Think BIGGER

DEEP DIVE FÜR FACHLEUTE

BUSINESS INTELLIGENCE

Architekturebenen, Werkzeuge und Hersteller im Überblick

Autor	Erhard Rainer
Veröffentlicht	15.03.2014
Aktualisiert	07.07.2026
Umfang	ca. 33 Min. Lesezeit
Online	www.thinkbigger.at/wissen/business-intelligence/

thinkbigger.at

Business Intelligence · Machine Learning · KI-Agenten · Schulungen

Erhard Rainer · Schönaugasse 106 · 8020 Graz · info@think-bigger.at

Inhalt

Business Intelligence – Landkarte 2
Rollen im BI-Analytics-Cluster 2
Begriffsbestimmungen 2
Architektur einer BI-Plattform 3
Ebene 1 – Quellsysteme 3
Ebene 2 – Datenintegration: ETL / ELT 4
Ebene 3 – Data Warehouse / Data Lake / Lakehouse 6
Ebene 4 – Semantische Schicht / OLAP / Data Marts 8
Ebene 5 – Reporting & Dashboards 10
Ebene 6 – Advanced Analytics / Data Science / KI 12
Marktüberblick nach Hersteller 13
Datenbankmodellierung im DWH 15
Marktentwicklung – Gartner Magic Quadrant 15
Typische BI-Stacks in Großunternehmen 19
Entscheidungshilfe: Welcher Stack 2026? 24
Cloud Data Warehouses – Demo-Ressourcen 25

Business Intelligence ist in diesem Cluster der Architektur-Hub: Der Artikel ordnet Datenquellen, Integration, Data Warehouse, semantische Schicht, Reporting und Plattformstrategie. Einzelne Produkt- und Begriffsfragen gehören in die Detailartikel, besonders Begriffsbestimmungen rund um Data für Terminologie und Power BI für Microsofts Reporting- und Semantic-Model-Schicht.

Business Intelligence – Landkarte

Business Intelligence (BI) bezeichnet alle Technologien, Prozesse und Werkzeuge, die Rohdaten aus operativen Systemen in entscheidungsrelevante Information verwandeln. Die nachfolgende Landkarte gliedert die BI-Welt nach **Architekturebenen** – von der Datenquelle bis zum Dashboard. Die Ebenen sind teilweise austauschbar; Hersteller bieten oft Werkzeuge auf mehreren Ebenen an.

Rollen im BI-Analytics-Cluster

Artikel	Rolle	Wann zuerst lesen?
Business Intelligence	Architektur-Hub für Datenquellen, Integration, DWH, Lakehouse, Reporting und Plattformwahl	Wenn der Gesamtstack oder eine BI-Zielarchitektur sortiert werden soll
Begriffsbestimmungen rund um Data	Begriffsartikel für BI, Data Science, Big Data und Data Management	Wenn Begriffe im Team uneinheitlich verwendet werden
Power BI	Power-BI-Hub für Modell, Report, DAX, Betrieb und Troubleshooting	Wenn Microsoft Power BI konkret eingefuehrt oder betrieben wird
Power Query (M) vs. DAX	Brueckenartikel zwischen Datenaufbereitung und Modelllogik	Wenn unklar ist, ob Logik in Power Query oder DAX gehört
DAX	Detail-Hub für Kennzahlenlogik und tabellarische Modelle	Wenn Measures, Filterkontext oder Zeitintelligenz im Vordergrund stehen
Microsoft Fabric	Plattform-Hub für OneLake, Lakehouse, Warehouse, Pipelines und Semantic Models	Wenn Power BI Teil einer größeren Fabric-Datenplattform ist
SQL Server	Datenbank- und Betriebs-Hub im Microsoft-Stack	Wenn Datenhaltung, T-SQL, SSMS, Backup oder Performance das Problem sind

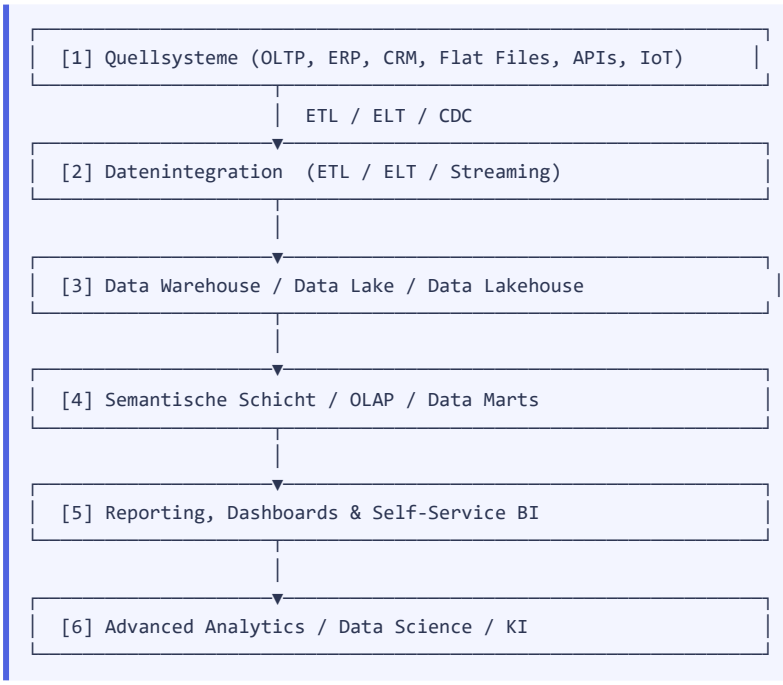
Begriffsbestimmungen

Begriff	Kurzdefinition
BI (Business Intelligence)	Sammelbegriff für Technologien und Methoden zur Analyse von Unternehmensdaten
OLTP	Online Transaction Processing – operative Datenbank, optimiert für Lese-/Schreibzugriffe
OLAP	Online Analytical Processing – analytische Datenbank, optimiert für komplexe Leseabfragen
ETL	Extract, Transform, Load – Daten werden transformiert, bevor sie ins Zielsystem geladen werden
ELT	Extract, Load, Transform – Daten werden erst geladen, dann im Zielsystem transformiert
Data Warehouse (DWH)	Zentrales, historisiertes Datenlager für Analyse und Reporting

Begriff	Kurzdefinition
Data Mart	Themenspezifischer Ausschnitt aus dem DWH (z. B. Vertrieb, Finanzen)
Data Lake	Rohdatenspeicher ohne festes Schema – strukturiert, semi- und unstrukturiert
Data Lakehouse	Hybrides Konzept: offenes Dateiformat (Parquet/Delta) + SQL-Abfrageebene
Semantische Schicht	Geschäftsdefinitionen über physischen Daten – Metriken, KPIs, Dimensionen
Cube / OLAP-Würfel	Multidimensionale Datenstruktur für schnelle Aggregationsabfragen
Dimensionsmodell	Star Schema / Snowflake Schema – Fakten- und Dimensionstabellen
Data Science	Statistik, ML und Programmierung zur Mustererkennung und Vorhersage
Big Data	Datenvolumen, das klassische Werkzeuge überfordert (Volume, Velocity, Variety)
Gartner Magic Quadrant	Jährliches Markt-Positioning-Modell für BI- und Analytics-Plattformen

---

Architektur einer BI-Plattform



Ebene 1 – Quellsysteme

Ausgangspunkt aller BI-Initiativen. Die Daten stammen aus operativen Systemen.

Typ	Beispiele
ERP	SAP S/4HANA, Oracle EBS, Microsoft Dynamics 365
CRM	Salesforce, SAP CRM, Microsoft Dynamics CRM
Relationale Datenbanken	SQL Server, Oracle DB, PostgreSQL, MySQL
Dateisysteme	CSV, Excel, XML, JSON
APIs / Webservices	REST, SOAP, GraphQL

Typ	Beispiele
Streaming / IoT	Kafka, MQTT, Azure Event Hub, AWS Kinesis
SaaS-Quellen	Shopify, HubSpot, Google Analytics

OLTP-Datenbanken sind für schnelle Inserts/Updates optimiert – nicht für analytische Abfragen über große Zeiträume.

Ebene 2 – Datenintegration: ETL / ELT

ETL (Extract–Transform–Load) ist die klassische Variante: Daten werden außerhalb des Zielsystems transformiert. Bei ETL übernimmt eine dedizierte Engine die Logik; bei ELT – zunehmend Standard in Cloud-Umgebungen – wird die Rohdatenlast direkt ins Zielsystem geschrieben und erst dort mit SQL oder Spark verarbeitet. ELT setzt die hohe Parallelrechenleistung moderner Cloud-DWH voraus und ermöglicht kürzere Ladewege. CDC (Change Data Capture) ergänzt beide Ansätze: statt ganzer Tabellen werden nur geänderte Datensätze übertragen, was Latenz und Datenmenge deutlich reduziert.

On-Premises ETL

On-Prem-ETL-Werkzeuge sind etabliert in gewachsenen Infrastrukturen und bieten tiefe Integration in bestehende Systeme. Sie laufen auf eigenen Servern, sind oft lizenzpflichtig und werden in Projekten mit hohen Governance- oder Datenschutzanforderungen bevorzugt.

