



NUCLEUS
RESEARCH

DATA WAREHOUSE TECHNOLOGY VALUE MATRIX 2022

ANALYST

Alexander H. Wurm

THE BOTTOM LINE

Organizations of all sizes continue to embrace large-scale cloud transformations to become more agile, resilient, and data-driven. The adoption of modern data warehouse technology has accelerated in parallel as companies increase the scale of data ingest and analytics. Data warehouse performance and flexibility have also improved as open-source software frameworks like Kubernetes and Apache have matured through vendor investment, enabling big-data and in-memory analytics. As we progress in 2022, Nucleus anticipates that customers will prioritize a solution's respective scalability, serverless capabilities, multi-region and multi-cloud completeness, and performance optimization for big-data and highly parallel tasks when selecting a data warehouse solution.



OVERVIEW

Data warehouse technology has served as the backbone of most modern cloud applications by promising constant availability alongside data durability and correctness. Although data warehouses typically bring to mind large archaic units of specialized hardware, the mass vendor adoption of open-source technologies, such as Kubernetes, Apache Hadoop, and Apache Spark, have enabled multi-tenant elastic services yielding incredible flexibility and

scalability. Now, data warehouse technologies operate at the confluence of data science and networking innovation by housing an organization's data in a flexible cloud-enabled global network while maintaining guarantees of availability, data durability, and correctness.

Nucleus defines a data warehouse as a solution that extracts and consolidates data from various data sources, including transaction, operational, and external systems to support various data-intensive SaaS applications. Select data platform and data lake offerings that meet these criteria and regularly compete with data warehouse solutions are also considered in this Value Matrix. The data warehouse market includes solutions of various scopes and specialties for use cases across select environments or data use-cases. Nucleus has witnessed organizations of all sizes increasingly adopt data warehouse solutions: SMBs favor cloud-native solutions with intuitive experiences, secure data sharing, and fully managed services to accelerate time-to-insight; Large enterprises prioritize the most performant platforms optimized to drive low latency with in-memory capabilities for real-time insights and petabyte-scale data processing leveraging massively parallel processing (MPP) architecture for big data tasks.

Over the past year, vendors have focused development in five primary areas to differentiate their data warehouse offerings: Efficient cluster provisioning and assignment leveraging proprietary ML algorithms, fully managed and serverless capabilities to reduce administrative oversight and associated costs, multi-cloud and multi-region completeness to maximize customer flexibility and reduce the risk of cloud and region-specific outages, and service offerings to streamline implementations and mitigate misconfigured deployment risks (Nucleus Research W8 – *Database and data warehouse trends for 2022* – January 2022).

Because data warehouses function as a single source of truth for an organization's data, data warehouses are also at the forefront of innovation for machine learning (ML) and artificial intelligence (AI) - based analytics and modeling. As such, vendors compete to deliver self-service data-science capabilities with select vendors boasting native integration with specialized data science services and others offering collaborative services where users can securely share third-party data sets and notebooks to accelerate analytics and model development.

In this Value Matrix, vendors are positioned according to the relative usability and functionality of their respective solutions, as well as the value that customers realized from each product's capabilities (Nucleus Research V67 – *Understanding the Value Matrix* – April 2021) and presented as a snapshot of the current market rather than an empirical ranking of vendors. The arrows indicate perceived momentum in the indicated direction with respect to usability and functionality. Positioning and momentum are informed primarily by conversations with end-users, along with the most recently released capabilities/features and areas of vendor investment.

LEADERS

Leaders in this year's Value Matrix include Amazon Redshift, Microsoft Azure Synapse Analytics, Oracle Autonomous Data Warehouse, and Teradata Vantage.

AMAZON REDSHIFT

Amazon Redshift is a leader in the 2022 edition of the Data Warehouse Technology Value Matrix. Amazon Redshift can analyze structured and semi-structured data stored across data warehouses, data lakes, operational databases, and third-party data sets. As a cloud-native data warehouse, Amazon Redshift is known for its ease of use and performance at scale with widespread data accessibility to promote visibility and fast implementations to accelerate return on data. Customers leverage Amazon Redshift to shorten time-to-insight with simplified, managed data access, ingest, and egress, enhancing developer and analyst productivity.

