



Image credit: Vladimir\_Timofeev



**Andrew J. Brust**  
Aug 6, 2021

# **GigaOm Radar for Data Warehouses** v2.0

Data Warehouse

# GigaOm Radar for Data Warehouses

## Table of Contents

- 1 Summary
- 2 Key Criteria Comparison
- 3 GigaOm Radar
- 4 Vendor Insights
- 5 Analyst's Take
- 6 About Andrew Brust
- 7 About GigaOm
- 8 Copyright

# 1. Summary

Data warehouse platforms are an established part of the enterprise technology stable and a go-to solution for managing data at scale within an organization. But recently, features are emerging that transform data warehouses from serving their more traditional technology role of the past into the modern, flexible, cloud-based big data platforms they are today.

These innovations include advancements to the platforms themselves, as well as improvements that allow them to integrate with a wide ecosystem of external tools. Longstanding breakthroughs include massively parallel processing, columnar storage, and vector processing. More recently, data warehouses have accumulated several advantages that the cloud brings, including scalability derived from separated compute and storage, ease of setup, cost effectiveness, and optimizations for concurrency. Integrations include those with business intelligence (BI) tools, machine learning (ML) tools, security and governance capabilities, and other elements of vendors' cloud application suites.

As data warehouses are a core component of many data management strategies, choosing one is a weighty decision. This report is intended to inform and assist organizations in that decision-making process, to help them make the best-informed decisions they can about the platform that best suits their needs.

In our [Key Criteria Report for Evaluating Data Warehouses](#) we articulated a number of key findings in the sector. They are:

- Data warehouses are widely known and trusted solutions that enterprises have been using for many years to manage and analyze huge amounts of data within their organizations, and they are a core component of an enterprise's data management strategy.
- The cloud is a major innovation catalyst in the data warehousing category, introducing concepts such as auto-scaling and elasticity, concurrency, freedom from physical appliance, and usage-based pricing. All data warehouse vendors have some cloud option. Some platforms are cloud-native, some are designed to encourage customers to migrate their existing systems to the cloud, and some support a hybrid approach that includes both cloud and on-premises or implementations across multiple public clouds. Most of the category's future focus and new development is now in the cloud, and we think this trend is irreversible.
- Vendors are developing and including more sophisticated built-in and integrated machine learning, business intelligence and analysis functions, and other tools. Vendors, including cloud providers, are also integrating their data warehouse products with other elements of their enterprise suites to create overarching unified data platforms. We expect this integration to continue as vendors work to remain competitive.
- Integration with data lakes for in-place querying of structured and unstructured data is becoming more widespread.
- Some vendors are including features that make their platforms more appealing and usable by

business users, including semantic business layers and role-based user experiences, another trend we expect will continue.

This GigaOm Radar Report for Data Warehouses expands on the decision-making framework presented in the Key Criteria report to provide a detailed assessment of vendor solutions and how they can impact an organization.

## HOW TO READ THIS REPORT

This GigaOm report is one of a series of documents that helps IT organizations assess competing solutions in the context of well-defined features and criteria. For a fuller understanding consider reviewing the following reports:

**Key Criteria report:** A detailed market sector analysis that assesses the impact that key product features and criteria have on top-line solution characteristics—such as scalability, performance, and TCO—that drive purchase decisions.

**GigaOm Radar report:** A forward-looking analysis that plots the relative value and progression of vendor solutions along multiple axes based on strategy and execution. The Radar report includes a breakdown of each vendor's offering in the sector.

**Solution Profile:** An in-depth vendor analysis that builds on the framework developed in the Key Criteria and Radar reports to assess a company's engagement within a technology sector. This analysis includes forward-looking guidance around both strategy and product.

## 2. Key Criteria Comparison

Building on the findings from the GigaOm report, “Key Criteria for Evaluating Data Warehouses,” **Table 1** summarizes how each vendor included in this research performs in the areas that we consider differentiating and critical in this sector. The objective is to give the reader a snapshot of the technical capabilities of different solutions and define the perimeter of the market landscape.

Table 1. Key Criteria Comparison

	KEY CRITERIA				
	Optimizations for Concurrency	Security and Governance	Managed Services Offerings	Data Lake Integrations	BI Capabilities and Integrations
AWS Redshift	+++	++	+++	++	++
Azure Synapse	++	++	++	+++	+++
Cloudera	+++	+++	++	+++	+++
Google BigQuery	++	+++	+++	+++	+++
Oracle	++	+++	+++	++	+++
SAP	+	+++	+++	+++	++
Snowflake	+++	+++	+++	–	++
Teradata	+++	+	++	+++	++
Vertica	++	+++	++	+++	+++
Yellowbrick	++	+++	+++	+++	++

+++ Exceptional: Outstanding focus and execution  
 ++ Capable: Good but with room for improvement  
 + Limited: Lacking in execution and use cases  
 – Not applicable or absent