Werkzeug	Hersteller	Notiz
SSIS (SQL Server Integration Services)	Microsoft	Standard im MS-Umfeld; visueller Dataflow-Designer in Visual Studio / SSDT
SAP Data Services	SAP	ETL für SAP BW und Nicht-SAP-Quellen
Oracle Data Integrator (ODI)	Oracle	ELT-Ansatz; Wissensmodule je Technologie
IBM DataStage	IBM	Enterprise ETL; Teil des InfoSphere-Portfolios
Informatica PowerCenter	Informatica	Marktführer im Enterprise-ETL-Segment
Talend Open Studio	Talend (Qlik)	Open-Source-Basis; Java-Codegenerierung
Pentaho Data Integration (PDI)	Hitachi Vantara	Auch bekannt als „Kettle"; Open-Source-Kern
Apache NiFi	Apache	Datenfluss-Orchestrierung; geeignet für Streaming

SSIS ist das Standard-ETL-Werkzeug im Microsoft-Umfeld. Pakete werden in Visual Studio entwickelt, auf einem SQL-Server-Agent deployed und visuell als Datenfluss-Graphen modelliert. Es ist ausgereift, aber nicht für Cloud-Szenarien ausgelegt.

Oracle Data Integrator (ODI) verfolgt konsequent den ELT-Ansatz: Die Transformationslogik wird als natives SQL im Zieldatenbank-Dialekt generiert. Das reduziert Middleware-Overhead, setzt aber SQL-Kenntnisse voraus.

Informatica PowerCenter ist der traditionelle Marktführer für unternehmenskritische ETL-Prozesse. Es überzeugt durch Metadatenverwaltung, Datenlineage und Konnektivität zu hunderten Quellen – zu einem entsprechend hohen Lizenzpreis.

IBM DataStage (InfoSphere) ist auf parallele Verarbeitung großer Datenvolumen ausgelegt und findet sich vorwiegend in IBM-lastigen Enterprise-Umgebungen.

Talend Open Studio hat einen Open-Source-Kern und generiert Java-Code aus grafischen Transformationsdefinitionen. Die Community-Edition deckt viele Anwendungsfälle ab; Enterprise-Features erfordern eine kommerzielle Lizenz.

Apache NiFi ist weniger klassisches ETL als ein visueller Datenfluss-Router. Besonders geeignet für IoT-Daten, Streaming und heterogene Quellen – weniger für komplexe Transformationslogik.

Cloud-native ETL / ELT

Cloud-ETL-Dienste werden als verwalteter Service betrieben – kein eigener Server, automatisches Scaling, Abrechnung nach Laufzeit oder Datenvolumen. Der Trend geht klar Richtung ELT mit dbt als Transformationsschicht und spezialisierten Konnektordiensten für die Ingestion.

Werkzeug	Hersteller	Notiz
Azure Data Factory (ADF)	Microsoft	Verwalteter ETL/ELT-Dienst in Azure
MS Fabric Dataflows Gen2	Microsoft	Low-Code ELT im Microsoft Fabric Ökosystem
AWS Glue	Amazon	Serverless ETL auf Basis von Spark
Google Dataflow	Google	Apache Beam-basiert; Batch + Streaming
SAP BTP Integration Suite	SAP	Cloud-Integration für SAP-Landschaften
Fivetran	Fivetran	Vollautomatisierte Konnektoren zu 300+ Quellen
Airbyte	Airbyte	Open-Source-Alternative zu Fivetran
dbt (data build tool)	dbt Labs	ELT-Transformationsschicht als SQL + Versionskontrolle
Matillion	Matillion	Cloud-nativer ETL; Konnektoren für Snowflake/Redshift

Azure Data Factory (ADF) ist das zentrale ETL/ELT-Werkzeug in Azure. Es orchestriert Pipelines zwischen hunderten Quell- und Zielsystemen, unterstützt Spark-Transformation über Azure Databricks und integriert nahtlos in Synapse und Fabric.

dbt (data build tool) hat die Art, wie Data-Engineering-Teams arbeiten, grundlegend verändert. Es behandelt SQL-Transformationen wie Softwarecode: Versionskontrolle mit Git, Tests, Dokumentation, Abhängigkeitsgraph. dbt übernimmt ausschließlich die T in ELT – die Ingestion muss separat gelöst werden.

Fivetran löst das Ingestion-Problem: vollautomatische, wartungsfreie Konnektoren zu über 300 SaaS-Quellen, Datenbanken und APIs. Sobald ein Konnektor konfiguriert ist, läuft er ohne Eingriff. Teuer, aber der Zeitaufwand für selbst entwickelte Pipelines ist oft teurer.

Airbyte ist die Open-Source-Alternative zu Fivetran mit ähnlichem Konzept, aber selbst gehostet oder als Cloud-Dienst verfügbar. Die Konnektorbibliothek ist groß und community-gepflegt.

AWS Glue ist der serverlose ETL-Dienst von AWS, der auf Apache Spark läuft. Er eignet sich für große Datenmengen im AWS-Ökosystem, erfordert aber Spark-Kenntnisse für komplexere Transformationen.

CDC (Change Data Capture): Debezium, Oracle GoldenGate, AWS DMS – erfassen nur geänderte Datensätze in Echtzeit und reduzieren so Latenz und Ladevolumen erheblich.

Ebene 3 – Data Warehouse / Data Lake / Lakehouse

Das zentrale Datenlager ist das Herzstück jeder BI-Architektur. Es aggregiert Daten aus allen Quellsystemen, historisiert sie und stellt sie konsistent für Analysen bereit. Bill Inmon prägte vier Eigenschaften des klassischen DWH: themenorientiert, integriert, nicht-flüchtig und zeitvariabel. In der Praxis haben sich drei Paradigmen herausgebildet: das relationale Data Warehouse (strukturiert, stark modelliert), der Data Lake (alle Daten, rohes Format, Schema-on-Read) und das Lakehouse (Kombination beider Ansätze mit offenem Dateiformat und SQL-Oberfläche).

Relationale Data Warehouses (on-premises)

On-Prem-DWH-Systeme sind in vielen Unternehmen seit Jahrzehnten im Einsatz. Sie bieten maximale Kontrolle über Hardware, Sicherheit und Lizenzkosten, erfordern aber eigenen Betrieb und Skalierung durch zusätzliche Hardware. Spaltenbasierte Speicherung (Columnstore) ist der entscheidende Performance-Hebel für analytische Abfragen.

System	Hersteller	Notiz
SQL Server DW	Microsoft	Columnstore-Indizes seit SQL Server 2012
SAP BW/4HANA	SAP	Moderne Neuauflage des klassischen SAP BW; läuft auf HANA
Oracle Autonomous DW	Oracle	Selbstverwaltend; on-prem und Cloud
Teradata Vantage	Teradata	Enterprise-DWH; MPP-Architektur
IBM Db2 Warehouse	IBM	MPP-DWH; Columnstore
Exasol	Exasol	In-Memory-DWH; hohe Query-Performance
Greenplum	VMware	Open-Source MPP auf PostgreSQL-Basis

SQL Server ist der Einstieg ins DWH für Microsoft-Umgebungen. Columnstore-Indizes ermöglichen analytische Abfragen auf relationalen Tabellen ohne separate OLAP-Engine. Für mittlere Datenmengen und Teams, die bereits mit T-SQL vertraut sind, oft ausreichend.

SAP BW/4HANA ist das vollständig überarbeitete SAP Data Warehouse, das ausschließlich auf SAP HANA läuft. Im Vergleich zum klassischen BW wurden viele ältere Objekte gestrichen; der Fokus liegt auf HANA-nativen Modellen und Integration mit SAP Analytics Cloud.

Teradata Vantage ist der Klassiker für sehr große Datenvolumen in Enterprise-Umgebungen. Die MPP-Architektur (Massively Parallel Processing) verteilt Abfragen auf viele Knoten. Teradata ist teuer, aber bei extremen Volumen und Konzernumgebungen mit hohem SLA-Anspruch eine bewährte Wahl.

Exasol fällt durch seine In-Memory-Architektur auf: Abfragen sind oft signifikant schneller als bei vergleichbaren Systemen, weil der gesamte heiße Datensatz im RAM gehalten wird. Beliebt als analytisches Layer über bestehenden Quellsystemen.

Cloud Data Warehouses

Cloud Data Warehouses trennen Storage von Compute: Daten liegen günstig auf Object-Storage (S3, Azure Blob), während Rechenkapazität nach Bedarf zugeschaltet und wieder freigegeben wird. Das ermöglicht elastische Skalierung ohne Vorausinvestition in Hardware und hat die DWH-Landschaft seit Mitte der 2010er-Jahre grundlegend verändert. Snowflake, BigQuery und Redshift haben den Markt definiert; Microsoft und SAP haben mit Synapse/Fabric bzw. Datasphere nachgezogen.

System	Hersteller	Notiz
Azure Synapse Analytics	Microsoft	Kombiniert DWH + Data Lake + Spark in Azure
Microsoft Fabric (Warehouse)	Microsoft	Nächste Generation; vereint OneLake + Synapse
Google BigQuery	Google	Serverless; Trennung von Storage und Compute
Amazon Redshift	AWS	Spaltenbasiertes Cloud-DWH
Snowflake	Snowflake	Multi-Cloud; elastische Compute-Cluster; breite Adoption
SAP Datasphere	SAP	Cloud-DWH der SAP; Nachfolger von SAP Data Warehouse Cloud
Oracle Autonomous DW	Oracle	Selbstverwaltend; Cloud-nativ
Databricks SQL Warehouse	Databricks	Lakehouse-Ansatz auf Delta Lake

Azure Synapse Analytics ist Microsofts integrierte Analyseplattform: es vereint ein dediziertes SQL-DWH (früher Azure SQL DW), serverlose SQL-Abfragen auf dem Data Lake, Apache Spark und Power BI in einer gemeinsamen Oberfläche. Wird zunehmend durch Microsoft Fabric abgelöst, existiert aber weiterhin parallel.