Over the past year, Amazon delivered various new capabilities to extend the value of its data warehouse solution.

- In March 2021, Amazon announced the general availability of Amazon Redshift Data Sharing. Amazon Redshift data sharing lets customers extend the ease of use, performance, and cost benefits that Amazon Redshift offers in a single cluster to multi-cluster deployments while being able to share data. It enables granular and fast data access across Amazon Redshift clusters without the need to copy or move data around. Customers can securely share live data with Amazon Redshift clusters in the same or different AWS accounts within the same region or across regions.
- In May 2021, Amazon integrated Redshift ML with Amazon SageMaker to create, train, and deploy machine learning models using SQL statements for use cases such as churn prediction modeling and fraud risk scoring within user queries and reports. Amazon Redshift ML can also leverage SageMaker Autopilot to automatically discover and tune the appropriate regression, binary, or multi-class classification model. Users can also select specific models, including Xtreme Gradient Boosted tree or Multilayer perceptron, and edit model hyperparameters.
- In September 2021, Amazon launched Amazon Redshift Query Editor V2 which is a web-based tool for data analysts, data scientists, and database developers to explore, analyze, and collaborate on data in Redshift data warehouses and data lake. Users can use Query Editor's visual interface to create and browse schema and tables, load data, author SQL queries and stored procedures, and visualize query results with charts.

- In October 2021, Amazon extended its collaborative capabilities with AWS Data Exchange, enabling users to accelerate time to insight by subscribing to third-party data sets to accelerate the development of queries and analytics. Additionally, AWS Data Exchange integration also provides entitlement, billing, and payment management to help data providers monetize their data.
- In November 2021, Amazon introduced Amazon Redshift Serverless, which automates cluster provisioning and scaling, enhancing IT productivity while driving cost efficiency with usage-based pricing.

Looking ahead, Nucleus expects Amazon Redshift to gain market share as Amazon further enhances the intuitiveness of its machine learning capabilities and Redshift Serverless comes out of preview.

MICROSOFT AZURE SYNAPSE ANALYTICS

Microsoft Azure Synapse Analytics, formerly Azure SQL Data Warehouse, is a leader in the 2022 Data Warehouse Technology Value Matrix. With end-to-end capabilities for data integration, data warehousing, and big data analytics, Azure Synapse Analytics offers a managed elastic data platform for analytics over public, operational, and historic data. Azure Synapse customers can tailor their analytics infrastructure to their processing needs with dedicated and serverless, usage-based pricing options. Azure Synapse Analytics also boasts native integration with Azure Machine Learning, which also employs a usage-based pricing model. This managed environment extends to the platform's integration with Apache Spark and eliminates the complexity of cluster provisioning and tuning. Azure Synapse Analytics uses an enhanced shuffle service with its implementation of Apache Spark to align data and improve query performance. Azure Synapse further reduces processing latency with dynamic partition pruning to remove unnecessary data during task execution. Microsoft's cloud-native hybrid transaction/analytical processing (HTAP) solution, Azure Synapse Link, drives further speed improvements, enabling continuous analytics over operational data in Azure Cosmos DB without interfering with operational or application workloads.

To accelerate time-to-insight, customers can also leverage the Synapse Studio Knowledge Center to link sample datasets and add code in the form of sample SQL scripts, notebooks, and pipelines. Customers can expect a faster return on data as the Knowledge Center expands with more pipeline, SQL script, and notebook templates. Looking ahead, Nucleus expects further adoption of Microsoft PowerBI and Azure Synapse Analytics as the vendor develops and integrates these AI-based data management and analytic capabilities into the vendor's cloud application suite.