Source: GigaOm 2021

**Table 2** compares the vendors in terms of the evaluation metrics relevant in this sector.

*Table 2. Evaluation Metrics Comparison*

	EVALUATION METRICS				
	Scalability	Autonomy	Extensibility	Deployment Flexibility	Support for Collaboration
AWS Redshift	+++	+++	+++	++	++
Azure Synapse	+++	++	+++	+++	+++
Cloudera	+++	++	+++	+++	+
Google BigQuery	+++	+++	+++	+++	++
Oracle	+++	+++	+++	+++	-
SAP	++	++	+++	+++	++
Snowflake	+++	+++	++	+++	+++
Teradata	+++	++	++	+++	+
Vertica	++	++	+++	+++	++
Yellowbrick	++	++	+++	+++	-

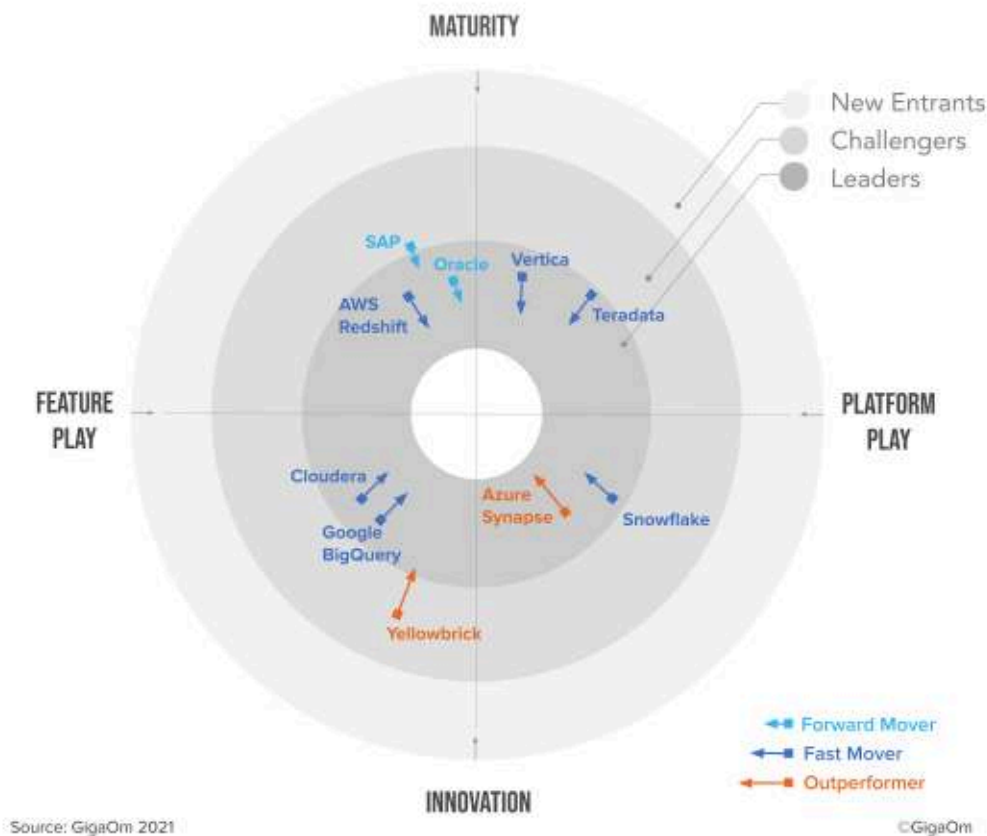
+++ Exceptional: Outstanding focus and execution  
 ++ Capable: Good but with room for improvement  
 + Limited: Lacking in execution and use cases  
 - Not applicable or absent

Source: GigaOm 2021

By combining the information provided in the tables above, the reader can develop a clear understanding of the technical solutions available in the market.

### 3. GigaOm Radar

This report synthesizes the analysis of key criteria and their impact on evaluation metrics to inform the GigaOm Radar graphic in **Figure 1**. The resulting chart is a forward-looking perspective on all the vendors in this report, based on their products' technical capabilities and feature sets.



*Figure 1. GigaOm Radar for Data Warehouses*

The GigaOm Radar plots vendor solutions across a series of concentric rings, with those set closer to center judged to be of higher overall value. The chart characterizes each vendor on two axes—Maturity versus Innovation, and Feature Play versus Platform Play—while displaying an arrow that projects each solution's evolution over the coming 12 to 18 months.

As you can see in the Radar chart in **Figure 1**, the data warehouse space shows many of the characteristics of a mature sector, with several solutions present across multiple quadrants of the Leaders circle. Notably, there is an even distribution of vendors attacking the market from a Maturity and an Innovation standpoint, with a similar parity displayed between the Platform and Feature Play hemispheres of the chart.