Microsoft Fabric ist die neue Unified-Analytics-Plattform von Microsoft (ab 2023). Sie vereint Data Engineering, Data Warehouse, Data Science, Real-Time Intelligence und Power BI unter einem Dach – mit OneLake als gemeinsamen Datenspeicher im Delta-Parquet-Format. Fabric ist der strategische Nachfolger von Synapse und Azure Analysis Services.

Google BigQuery ist vollständig serverlos: es gibt keine Cluster, keine Konfiguration von Rechenkapazität. Abfragen werden nach verarbeitetem Datenvolumen abgerechnet. BigQuery überzeugt durch extrem einfache Bedienung, native ML-Integration (BigQuery ML) und Stärke bei Ad-hoc-Analysen über sehr große Datasets.

Amazon Redshift war das erste weit verbreitete Cloud-DWH und ist tief ins AWS-Ökosystem integriert. Redshift Spectrum erlaubt SQL-Abfragen direkt auf S3-Daten, ohne sie vorab zu laden. Mit Redshift Serverless entfällt die Clusterkonfiguration.

Snowflake hat durch seinen Multi-Cloud-Ansatz (läuft auf AWS, Azure und GCP) und das Konzept des Daten-Sharings zwischen Organisationen eine starke Marktposition aufgebaut. Compute-Cluster (Virtual Warehouses) lassen sich sekundengenau starten und stoppen. Snowflake ist de facto Industriestandard für neue DWH-Projekte ohne bestehende Cloud-Bindung.

SAP Datasphere (vormals SAP Data Warehouse Cloud) ist SAPs Antwort auf die Cloud-DWH-Welt. Besondere Stärke: die enge Integration mit SAP S/4HANA und anderen SAP-Quellen, Data Access über virtuelle Modelle (kein Datenkopieren nötig) sowie eingebautes Business-Semantik-Layer.

Databricks SQL Warehouse nutzt Delta Lake als offenes Tabellenformat und positioniert sich als Lakehouse – ein DWH auf Basis eines Data Lakes. Besonders attraktiv für Teams, die ohnehin mit Databricks für Data Engineering und ML arbeiten und keine separate DWH-Schicht einführen wollen.

Data Lakes & Lakehouse

Ein Data Lake speichert Rohdaten in ihrem ursprünglichen Format – strukturiert, semistrukturiert und unstrukturiert. Kein fixes Schema bei der Einlagerung (Schema-on-Read), dafür maximale Flexibilität. Das Lakehouse-Konzept ergänzt den Lake um ACID-Transaktionen, Schema-Enforcement und SQL-Abfragbarkeit durch offene Tabellenformate wie Delta Lake, Apache Iceberg oder Apache Hudi. Damit wird aus einem rohen Datensee eine belastbare Analyseplattform.

System	Hersteller	Notiz
Azure Data Lake Storage (ADLS)	Microsoft	Object-Storage auf Basis Azure Blob
MS Fabric OneLake	Microsoft	Unified Lakehouse; Delta Parquet; ein Lake, alle Dienste
AWS S3 + Athena	Amazon	S3 als Lake; Athena für serverlose SQL-Abfragen
Google Cloud Storage + BigLake	Google	Offenes Dateiformat; einheitliche Governance
Databricks Lakehouse (Delta Lake)	Databricks	ACID-Transaktionen auf Parquet; Open Source
Apache Iceberg / Apache Hudi	Apache	Open-Source Tabellenformate; Multi-Engine-Unterstützung
Hadoop HDFS	Apache	Älteres Dateisystem; wurde durch Object-Storage verdrängt

Azure Data Lake Storage Gen2 (ADLS) kombiniert die Kosteneffizienz von Azure Blob Storage mit hierarchischer Namespace-Struktur für performante Dateisystemoperationen. Es ist die Datenbasis für Azure Synapse und Microsoft Fabric.

Microsoft Fabric OneLake ist der zentrale Datenspeicher in Microsoft Fabric – ein einziger logischer See für alle Fabric-Dienste. Daten liegen im Delta-Parquet-Format; verschiedene Workloads (SQL, Spark, Power BI) greifen auf denselben Datensatz zu, ohne ihn zu kopieren.

Delta Lake (Databricks, Open Source) ist das bekannteste Lakehouse-Tabellenformat. Es fügt ACID-Transaktionen, Schema-Evolution und Time Travel (Abfrage historischer Zustände) auf Parquet-Dateien hinzu. Auch von Spark, Flink und anderen Engines direkt unterstützt.

Apache Iceberg ist das Tabellenformat, das zunehmend als Industriestandard für offene Lakehouse-Architekturen gesehen wird. Unterstützung durch BigQuery, Snowflake, AWS, Spark und viele weitere – besonders geeignet, wenn mehrere Engines auf denselben Datensatz zugreifen sollen.

Hadoop HDFS war der ursprüngliche Standard für Big-Data-Speicher, wird aber in neuen Projekten kaum noch eingesetzt. Object-Storage (S3, ADLS, GCS) ist günstiger, einfacher zu betreiben und mindestens genauso performant bei analytischen Workloads.

Ebene 4 – Semantische Schicht / OLAP / Data Marts

Die semantische Schicht ist die Übersetzungsebene zwischen technischen Datenstrukturen und fachlichen Begriffen. Sie definiert Metriken, KPIs, Dimensionen und Geschäftsregeln an einem zentralen Ort – sodass alle Berichte und Dashboards konsistente Zahlen liefern. OLAP-Engines (Cubes) gehen einen Schritt weiter: sie aggregieren Daten vorab und ermöglichen so sekundenschnelle Pivotanalysen über Millionen von Zeilen.

Die Entscheidung zwischen einem echten OLAP-Cube und einem tabellarischen In-Memory-Modell hängt von Abfragekomplexität, Datenmenge und Team-Know-how ab. In modernen Architekturen übernimmt oft ein zentrales semantisches Modell (z. B. Power BI Semantic Model oder dbt Semantic Layer) die Rolle der klassischen Cube-Engine.

OLAP-Engines (Cubes, tabellarische Modelle)

System	Hersteller	Notiz
SSAS Multidimensional	Microsoft	Klassischer MDX-Cube; on-prem; stabil, ausgereift
SSAS Tabular	Microsoft	Spaltenbasiertes In-Memory-Modell; DAX-Abfragesprache

System	Hersteller	Notiz
Azure Analysis Services	Microsoft	Tabular als PaaS in Azure
MS Fabric Semantic Model (Power BI)	Microsoft	Zentrales Modell in Fabric; ersetzt langfristig AAS
SAP HANA (Calculation Views)	SAP	In-Memory OLAP nativ in SAP HANA
SAP BW InfoProvider	SAP	Klassisches BW-Modell: DSO, InfoCubes, CompositeProvider
Oracle Essbase	Oracle	Älteste OLAP-Engine; Hyperion-Ursprung; noch im Einsatz
IBM Planning Analytics (TM1)	IBM	Cube-basiert; stark im Planungsbereich
AtScale	AtScale	Universelle semantische Schicht; OLAP-Caching
Cube.dev (Semantic Layer)	Cube	Open-Source headless BI; API-first
dbt Metrics / dbt Semantic Layer	dbt Labs	Metriken als Code; Git-versioniert

SSAS Multidimensional ist die klassische Microsoft OLAP-Engine mit MDX-Abfragesprache und vorab aggregierten Cubes. In Bestandsarchitekturen noch weit verbreitet; neue Projekte setzen fast ausschließlich auf Tabular.

SSAS Tabular hat das multidimensionale Modell in den meisten neuen Projekten abgelöst. Das Modell lädt Daten als komprimierte Spaltenspeicher in den RAM (VertiPaq-Engine), Abfragen erfolgen in DAX. Power BI basiert intern auf derselben Engine.

Azure Analysis Services ist SSAS Tabular als verwalteter Cloud-Dienst. Wird langfristig durch das Power BI / Fabric Semantic Model abgelöst, ist aber in vielen Unternehmensarchitekturen noch das zentrale Enterprise-Semantikmodell.

SAP HANA Calculation Views sind das native OLAP-Konzept in SAP HANA. Daten werden nicht vorab aggregiert, sondern bei Abfrage in-memory berechnet – sehr schnell, aber ohne explizites Aggregations-Caching wie in klassischen Cubes.

Oracle Essbase ist eine der ältesten OLAP-Engines überhaupt, ursprünglich von Hyperion entwickelt und 2007 von Oracle übernommen. Stark im Finanzplanungsbereich (Budgetierung, Konsolidierung), aber Neuprojekte sind selten.

IBM Planning Analytics (TM1) ist ein Cube-basiertes System mit Turbo Integrator als ETL-Sprache. Besonders stark für Finanzplanung und Budgetierung; wird oft als Excel-Add-in (PAX) genutzt.

dbt Semantic Layer ist ein neuerer Ansatz: Metriken werden als Code (YAML + SQL) direkt im dbt-Projekt definiert und versioniert. Verschiedene BI-Tools greifen per standardisierter API auf dieselben Metrikdefinitionen zu.

Dimensionsmodellierung

Das Dimensionsmodell ist das Fundament eines jeden DWH-Berichts. Es trennt messbare Fakten (Umsatz, Menge, Kosten) von beschreibenden Dimensionen (Zeit, Produkt, Kunde, Region).