ORACLE AUTONOMOUS DATA WAREHOUSE

Nucleus recognizes Oracle Autonomous Data Warehouse as a leader in the 2022 edition of the Data Warehouse Technology Value Matrix. Autonomous Data warehouse is available for deployments in Oracle's cloud and on-premises in customer data centers. As its name may imply, the Autonomous Data Warehouse offloads the operational complexities involved in managing a data warehouse, including provisioning, configuring, securing, performance tuning, patching, and autoscaling processes. This automation enables customers to run various analytics tasks without downtime, human intervention, or over-provisioning. The Oracle Autonomous Data Warehouse is also the only solution that uses a converged database providing built-in support for multimodal data and multiple parallel workloads. Customers can further leverage Oracle's Hybrid Columnar Compression to achieve maximal data storage savings. Furthermore, with Oracle's no-code self-service tools for data loading, data transformation, and data insights, business analysts, data scientists, and developers can accelerate common workloads and improve productivity. With Oracle's in-database algorithms and model development leveraging an AutoML UI, users can easily analyze and model data at its source. Oracle complements these functionalities with pervasive capabilities for data security, including data encryption at rest and in motion, regulated data protection, security patching, and threat detection. For additional security, customers can use Oracle Data Safe to conduct continuous monitoring over sensitive data and privileged user access.

Oracle plans to use its Autonomous Data Warehouse and Autonomous Database solutions to connect analytics across the vendor's suite of cloud applications (Nucleus Research W2 – *Oracle apps unite via AI-driven analytics* – January 2022). To deliver on this cross-departmental and cross-functional vision, Oracle continues to pour resources into the Autonomous Database to differentiate Oracle Analytics. As this AI-based data management and analytics becomes developed and integrated into the vendor's cloud application suite, Nucleus expects increasing adoption with especially strong consideration in competitive deals for organizations with high processing volumes.

TERADATA VANTAGE

Teradata is a leader in the 2022 edition of the Data Warehouse Technology Value Matrix. Teradata is an enterprise-grade data warehouse with mature workload management, integrated advanced analytics, security, and data access components. Teradata delivers seamless access to data across multi-cloud and hybrid environments with support for AWS, Google, and Microsoft public clouds. Teradata Vantage is also available on virtualized software using VMware and with Teradata's IntelliFlex hardware. In providing this cohesive approach, customers can manage multiple deployments as a single ecosystem, minimizing costs associated with data movement or replication across environments. Customers can

easily migrate between deployment and pricing options with license portability to ensure flexibility and avoid cloud vendor lock-in. With an architecture designed for multidimensional linear scalability, Teradata enables enterprise customers to handle massive data workloads without any loss to performance. Teradata provides the ability to manage performance with advanced workload management software to meet SLAs rather than scaling performance with hardware at additional and often unpredictable cost.

Teradata Vantage is a full analytics platform featuring advanced analytics, graph and machine learning engines, support for R and Python, Vantage Analyst, and the ability to bring your own model to simplify access to complex modeling and analytics. Teradata Vantage is also a highly integrated analytics platform, enabling customer organizations to utilize and operationalize their analytics simply within their enterprise applications. By integrating with various SaaS applications and BI systems, Vantage enables reporting, static queries, ad hoc analysis, active data feeds, operational data warehousing, and complex analysis in one environment.

Over 2021, Teradata delivered various new capabilities and strategic partnerships to enhance the value of Teradata Vantage including:

- BigQuery connectivity to QueryGrid allowing users to query directly from Teradata Vantage to Google BigQuery.
- AnalyticOps to productize, manage and govern any Vantage Analytic Model whether it be R, Python, or micro-models based upon SCRIPT table operator, models or micro-models from Vantage Analytics Library or in-Database SQL-MR, and models built externally to Vantage via Bring Your Own Model (BYOM).
- Cliquing in the Azure Cloud for new as-a-service deployments to significantly reduce downtime from node failures. Teradata also extended this strategic partnership with Microsoft to integrate Teradata with Power BI.
- Cloud partner network expansion featuring over 40 new partners including Alteryx, Dataiku, Domino Data, dotData, H2O.ai, IBM-Cognos, MicroStrategy, SAS, and Tableau.
- Plugins for Dataiku allowing its users to create full end-to-end workflows using Vantage in-database R & Python, Teradata's Advanced SQL Engine, or Dataiku's native analytic modeling capabilities.

Nucleus expects Teradata to improve its placement in future editions of the Data Warehouse Technology Value Matrix as the vendor continues to invest and develop cloud capabilities to pair with its strong on-premises following. Furthermore, Nucleus expects increasing adoption of Teradata Vantage as organizations realize the relative cost inefficiency of serverless and usage-based pricing models offered by cloud native data warehouse vendors who prioritize scalability over resource efficiency.

