A Buyer's Guide to Data Warehousing in the Cloud

EMA™

IT AND DATA MANAGEMENT
RESEARCH | INDUSTRY ANALYSIS | CONSULTING

Table of Contents

EMA

IT AND DATA MANAGEMENT
RESEARCH | INDUSTRY ANALYSIS | CONSULTING

EXECUTIVE SUMMARY

The digital revolution has spawned a rapid expansion of cloud data warehousing. In the cloud, there are two viable options for organizations moving toward analytic dominance: data warehouse as a service (DWaaS) and bring your own license (BYOL) to the cloud. Hyper-marketing around DWaaS has swayed opinions, but wise leaders are carefully considering both viable options before making a decision. This paper explores cloud data warehouse options to assist leaders in their selection of the right technology.

AN INTRODUCTION TO CLOUD DATA WAREHOUSING

Digital transformation changes the requirements for organizations pursuing insight-driven strategies. Mobile applications, the Internet of Things (IoT), and digital customer interaction are producing new types of data at unfathomable rates. Audience engagement and personalization push the need for data and analytics to real time. The speed of innovation puts the rate of change into overdrive.

The traditional data warehouse and cloud data warehouse (CDW) were both originally built as technical solutions to address the need to understand static, historical data. This focus required data processing for quality and consistency and data structuring for analysis across many different dimensions. A look at the history of cloud data warehousing reveals that the days of the rigid, monolithic data warehouse are gone.

Traditional Data Warehousing

Traditional data warehouses were built on software-only solutions, run on self-owned hardware, and managed by skilled workers employed or contracted by organizations. This was a break from the days when reports were run on operational systems and required new skillsets for designing, building, and operating these new data stores. While this on-premises approach required a concerted effort, initially, there were no other options.

Appliances and Managed Services

In the pre-cloud era, organizations were deploying their data warehouses on appliances, specialized hardware, and software combinations. Appliance vendors had heavy lock-in since their data warehouses wouldn't work on any other hardware. They offered their solutions with a managed service, enhancing lock-in by ensuring only their service people understood how to keep the appliance running. To the owners of the data warehouse, the appliance was largely a black box.

Cloud Data Warehousing (CDW)

A critical area of modernization is the move to the cloud. In the most recent EMA big data research, 87.4% of respondents indicated that they were adopting cloud implementation strategies. There are a few different types of data warehousing in the cloud, mainly bring your own license (BYOL) to the cloud and DWaaS.

Bring Your Own License (BYOL)

With growing cloud popularity, many organizations purchased infrastructure in the cloud and run their data warehouse with software licensed from on-premises data warehouse platform vendors. This approach has proven popular because organizations can move to the cloud and retain all the expected enterprise data warehouse functionality of their preferred software technology. Not all data warehouse vendors offer this option. It removes much of the pain of infrastructure lock-in, allowing users considerable flexibility in deployment options.

Data Warehouse as a Service

DWaaS appeared in the early 2010s and gained market share quickly with basic cloud functionality like rapid provisioning, elasticity, and subscription pricing. Market adoption soared, but first-to-market solutions lacked performance and functionality for the emerging data landscape and analytic tsunami. New DWaaS platforms, built specifically for the cloud, appeared in the mid-2010s with a focus on making adoption simpler for business users.

Because DWaaS focuses on quick implementation and ease of use, business users can define and build their own analytic solutions and, in some cases, get them up and running without IT. However, many enterprise data warehouse features that long-time users expect have been eliminated to provide a simpler user experience.

Making the Right Decision

Adoption of analytics technologies in the cloud continues to accelerate. Innovations in both BYOL and DWaaS offerings include features like automated indexing and the creation of structure-on-read, which provides support for semi-structured data. While all the marketing hype and momentum points toward the DWaaS option, EMA believes that a move to cloud data warehousing requires careful consideration.

It is important to understand an organization's business, technical, and financial requirements to make the right selection. One size does not fit all. DWaaS, while well-suited to many use cases, does not meet all the needs of every organization. The current marketplace includes viable options for both DWaaS and BYOL options.

WHAT COMPANIES WANT IN A CLOUD DATA WAREHOUSE

SaaS, mobile, and the cloud create new expectations for business professionals who are used to applications that are intuitive, boost productivity, and respond quickly. To meet these emerging expectations, insight-driven organizations need a data warehouse that provisions quickly, rapidly produces value, scales up and down as needed, supports varying use cases, connects easily to new systems, and interoperates with a broad spectrum of business applications.

