analytics Library (PAL). SAP Databricks nutzt hingegen eine große Bandbreite an Open-Source-Bibliotheken. Damit öffnen sich SAP-Systeme für KI-Anwendungen wie Anomalieerkennung, Risikobewertung oder Nachfrageprognosen. Innovativ ist außerdem, dass Feature Engineering über strukturierte und semistrukturierte Quellen hinweg möglich ist und Delta Lakes als Feature Store genutzt werden können. Zudem liefert SAP Databricks mit Mosaic AI eine integrierte Plattform für generative und klassische KI.

Neuerungen zu SAP Databricks in den Release Notes

Für SAP Databricks finden Sie die kommenden Releases nicht in der Roadmap. Zukünftige und bereits bearbeitete Neuerungen können Sie stattdessen den Release Notes entnehmen: *http://s-prs.de/v1134106*.

Mit Mosaic AI ist ein effizientes Fine-Tuning von generativen Modellen auf eigenen SAP-Daten möglich, während die Daten vollständig unter Kontrolle bleiben. Mit dem Mosaic AI Agent Framework eröffnen sich Ihnen weitere innovative Optionen. So wäre es bspw. möglich, KI-gestützte Unternehmensagenten einzusetzen, die automatisiert auf SAP-Daten zugreifen. Darüber hinaus kann generative KI im Unternehmenskontext eingesetzt werden, um bspw. Text- oder Code-Generierung zu unterstützen. Mit Mosaic AI Agent Frameworks wird eine Plattform geboten, die zur Erstellung und Orchestrierung von KI-Agenten eingesetzt werden kann.

Unternehmensagenten

Es können intelligente Unternehmensagenten erstellt werden, die auf die SAP-Daten zugreifen, versuchen, die dokumentierten Geschäftsprozesse zu verstehen, u. a. Datenprodukte konsumieren und auf Basis dieser Informationen schließlich sichere und kontrollierbare Entscheidungen treffen.

Abbildung 4.2 veranschaulicht die Fundamente von SAP Databricks. Die Kombination dieser Merkmale schafft einen intelligenten Datenraum, in dem Daten offen, teilbar und gemeinsam nutzbar sind. Datenprodukte helfen dabei, die Daten im Unternehmen nutzbar zu machen, während die serverlose Technologie die Entwicklung erleichtert. Mit KI, insbesondere mit Mosaic AI, können Unternehmensdaten in KI-Modelle eingebunden werden. *Mosaic AI* ist eine KI-Plattform von SAP Databricks, mit deren Hilfe Unternehmen eigene KI-Modelle entwickeln, anpassen oder einsetzen können. Auf Basis der Unternehmensdaten kann die KI bei Entscheidungen unterstützen.

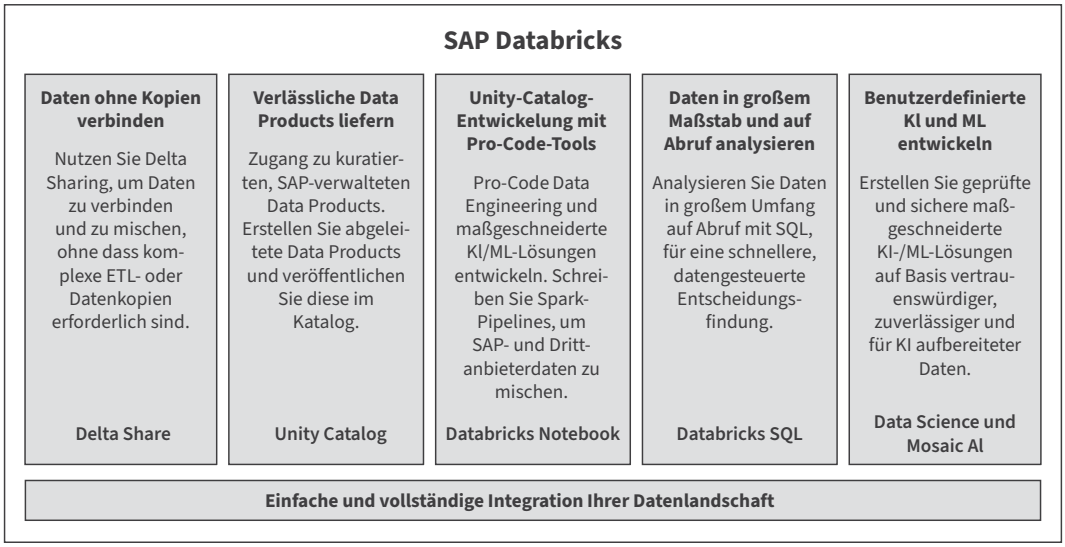


Abbildung 4.2: Die Bausteine von SAP Databricks

4.1.4 Berechtigungen

Berechtigungsmodell

Ein zentrales Element der SAP BDC ist die kontrollierte und sichere Nutzung von Daten, Analysen und KI-Funktionen. SAP Databricks setzt hierfür auf ein feingranulares, rollenbasiertes Berechtigungsmodell. Dieses Modell stellt sicher, dass Nutzerinnen und Nutzer, Anwendungen und KI-Agenten nur auf Daten und Funktionen zugreifen können, für die sie autorisiert sind.

Das Berechtigungskonzept in SAP Databricks folgt den folgenden Regeln:

- Jede Rolle erhält nur die minimale Anzahl an erforderlichen Rechten.
- Die Verwaltung der Rollen erfolgt im Unity Catalog.
- Die Vergabe der Rollen ist einheitlich und gilt somit auch für Notebooks, Data Science, KI etc.
- Jede Zugriffsentscheidung ist zu jeder Zeit auditierbar.

Weiterführende Informationen zu den Berechtigungen innerhalb von SAP Databricks finden Sie unter *http://s-prs.de/v1134107*.

Revisionssicherheit

Diese Prinzipien sind besonders für Unternehmen relevant, in denen Revisionssicherheit essenziell ist.

Nach der Bereitstellung von SAP Databricks müssen den Benutzern bestimmte Berechtigungen zugewiesen werden. Anstatt Rechte direkt zuzuweisen, können Sie diese auch Gruppen zuordnen. Die Benutzer erben die Rechte automatisch von ihrer Gruppenzugehörigkeit. Wir empfehlen dieses Vorgehen, um bspw. fachliche oder organisatorische Rollen erstellen zu können.

Governance und Compliance

Ein weiterer klarer Mehrwert ist die Durchsetzung von Governance und Compliance. Mithilfe der Gruppen können Sie Governance-Regeln konsistenter durchsetzen, um z. B. die Systeme Entwicklung, Test und Produktion zu trennen. Insbesondere KI-Agenten sollten nur die minimal notwendigen Rechte zugewiesen bekommen.

Berechtigungen werden objektbasiert vergeben und unterliegen der folgenden Hierarchie:

- Catalog
- Schema
- Tabelle/View
- Model

Die Verwaltung und das Teilen von Daten und Notebooks können direkt in SAP Databricks realisiert werden. Benutzer mit der Rolle **BD-Administrator** können SAP-BDC-Benutzer erstellen und Rollen sowie Berechtigungen verwalten.