- **Star Schema:** Faktentabelle zentral, flache Dimensionstabellen drum herum – einfach, performant, gut für BI-Tools
- **Snowflake Schema:** normalisierte Dimensionen – speichereffizienter, aber komplexere Joins
- **Data Vault 2.0:** Rohdatenschicht (Hub/Link/Satellite) – auditierbar, flexibel erweiterbar, geeignet für agile DWH-Entwicklung
- **Anchor Modeling:** weitgehend normalisiertes Vault-Derivat, selten in der Praxis

---

Ebene 5 – Reporting & Dashboards

Die Präsentationsebene ist das, was Fachanwender täglich sehen. Hier trennen sich zwei Welten: **Enterprise Reporting** (pixel-perfect, druckfertig, für operative Berichte und regulatorische Anforderungen) und **Self-Service BI** (interaktive Dashboards, explorative Analyse, direkt durch den Fachanwender konfigurierbar). Beide Welten haben unterschiedliche Werkzeuge, unterschiedliche Nutzer und unterschiedliche Anforderungen – sie ergänzen sich in einer vollständigen BI-Architektur.

Enterprise Reporting (operational, pixel-perfect)

Pixel-perfect Reporting bedeutet: exakte Layoutkontrolle, Druckseitenumbrüche, Kopf- und Fußzeilen, barcodierte Listen – für Rechnungen, Lagerinventare, regulatorische Meldungen. Diese Berichte werden von Fachsystemen getriggert oder zeitgesteuert ausgeführt und an Empfänger verteilt, oft als PDF oder Excel.

Werkzeug	Hersteller	Notiz
SSRS (SQL Server Reporting Services)	Microsoft	Pixel-perfect Reports; on-prem; RDL-Format
SAP Crystal Reports	SAP	Klassiker für Druckreports; stark im Mittelstand
SAP BusinessObjects (BO)	SAP	Enterprise-Suite: Web Intelligence, Crystal, Explorer
Oracle Analytics Publisher	Oracle	Vormals BI Publisher; Druckreports, PDF-Ausgabe
IBM Cognos Analytics	IBM	Vollständige Enterprise-BI-Suite
Jaspersoft	TIBCO	Open-Source-Basis; PDF/HTML-Reports
BIRT	Eclipse/OpenText	Open-Source Reporting Framework

SSRS ist das Standard-Reporting-Tool im Microsoft-Stack. Berichte werden als RDL-Dateien definiert, auf einem Report-Server deployed und können als Subscriptions (E-Mail, Dateiablage) geplant werden. Pixel-perfect, aber aufwändig zu entwickeln.

SAP Crystal Reports ist seit Jahrzehnten der Marktstandard für formatierte Druckreports im Mittelstand. Viele ERP-Systeme (auch Nicht-SAP) liefern Crystal-Reports-Integration mit. Gut für statische Vorlagen, weniger für explorative Analyse.

SAP BusinessObjects ist die Enterprise-Suite für größere SAP-Umgebungen: Web Intelligence für Ad-hoc-Reports, Crystal für Drucklayouts, Explorer für Exploration. In SAP-Umgebungen der Vorgänger von SAP Analytics Cloud.

IBM Cognos Analytics ist eine vollständige BI-Suite mit Report Studio, Dashboard, Query Studio und Planning Analytics-Integration. In IBM-lastigen Organisationen und Großunternehmen mit langer Cognos-Geschichte nach wie vor verbreitet.

Self-Service BI & Dashboards

Self-Service BI ermöglicht Fachanwendern, ohne IT-Unterstützung Berichte und Dashboards zu erstellen. Der Schlüssel ist ein gutes semantisches Modell im Hintergrund – ohne dieses endet Self-Service in unkontrollierten „Shadow BI"-Insellösungen. Die Grenze zwischen Self-Service und Enterprise Reporting verschwimmt zunehmend; moderne Tools wie Power BI bedienen beide Welten.

Werkzeug	Hersteller	Notiz
Power BI	Microsoft	Marktführer im Self-Service-Segment; Desktop + Service + Embedded

Werkzeug	Hersteller	Notiz
SAP Analytics Cloud (SAC)	SAP	Cloud-native BI + Planung; nativer SAP-Anschluss
Tableau	Salesforce	Starke Visualisierung; breite Nutzergemeinschaft
Qlik Sense	Qlik	Assoziatives Datenmodell; Nicht-SQL-Exploration
Oracle Analytics Cloud (OAC)	Oracle	Vollständige Cloud-BI-Suite; ML integriert
Looker / Looker Studio	Google	LookML als semantische Schicht; eingebettet in Google Cloud
Amazon QuickSight	AWS	Serverless BI; SPICE In-Memory-Engine
MicroStrategy	MicroStrategy	Enterprise BI; starke Sicherheits- und Governance-Features
Sisense	Sisense	Embedded Analytics; White-Label
ThoughtSpot	ThoughtSpot	Search-basierte BI; NLP-Abfragen

Power BI ist der klare Marktführer im Self-Service-Segment. Power BI Desktop (kostenlos) für die Entwicklung, Power BI Service (SaaS) für die Verteilung, Power BI Embedded für die Integration in eigene Applikationen. Zentrale Stärke: enge Integration mit dem Microsoft-Ökosystem (Excel, Teams, Fabric) und sehr aktive Community. Die DAX-Sprache für Berechnungen hat eine steile Lernkurve.

SAP Analytics Cloud (SAC) vereint BI, Planung und Predictive in einer Cloud-Plattform. Besondere Stärke: SAP-native Konnektoren (direkte Live-Verbindung zu SAP HANA, BW, S/4) ohne Datenkopierung. SAC ist SAPs strategische Antwort auf Power BI und Tableau und löst BusinessObjects langfristig ab.

Tableau hat den Self-Service-BI-Markt mitgeprägt. Stärke liegt in der Visualisierungsqualität und der Drag-and-Drop-Oberfläche, die auch komplexe Analysen ohne SQL ermöglicht. Nach der Übernahme durch Salesforce enger ins CRM-Ökosystem integriert. Tableau Prep als ETL-Ergänzung.

Qlik Sense arbeitet mit einem assoziativen Datenmodell: alle Felder aller Tabellen sind jederzeit filterbar und miteinander verknüpft, ohne expliziten Join. Das ermöglicht Exploration, die SQL-basierte Tools nicht bieten – erfordert aber Umdenken beim Modelldesign.

Looker setzt auf LookML als zentrale semantische Definitionssprache. Alle Metriken und Dimensionen werden einmalig in LookML definiert; Dashboard-Autoren wählen nur noch aus diesen Bausteinen. Sehr konsistente Zahlen, aber steile Einstiegshürde. Looker Studio (früher Data Studio) ist die kostenlose, vereinfachte Variante.

Amazon QuickSight ist der serverlose BI-Dienst von AWS mit SPICE als In-Memory-Beschleuniger. Günstig per User-Session abgerechnet, tief ins AWS-Ökosystem integriert, aber im Funktionsumfang gegenüber Power BI oder Tableau eingeschränkt.

ThoughtSpot erlaubt Nutzern, Fragen in natürlicher Sprache zu stellen ("Umsatz nach Region letztes Quartal") und beantwortet sie mit automatisch generierten Visualisierungen. Interessanter Ansatz für nicht-technische Nutzer; ergänzt klassische BI-Tools, ersetzt sie selten vollständig.

Open Source / Community BI

Open-Source-BI-Werkzeuge sind in Data-Engineering-Teams und technisch geprägten Organisationen beliebt. Sie sind kostenlos einsetzbar, erfordern aber Eigenleistung beim Betrieb, Hosting und Sicherheit.

Werkzeug	Notiz
Apache Superset	SQL-Lab + Dashboards; stark in Data-Engineering-Teams
Metabase	Einfache Installation; für technisch-affine Fachanwender

Werkzeug	Notiz
Redash	Query-zentriert; geeignet für Entwickler
Grafana	Primär Monitoring/Observability; zunehmend BI-Einsatz

Apache Superset (Airbnb, Apache) bietet einen SQL-Editor, Visualisierungsbibliothek und Dashboards. Weit verbreitet in Data-Engineering-Umgebungen, die ohnehin mit Airflow, dbt und Spark arbeiten.

Metabase ist einfacher zu installieren und zu bedienen als Superset. Keine SQL-Pflicht für einfache Abfragen; gut für kleinere Teams mit gemischten technischen Kenntnissen.

Grafana kommt aus dem Monitoring-Bereich (Prometheus, InfluxDB), wird aber zunehmend auch für Geschäftsdaten eingesetzt. Stärke liegt in Zeitreihen und Echtzeit-Daten.

Ebene 6 – Advanced Analytics / Data Science / KI

Advanced Analytics geht über das Beschreiben der Vergangenheit (deskriptiv) und das Erklären von Zusammenhängen (diagnostisch) hinaus: prädiktive Modelle sagen zukünftige Ereignisse vorher, präskriptive Modelle empfehlen Handlungen. Diese Ebene nutzt statistische Methoden, Machine Learning und zunehmend generative KI. Die Grenzen zwischen BI und Data Science verschwimmen: moderne BI-Plattformen integrieren ML-Funktionen direkt, während Data-Science-Plattformen SQL-basierte Abfragen ermöglichen.

Plattformen

Plattform	Hersteller	Notiz
Azure Machine Learning	Microsoft	MLOps-Plattform; AutoML; Integration mit Fabric
Azure AI Foundry	Microsoft	KI-Entwicklung, Deployment und Governance
MS Fabric (Data Science)	Microsoft	Notebooks, Spark ML, MLflow in Fabric
SAP Analytics Cloud (Predictive)	SAP	Predictive-Szenarien nativ in SAC
Google Vertex AI	Google	Vollständige MLOps-Plattform
AWS SageMaker	Amazon	End-to-End ML; Auto-Scaling
Databricks ML + MLflow	Databricks	Lakehouse + ML in einer Plattform
IBM Watson Studio	IBM	Enterprise ML; AutoAI

Azure Machine Learning ist Microsofts MLOps-Plattform: Experimente, Modelltraining, Deployment als REST-Endpunkt, Monitoring. AutoML trainiert automatisch Modelle auf strukturierten Daten ohne manuelle Feature-Engineering-Schritte. Integriert in Microsoft Fabric.