For IT decision-makers grappling with a data warehouse decision, the news is both good and bad.



The good news: There are multiple, quality options available, whether your organization seeks access to cutting-edge features and performance or prioritizes the assurance that comes from an established provider working with proven technologies. The bad news? There are multiple, quality options available, and narrowing the field of candidates could prove daunting.

One decision factor is the role of platform leverage in data warehouse solutions. While our chart shows Google BigQuery slightly behind competitors like AWS Redshift and Microsoft Azure Synapse Analytics, the fact is that a shop aligned on Google Cloud Platform likely will value BigQuery more highly. The opportunities for stack integration and skill reuse are often too valuable to pass up.

Ultimately, our review of 10 data warehouse vendors found that Microsoft Azure Synapse Analytics stood out. While it sits today on par with AWS Redshift—and not far ahead of Teradata, Snowflake, Vertica, and Google BigQuery—Microsoft's pace of innovation makes it the lone outperformer in the Leaders circle. And certainly for organizations aligned on Microsoft and Azure, its value is hard to beat.

## INSIDE THE GIGAOM RADAR

The GigaOm Radar weighs each vendor's execution, roadmap, and ability to innovate to plot solutions along two axes, each set as opposing pairs. On the Y axis, **Maturity** recognizes solution stability, strength of ecosystem, and a conservative stance, while **Innovation** highlights technical innovation and a more aggressive approach. On the X axis, **Feature Play** connotes a narrow focus on niche or cutting-edge functionality, while **Platform Play** displays a broader platform focus and commitment to a comprehensive feature set.

The closer to center a solution sits, the better its execution and value, with top performers occupying the inner Leaders circle. The centermost circle is almost always empty, reserved for highly mature and consolidated markets that lack space for further innovation.

The GigaOm Radar offers a forward-looking assessment, plotting the current and projected position of each solution over a 12- to 18-month window. Arrows indicate travel based on strategy and pace of innovation, with vendors designated as Forward Movers, Fast Movers, or Outperformers based on their rate of progression.

Note that the Radar excludes vendor market share as a metric. The focus is on forward-looking analysis that emphasizes the value of innovation and differentiation over incumbent market position.



## 4. Vendor Insights

### Amazon Redshift

Amazon is the pioneer in cloud-native data warehousing. Its Amazon Redshift offering is a cloud-native, fully managed, enterprise data warehousing service.

Amazon Redshift is optimized for high performance through its use of data warehousing must-haves, including columnar storage, massively parallel processing, data compression, results caching, and a query optimizer. ML algorithms help manage performance dynamically by classifying incoming queries based on their run times and resource requirements. Amazon Redshift workload management (WLM) flexibly manages priorities within workloads so that shorter, fast-running queries do not languish behind long-running queries. The solution also automates provisioning, backups, and even table design and optimization.

Users can size their data warehouse according to the query performance required. Different types of nodes accommodate different amounts of data, and users can switch among the types easily as their workloads change. Concurrency scaling allows consistent service levels across nearly unlimited concurrent users and queries.

Amazon Redshift supports connections with many services, including BI tools, ETL, reporting, data, and analytics tools. It also integrates with other AWS services and the AWS analytics ecosystem, resulting in smooth end-to-end analytics. Amazon Redshift allows users to both query and export data to and from their data lakes. Users can query open file formats such as Parquet, ORC, JSON, Avro, and CSV directly in S3 object storage using Redshift Spectrum. Amazon Redshift data sharing, in preview, enables sharing live data across Redshift clusters, improving scalability and agility through instant, granular access to data inside any Redshift cluster without copying or moving it.

The new federated query capability allows querying into operational relational databases such as Amazon RDS and Aurora databases, now enabling users to join data across data warehouses, data lakes, and operational databases. Redshift ML enables users to create, train, and deploy Amazon SageMaker models on their data in Amazon Redshift using SQL, and then use those models for predictions—such as churn detection and risk scoring—directly in their queries and reports.

**Strengths:** Concurrency scaling, automation, elasticity, data lake integration, tool integration, and integration with the AWS data services.

**Challenges:** Amazon pioneered cloud data warehouses and will want to maintain its position even as other offerings continue to grow and develop. The other cloud hyperscalers and Snowflake are becoming exceedingly competitive and Redshift, though hardly standing still, will need to increase its velocity of innovation in order to stay in the top tier.

## Cloudera DW

The Cloudera Data Warehouse (CDW) provides high-performance data warehousing either on-premises or as a cloud service. CDW is a service within the Cloudera Data Platform (CDP), the vendor's framework for enterprise data cloud services.