EXPERTS

Experts in this year's Value Matrix include Databricks Lakehouse, IBM Db2 Warehouse, and Yellowbrick Data.

DATABRICKS LAKEHOUSE

Databricks is placed as an expert in the 2022 edition of the Data Warehouse Technology Value Matrix as recognized for the Databricks Lakehouse Platform. The Databricks Lakehouse Platform offers the data management and performance capabilities of data warehouses with the flexible object stores of data lakes. This platform also simplifies analytics infrastructures by bringing analytics, data science, and machine learning workloads together under one cohesive environment. A key differentiator of the Databricks Lakehouse involves its Delta Lake foundation which provides a unified ecosystem for management and governance of structured, semi-structured, and unstructured data optimized for cost efficiency and scalability. Customers can also leverage Delta Lake to perform open protocol secure data sharing with Delta Sharing. This significantly enhances collaboration within the user ecosystem as organizations can securely share data assets across their partner network. Because of Databrick's focus on open source and open protocol, customers can benefit from a collaborative developer ecosystem and maintain flexibility with escape hatches to avoid vendor lock-in. With the vendors DNA in Apache Spark, Databricks is well suited for large scale data analytics with robust capabilities for data engineering and automated machine learning capabilities leveraging Managed MLflow. Databricks also features "glass box" AutoML functionality allowing users to quickly generate baseline models and notebooks with a low-code approach.

Databricks has continued to invest in its Data Lakehouse Platform over the past year.

- In August 2021, Databricks announced \$1.6 billion in Series H funding to accelerate adoption of data lakehouse architectures.
- In October 2021, Databricks acquired German data science startup 8080 labs to enable rapid and intuitive data exploration and transformation with the startup's low/no-code data science tool, bamboolib.
- In January 2022, Databricks launched its data lakehouse for retail and consumer goods to address common problems for customers in those industries.

Nucleus anticipates Databricks will capture market share from other data warehouse vendors for clients prioritizing data science functionality. As Databricks improves the intuitiveness of its data science with further investments into low/no-code capabilities and industry-specific data lakehouses, customers can expect increasing value from their Databricks deployments as represented by improved positioning along both the usability axis.

IBM DB2 WAREHOUSE

IBM Db2 is placed as an expert in the 2022 edition of the Data Warehouse Technology Value Matrix. With its Db2 Warehouse offering, IBM delivers a preconfigured data warehouse deployable to private clouds and other Docker-container-based infrastructure. IBM Db2 Warehouse is ideal for customers enlisting hybrid cloud approaches as it pairs the flexibility of a cloud environment with thorough data and security controls. This focus enables customers to progressively transition sensitive data to the cloud with increasing consideration in competitive deals as the security requirements and complexity of the customer's cloud transformation scale. IBM Db2 Warehouse boasts a wide range of functionality with support for both SMP and MPP system designs, built-in analytics and machine learning, and automated elastic scaling. IBM Db2 Warehouse is also highly integrated with Apache Spark, further streamlining data analytics and machine learning at scale.

Over the past year, IBM has delivered various new and extended capabilities to the Db2 Warehouse to enhance customer value.

- In July 2021, IBM integrated Db2 Warehouse with IBM Key Protect to help users manage their keys, improving control over data encryption within the system.
- In September 2021, IBM released a panel detailing the status of activity, locking, utility, and statistics event monitors for a given database connection. IBM also added support for workload customization on activity and statistic event monitors.
- In January 2022, IBM extended support for indexes, table spaces, and buffer pools and increased storage capacity to 240 TB.

With this focus on data controls and security IBM has built confidence in its cloud and Db2 Warehouse, unlocking a new area of cloud migrations traditionally limited by data security requirements or risk thresholds. Furthermore, IBM Db2 Warehouse's differentiated commitment to hybrid cloud customers allows the vendor to build a moat around its position as the leading provider of solutions and services geared towards hybrid cloud approaches.