Berechtigungsfelder

Die folgenden Berechtigungsfelder können aktiviert werden:

- **Create Catalog**
  Für das Anlegen neuer Catalogs. *Catalogs* dienen im Unity Catalog als organisatorische Einheit für Datenprodukte etc.
- **Create Share**
  Mit dieser Berechtigung können Benutzer Daten gemäß dem Delta-Sharing-Protokoll freigeben. Mit den Shares wird definiert, welche Tabellen oder Views für Konsumenten bereitgestellt werden.
- **Set Share Permission**
  Mit diesem Feld können Freigaben so konfiguriert werden, dass Konsumenten auf einer erweiterten Basis verwaltet werden können. Dieses Szenario ist insbesondere bei Multi-Tenant-Szenarien sehr hilfreich.
- **Use Provider**
  Mit diesem Feld kann auf weitere Datenanbieter zurückgegriffen werden.

- **Use Recipient**
  Empfänger definieren technische Endpunkte, an die Daten über Delta Sharing ausgeliefert werden.
- **Use Share**
  Mit dieser Berechtigung können Freigaben aktiv konsumiert und auf die darin bereitgestellten Tabellen zurückgegriffen werden.

4.1.5 SAP Databricks Lakehouse

Während SAP-Systeme für strukturierte Daten bekannt sind, verfolgt SAP Databricks einen anderen Ansatz und setzt auf eine Lakehouse-Architektur. Aus technologischer Sicht verbindet das Lakehouse-Konzept SAP und Databricks miteinander.

Data Warehouses

Bevor das Lakehouse-Konzept an Popularität gewann, waren die etablierten Architekturen Data Warehouse und der Data Lake. *Data Warehouses* wie SAP BW/4HANA oder SAP Datasphere fokussieren sich auf die Datenqualität, die Datenmodellierung und das anschließende Bereitstellen von Datenmodellen bspw. für betriebswirtschaftliche Zwecke. Allerdings bringen Data Warehouses auch Nachteile mit sich. Sobald sich ein Geschäftsprozess ändert oder neue Anforderungen aus den Fachbereichen entstehen, müssen die Datenmodelle angepasst werden. Außerdem sind insbesondere In-Memory-basierte Systeme in der Regel teuer. Bisher bestand ein wesentlicher Nachteil darin, dass Data Warehouses zwar auf Reporting optimiert wurden, der Fokus jedoch nicht auf ML und unstrukturierten Daten lag.

Data Lakes

Daher setzen viele Unternehmen im Big-Data-Umfeld sogenannte *Data Lakes* ein, um große Datenmengen kostengünstiger speichern zu können. Data Lakes bieten noch weitere Vorteile, wie bspw. die Persistierung verschiedenster Datenarten. In der Praxis haben sie sich jedoch nicht durchgesetzt, da insbesondere die mangelnde Datenqualität und die fehlenden Governance-Strukturen problematisch sind. Deshalb wurde nach einer Lösung gesucht, die sowohl die Governance und Datenqualität eines Data Warehouses als auch die Flexibilität und Skalierbarkeit eines Data Lakes bietet.

Lakehouse

Das *Lakehouse* kombiniert diese Funktionalitäten in einem Ansatz. SAP Databricks fokussiert sich auf dieses Modell und nutzt dafür den *Delta Lake*. Innerhalb des Lakehouses werden strukturierte und unstrukturierte Daten persistiert und es wird auf die Konsistenz sogenannter *ACID-Transaktionen* (Atomarität, Konsistenz, Integrität und Dauerhaftigkeit) geachtet. So können transaktionale Garantien erzeugt werden. Darüber hinaus ist ein direkter Zugriff für BI, Data Science und KI-/ML-Anwendungen möglich. Eine einheitliche Governance soll dabei ebenfalls vorhanden sein. Für SAP-Kunden bedeutet dies, dass sowohl

Rohdaten aus SAP-Systemen als auch Datenprodukte im selben Format abgelegt werden können.

Trennung von Speicher und Rechenleistung

Ein Lakehouse trennt Speicher und Rechenleistung voneinander, sodass Workloads, wie bspw. SQL-Abfragen oder das Training von KI-Modellen, eigene Compute-Ressourcen nutzen, ohne dass zuvor Daten persistiert und kopiert werden müssen. In der SAP BDC bedeutet dies, dass eine elastische Skalierung komplexer Anfragen möglich ist, Daten parallel verarbeitet werden können und auf User-Seite eine klare Kostenkontrolle gewährleistet ist, da die Compute-Ressourcen bedarfsorientiert genutzt werden.

Governance-Schicht

Eine weitere Stärke des Lakehouse-Ansatzes liegt in der konsolidierten Governance-Schicht. Aus einem technischen Speicherformat wird ein kontrollierbarer Unternehmensdatenraum. Der Unity Catalog ermöglicht eine einheitliche Zugriffskontrolle über alle Workspaces hinweg und bringt zentralisierte Rollen und Berechtigungen mit sich.

Workspaces

Workspaces dienen in SAP Databricks als zentrale, kollaborative Entwicklungsumgebung, in der Teams gemeinsam Datenanalysen, Machine-Learning-Modelle und Datenpipelines erstellen, verwalten und ausführen können.

Daten in Echtzeit einbinden

Mithilfe der Lineage kann verfolgt werden, wie sich Daten über Tabellen, Pipelines und Modelle hinweg verhalten. Daten aus SAP-Databricks-Delta-Lake-Tabellen können direkt über eine Echtzeitverbindung in virtuelle Remote-Modelle von SAP Datasphere eingebunden werden. Die eingebundenen Daten werden wiederum direkt in semantische Modelle integriert. So sind effiziente Echtzeitanalysen über Systemgrenzen hinweg möglich.

Der Integrationsprozess erfordert zunächst den Verbindungsaufbau. Zuerst muss mithilfe von Konnektoren (JDBC-Konnektivität mit Camel JDBC-Adapter) sowie Authentifizierungsmechanismen eine sichere Verbindung zwischen SAP Datasphere und SAP Databricks hergestellt werden. Im Anschluss müssen in SAP Datasphere virtuelle Tabellen konfiguriert werden, die auf die Live-Daten in SAP Databricks Delta Lake verweisen. Hierbei ist zu beachten, dass die Daten nicht physisch verschoben werden, sondern nur zur Laufzeit auf die Daten in SAP Databricks zugegriffen wird. Ein bestehendes Datenmodell kann mit diesen Daten angereichert werden. Sobald das Datenmodell fertiggestellt ist, kann es in SAP Analytics Cloud zur Erstellung von Dashboards und Berichten genutzt werden.

4.1.6 End-to-End-Integration: SAP Databricks, SAP Datasphere und SAP Analytics Cloud

End-to-End Integration

In diesem Abschnitt demonstrieren wir, wie SAP Databricks, SAP Analytics Cloud und SAP Datasphere zusammenarbeiten, um eine Business-Data-Fabric zu bilden. Vor dem Zusammenschluss von SAP und Databricks wurde bspw. auf die Bibliotheken APL (Automated Predictive Library) und PAL (Predictive Analytics Library) zurückgegriffen, um in Notebooks über einen HANA-Client KI-Modelle zu erstellen. Dieser Workaround kann jetzt durch die Integration von Databricks in die SAP BDC vermieden werden. Dafür kann ein Data-Provisioning-Agent (DP-Agent) eingesetzt werden.