The Cloudera Data Warehouse is based on open source technologies. CDW is composed of self-service virtual warehouses that can be set up easily with template-based deployments. The virtual warehouses run on data in cloud object stores including Amazon S3, Microsoft ADLS, Apache Kudu, and HDFS. Data is organized into database catalogs, which (potentially multiple) virtual warehouses may access. The resulting auto-scaling, workload isolation, and workload optimization support allows concurrent users and concurrent workloads to be run. Automated suspending and resuming of compute resources lets users pay only for what is actually used. CDW's query processing is powered by SQL engines, including Impala and Hive LLAP. CDW supports deployment on the AWS and Azure clouds, on-premises, or in a hybrid of both.

As one part of the CDP, the CDW integrates with the suite of other Cloudera products and tools. Machine learning and predictive analytics are enabled through integration with Cloudera Machine Learning. Integration with CDP Data Visualization enables users to build and publish custom dashboards and analytic applications. Integrations with Hue and Workload XM provide querying and workload management optimization. The Cloudera Data Warehouse also automatically inherits the same security features from the Cloudera Data Platform.

**Strengths:** Flexible deployment options, automatic scaling capabilities, optimizations to handle concurrency, integration with CDP security features, and CDP's strong built-in analytic, ML, and data visualization tools.

**Challenges:** CDW is not a standalone product but rather a part of the Cloudera Data Platform suite of services, and therefore may not immediately appeal to companies lacking an existing Cloudera infrastructure or looking for a standalone product.

## Google BigQuery

Google BigQuery is a fully-managed, cloud-native data warehouse solution. It is serverless and there is no infrastructure to set up, so resource provisioning is done behind the scenes and administration and maintenance are low-effort, resulting in improved ease of use and cost effectiveness.

For Google BigQuery, a key feature is BigQuery ML, which provides built-in ML capabilities, and allows ML models to be built using simple SQL over structured or semi-structured data directly inside the BigQuery data warehouse. ML and AI capabilities are also available through integration with TensorFlow and Google's Vertex AI. The built-in BigQuery BI engine is an in-memory analytics service that lets users perform complex analytics quickly. The engine integrates with Google Data Studio and Connected Sheets (discussed below), as well as Google's Looker and third-party BI tools via ODBC/

JDBC. BigQuery GIS provides built-in geospatial analysis capabilities, and Connected Sheets allows users to analyze live BigQuery data in Google Sheets without any prerequisite SQL knowledge.

Google BigQuery includes a storage API, which enables the convergence of data warehouses and data lakes. A federated query capability can process data from data sources in object storage for the Parquet and ORC open-source file formats, transactional databases, or spreadsheets in Google Drive. The BigQuery Data Transfer Service allows transfers of data from other Google ecosystem sources, such as Marketing Platform, Ads, YouTube, and partner SaaS applications, as well as from Teradata and Amazon S3. BigQuery also provides integration with the big data ecosystem through Dataproc and Dataflow. Finally, BigQuery offers fine-grained security and governance controls through integration with Google Identity and Access Management (IAM).

Capabilities currently in preview, and expected to arrive soon, include natural language processing with Data QnA and BigQuery Omni, which allows analysis of data across clouds via Google's Anthos technology using standard SQL from within the BigQuery interface.

**Strengths:** A serverless, low-maintenance, easy-to-set-up platform with high scaling capabilities, integration with other Google applications and tools, and the strong built-in capabilities of BigQuery ML and BigQuery BI engines. BigQuery will appeal to users with a wide range of technical backgrounds because of the forthcoming natural language query processor, integration with Google Sheets, and ML capabilities with simple SQL.

**Challenges:** BigQuery's serverless model makes the architecture and management a bit of a black box. For many customers, this may be preferable and provide an easier on-ramp to cloud data warehousing, but data warehouse veterans may find this approach unwieldy. The addition of features that make BigQuery resources more configurable and tunable, on an optional basis, might broaden the appeal of the platform without disrupting individual customers who prefer the low-maintenance approach.

## Microsoft Azure Synapse Analytics

Microsoft Azure Synapse Analytics is a cloud-native platform that integrates data analytics, data lakes, and data warehousing into a single managed solution. It is built on Microsoft's previous data warehouse solution, Azure SQL Data Warehouse, but goes well beyond what that platform offered.

Azure Synapse Analytics is structured with massively parallel processing and columnar storage. Compute resources and storage capacity are separate, allowing each to be scaled independently. A rearchitected query processing framework features a new cloud-native, distributed SQL engine that improves the ability to scale queries depending on the customer's workload needs.

Above this base structure, Azure Synapse Analytics integrates technologies used in enterprise data lake, data integration, and other Azure services to present a unified experience and meet the needs of modern, high-scale data workloads. Integrated Microsoft services include Power BI, Cosmos DB, Azure Machine Learning (Azure ML), Azure Data Share, and Azure Purview. Synapse also offers a

native implementation of the data pipeline functionality in Azure Data Factory.