Databricks ML + MLflow ist besonders attraktiv, wenn das Unternehmen ohnehin Databricks für Data Engineering nutzt. MLflow (Open Source, von Databricks entwickelt) ist der De-facto-Standard für Experiment-Tracking und Modell-Registry. Feature Store und AutoML ebenfalls integriert.

Google Vertex AI vereint Training, Hyperparameter-Tuning, Serving und Monitoring auf der Google-Cloud-Infrastruktur. Enge Integration mit BigQuery ML, das SQL-Nutzer ML-Modelle direkt per SELECT-Anweisung trainieren und anwenden lässt.

AWS SageMaker ist die umfangreichste ML-Plattform im AWS-Ökosystem mit einer Vielzahl spezialisierter Dienste (SageMaker Studio, Autopilot, Feature Store, Ground Truth für Labeling). Sehr mächtig, aber auch komplex im Einstieg.

SAP Analytics Cloud Predictive erlaubt Planern und Analysten, direkt in der SAC-Oberfläche Vorhersagemodelle zu erstellen – ohne Data-Science-Kenntnisse. Basiert auf SAP HANA Predictive Analysis Library (PAL).

Sprachen & Frameworks

Tool	Notiz
Python	Pandas, NumPy, scikit-learn, PyTorch, TensorFlow
R	Statistik-orientiert; tidyverse, ggplot2
SQL (erweitert)	Window-Funktionen, ML-Erweiterungen (BigQuery ML, Redshift ML)
Apache Spark (PySpark)	Verteilte Datenverarbeitung; Standard für Big Data
Jupyter Notebooks	Interaktive Entwicklungsumgebung

Python ist die Lingua franca der Datenwissenschaft. Die Bibliotheken Pandas (Datenmanipulation), scikit-learn (klassisches ML), PyTorch und TensorFlow (Deep Learning) decken nahezu alle Anwendungsfälle ab.

R ist statistisch mächtiger als Python in manchen Nischen (z. B. Zeitreihenanalyse, Biometrie), hat aber in Data-Engineering-Umgebungen gegenüber Python an Boden verloren. Tidyverse und ggplot2 sind für explorative Datenanalyse exzellent.

Apache Spark (PySpark) ist der Standard für verteilte Datenverarbeitung auf Millionen von Zeilen. Fast alle Cloud-DWH- und Lakehouse-Plattformen haben Spark integriert (Fabric, Databricks, Synapse, EMR).

Marktüberblick nach Hersteller

Microsoft

Produkt	Ebene	Deployment
SSIS	ETL	On-Prem
Azure Data Factory	ETL/ELT	Cloud
MS Fabric Dataflows	ELT	Cloud
SQL Server DW	DWH	On-Prem
Azure Synapse Analytics	DWH	Cloud
Microsoft Fabric (Warehouse)	DWH / Lakehouse	Cloud
SSAS (Multidimensional/Tabular)	OLAP	On-Prem
Azure Analysis Services	OLAP	Cloud
Fabric Semantic Model	OLAP	Cloud
SSRS	Reporting	On-Prem
Power BI	Self-Service BI	Cloud + Desktop

Produkt	Ebene	Deployment
Azure ML / AI Foundry	Data Science	Cloud

SAP

Produkt	Ebene	Deployment
SAP Data Services	ETL	On-Prem / Cloud
SAP BTP Integration Suite	ETL/Integration	Cloud
SAP BW/4HANA	DWH	On-Prem / Cloud
SAP Datasphere	DWH	Cloud
SAP HANA (Calc. Views)	OLAP / Semantik	On-Prem / Cloud
SAP BusinessObjects	Reporting	On-Prem
SAP Crystal Reports	Reporting	On-Prem
SAP Analytics Cloud (SAC)	Self-Service BI	Cloud

Oracle

Produkt	Ebene	Deployment
Oracle Data Integrator (ODI)	ETL	On-Prem / Cloud
Oracle Autonomous Data Warehouse	DWH	Cloud
Oracle Essbase	OLAP	On-Prem / Cloud
Oracle Analytics Publisher	Reporting	On-Prem / Cloud
Oracle Analytics Cloud (OAC)	Self-Service BI	Cloud

Amazon Web Services (AWS)

Produkt	Ebene
AWS Glue	ETL/ELT
Amazon S3	Data Lake
Amazon Redshift	DWH
Amazon Athena	Query Engine
Amazon QuickSight	Reporting/BI
AWS SageMaker	Data Science

Google Cloud

Produkt	Ebene
Google Dataflow	ETL/Streaming
Google BigQuery	DWH
Google BigLake	Data Lake
Looker / Looker Studio	BI / Semantik

Produkt	Ebene
Google Vertex AI	Data Science

Unabhängige Anbieter (Best-of-Breed)

Hersteller	Kernprodukt	Stärke
Snowflake	Cloud DWH	Multi-Cloud; Daten-Sharing
Databricks	Lakehouse (Delta Lake)	ML + Data Engineering
Informatica	PowerCenter / IDMC	Enterprise ETL + Governance
Fivetran	Cloud Konnektoren	Zero-Maintenance Pipelines
dbt Labs	dbt	ELT-Transformation als Code
Talend (Qlik)	Open Studio / Cloud	Open Source + Enterprise
Tableau (Salesforce)	Tableau	Visualisierung
Qlik	Qlik Sense	Assoziatives Modell
MicroStrategy	MicroStrategy	Enterprise BI
IBM	Cognos / Planning Analytics	Enterprise BI + Planung
Teradata	Vantage	MPP-DWH

Datenbankmodellierung im DWH

Transaktionsdatenbanken (OLTP) sind auf schnelle Inserts/Updates normalisiert. Data Warehouses (OLAP) nutzen denormalisierte Strukturen für schnelle Leseabfragen.

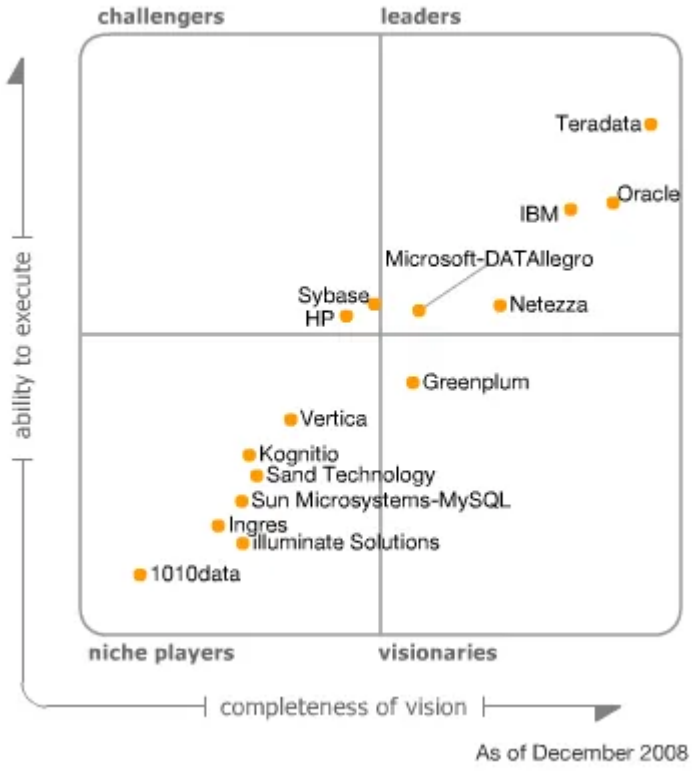
- **Star Schema:** Faktentabelle + flache Dimensionen → einfach, performant
- **Snowflake Schema:** Normalisierte Dimensionen → speichereffizienter
- **Data Vault 2.0:** Rohdatenschicht (Hub/Link/Satellite) → auditierbar, flexibel erweiterbar

---

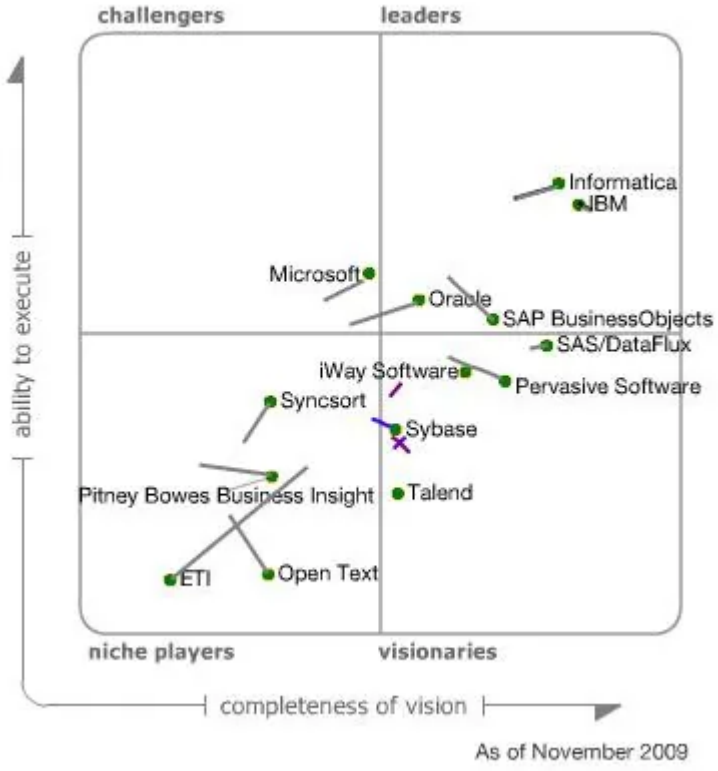
Marktentwicklung – Gartner Magic Quadrant

Der Gartner Magic Quadrant for Analytics and BI Platforms erscheint jährlich und positioniert Hersteller nach Vollständigkeit der Vision und Fähigkeit zur Umsetzung. Aktuell dominieren Microsoft (Power BI), Tableau und Qlik das Leader-Segment.