Live-Daten verbinden

Folgen Sie dazu dieser Anleitung:

1. Installieren Sie den DP-Agenten.
2. Installieren Sie den SAP-Databricks-JDBC-Treiber.
3. Starten Sie den DP-Agenten neu.
4. Wählen Sie den Camel-JDBC-Adapter in SAP Datasphere aus und erstellen Sie eine Generic JDBC Connection. Eine Remote-Tabelle in SAP Datasphere zeigt schließlich auf eine SAP-Databricks-Tabelle.
5. Stellen Sie mithilfe der Datenvorschau sicher, dass der Datenzugriff erfolgreich ist.
6. Verbinden Sie die Live-Daten in einem Fact Model in SAP Datasphere.

Datenfluss

Die Architektur hierzu wurde bereits in Kapitel 2, »SAP Datasphere als Kern der SAP Business Data Cloud«, vorgestellt. Der zugehörige Datenfluss sieht folgendermaßen aus (siehe auch Abbildung 4.3):

❶ Zunächst muss eine Intelligent Application oder ein Data Package aktiviert werden.

❷ Mittels Datenreplikation werden die Datenprodukte in den BDC Object Store übernommen.

❸ Ein aktives Datenprodukt kann für SAP Databricks freigegeben werden und ist über eine Delta-Tabelle im Unity Catalog verfügbar.

❹ In SAP Databricks kann mithilfe von KI und ML ein angepasstes Datenprodukt erstellt werden. Dieses Datenprodukt kann über Delta Share ans SAP BDC Cockpit zurückgeteilt werden.

❺ Im Folgenden kann das neue Datenprodukt von SAP Datasphere konsumiert werden.

❻ In SAP Datasphere kann es in beliebige Datenmodelle eingebunden werden.

❼ Sie können die erstellten Datenmodelle zur Visualisierung bzw. weiteren Verarbeitung an SAP Analytics Cloud teilen.

In SAP Analytics Cloud können Sie diese dann verwenden, um bspw. Reports zu erstellen.

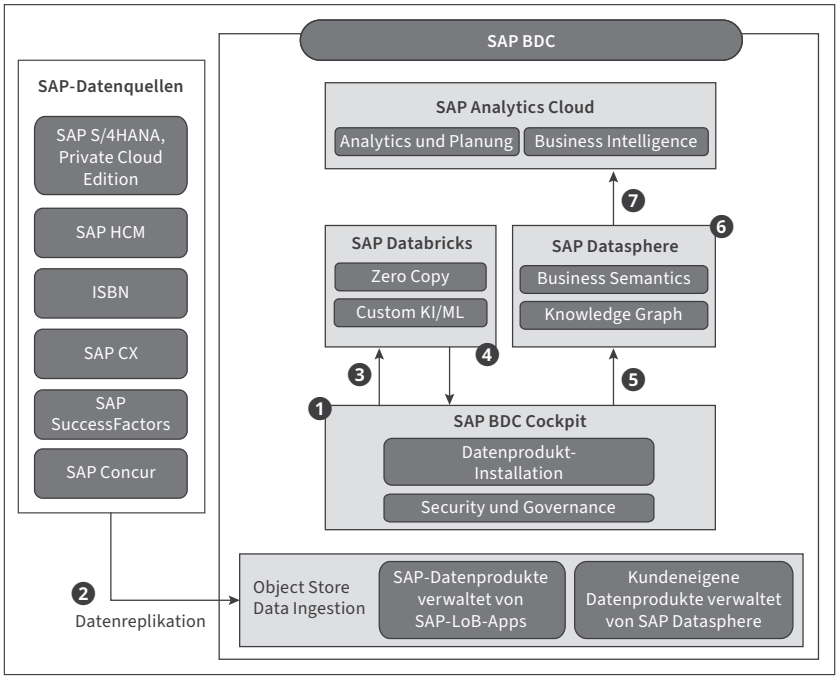


Abbildung 4.3: Datenfluss in SAP BDC

Weiterführende Informationen finden Sie unter *http://s-prs.de/v1134108*.

4.1.7 Abgrenzung von SAP Datasphere und SAP Databricks

Angesichts der Architektur der SAP BDC könnte die Frage aufkommen, wie sich SAP Datasphere und SAP Databricks voneinander abgrenzen lassen. SAP Datasphere und SAP Databricks sind beide leistungsstarke Plattformen für das Datenmanagement und die Datenanalyse. In diesem Abschnitt grenzen wir die beiden Plattformen voneinander ab.

SAP Datasphere

In den vorherigen Kapiteln dieses Buches wurde SAP Datasphere als Data Warehouse vorgestellt, das Unternehmensdaten integriert, über das Schichtkonzept hinweg verknüpft und aufbereitet. Ein wesentliches Ziel besteht dabei in der Bewahrung der Semantik. Insbesondere der Self-Service-Zugriff und die Self-Service-Modellierung von Daten im betriebswirtschaftlichen Kontext standen im Fokus von SAP.

Unterschiedliche Zielgruppen

Während SAP Datasphere sich an mehrere Benutzergruppen, wie bspw. End User und Data Engineers, richtet, fokussiert sich SAP Databricks auf eine schmale Anwendergruppe. SAP Databricks wird weitestgehend von Personen

verwendet, die bereits Erfahrung im Umgang mit KI- und ML-Lösungen haben. In SAP Databricks besteht die Möglichkeit, eine Vielzahl von Datenarten zu integrieren, um daraus benutzerdefinierte KI- und ML-Lösungen zu entwickeln oder SQL-Ad-hoc-Analysen auszuführen. Dies erfolgt in einer Pro-Code-Entwicklungsumgebung.

4.2 Delta Share

Zero-Copy-Prinzip

Ein technologischer Baustein von SAP Databricks ist das Delta-Sharing-Protokoll, das auf dem *Zero-Copy-Prinzip* basiert. Dieses Prinzip ermöglicht es, lesenden Zugriff auf Daten zu erlangen, ohne diese physisch zu replizieren. Technisch gesehen handelt es sich um ein offenes cloudübergreifendes Datenfreigabeprotokoll, das einen Live-Datenaustausch zwischen Systemen und Plattformen ermöglicht. Sobald ein Datensatz in SAP Databricks bereitgestellt wurde, kann er in der SAP BDC für Analysen oder KI-Use-Cases verwendet werden. Einen Mehrwert liefern außerdem die zentrale Verwaltung genutzter Daten, die Erstellung von Datenprodukten (z. B. KI-Modellen) und deren Vertrieb sowie die datenschutzkonforme Datenhaltung in der Cloud.

Delta-Tabellen

Die Datenkonsumenten – bspw. die verschiedenen SAP-Systeme – greifen über Endpunkte direkt auf die Delta-Tabellen im Cloud-Speicher zu. Abfragen erfolgen dementsprechend live, was wiederum bedeutet, dass die Daten weder verschoben noch transformiert werden müssen. Relevante Daten können somit ohne Replikation oder Schnittstellen mit anderen Systemen geteilt werden. Auch bei SAP Databricks wurde dieses Konzept beibehalten. Diese Integration gibt Unternehmen die Möglichkeit, SAP-Daten aus Quellsystemen wie SAP S/4HANA oder SAP BW mithilfe von KI-Tools in SAP Databricks zu konsumieren.

Delta Sharing

Delta Sharing funktioniert folgendermaßen:

1. Ausgewählte Datenprodukte werden mit dem Zielsystem geteilt und können über das Delta-Sharing-Protokoll in Echtzeit verwendet werden.
2. Da SAP Databricks Delta Sharing nativ unterstützt, werden die Daten in einem Delta Lake live gehalten und können direkt für Analysen, ML-Modelle usw. verwendet werden.