A new Power BI performance accelerator, in preview, automates the creation and optimization of materialized views to enhance Power BI report performance. Azure Synapse Link for Cosmos DB, also in preview, is a cloud-native, hybrid transactional-analytical processing (HTAP) solution that enables continuous analytics over operational data in Cosmos DB. Integration with Azure ML enables users to apply ML models automatically to Azure Synapse for predictive analytics.

Integration of Apache Spark and Delta Lake with Azure Synapse's new serverless T-SQL engine enables productivity and collaboration among data professionals on the open source and Microsoft T-SQL sides of the industry. In addition to T-SQL, Azure Synapse data can be queried and processed from a variety of programming languages, including Python, Scala, Java, and C#/.NET.

New features also include Synapse Studio—a web-native, unified environment for building complete solutions end-to-end—and Knowledge Center, where users can access pipeline templates, sample scripts for analytics, automation, notebooks, and Azure open datasets. Azure Synapse offers both no-code and code-first experiences for data ingestion, preparation, and transformation.

Azure Synapse offers both serverless and dedicated resource models for different consumption needs. For predictable performance and cost, dedicated SQL pools can be created to reserve processing power for data stored in SQL tables. To accommodate unplanned or fluctuating workloads, always-available serverless SQL clusters can be used as well.

**Strengths:** Strong integrations with other Azure offerings and Apache Spark; serverless and dedicated resource models that accommodate different consumption needs.

**Challenges:** In comparison to the other offerings in this report, Microsoft Azure Synapse Analytics is not as strong in collaborative features or concurrency optimizations. However, we suspect there are several new capabilities, not yet generally available, with the potential to change Microsoft's position here.

## Oracle Autonomous Data Warehouse

Oracle Autonomous Data Warehouse (ADW) is a cloud-native solution that operates on a fully autonomous basis. It is built on the well-known Oracle Database and runs on the Oracle Exadata platform.

ADW includes columnar storage, data compression, and parallel execution of SQL queries. The data warehouse is fully managed, featuring automated elastic scaling, provisioning, configuration, security, tuning, scaling, patching, back-up, and repair. Its automated capabilities aim to increase ease of use, eliminate manual tasks, and reduce the administrative costs of running the platform.

Optimizations for concurrency and multiple workloads include dynamic auto-scaling, which

instantaneously adjusts CPU and I/O resources based on workload requirements. Advanced security and governance features include Database Vault, which provides an extra layer of data protection, obscuring the most sensitive data even from database administrators; and Data Safe, which provides risk assessment, activity auditing, sensitive data discovery, and data masking.

ADW integrates tightly with Oracle Analytics Cloud, providing core BI functionality and tapping into the built-in ML algorithms within the warehouse. Within ADW there is a feature called “analytic views” that provides business-level metadata—hierarchies, dimensions, measures—mapping the metadata layer of Oracle Analytics Cloud onto these views. The ADW platform is also open to integration with third-party tools including Tableau and PowerBI.

The solution also includes built-in self-service data management via a web-based GUI that enables easy loading of data from files, databases, and cloud storage, as well as built-in machine learning through Oracle ML. A Graph Studio console is built in now also, and this integration provides comprehensive tools for working with graphs, including a graph modeling tool to map relational data to graphs, browser-based notebooks for interactive analysis and collaboration, and integrated graph visualization.

Refreshable cloning allows customers to spread the cost of database usage across multiple business units. GoldenGate Capture can capture changes and replicate them to any target database or platform that Oracle GoldenGate supports, including another Oracle Autonomous Data Warehouse. Now users can query data in Hadoop clusters as well.

This solution can be deployed to either a shared or dedicated Exadata infrastructure, which both run on the public cloud. It can be deployed also to Oracle Cloud@Customer, which runs on the customer’s datacenter.

Because of its integration with other Oracle applications, this solution will likely best suit organizations already running other Oracle database and enterprise solutions. As the platform operates autonomously, it may appeal especially to database administrators (DBAs) because it enables them to manage a greater number of implementations and/or frees up their time to devote to higher-value activities. It may appeal as well to organizations that lack dedicated DBA resources.

**Strengths:** The hands-off nature of this solution is its flagship characteristic and a big source of its appeal. Other strengths include its elastic scaling capability, built-in ML and analytics tools, and integration with other elements of the Oracle suite.

**Challenges:** ADW’s basis in Oracle Database and Exadata, and its integration with other elements of the Oracle suite, may lessen its appeal to customers who are not already running other Oracle solutions.

# SAP Data Warehouse Cloud, SAP BW/4HANA & SAP HANA for SQL Data Warehousing

SAP offers multiple data warehousing solutions, with SAP Data Warehouse Cloud as its primary cloud offering. This is a managed cloud-native solution that combines data management capabilities with analytics. SAP BW/4HANA and SAP HANA for SQL data warehousing are the vendor's on-premises data warehouse solutions. SAP Data Warehouse Cloud, SAP BW/4HANA, and SAP HANA for SQL Data Warehousing are based on SAP's in-memory HANA database technology.