YELLOWBRICK DATA

Yellowbrick Data is an expert in this year's edition of the Data Warehouse Technology Value Matrix. Yellowbrick Cloud Data Warehouse leverages PostgreSQL alongside MPP architecture to provide real-time insights from data at any scale. Yellowbrick markets itself as a Data Warehouse for Distributed Data Clouds and, as such, can be deployed anywhere, including private data centers, public clouds, and container environments like Kubernetes at the edge. Yellowbrick serves a diverse array of industries, including financial services,

insurance, telecommunications, manufacturing, healthcare, and retail. Yellowbrick manages data in a hybrid row and column store to provide optimal performance for low-latency tasks and ad-hoc queries. Yellowbrick further appeals to customers leveraging private cloud environments with its proprietary Andromeda architecture, featuring AMD 64-core processors and “Kalidah” scan accelerators to optimize performance and resource consumption efficiency. All data warehousing is managed by Yellowbrick Data Manager, a “single pane of glass” for unified data control across distributed cloud environments. Yellowbrick Data Manager also allows users to monitor deployments, browse and load data through a web browser, and write/edit SQL. Yellowbrick customers noted several key drivers of adoption, including the platform’s capabilities for live processing of streaming data and efficient batched processing, and flexibility between cloud and on-premises usage.

Throughout the past year, Yellowbrick enhanced customer value with new partnerships and investments into new capabilities.

- In April 2021, Yellowbrick Data introduced its data warehouse designed for distributed clouds to support emerging IoT and edge analytics use cases while seamlessly connecting private data centers and public clouds.
- In November 2021, Yellowbrick Data announced series C1 funding of \$75 million to accelerate the enterprise adoption of its cloud data warehouse and fund innovation in a highly iterative space.
- In February 2022, Yellowbrick Data announced its partnership with Nippon Information and Communication Corporation (NI+C) in Japan to continue global expansion, as well as to underscore the value of NI+C’s data to serve telecommunication, transportation, and banking institutions.

Nucleus expects Yellowbrick Data to improve its positioning in future editions of the Data Warehouse Technology Value Matrix as Yellowbrick Data becomes further integrated with various public cloud providers. Nucleus also anticipates customer adoption to accelerate as Yellowbrick Data begins offering serverless pricing options to extend the competitive advantage of Yellowbrick Data’s differentiated architecture and networking technologies.

FACILITATORS

Facilitators in this year’s Value Matrix include Google BigQuery, SAP Data Warehouse Cloud, and Snowflake.

GOOGLE BIGQUERY

Nucleus places Google BigQuery as a facilitator in the 2022 edition of the Data Warehouse Technology Value Matrix. With Google BigQuery, users can analyze structured and semi-structured data using SQL, leveraging a serverless approach within a fully managed multi-cloud environment. Google BigQuery serves a variety of industries, including retail, manufacturing, automotive, energy, healthcare, media, telecommunications, government and public sector, and financial services. Customers can ingest data into BigQuery through various modalities, including batch loading, data streaming, and live connectors with third-party applications. Customers can also run advanced analytics leveraging BigQuery ML, including regression, K-means clustering, matrix factorization, gradient boosted decision trees, and deep neural networks. Google pairs these capabilities for petabyte-scale analytics with BigQuery BI Engine's capabilities for in-memory analytics to enable latency-optimized query processing integrated with popular BI solutions, including Looker, Tableau, Google Data Studio, and Power BI. Google BigQuery also supports geospatial analytics with BigQuery GIS allowing users to integrate location data for a more comprehensive understanding on how geography impacts operations. Google BigQuery is central to Google's vision of supporting customers' Data Clouds by unifying their data lifecycle across databases, data analytics, BI, and AI use cases.

Throughout 2021, Google delivered various new and extended functionalities to enhance the value of BigQuery.

- In May 2021, Google announced Analytics Hub for data access and sharing. Google also announced Dataplex, an intelligent data mesh that unifies data management and governance across customer data lakes, data warehouses, and data marts.
- In July 2021, Google made BigQuery Omni generally available, allowing customers to perform multi-cloud analytics without having to move or replicate data stored within AWS S3 or Azure blob storage.
- In August 2021, Google introduced Cloud Spanner federated queries with BigQuery allowing users to query data held within Spanner in real-time without the need for data replication or new ETL pipelines to move data.
- In October 2021, Google announced support for Apache Spark on Google Cloud featuring serverless autoscaling to accelerate big data analytics, deliver usage-based pricing and save IT time previously spent managing Spark clusters.
- In January 2022, Google announced the general availability of Explainable AI in BigQuery to help users understand the results of their machine-learning models.