Delta Sharing vs. Sharing innerhalb von SAP Datasphere

Das Prinzip des Delta Sharings ist nicht mit dem des Sharings innerhalb von SAP Datasphere gleichzusetzen.

Vielleicht haben Sie bemerkt, dass Delta Sharing somit die wesentlichen Anforderungen moderner Datenökosysteme erfüllt. Dazu zählt zunächst die Interoperabilität zwischen den verschiedenen SAP-Systemen und SAP Databricks. Bei

Bedarf können zudem andere externe Systeme einbezogen werden. SAP Databricks bringt außerdem eine gewisse Offenheit mit sich, da die Nutzung offener Speicherformate möglich ist. Die wichtigste erfüllte Hauptanforderung ist jedoch die minimierte Datenredundanz, da keine Kopiervorgänge nötig sind.

Vorteile des Delta Sharings

Delta Sharing bringt einige Vorteile mit sich. So ermöglicht es die Kombination der SAP-Daten mit SAP Databricks, insbesondere große Datenmengen effektiv zu analysieren. Da die Datenbewegung so gering wie möglich gehalten wird, findet eine Kostenreduktion statt, wodurch die Effizienz erhöht werden kann. Durch die Nutzung moderner Sicherheitsprotokolle beim Delta Sharing wird sichergestellt, dass nur autorisierte Personen Zugriff auf die Daten erhalten. Ein weiterer Vorteil ergibt sich bei der Zusammenarbeit. Gerade in interdisziplinären Teams können verschiedene Daten geteilt werden, ohne dass aufwendige Exportmechanismen implementiert werden müssen. Der Zugriff auf Daten aus verschiedenen Unternehmenssparten ermöglicht die Erstellung vollkommen neuer Datenanalysen und KI-Modelle. Ein entscheidender Faktor ist außerdem die Bereitstellung von Echtzeitdaten. Dadurch wird die Zeitspanne zwischen der Datengenerierung und der Verfügbarkeit in den Analysen so kurz wie möglich gehalten.

4.2.1 Benötigte Rollen beim Delta Sharing

Delta Sharing verwendet ein strukturiertes Berechtigungsmodell. Dieses Modell erlaubt es Unternehmen, Datenprodukte intern zu kontrollieren und bei Bedarf auch extern bereitzustellen. So wird sichergestellt, dass Governance-, Compliance- und Audit-Anforderungen erfüllt werden.

Datenprodukte

Damit Datenprodukte aus dem SAP BDC Catalog für SAP Databricks freigegeben werden können, werden bestimmte Berechtigungen benötigt. Für die Suche und das Teilen von Datenprodukten ist eine globale Rolle erforderlich, die folgende Privilegien garantiert (der Bindestrich (-) dient an dieser Stelle als Platzhalter für eine nicht benötigte Berechtigung):

- BDC Data Packages (-R-----) für den Zugriff auf die SAP BDC.
- Catalog Asset (-R-----) für den Zugriff auf den Catalog und um Objekte in den Assets und in der Datenprodukt-Sammlung anschauen zu dürfen.
- Cloud Data Product (-----S-) für das Teilen von Datenprodukten mit den Zielsystemen.

4.2.2 Verwendung von Datenprodukten in SAP Databricks

Sobald einem Benutzer die notwendigen Berechtigungen zugeteilt wurden, kann er Datenprodukte in SAP Databricks verwenden, abfragen, verarbeiten und transformieren. Neben den typischen Programmier- und Abfragesprachen

wie SQL und Python können bspw. auch Scala und R genutzt werden, um verschiedenste KI-/ML-Projekte in Notebooks umzusetzen.

Administratorrolle zuweisen

Zunächst müssen einige Vorbereitungen getroffen werden, um die Datenprodukte in SAP Databricks verarbeiten zu können. Als Hauptbedingung muss der Benutzer die Administratorrolle besitzen. Führen Sie dazu die folgenden Schritte durch:

1. Zunächst müssen Sie den SAP Databricks Tenant zu SAP BDC Formation hinzufügen.
2. Anschließend können ein oder mehrere Data Packages aktiviert werden.
3. Nun erhalten Sie Zugriff auf den SAP Databricks Tenant.

Im Anschluss stehen Benutzern eine Vielzahl an Optionen in SAP Databricks zur Verfügung. Sie können auf Datenprodukte zugreifen, die aus dem Catalog heraus geteilt wurden. Dabei stellt SAP Databricks ein breites Toolset zur Verfügung. Sie können wiederum auch Datenprodukte an den SAP Datasphere Catalog verteilen. Im Folgenden sehen wir uns die Hauptfunktionen des Delta Sharings an.

Wie Sie in Abbildung 4.4 sehen, stellt ein SAP-System Daten bereit, die dem im SAP BDC Cockpit enthaltenen Catalog hinzugefügt werden. Mit diesem Catalog können der SAP Datasphere Catalog oder der SAP Databricks Unity Catalog gefüllt werden. Der Unity Catalog wird in Abschnitt 4.4 näher behandelt. Anhand des Symbols wird ersichtlich, in welche Richtung das Sharing umgesetzt werden kann.

Die Daten aus verschiedenen Quellsystemen befüllen den Catalog, der im SAP BDC Cockpit enthalten ist. Auf Basis dieses Catalogs können der SAP Databricks Unity Catalog sowie der SAP Datasphere Catalog befüllt werden.

Die aktivierten Packages können in SAP Databricks verwendet werden. Auf den Unity Catalog greifen im Anschluss die SAP Databricks Toolsets zu.

Hauptfunktionen des Delta Sharings

Mit diesem Wissen können wir die Hauptfunktionen des Delta Sharings in SAP BDC wie folgt zusammenfassen:

- Datenprodukte können mit SAP Databricks geteilt werden.
- In SAP Databricks erstellte Modelle und Analysen können in SAP-Systeme zurückgeführt werden.
- Die Daten können zwischen unterschiedlichen Cloud-Providern ausgetauscht werden.
- Bei Bedarf lassen sich Data-Mesh-Architekturen abbilden, da Domänen potenziell in der Lage sind, eigenständige Datenprodukte zu veröffentlichen.