SAP Data Warehouse Cloud is a data warehouse-as-a-service solution that is fully managed, cloud native, and designed for both business and technical users. It also helps customers extend their existing data warehousing investment through hybrid cloud deployment. Storage is scalable independently of compute, enabling cost control while supporting changing workload needs. Automated provisioning and resource allocation provide ease of use for administrators.

Users of SAP Data Warehouse Cloud can enable the SAP HANA Cloud machine learning libraries, SAP HANA Automated Predictive Library, and SAP HANA Predictive Analysis Library. A semantic business layer can be applied, which improves the user interface for business users and enables self-service modeling capabilities. Collaboration between business and technical users is encouraged through the use of "spaces," isolated virtual work environments to which users or groups can be assigned access.

SAP Data Warehouse Cloud includes native connection to SAP data sources and allows data stored in non-SAP sources to be queried as well. Users can extract, replicate, or federate data from different sources. The solution includes an integrated graphical editor and SQL capabilities for data integration and modeling, and it also integrates with external visualization tools. Data security features include granular, role-based access controls and data encryption. Real-time data anonymization preserves personal data privacy while allowing the original, non-anonymized data to be used for analytics.

SAP also offers SAP BW/4HANA, which is a packaged data warehouse built natively on SAP HANA. This is mainly an on-premises solution, but cloud deployment is also available. Features include integration with data from SAP and third-party applications, automated administrative tasks, analytical capabilities, and ML-powered workflows.

**Strengths:** Features that appeal to both business and technical users, fully managed services, collaboration features, self-service modeling and analytics, data security and governance features, integration with SAP applications, and native connection to SAP data sources.

**Challenges:** SAP seems to be moving new users toward its SAP Data Warehouse Cloud solution, while still providing standard maintenance support and new releases for existing users of SAP BW/4HANA through 2040. A challenge the vendor will need to address is whether to encourage migration to the cloud solution or commit to supporting on-premises implementations indefinitely.



# Snowflake

Snowflake is a cloud-based data warehousing and analytics platform offered by Snowflake Inc. It presents a powerful, flexible, and elastic solution that allows organizations to load, integrate, and analyze data, with a per-second pricing model for cost management. Snowflake made headlines when it went public in Fall 2020 by having the largest software IPO to date.

Snowflake markets its platform as a single-storage solution for an organization's structured and semi-structured data, thereby obviating the need for a separate data lake. The platform makes use of a multi-clustered, shared data architecture, automatically adding or pausing clusters to support concurrent users and changing workloads. Snowflake includes automatic query caching and optimization. Its data storage is structured with micro-partitions, which are contiguous units of storage that each contain compressed data organized in a columnar fashion.

Compute and storage are separate, and multiple virtual warehouses can be run on the same data. These virtual warehouses can be sized up or down based on service and performance needs. The automation of performance tuning, storage allocation, capacity planning, resource scaling, encryption, backups, and upgrades results in a low-maintenance platform, ostensibly freeing up database administrators' time to spend on higher-value activities.

Snowflake can be run on AWS, Azure, and Google Cloud Platform. Its list of compatible tools and technologies includes an array of well-known machine learning, business intelligence, ETL, and data governance platforms. In addition, Snowflake's built-in data ingestion service, Snowpipe, allows for continuous data ingestion.

A unique feature of Snowflake is its Secure Data Sharing, which enables the sharing of governed data that has been designated as shareable. Data sharing is available via three sources: Direct Share, Snowflake Data Marketplace, and Data Exchange. Direct Share enables account-to-account sharing of data. The Snowflake Data Marketplace provides a single location where users can access a variety of third-party data, query it without transformation, and join it with their own data. Users also can publish their data in the Data Marketplace. The Data Exchange is a centralized hub where users can share data and collaborate with a select group of members. Users invite members and specify whether they may access data, publish data, or both.

**Strengths:** A system architecture and "micro-partitioning" storage strategy that provides strong support for concurrent workloads and multiple virtual warehouses, highly-efficient querying, high elasticity, unique governed data sharing options, and an easy-to-use platform.

**Challenges:** While Snowflake's per-second pricing model provides very good cost effectiveness to accommodate frequently fluctuating workloads, its model may not be preferred by customers with applications featuring more consistent workloads. The addition of a flat pricing option may be required to stay competitive in the long-term.



## Teradata

Teradata made history by being the first company to specialize in massively parallel processing. Today, Teradata Vantage, which grew out of the former Teradata Database, is the vendor's flagship data warehouse and analytics platform.