- Google also announced support for new governance capabilities, such as field-level encryption, SQL workload management capabilities, and automatic data loss prevention (DLP) with Google Cloud DLP.

Looking ahead, Nucleus expects improved positioning of Google BigQuery, as the vendor continues to extend data science and governance capabilities alongside investment into new networking technologies and secured lakehouse architecture.

SAP DATA WAREHOUSE CLOUD

SAP Data Warehouse Cloud is placed as a facilitator in the 2022 edition of the Data Warehouse Technology Value Matrix. With SAP Data Warehouse Cloud, the vendor delivers a multi-cloud solution with data integration, database, data warehouse, and analytics capabilities built on the SAP HANA Cloud. SAP represents a comprehensive data warehouse solution with capabilities for governance, data modeling, business modeling, repository, and security services. SAP offers multiple on-premises, hybrid, and cloud deployment options with a choice of public and private environments for cloud deployments. Customers can also connect these environments across multi-cloud and on-premise data repositories to preserve business context across systems. Customers also gain intuitive ease-of-use with SAP's virtual workspace and no-code environment for data connection, modeling, visualization, and secure sharing. To enable real-time analytics, SAP supports in-memory processing of structured, unstructured, and geospatial data. SAP also offers a pre-integrated database, data warehouse, data lake, and analytics to accelerate implementations. SAP Data Warehouse Cloud is also a fully managed service minimizing ongoing complexity for business and IT post-deployment.

SAP continued to develop new functionalities for its SAP Data Warehouse Cloud throughout the past year, including:

- SAP BW Bridge, allowing users to import BW Bridge data models into SAP Data Warehouse Cloud and benefit from other BW functionality within a flexible public cloud environment.
- Intelligent Lookup to join any two entities with some sort of semantic relationship using a no-code UI.
- Extended flexibility in the Data Flow to import excel files from remote file storage and configure the worksheet to be used.

Nucleus anticipates SAP Data Warehouse Cloud to continue to grow, leveraging the platform's integrated capabilities to power hybrid and complex multi-cloud environments. Looking forward, we expect SAP to further invest in the usability of its Data Warehouse Cloud with capabilities for serverless pricing and secure data sharing.

SNOWFLAKE DATA CLOUD

Nucleus recognizes Snowflake as a facilitator in the 2022 edition of the Data Warehouse Technology Value Matrix. Snowflake sets its ambitions beyond multi-cloud, aspiring for a fully cross-cloud data platform. Snowflake achieves this by leveraging Secure Data Sharing to manage data access through Snowflake's unique services layer and metadata store to provide internal departments, customers, and business partners with a read-only database containing shared data that does not contribute to a client's data storage charges. This capability also extends to live governed data enabling collaborative insights in real-time. Snowflake Data Cloud offers a completely serverless architecture allowing customers to easily scale up or scale out their data warehouse instances and pay based on usage. These serverless capabilities also save IT time by internally managing capacity planning, resource provisioning, and server maintenance. Although this serverless approach grants usage-based pricing, Snowflake does not scale computational resources in linear increments, rather opting for exponential increments (2^n). For example, if a compute task required five GPUs, Snowflake would assign the user eight GPUs. Snowflake's usage-based pricing model can lead customers to accrue usage costs that exceed the customer's resource utilization.

Snowflake has continued to invest in its platform's ease of use and functional depth. Recent product updates and announcements include:

- Support for International Traffic in Arms Regulations (ITAR) Compliance on Microsoft Azure Government and AWS GovCloud, improving Snowflake's appeal to customers operating within regulated industries.
- In September 2021, Snowflake announced a strategic partnership with the Python-based data science platform Anaconda to extend access and management of open-source libraries. The vendor also announced native support for Python within Snowflake as a part of the vendor's developer framework, Snowpark.
- In September 2021, Snowflake introduced the Financial Services Data Cloud to facilitate data-driven innovation amongst financial service institutions.
- In October 2021, Snowflake launched its Media Data Cloud for facilitating data collaboration in media and advertising user networks.