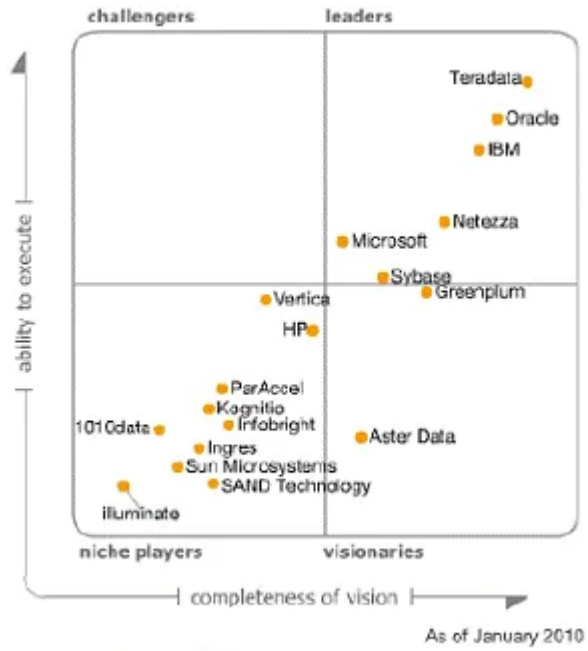
Historische Reports (2008–2013) zeigen die Konsolidierung des Markts: aus vielen Nischenanbietern wurden wenige Plattformführer.



As of December 2008

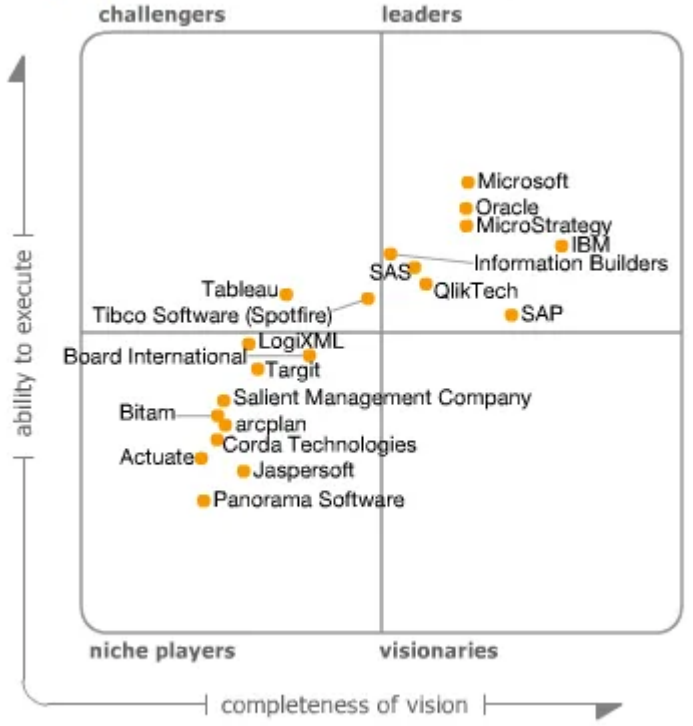


As of November 2009

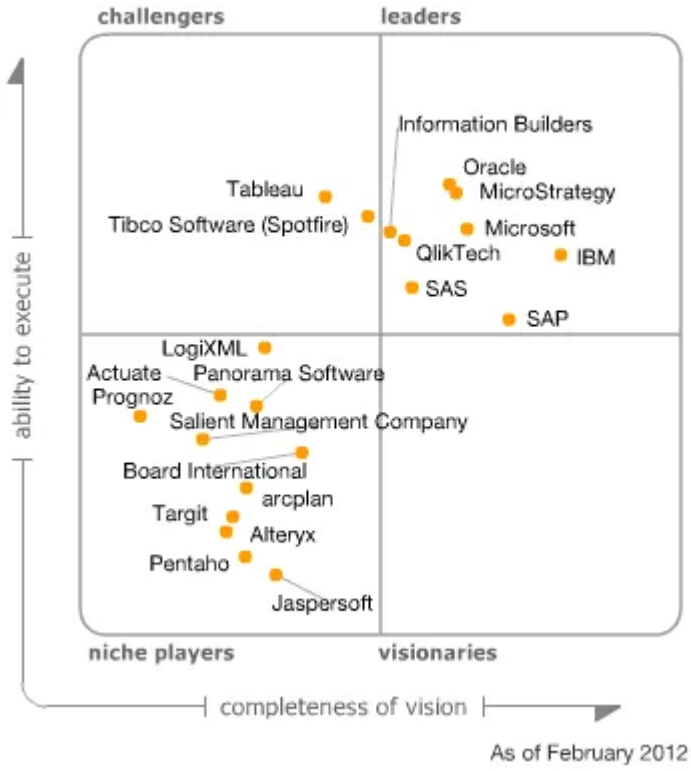


Source: Gartner (January 2010)

Magic Quadrant for Business Intelligence Platforms



Source: Gartner (January 2011) As of January 2011



As of February 2012

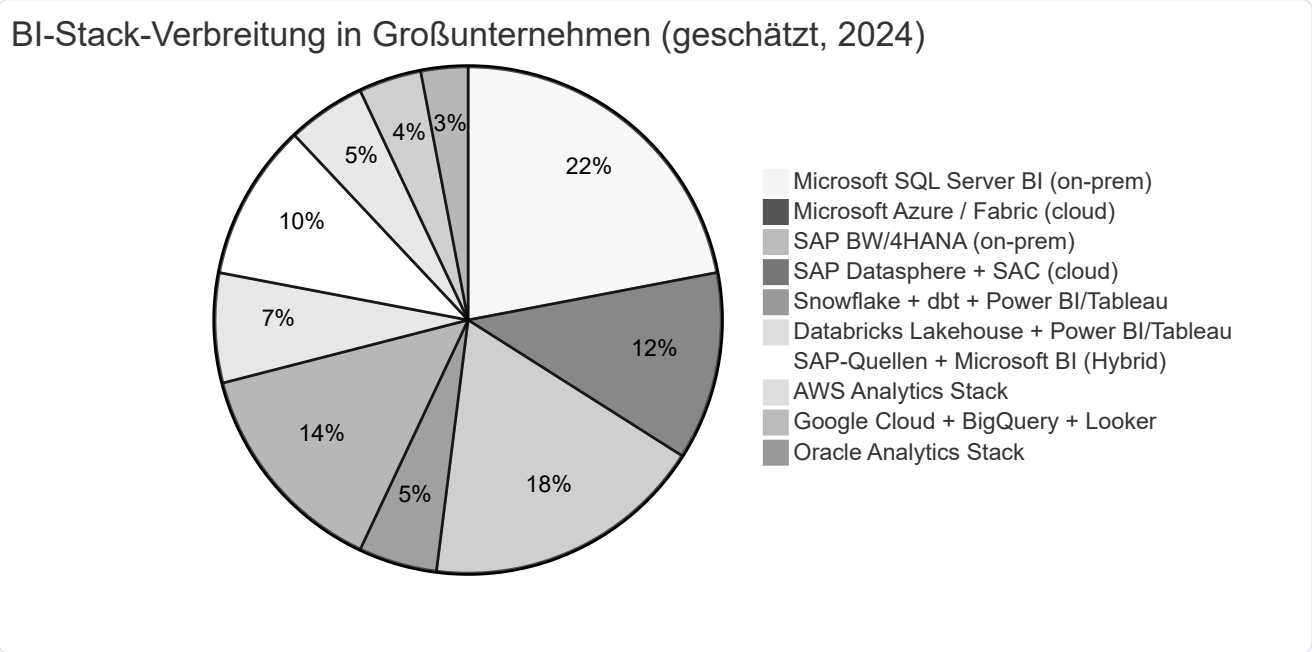


As of January 2013

Typische BI-Stacks in Großunternehmen

In der Realität setzen die wenigsten Konzerne einen einzigen homogenen BI-Stack durch – historisch gewachsene Systeme, Unternehmenskäufe und abteilungsspezifische Anforderungen führen fast immer zu Mischlandschaften. Trotzdem lassen sich wiederkehrende Muster erkennen.

Der BI-Plattformmarkt (Gartner, IDC, 2024) ist in Großunternehmen nach installierter Basis ungefähr so verteilt: SAP dominiert in ERP-lastigen Konzernen (besonders DACH/Europa) mit BW und BusinessObjects, Microsoft hat durch SQL Server, Power BI und Azure die größte Breitenabdeckung über alle Unternehmensgrößen, Oracle ist stark in Finance/ERP-Umgebungen mit Oracle-Datenbankbasis. Der „Modern Data Stack" (Snowflake/dbt/Tableau oder Power BI) wächst seit 2020 am schnellsten – getrieben von Cloud-first-Strategien und Data-Engineering-Teams. Reine AWS- und Google-Cloud-Stacks sind eher in Tech-Unternehmen und Start-ups verbreitet.