Teradata Vantage is a highly scalable, unified analytic and data warehouse platform. The architecture consists of an Advanced SQL Engine, an ML engine, a graph engine and a cross-engine orchestration layer that assigns requests to a particular engine as appropriate. Languages supported include SQL, Python, R, and Java. There is a persona-based user experience with available roles such as analyst, developer, data scientist, and business user.

The Vantage App Launcher provides access to Teradata applications and other analytical tools including Editor, where users can run SQL queries, upload data, download results, and browse object stores; Vantage Analyst, where users can create analytic workflows and define rules for repeatable and operationalized processes; and Console, where users can view site information, adjust system elasticity, start and stop the system, and open support incidents. Analytic tools supported include Jupyter, Teradata Studio, and RStudio. Teradata Vantage provides support for external object stores including Amazon S3, Google Cloud Storage, Microsoft Azure Blob storage, and Microsoft Azure Data Lake Storage Gen2.

Teradata Vantage can be deployed in the cloud on AWS, Azure, and Google Cloud Platform. It can be deployed also in hybrid- and multi-cloud environments, on-premises with Teradata IntelliFlex, or on virtualized infrastructure with VMware.

**Strengths:** In-database analytics provided through the capabilities of its multiple analytic engines and through Teradata applications and external analytical tools. Multiple flexible deployment options.

**Challenges:** Teradata was essentially the first data warehouse on the block, specializing in the technology since the early days, and it now must transition into the modern era of cloud data warehouses. While it is well on its way, the remaining work is non-trivial, and carries with it a certain amount of risk.

## Vertica

Now owned by Micro Focus after several acquisitions and mergers, Vertica is a database management company whose data warehouse offering is the Vertica Analytics Platform, a fast and flexible data warehouse and analytics solution.

The Vertica Analytics Platform is optimized for large volumes of data, with the requisite column-oriented storage, data compression, and massively parallel processing. It also possesses built-in advanced SQL-based analytical functions, including geospatial, event series, time series, pattern matching, and regression analytics, among others.

The Vertica Analytics Platform boasts a number of built-in ML capabilities. These include ML algorithms that support classification, clustering, and predictive analytics: linear regression, logistics regression, k-means, naïve Bayes, support vector machines, and random forest. Additionally, ML functions are included for data prep, evaluation, prediction, and supervised and unsupervised learning. The platform offers support for R, Java, C/C++, Python, and SQL for the development of user-defined extensions.

The Vertica Analytics Platform includes integration with a wide array of ETL, BI, visualization, and data transformation products and technologies, including Informatica, Talend, Pentaho, Power BI, Tableau, MicroStrategy, Spark, and Kafka. The platform offers portability across multi-cloud, on-premises infrastructure, and Hadoop clusters, and it provides the ability to query externally stored data: Users can query data directly in either the Hadoop Distributed File System (HDFS) or Amazon S3 object storage. Security integrations include authentication technologies like LDAP and Kerberos, encryption technologies like Voltage SecureData, and compliance with standards such as FIPS.

In a move towards automation, and to help increase ease of use for non-technical users, the platform includes Vertica Database Designer, a tool that helps design “projections” (query-specific table subsets and pre-aggregations) for faster queries. It does so by analyzing the logical schema, sample data, and any sample queries, then creates a physical schema design (a set of projections) that can be deployed automatically or manually.

Cloud deployment options for the Vertica Analytics Platform include Microsoft Azure, AWS, and the Google Cloud Platform. Two options are available to accommodate users’ workloads: Vertica in Enterprise Mode, with its coupled storage and compute resources, is the option that best supports stable workloads and regular queries. Vertica in Eon Mode, with separate compute and storage resources that can be scaled independently, is best for management of dynamic workloads.

**Strengths:** An overall robust platform, the particular strengths of Vertica’s offering lie in its ecosystem integration, built-in analytics and ML capabilities, security integrations, flexible deployment models, and availability across cloud providers.

**Challenges:** Vertica’s offering is moving toward fully autonomous operation, specifically with its option for automatic scaling of compute and storage and with its Vertica Database Designer. However, it still has some way to go.

## Yellowbrick

Yellowbrick offers a data warehouse solution for the distributed cloud. The vendor is a relatively new one in the data warehouse landscape and has defined itself as an enabler of data warehouse modernization, emerging as a strong competitor to more established solutions in the areas of cost, performance, and flexibility.

Yellowbrick Data Warehouse is structured with massively parallel processing and is designed to run on both physical and virtualized infrastructure. On top of this base structure is a PostgreSQL database

interface, intended for compatibility with a broad ecosystem of tools and to ease migration from other legacy platforms. The hybrid cloud nature of the solution allows users the flexibility to natively run mixed workloads on-premises, in private clouds, or on public cloud platforms.