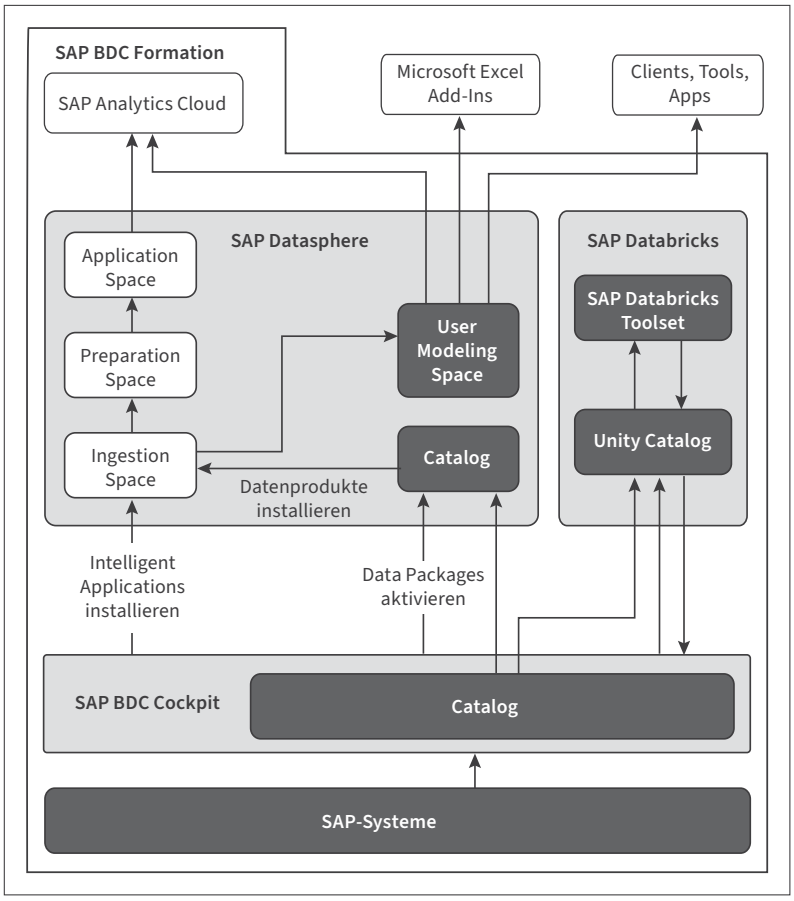


Abbildung 4.4: SAP Business Data Cloud – Überblick

4.2.3 Umsetzung des Sharings

In den vorherigen Abschnitten wurde die Architektur hinter dem Sharing erläutert. In diesem Abschnitt zeigen wir Ihnen, wie das Sharing umgesetzt wird. Damit Datenprodukte geteilt werden können, muss ein User über die notwendigen Berechtigungen verfügen. Die Datenprodukte können entweder direkt aus dem SAP BDC Catalog mit SAP Databricks geteilt werden oder es wird alternativ in SAP Databricks ein Catalog aus dem Delta Share erstellt, der mit den Datenprodukten verbunden ist. Wir beschreiben beide Möglichkeiten und beginnen mit der ersten.

Möglichkeit 1: SAP BDC Catalog

Gehen Sie folgendermaßen vor, um die Datenprodukte aus dem SAP BDC Catalog heraus mit SAP Databricks zu teilen:

1. Suchen Sie zunächst im SAP BDC Catalog nach dem Datenprodukt (siehe Abbildung 4.5). Sie können dabei Filter verwenden. Wichtig ist, dass das Datenprodukt geteilt werden kann. Dafür muss es den Status Shareable haben.

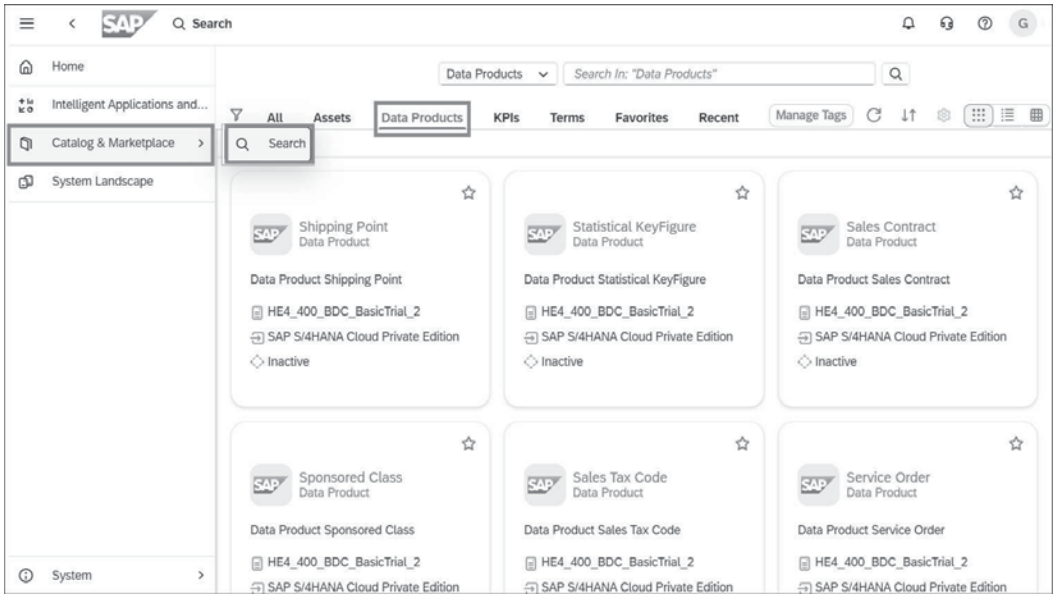


Abbildung 4.5: Ansicht im SAP BDC Catalog

2. Öffnen Sie die Detail-Seite des Datenprodukts.
3. Öffnen Sie mit einem Klick auf den **Share**-Button den Manage-Share-Access-Dialog (siehe Abbildung 4.6).

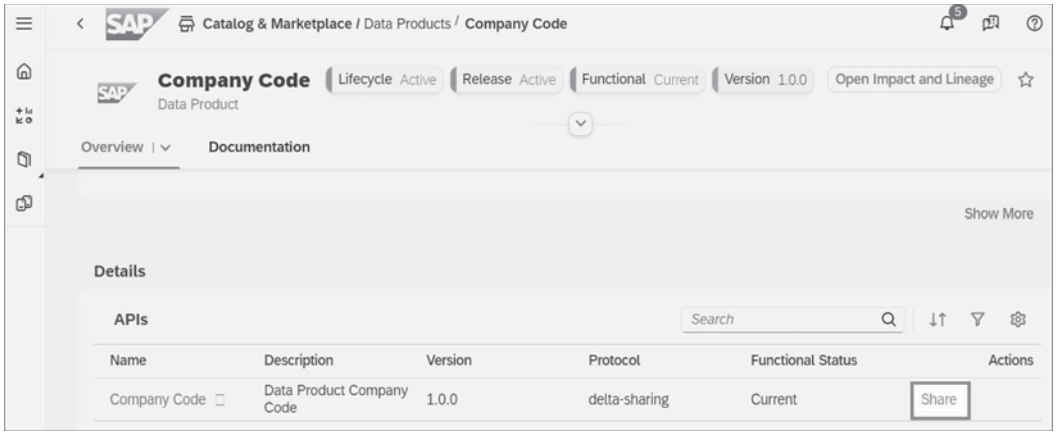


Abbildung 4.6: Den Manage-Share-Access-Dialog öffnen

4. Wählen Sie bei **Target System** das passende System aus. Klicken Sie auf den **Share**-Button (siehe Abbildung 4.7).

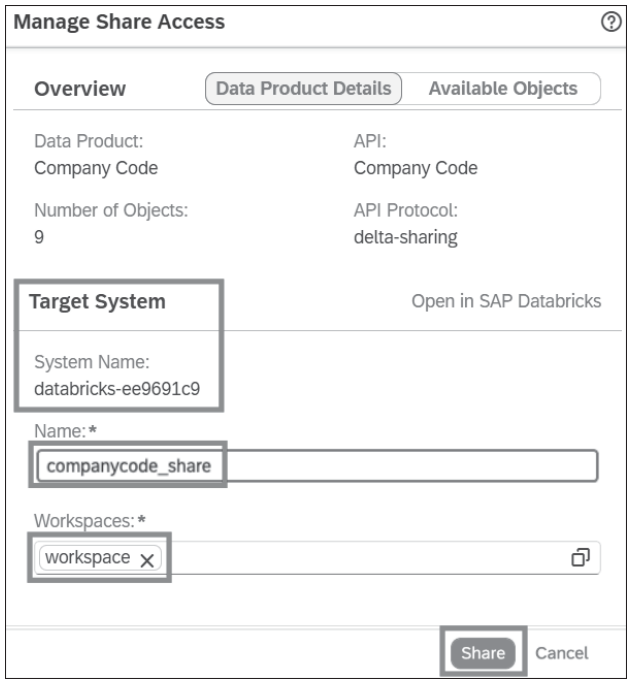


Abbildung 4.7: Auswahl für Share

Das Datenprodukt ist jetzt ausgewählt und der Status wurde aktualisiert. Der festgelegte Name kann nun nicht mehr verändert werden. Soll der Name im Nachhinein geändert werden, muss der Vorgang wiederholt werden.