Hinweis: Schätzwerte auf Basis von Gartner Magic Quadrant, IDC-Berichten und Marktbeobachtungen. Mehrfachzählungen möglich, da viele Konzerne parallel mehrere Stacks betreiben.

Stack 1 – Microsoft SQL Server BI (on-premises)

Verbreitung: Sehr hoch; Standard in Mittelstand und Konzernen ohne SAP-Backbone, historisch gewachsen seit den 2000ern.

Der klassische Microsoft-Stack deckt die gesamte BI-Pipeline mit eng verzahnten Komponenten ab. On-Prem bedeutet volle Kontrolle über Hardware und Lizenz, aber eigenverantwortlicher Betrieb und begrenzte Skalierbarkeit. In den meisten Installationen wurde SSRS als Reporting-Frontend inzwischen durch Power BI ersetzt oder ergänzt.

Ebene	Werkzeug
Datenintegration	SSIS
Data Warehouse	SQL Server (Columnstore)
OLAP / Semantik	SSAS Tabular oder Multidimensional

Ebene	Werkzeug
Reporting	SSRS, Power BI

- **SSIS** – Visueller Dataflow-Designer in Visual Studio; bewährt für On-Prem-Integrationen, nicht für Cloud-native Szenarien geeignet.
- **SSAS** – Tabular (DAX, VertiPaq-Engine) für neue Projekte; Multidimensional (MDX-Cubes) in Bestandsarchitekturen. Power BI basiert intern auf derselben Engine.
- **SSRS** – Pixel-perfect Reports per Subscription; wird in vielen Organisationen noch parallel zu Power BI betrieben.

Installation: Installing SSIS, SSRS and SSAS with Visual Studio 2019 | SQL Server Integration Services (SSIS) | SSAS Cubes

---

Stack 2 – Microsoft Azure / Fabric (cloud)

Verbreitung: Stark wachsend; strategische Zielarchitektur für Microsoft-Kunden, die on-prem ablösen wollen.

Microsoft Fabric (ab 2023) vereint alle bisherigen Azure-Datendienste unter einer Plattform mit **OneLake** als gemeinsamem Datenspeicher im Delta-Parquet-Format. Alle Workloads – ETL, DWH, Spark, Power BI, ML – greifen auf denselben Lake zu, ohne Daten zu kopieren. Fabric löst Azure Synapse, ADF und Azure Analysis Services langfristig ab.

Ebene	Werkzeug
Datenintegration	Data Factory, Dataflows Gen2
Data Warehouse / Lakehouse	Fabric Warehouse, Fabric Lakehouse
OLAP / Semantik	Fabric Semantic Model (DAX)
Reporting	Power BI
Data Science	Fabric Notebooks, MLflow

Komponenten im Überblick

Komponente	Entspricht (alt)	Funktion
Data Factory	Azure Data Factory	ETL/ELT-Pipelines; 150+ Konnektoren
Dataflows Gen2	Power Query / ADF Dataflows	Low-Code ELT; visuell, Power-Query-Syntax
Data Engineering (Notebooks)	Azure Databricks / Synapse Spark	PySpark-Notebooks; Lakehouse-Bearbeitung
Data Warehouse	Azure Synapse SQL Pool	Vollständiges SQL-DWH auf OneLake; T-SQL
Lakehouse	Azure Data Lake + Synapse	Delta-Lake-Tabellen + SQL-Endpunkt
Real-Time Intelligence	Azure Stream Analytics	Streaming-Ingestion; Eventstream; KQL
Semantic Model	Azure Analysis Services	Zentrales Power-BI-Datenmodell (DAX)
Power BI	Power BI Service	Dashboards und Reports; direkt in Fabric integriert
Data Science	Azure ML / Synapse ML	Notebooks, MLflow, Spark ML
Data Activator	–	Ereignisbasierte Alerts auf Echtzeit-Daten

OneLake – ein einziger logischer Data Lake pro Tenant; Shortcuts binden S3, GCS oder ADLS ein, ohne Daten physisch zu kopieren. Lizenziert über Capacity Units (F-SKUs F2–F2048); Pause/Resume reduziert Kosten. Wettbewerber: Databricks, Snowflake, Google BigQuery.

Stack 3 – SAP BW/4HANA (on-premises / cloud)

Verbreitung: Sehr hoch in SAP-Konzernen; besonders stark in DACH, automotive, Chemie, Pharma, Handel.

SAP BW/4HANA ist das vollständig überarbeitete SAP Data Warehouse, das ausschließlich auf SAP HANA läuft. Es ist die modernisierte Ablösung des klassischen SAP BW (NetWeaver) und läuft entweder on-premises oder in der RISE-with-SAP-Cloud-Umgebung. Charakteristisch sind die spezifischen SAP-Datenstrukturen, die enge Verzahnung mit SAP-Quellsystemen und der Einsatz von BEx-Tools oder Analysis for Office als Reporting-Frontend.

Ebene	Werkzeug
Datenintegration	SAP Data Services, BW Process Chains
Data Warehouse	SAP BW/4HANA auf SAP HANA
OLAP / Semantik	InfoProvider, HANA Calculation Views
Reporting	BEx Query Designer, Analysis for Office, SAP BusinessObjects

Zentrale Objekte: **InfoObjects** (Merkmale/Kennzahlen), **DSO** (DataStore Object), **CompositeProvider**, **HANA-optimierte InfoCubes**, **Process Chains** für ETL-Steuerung.

Zertifizierung: SAP Certified Application Associate – SAP BW/4HANA 2.x

---

Stack 4 – SAP Datasphere + SAC (cloud)

Verbreitung: Wachsend; SAPs strategische Cloud-Zielarchitektur für neue Projekte und BW-Migrationen.

SAP Datasphere (vormals SAP Data Warehouse Cloud) ist SAPs Cloud-DWH mit eingebautem Business-Semantik-Layer. Besonderheit: virtuelle Modelle erlauben den Zugriff auf SAP-Quellen ohne physischen Datenkopiervorgang. SAP Analytics Cloud (SAC) ist das Frontend für BI, Planung und Predictive – und der strategische Ablöser von SAP BusinessObjects.

Ebene	Werkzeug
Datenintegration	SAP BTP Integration Suite, Data Flows
Data Warehouse	SAP Datasphere
OLAP / Semantik	SAP HANA Calculation Views, semantisches Modell in Datasphere
Reporting & Planung	SAP Analytics Cloud (SAC)

SAC vereint BI (Stories, Dashboards), integrierte Planung (Eingabeformulare, Versionen) und Predictive (Smart Predict) in einer Plattform. Direkte Live-Verbindung zu SAP HANA und BW ohne Datenkopierung ist die zentrale Stärke.

Zertifizierung: SAP Certified Application Associate – SAP Analytics Cloud

---

Stack 5 – Modern Data Stack: Snowflake + dbt + Power BI / Tableau

Verbreitung: Stark wachsend; Standard in neuen Cloud-first-Projekten ohne bestehende Plattformbindung.

Der „Modern Data Stack" hat sich seit ca. 2020 als de-facto-Standard für neue BI-Plattformen etabliert. Jede Komponente ist ein spezialisiertes Best-of-Breed-Produkt: Fivetran oder Airbyte für Ingestion, Snowflake als Cloud-DWH, dbt für Transformation als Code, Power BI oder Tableau für Visualisierung. Die Stärken sind hohe Agilität, Git-Versionierung aller Transformationen und einfache Skalierung – die Schwäche ist die Komplexität durch viele Einzelprodukte von verschiedenen Anbietern.

Ebene	Werkzeug
Datenintegration	Fivetran / Airbyte (Ingestion) + dbt (Transformation)
Data Warehouse	Snowflake
OLAP / Semantik	dbt Semantic Layer, oder Cube.dev
Reporting	Power BI oder Tableau

Snowflake punktet mit Multi-Cloud-Flexibilität (AWS/Azure/GCP), elastischen Virtual Warehouses und Daten-Sharing zwischen Organisationen. dbt behandelt SQL-Transformationen wie Softwarecode: Tests, Dokumentation, Lineage, Git-History – und hat die Art, wie Data Teams arbeiten, grundlegend verändert.

Stack 6 – Databricks Lakehouse + Power BI / Tableau

Verbreitung: Mittel, wachsend; besonders stark in Unternehmen mit hohem ML-Bedarf und großen Datenmengen.

Databricks kombiniert Data Engineering (Spark), Data Warehouse (SQL Warehouse auf Delta Lake) und Machine Learning (MLflow) in einer Plattform. Der Lakehouse-Ansatz auf Delta Lake ist besonders attraktiv, wenn ML-Modelle und BI-Reports auf denselben Daten operieren sollen – ohne separate Datenkopie in ein DWH. Power BI oder Tableau werden als Frontend angebunden.

Ebene	Werkzeug
Datenintegration	Databricks Autoloader, DLT (Delta Live Tables)
Data Warehouse / Lakehouse	Databricks SQL Warehouse auf Delta Lake
OLAP / Semantik	Power BI Semantic Model (via DirectQuery)
Reporting	Power BI, Tableau
ML / Data Science	Databricks ML, MLflow

Delta Lake liefert ACID-Transaktionen, Schema-Evolution und Time Travel auf Parquet-Dateien. Besonders geeignet für Unternehmen, die bereits mit Spark arbeiten und keinen separaten DWH-Layer einführen wollen.

Stack 7 – Hybrid: SAP-Quellen + Microsoft BI

Verbreitung: Sehr hoch; einer der häufigsten Mischstacks in SAP-Konzernen, besonders in DACH.