Advanced workflow management optimizes the solution for fast query execution and numerous concurrent users and multiple workloads. BI tool compatibility includes Tableau, Power BI, Microstrategy, and SAS. Data integration tool support includes Informatica and Talend, while multiple data virtualization platforms, including Denodo, are supported. There is support as well for multiple programming languages, including Python and R.

Yellowbrick also works “out of the box,” as the company’s literature puts it, with common BI and ML tools. Security and privacy are emphasized within the offering through “private by default” data management and role-based access control for all database objects. Data lake augmentation is another core differentiator of Yellowbrick’s solution. Yellowbrick supports data ingestion from any data lake source in bulk or through streaming, and it allows users to query data in multiple formats.

The vendor emphasizes the ability of its solution to be deployed across distributed clouds, making it ideal for companies for which centralizing data in a single location is not an option. Deployment options are extremely flexible and include the customer’s data center, private clouds, public clouds, and a distributed cloud architecture. For cloud deployments, private link support is provided for AWS, Azure, and Google Cloud Platform. Yellowbrick Manager is a single, unified point of control for users to manage all their distributed warehouse infrastructure.

Yellowbrick offers a unique three-tier fixed-rate pricing model, with options called Standard, Enterprise, and Enterprise+. As a result of its fixed-rate service plan, the more queries that are run, the lower the cost per query. Its pricing model offers a contrasting alternative to the usage-based pricing offered by many other vendors described in this report.

**Strengths:** Yellowbrick is a comparatively new entrant to the market that has risen in prominence and possesses a strong and robust platform. Differentiating features of Yellowbrick’s data warehouse offering include its wide partner ecosystem, security features, flexibility of deployment, high performance, and fixed-rate subscription pricing model.

**Challenges:** Compared to other offerings in this report, Yellowbrick’s data sharing and collaboration capabilities are not as strong as those of many of its competitors.

## 5. Analyst's Take

This report provides an overview of the fundamentals of data warehouse technologies, highlights key criteria and evaluation metrics to assist consumers in determining the best solution for their needs, examines the leading platforms in the data warehouse landscape today, rates them against the criteria and evaluation metrics, and identifies emerging technologies to keep an eye on in the future.

Data warehousing is an established and mature category, as can be seen by the number of leaders present in the Leaders circle of our GigaOm Radar chart. Still, there is room in the landscape for exciting developments, with vendors constantly seeking to improve their offerings. While the underlying technology—MPP and columnar storage—is solid and mature, vendors are building on their improvements, introducing even faster performance accelerators, more powerful analytics capabilities, and more flexible scaling, while at the same time broadening the scope of the data warehouse by opening it up to ever more integrations.

All the vendors in the landscape provide robust solutions, and all have their differentiating characteristics in which they excel. This report is intended to leverage the decision-making framework developed in the Key Criteria Report for Data Warehouses to help assess and identify the solutions that best fit the needs of your organization.

## 6. About Andrew Brust

Andrew Brust has held developer, CTO, analyst, research director, and market strategist positions at organizations ranging from the City of New York and Cap Gemini to Gigaom and Datameer. He has worked with small, medium, and Fortune 1000 clients in numerous industries and with software companies ranging from small ISVs to large clients like Microsoft. The understanding of technology and the way customers use it that resulted from this experience makes his market and product analyses relevant, credible, and empathetic.

Andrew has tracked the Big Data and Analytics industry since its inception, as Gigaom's Research Director and as ZDNet's lead blogger for Big Data and Analytics. Andrew co-chairs Visual Studio Live!, one of the nation's longest-running developer conferences. As a seasoned technical author and speaker in the database field, Andrew understands today's market in the context of its extensive enterprise underpinnings.

## 7. About GigaOm

GigaOm provides technical, operational, and business advice for IT's strategic digital enterprise and business initiatives. Enterprise business leaders, CIOs, and technology organizations partner with GigaOm for practical, actionable, strategic, and visionary advice for modernizing and transforming their business. GigaOm's advice empowers enterprises to successfully compete in an increasingly complicated business atmosphere that requires a solid understanding of constantly changing customer demands.

GigaOm works directly with enterprises both inside and outside of the IT organization to apply proven research and methodologies designed to avoid pitfalls and roadblocks while balancing risk and innovation. Research methodologies include but are not limited to adoption and benchmarking surveys, use cases, interviews, ROI/TCO, market landscapes, strategic trends, and technical benchmarks. Our analysts possess 20+ years of experience advising a spectrum of clients from early adopters to mainstream enterprises.

GigaOm's perspective is that of the unbiased enterprise practitioner. Through this perspective, GigaOm connects with engaged and loyal subscribers on a deep and meaningful level.

## 8. Copyright

© [Knowingly, Inc.](#) 2021 "*GigaOm Radar for Data Warehouses*" is a trademark of [Knowingly, Inc.](#). For permission to reproduce this report, please contact [sales@gigaom.com](mailto:sales@gigaom.com).