Möglichkeit 2: Datenprodukt

Gehen Sie folgendermaßen vor, um den SAP Databricks Delta Sharing Catalog von einem Datenprodukt aus zu befüllen:

1. Stellen Sie zunächst sicher, dass Sie die Data-Product-Admin-Rolle besitzen und im SAP Databricks Metastore die Berechtigungen **USE PROVIDER** und **CREATE_CATALOG** besitzen.
2. Greifen Sie auf den SAP Databricks Workspace zu und öffnen Sie den Catalog Explorer.
3. Klicken Sie auf das Gear-Icon und wählen Sie **Delta Sharing** aus (siehe Abbildung 4.8).
4. Unter dem Reiter **Share with me** müssen Sie nun den passenden Provider zu dem Datenprodukt suchen und auswählen.
5. Wählen Sie **Create Catalog** aus und vergeben Sie einen Namen für den Catalog.

Sobald der Catalog erstellt wurde, wird er im Hauptmenü des Catalog Explorers in der **Shared**-Auswahl gelistet.

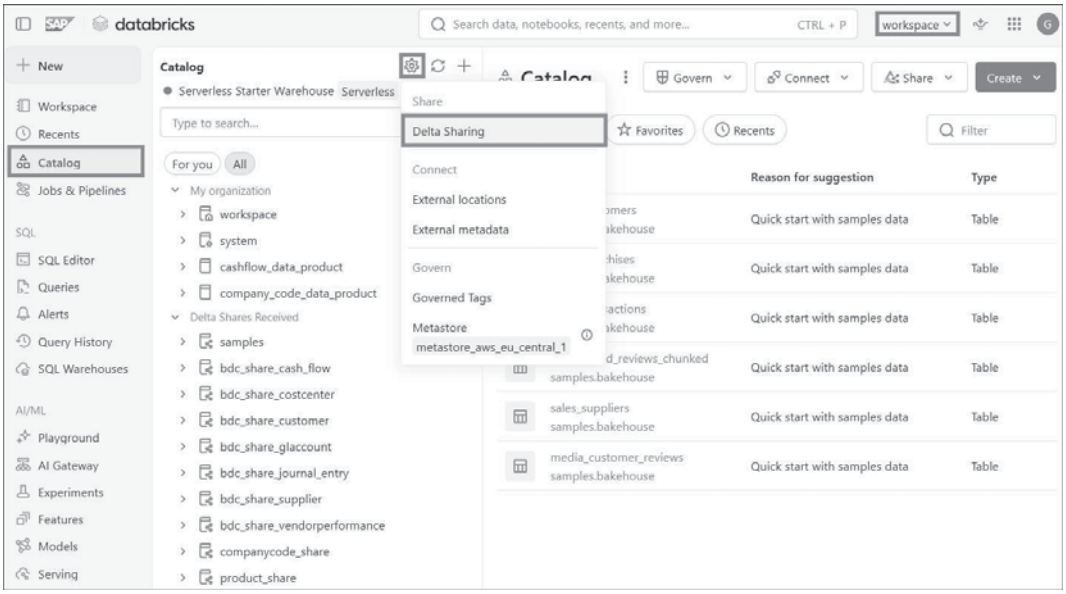


Abbildung 4.8: Delta Sharing Catalog

Sobald ein Delta Sharing Catalog auf Basis eines geteilten Datenprodukts erstellt wurde, können die Daten konsumiert, abgefragt, verarbeitet und transformiert werden. Dies erfordert jedoch bestimmte Berechtigungen. Um Daten eines Datenprodukts lesen zu können, muss der Benutzer vom Catalog Owner die Berechtigungen **USE CATALOG**, **BROWSE** und **SELECT** erhalten haben.

Eine Möglichkeit für die Verarbeitung der Daten besteht darin, eine Abfrage über die Daten in einem Python-Notebook zu realisieren. Dafür müssen Sie zunächst ein neues Notebook im SAP Databricks Workspace erstellen und Python als Programmiersprache auswählen. Anschließend können Sie **Serverless-Compute** aktivieren. Gerade für Python- oder R-Entwicklerinnen und -Entwickler ist es praktisch, dass die Daten direkt in SAP Databricks visualisiert werden können.

4.2.4 Zurückteilen der Daten in der SAP BDC

Die Integration von SAP Databricks in die SAP BDC ermöglicht nicht nur den Zugriff auf SAP-Daten für Analytics, Data Engineering und KI, sondern auch den Rückfluss von Daten, Modellen und Erkenntnissen zurück in SAP-Domänen. Dazu kann ein benutzerdefiniertes Datenprodukt erstellt und in SAP Datasphere für die Erstellung eines Datenmodells genutzt werden. Dieses neue Datenmodell kombiniert schließlich die SAP-Standarddatenprodukte mit dem benutzerdefinierten Datenprodukt.

Ergebnisse in Delta-Tabelle teilen

Im ersten Schritt müssen die Ergebnisse in einer Delta-Tabelle gespeichert werden. Sobald diese Delta-Tabelle im SAP Databricks Catalog enthalten ist, kann sie über Delta Share geteilt werden.

Sonderdateien

Wenn Sie ein Datenprodukt anlegen und in der SAP BDC publizieren möchten, müssen Sie zwei Sonderdateien mit Python-Code anlegen. Zunächst muss eine Open-Ressource-Discovery-Datei (ORD) generiert werden, um die Ressource und Funktionen des Datenprodukts zu beschreiben. Dieses von SAP entwickelte Protokoll kann von Datenanbietern verwendet werden, um Datenprodukte sowohl bereitzustellen als auch aufzusuchen. Zusätzlich ist eine Core-Schema-Notation-Datei (CSN) erforderlich, um die Datenproduktstruktur, d. h. Eigenschaften, Datentypen, Namen und Beziehungen, zu beschreiben.

Zugriff

Sobald Sie das neue Datenprodukt in SAP Databricks geteilt haben, können Sie in der SAP BDC darauf zugreifen und es in SAP Datasphere installieren und verwenden. Abbildung 4.9 zeigt, wie die Delta-Sharing-Maske in SAP Databricks aussieht. Sie müssen sich im Menüpunkt **Catalog** befinden und erhalten im **Catalog Explorer** eine Übersicht der geteilten oder erhaltenen Datenprodukte. Ebenso können Sie von hier aus neue Delta-Sharing-Berechtigungen anfordern oder Daten aktiv teilen (über den Button **Share data**).