Viele Konzerne betreiben SAP als operatives Rückgrat (S/4HANA, ECC), wollen aber nicht das gesamte Reporting in SAP-Tools abbilden. Typische Lösung: SAP liefert die Daten, ein Microsoft-Stack übernimmt DWH und Visualisierung. Die Integration erfolgt über SAP-Konnektoren (OData, BAPI, SAP BW DirectQuery)

oder durch Replikation via Azure Data Factory / SSIS.

Ebene	Werkzeug
Quellsystem	SAP S/4HANA oder SAP ECC
Datenintegration	Azure Data Factory, SSIS oder SAP Data Services
Data Warehouse	Azure Synapse / Fabric Warehouse oder SQL Server DW
OLAP / Semantik	SSAS Tabular oder Fabric Semantic Model
Reporting	Power BI

Der Grund für diesen Mischstack: Power BI hat bessere Self-Service-Fähigkeiten und eine breitere Nutzerbasis als SAP-native Tools; gleichzeitig bleiben SAP-Transaktionsdaten der Single Point of Truth. Risiko: zwei Governance-Strukturen (SAP und Microsoft), die synchron gehalten werden müssen.

Stack 8 – AWS Analytics Stack

Verbreitung: Mittel; Standard in AWS-first-Unternehmen, besonders in USA und Tech-Branche.

AWS bietet einen vollständigen Analytics-Stack aus eigenen Diensten. Die Stärke liegt in der tiefen Integration mit anderen AWS-Diensten (S3, Lambda, SageMaker) und dem Pay-per-Use-Modell. Die Schwäche ist die Fragmentierung: viele Einzeldienste, die manuell integriert werden müssen.

Ebene	Werkzeug
Datenintegration	AWS Glue (Spark ETL), AWS DMS, Kinesis (Streaming)
Data Lake	Amazon S3
Data Warehouse	Amazon Redshift
Query Engine	Amazon Athena (serverlos auf S3)
Reporting	Amazon QuickSight, Tableau oder Power BI
ML	AWS SageMaker

Redshift ist spaltenbasiert und tief mit S3 integriert (Redshift Spectrum für direkte S3-Abfragen). QuickSight ist günstig und serverlos, im Funktionsumfang aber gegenüber Power BI und Tableau eingeschränkt – weshalb viele AWS-Kunden trotzdem Tableau oder Power BI als Frontend nutzen.

Stack 9 – Google Cloud: BigQuery + Looker

Verbreitung: Gering bis mittel; dominiert in Tech-Unternehmen mit Google-Affinität und Data-Science-Fokus.

Google Cloud Analytics basiert auf BigQuery als vollständig serverlosem DWH. Stärke ist die Skalierbarkeit auf Petabytes ohne Konfigurationsaufwand und die native ML-Integration (BigQuery ML trainiert Modelle per SELECT-Statement). Looker liefert ein zentrales semantisches Modell in LookML – alle Metriken werden einmalig definiert, alle Reports bauen darauf auf.

Ebene	Werkzeug
Datenintegration	Google Dataflow, dbt, Fivetran
Data Warehouse	Google BigQuery

Ebene	Werkzeug
Semantik	Looker (LookML)
Reporting	Looker, Looker Studio (kostenlos)
ML	Google Vertex AI, BigQuery ML

BigQuery rechnet per verarbeitetem Datenvolumen ab – ideal für Ad-hoc-Analysen, aber bei schlecht optimierten Abfragen schnell teuer. Looker Studio (früher Data Studio) ist die kostenlose, vereinfachte Variante für kleinere Teams.

Stack 10 – Oracle Analytics Stack

Verbreitung: Mittel; stark in Unternehmen mit Oracle-ERP-Basis (JD Edwards, Oracle E-Business Suite, PeopleSoft).

Oracle bietet einen vollständigen Stack von ETL bis Reporting, der besonders in Finanzdienstleistung, Versicherung und Großhandel mit Oracle-Datenbankbasis verbreitet ist. Essbase ist die älteste OLAP-Engine am Markt und im Finanzplanungsbereich (Budgetierung, Konsolidierung) nach wie vor stark vertreten.

Ebene	Werkzeug
Datenintegration	Oracle Data Integrator (ODI)
Data Warehouse	Oracle Autonomous Data Warehouse (ADW)
OLAP / Planung	Oracle Essbase
Reporting	Oracle Analytics Cloud (OAC)

Oracle ADW ist selbstverwaltend: Indizes, Partitionierung und Performance-Tuning werden automatisch durchgeführt. ODI verfolgt einen ELT-Ansatz – Transformationen werden als nativer SQL-Code im Zieldatenbank-Dialekt generiert, was Middleware-Overhead reduziert.

Entscheidungshilfe: Welcher Stack 2026?

Die Plattformwahl ist 2026 strategischer als je zuvor – denn ein einmal aufgebauter Stack zieht Datenmodelle, Qualifikationen und Governance-Strukturen nach sich, die über Jahre bestehen bleiben. Ein paar Leitfragen helfen bei der Orientierung:

Welches ERP/Quellsystem dominiert? Wer SAP S/4HANA als zentrales ERP betreibt, hat mit SAP Datasphere + SAC den geringsten Integrationsaufwand – besonders wenn Live-Zugriff auf SAP-Daten ohne Replikation gewünscht ist. Wer kein SAP im Kern hat, hat keine technische Notwendigkeit, den SAP-BI-Stack einzuführen.

Welche Cloud-Strategie gilt im Unternehmen? Azure-first-Organisationen mit Microsoft-365-Umgebung fahren mit Fabric am besten – der einheitliche Governance-Layer (Entra ID, Purview), die Power-BI-Integration und das Lizenzmodell greifen ineinandergreifend. AWS-first-Organisationen haben mit Glue + Redshift + QuickSight oder einem Snowflake-Overlay eine sinnvolle Ausgangsbasis. Ohne Cloud-Bindung ist der Modern Data Stack (Snowflake + dbt + Power BI) die flexibelste Option.

Wie hoch ist der ML- und Data-Engineering-Anteil? Wo Data Science und BI gleichwertig sind, lohnt sich Databricks als Unified Lakehouse – ein Stack für Feature Engineering, Modelltraining und SQL-Reporting. Reine BI-Organisationen ohne ML-Ambition brauchen keine Spark-Plattform.

Wie groß ist das Team? Kleine Teams profitieren von integrierten Plattformen (Fabric, Databricks) ohne viele Integrationspunkte. Größere Teams mit spezialisierten Rollen (Data Engineers, Analytics Engineers, Analysts) können Best-of-Breed-Stacks durch klare Schnittstellen besser aufteilen.

Meine persönliche Einschätzung (Erhard Rainer, 2026)

Ich setze aktuell konsequent auf **Microsoft Fabric as the only platform** – und halte das für die richtige Entscheidung für Microsoft-affine Umgebungen. Die Gründe:

- **Ein Stack, eine Governance:** Keine Brücken zwischen ADF, Synapse, AAS und Power BI mehr; alles teilt denselben OneLake und dasselbe Berechtigungsmodell.
- **Delta Lake als offenes Format:** Die Daten sind nicht in Fabric eingesperrt – andere Tools können direkt auf das Parquet-Format zugreifen.
- **Roadmap-Momentum:** Microsoft investiert massiv; das Feature-Tempo bei Fabric ist höher als bei jedem vergleichbaren Dienst der letzten Jahre.
- **Power BI als Frontend ist gesetzt:** Wer Power BI als Reporting-Standard hat, hat den stärksten Grund, auch den Rest des Stacks in Fabric zu halten – das Semantic Model lebt direkt im selben Ökosystem.
- **Reduktion auf das Wesentliche:** Weniger Werkzeuge, weniger Integrationspunkte, weniger Betriebsaufwand – das spart Zeit, die in Datenqualität und Modellierung investiert werden kann.

Der einzige echte Nachteil: wer nicht primär in Azure/Microsoft-Welt lebt, holt sich eine starke Vendor-Bindung ins Haus. In dem Fall ist Snowflake + dbt + Power BI eine faire Alternative, die ähnliche Prinzipien (offene Formate, Code-basierte Transformation, zentrales Semantic Model) mit Cloud-Unabhängigkeit kombiniert.

Cloud Data Warehouses – Demo-Ressourcen

- Google BigQuery – kostenloses Sandbox-Tier verfügbar
- Amazon Redshift – kostenloser ra3.xlplus-Knoten für 2 Monate
- Azure Synapse Analytics – serverlose Abfragen ab 5 TB/Monat kostenlos
- Snowflake – 30-Tage-Trial inkl. Credits
- Benchmark: Data Warehouse Cloud Benchmark (GigaOm)

Demo-Datenbanken für lokale Tests: Demodatenbanken

Über den Autor

Erhard Rainer berät kleine und mittlere Unternehmen in Österreich zu Business Intelligence, Machine Learning und KI-Agenten. Er hat 15 Jahre Erfahrung in Data Analytics, davon 10 Jahre in der Leitung von Teams und Departments in Konzernen, und ist Dozent für betriebswirtschaftliche und Data-Analytics-Fächer.

Sprechen Sie mit mir über Ihre Daten: Das Erstgespräch ist kostenlos.

https://www.thinkbigger.at/contact.html#anfrage

© 2026 Think BIGGER – Erhard Rainer. Dieses PDF gibt den Stand des Online-Artikels vom 07.07.2026 wieder; maßgeblich ist die Online-Fassung unter https://www.thinkbigger.at/wissen/business-intelligence/