Looking ahead, Nucleus anticipates that Snowflake will gain traction amongst customers in regulated industries looking to modernize while maintaining high levels of data security. As Snowflake further enhances the intuitiveness of its data science capabilities with specialized data clouds, customers can expect increasing value from their Snowflake deployments as represented by improved positioning along both the usability and functionality axes.

CORE PROVIDERS

Core Providers in this year's Value Matrix include the Cloudera Data Platform and the Panoply Cloud Data Platform.

CLOUDERA DATA PLATFORM

The Cloudera Data Platform (CDP) is placed as a core provider in the Data Warehouse Technology Value Matrix. The Cloudera Data Platform provides data management and security throughout the data lifecycle and is available for public, private, and multi-cloud deployments. By integrating these environments, Cloudera offers a "single pane of glass" to manage multi-cloud and hybrid environments. Cloudera Data Platform is available in two editions, CDP Public Cloud and CDP Private Cloud. The CDP Public Cloud allows users to create and manage secure data lakes and provides self-service analytics and machine learning services without the need to install and manage the data platform software. Catering to hybrid cloud customers, CDP Private Cloud seamlessly connects on-premise environments to public clouds featuring consistent, built-in security and governance. Cloudera is an elastic service with full self-service provisioning and administration to enable zero-touch administration that automatically scales and suspends with respect to client usage to control cloud costs. Cloudera provides sub-second query response times, with SQL engines like Apache Impala and Hive capable of handling datasets of up to 150PB. Cloudera supports semi-structured and unstructured data in addition to traditional structured datasets. Cloudera boasts plug and play compatibility for various third-party analytical tools to connect data and power the BI tools the company already uses. CDP Data Engineering also offers sophisticated capabilities for data engineering, built on Apache Spark to streamline ETL and facilitate orchestration and automation with Apache Airflow. Apache Airflow also grants Cloudera customers pipeline monitoring, visual debugging, and management tools.

- In April 2021, Cloudera announced a strategic partnership with NVIDIA to integrate the RAPIDS Accelerator for Apache Spark 3.0 with the Cloudera Data Platform.
- In June 2021, Cloudera acquired Datacoral and Cazena to expand the vendor's push into no/low-code self-service analytics. Datacoral streamlines ETL with fully managed services for data transformations and integrations, while Cazena accelerates AI/ML-based analytics.
- In August 2021, Cloudera announced an agreement to be acquired by Clayton, Dubilier & Rice (CD&R), and KRR in a deal worth \$5.3 billion to take Cloudera private. This move provides Cloudera with the resources to accelerate the innovation of Cloudera's hybrid cloud platform.

- Also, in August 2021, Cloudera introduced Cloudera DataFlow for the Public Cloud to process hybrid streaming workloads. Users can now automate complex data flow operations and improve resource efficiency with auto-scaling capabilities.

Nucleus expects continued adoption of CDP as customers increasingly look to Cloudera for integrated capabilities for hybrid and complex multi-cloud environments. We anticipate Cloudera will improve along both axes as the vendor develops extended cloud-native capabilities.

PANOPLY

Panoply is recognized as a core provider in the 2022 Data Warehouse Technology Value Matrix. Panoply provides fully managed ETL and cloud data warehousing to analyze big data using standard SQL. The Panoply automated data pipeline service offloads configuration, software updates, and infrastructure scaling reducing complexity for loading data into Panoply and enabling users to connect files, databases, and APIs without code. The Panoply Cloud Data Platform features connectors with over 40 data sources, including Amazon S3, PostgreSQL, Google Analytics, HubSpot, MongoDB, Shopify, and Salesforce. Panoply also supports a broad range of BI and data visualization tools, such as Count, Tableau, Power BI, Looker, cumul.io, Mode Analytics, Panintelligence, and Metabase. The Panoply Cloud Data Platform also features security and governance capabilities such as SOC 2 and HIPAA compliance to handle sensitive data and table-level user permissions for fine-grained control.

In December 2021, Panoply was acquired by SQream Technologies, a GPU-centric data acceleration platform. This acquisition should speed up SQL queries for Panoply customers and help deliver on SQream's vision of rapid analytics across cloud, on-premises, and edge environments. As SQream integrates its hybrid analytics platform with the Panoply Cloud Data Platform, Nucleus expects enhanced value for users of both products.