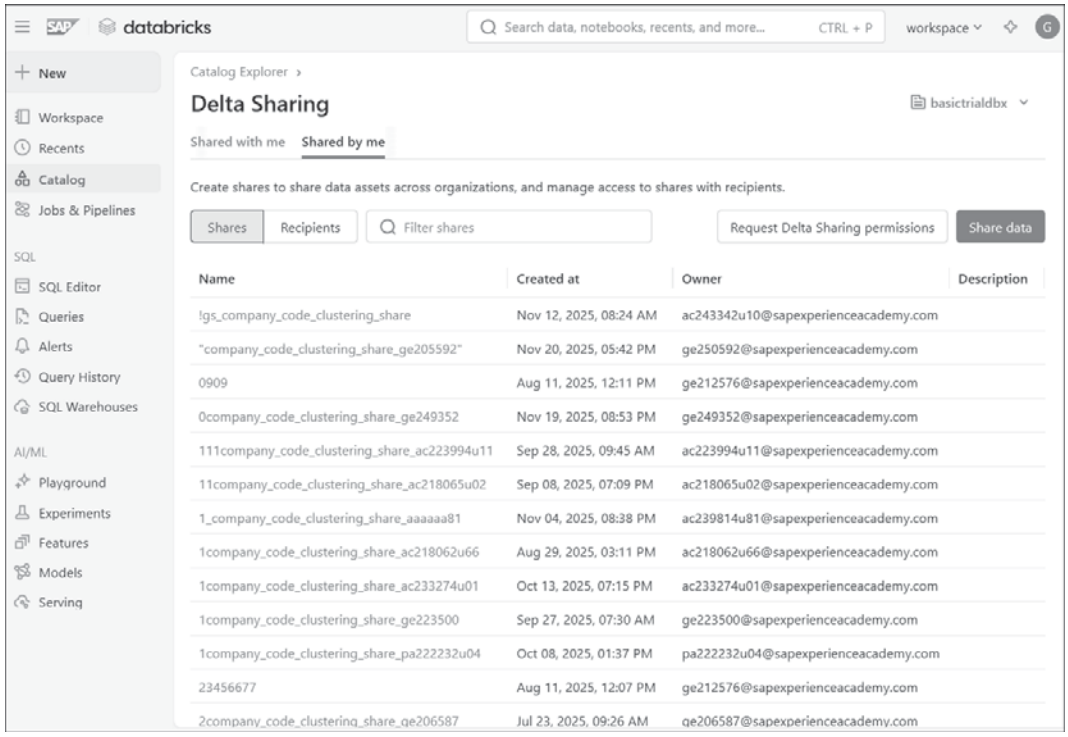


Abbildung 4.9: Catalog Explorer

4.3 Partnerplattformen mit der SAP BDC verbinden

Partnerintegration

Bevor wir uns weiter mit SAP Databricks beschäftigen, ist es für das Gesamtverständnis sinnvoll, zunächst den übergeordneten Kontext der Partnerintegration zu erläutern, da SAP BDC nicht als isolierte Einzellösung konzipiert ist, sondern als offenes Ökosystem, in dem strategische Technologiepartner eine integrale Rolle spielen. SAP verfolgt dabei bewusst einen Partner-First-Ansatz, um spezialisierte Plattformen nahtlos in die eigene Datenarchitektur einzubetten. Dies wird im Rahmen von SAP BDC Connect umgesetzt. *SAP BDC Connect* ist ein Service der SAP BDC, der sichere, bidirektionale Datenaustausche ohne Datenkopien zwischen der SAP BDC und Partnerplattformen wie Databricks Enterprise und Google BigQuery ermöglicht. SAP BDC Connect ist nicht für SAP Databricks erforderlich, da hier das Zero-copy-Sharing verwendet wird.

Im Rahmen dieser Partnerstrategie ist Databricks als strategischer Technologiepartner eng in die SAP BDC integriert. Wie bereits eingangs erwähnt, ist es das Ziel der Partnerschaft, moderne Data-Engineering-, ML- und KI-Workloads direkt mit geschäftsrelevanten SAP-Daten zu verbinden.

Architektur

Das Zusammenspiel erfolgt nicht als lose Kopplung, sondern als tief integrierte Architektur innerhalb der SAP BDC. Datenprodukte aus SAP-Systemen können kontrolliert und standardisiert in Databricks-Umgebungen genutzt werden, ohne dass die semantische Konsistenz oder Governance-Vorgaben verloren gehen. Gleichzeitig profitieren Unternehmen von offenen Standards und der Möglichkeit, Nicht-SAP-Daten in analytische Prozesse einzubinden.

SAP Analytics Cloud und SAP Datasphere

Die Integration mit den SAP-Komponenten erfolgt in der SAP BDC mit SAP Analytics Cloud und SAP Datasphere. In Abbildung 4.10 sehen Sie, wie SAP Analytics Cloud und SAP Datasphere in die SAP BDC eingebettet und die Daten mit den Partnersystemen verbunden sind. SAP Datasphere dient als semantisches Data Warehouse, das Datenmodelle, Datenprodukte und Metadaten zentral verwaltet und somit die Governance der Daten sicherstellt. SAP Analytics Cloud nutzt die Daten für Reporting, Dashboards, Planung und Vorhersagen und profitiert dabei von der zentralen semantischen Basis der SAP BDC. Die enge Verzahnung der SAP-Komponenten sorgt dafür, dass die Daten nicht nur gespeichert, sondern auch mit Geschäftssemantik angereichert und für analytische sowie operative Szenarien gleichermaßen nutzbar gemacht werden.

Prinzipien

Für SAP BDC Partner Connect sind folgende Prinzipien und Protokolle relevant, daher fassen wir sie hier schon einmal zusammen. Wir werden sie in den folgenden Abschnitten noch näher erläutern.

Zero Copy sorgt als Prinzip dafür, dass keine Datenkopien erzeugt werden müssen und stets mit aktuellen Daten gearbeitet werden kann. Dadurch fallen weniger Speicherkosten an und es erfolgt ein Zugriff auf Originaldaten.

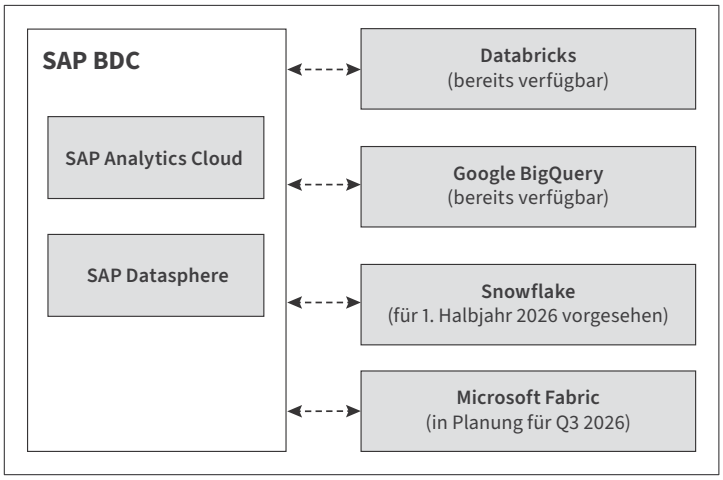


Abbildung 4.10: Integration von Partnern in der SAP BDC

Delta Share ist ein Protokoll für den sicheren Echtzeitdatenaustausch zwischen Systemen und Organisationen. Dadurch wird der Datenaustausch zentral steuerbar.

Open Resource Discovery (ORD) ist ein offenes Metadaten-Protokoll, mit dem sich Anwendungen und Services selbst beschreiben lassen. Es umgeht eine manuelle Dokumentation und schafft so mehr Überblick über vorhandene APIs und Datenprodukte. Die Metadaten umfassen Attribute wie Sichtbarkeit, Release-Status und Integrationsabhängigkeiten. In der Regel enthält das ORD-Protokoll einen Titel und eine Beschreibung des Datenprodukts.

Core Schema Notation (CSN) beschreibt die konkreten Metadaten, nicht aber die Daten selbst. So wird bspw. beschrieben, welche Tabellen es gibt, welche Felder sie haben und welche Typen sowie Beziehungen vorliegen.

Ein *Datenprodukt* setzt sich somit aus ORD und CSN zusammen (siehe Abbildung 4.11).

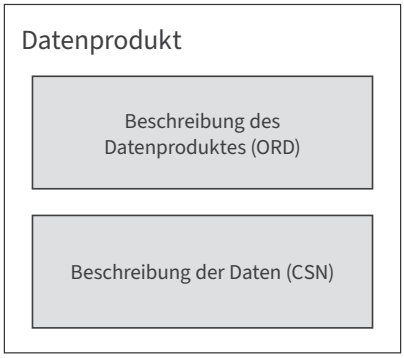


Abbildung 4.11: Zusammensetzung eines Datenprodukts

Standardisierte Integration

SAP BDC Partner Connect baut also auf diesen Prinzipien und Protokollen auf und bietet bidirektionale Zero-Copy-Sharing-Mechanismen an, um standardisierte Integrationen zu etablierten Cloud-Plattformen zu bieten.

Integrationsprozess

Je nach Partner-Tenant läuft der Integrationsprozess etwas anders ab. Zunächst müssen Sie in *SAP for Me* wesentliche Informationen über den Partner-Tenant pflegen, woraufhin ein Einladungslink generiert wird (❶). Diesen müssen Sie im Anschluss in den Partner-Tenant einfügen (❷). Letzterer kann dann zur Formation hinzugefügt werden (❸, siehe Abbildung 4.12).

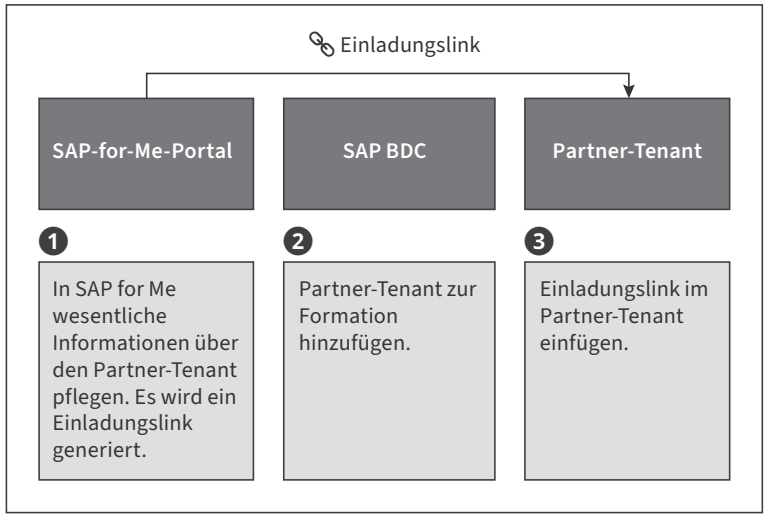


Abbildung 4.12: Integrationsprozess

Datenprodukte nutzen

Ziel ist es, die Datenprodukte nutzen zu können. Dafür müssen diese vom Partner-Tenant für die SAP BDC freigegeben werden. Hierzu müssen Sie diesen Schritten folgen:

1. Share (Open Resource Description) erstellen
2. CSN erstellen
3. Datenprodukt veröffentlichen

Nun sollten die Daten bidirektional zugänglich sein.

SAP Business Data Cloud Connect SDK

An dieser Stelle möchten wir *SAP BDC Connect SDK* erwähnen. SAP BDC Connect SDK ist eine Python-Bibliothek, die es externen Systemen – insbesondere SAP Databricks – ermöglicht, Datenprodukte mit der SAP BDC auszutauschen und freizugeben. Dazu müssen Sie lediglich einen SAP BDC Client erstellen und den Share (ORD) sowie die CSN-Datei erstellen bzw. updaten und das Datenprodukt freigeben oder stornieren. Die entsprechende Installation lautet: `pip install sap-bdc-connect-sdk`.

Die folgenden Partnerplattformen können Sie mithilfe des SAP BDC Partner-Connect Services mit der SAP BDC verbinden:

- *Databricks*
  Auf die Integration mit SAP Databricks und Databricks Enterprise gehen wir in Abschnitt 4.1.1, »Architektur und Verortung von SAP Databricks in der SAP BDC«, sowie in Kapitel 14, »Status Quo: Einsatz von Databricks Enterprise«, tiefer ein. Zusammenfassend lässt sich sagen, dass SAP-Datenprodukte direkt und ohne redundante Replikation über Zero Copy Sharing ausgetauscht werden können. Dadurch kann in Databricks auf semantisch harmonisierte SAP-Daten zugegriffen werden. Diese können für ML-Workloads, erweiterte Analysen oder Lakehouse-Szenarien genutzt werden.
- *GoogleBig Query*
  Über die SAP BDC Connect können semantisch reiche Daten aus der SAP BDC auch für Google BigQuery bereitgestellt werden. Wie bei den anderen Partnern wird dabei ein bidirektionaler Austausch ohne physische Kopien angestrebt, sodass BigQuery-Workloads direkt auf SAP-Daten zugreifen können. Dies ist insbesondere bei großen Analyse- und KI-Use-Cases von Vorteil.
- *Snowflake*
  Mit der Integration von Snowflake, die über Solution Extension realisiert wird, profitieren Kunden von einer skalierbaren Data Cloud. Mithilfe von SAP BDC Connect können Daten in Snowflake analysiert, kombiniert und für KI-Workloads genutzt werden, während gleichzeitig die zentrale Governance und die semantische Konsistenz erhalten bleiben. SAP BDC Connect für Snowflake ist auf AWS verfügbar, weitere Hyperscaler sollen folgen.
- *Microsoft Fabric*
  Ein weiterer strategischer Partner ist Microsoft Fabric. Über SAP BDC Connect for Microsoft Fabric wird ein bidirektionaler Zero-Copy-Datenaustausch ermöglicht, bei dem SAP-Datenprodukte direkt in Microsofts zentrale Datenplattform OneLake integriert werden können. Dadurch lassen sich SAP-Daten in Power BI, Foundry, Copilot und weiteren Fabric-Services nutzen, was insbesondere für BI- und KI-Szenarien wertvoll ist. Die allgemeine Verfügbarkeit ist für das dritte Quartal 2026 angekündigt.

SAP BDC Partner Connect schafft somit die Voraussetzung für ein modernes, skalierbares Daten- und KI-Ökosystem, das sowohl SAP-nativ als auch offen und multicloudfähig ist. Dies leistet einen großen Beitrag für die Entwicklung hin zu einer unternehmensweiten Data Fabric.