Ten years of cloud maturity and adoption produced some basic expectations for the CDW. At a basic level, forward-thinking organizations expect the following from a cloud data warehouse.

Value Creation

Ultimately, when organizations bring up a data warehouse, they expect to discover insight and create value quickly. They expect a rapid return on their investment. Since most organizations measure value creation in terms of time, money, resources, and speed of innovation, any new insight must be easily transferrable to decision makers and business sponsors.

Context and relevance are the keys to accelerating business value. Business users are looking for data and insights that are specifically relevant to their focus area. Business focus is typically provided in semantic layers designed in the data warehouse for specific users or within the business intelligence tools. Both approaches are legitimate. Data out of context does not help them make better decisions and they do not have time to forage through irrelevant data. For this reason, integration with visualization and dashboarding tools is essential.

Flexibility

The cloud data warehouse must provide business, technical, and financial flexibility. Since the CDW must support a number of business domains like sales, marketing, logistics, manufacturing, or human resources, it must meet a wide range of use cases from popup analytic marts to enterprise data warehouses that provide a historical perspective to forecast what might happen in the future.

At the highest level, analytic workloads can be broken down into data-intensive processing or compute-intensive processing. To address these needs, the CDW must provide a significant number of options to meet a full spectrum of processing requirements. The broader the range of analytic use cases, the more important it is to identify highly flexible deployment solutions.

Many organizations move to the cloud because of pricing flexibility, especially subscription or pay-as-you-go pricing models. However, it is imperative for businesses to balance this flexibility with the need for predictable cost. No one wants hidden costs, surprise costs, or unexpected spikes in cost.

Elasticity

Elasticity is the ability to scale up and down, quickly and simply, based on the needs of the business and use case. In the early days of cloud data warehousing, elasticity was focused almost entirely on data volumes. The modern CDW must be elastic for shifts in many different dimensions, including data volumes, number of users, new datasets, and concurrent query volume and query complexity. There is also a need to separate workloads so that a short-term intense workload, such as a data scientist training a new model, will not interfere with or slow down other workloads, such as business intelligence (BI) dashboards.

Without multidimensional elasticity, customers growing their cloud presence will begin to search for new offerings. This kind of flexibility is available in both the BYOL and DWaaS options. New pricing models, provisioning capabilities, and tuning features built into pure software vendor technology can often meet granular requirements for elasticity, and DWaaS offerings were built with those requirements in mind.

Analytics

Analytics is the number-one use case for data warehouses. While ANSI SQL has become the standard for analytic functions in a data warehouse, businesses increasingly rely on more complex analytic and data science use cases, and are requiring advanced functions to support them. Users now expect there to be either built-in advanced analytics or seamless integration with analytic and machine learning packages. In addition to accessibility, users expect analytic performance to be sufficient and scalable to handle both complexity of analytics and increasing concurrency of analytics. This is just as true for cloud data warehouses as on-premises.

Recent EMA research showed that 24% of participants want to keep all analytics projects on-premises, and 23% want to keep existing analytics project on-premises and move new analytics projects to the cloud. That is 47% who want to keep all or a significant portion of their analytics on-premises. That means that any cloud analytics project must be highly interoperable with on-premises data and analytics platforms. The inability to support these kinds of hybrid deployments can be a weakness of DWaaS, which is restricted to cloud only. In many cases, BYOL data warehouses in the cloud tend to have the advantage, especially when flexibility of deployment is a business requirement.

Security

Early cloud security concerns were replaced by public cloud security that often surpasses on-premises security for many organizations. Based on consumer demands and regulatory compliance, data warehouses have been required to meet stringent privacy and security since their inception. DWaaS and BYOL offerings must follow these same standards.

CLOUD DATA WAREHOUSING CAUTIONS AND CONSIDERATIONS

Any time there is a mass exodus to a new technology, caution is in order. The Hadoop phenomenon was a great example of unchecked, and often unwarranted, migration. Too many organizations get caught in the wave and make moves without systematically thinking through the implications. The massive move to the cloud is no exception.

Since a predominant deployment strategy is to keep some or all analytics on-premises, carefully consider these three factors:

1. Interactivity among cloud and on-premises data is vital.
2. Migration to the cloud must not interrupt existing analytic value creation.
3. Cloud implementation must be carefully aligned with business requirements.

In cloud data warehousing, the adoption trend leans toward the DWaaS model, a model that seems proven based on market acceptance. However, EMA recommends taking a close look at the DWaaS model. Consider the following points in your decision-making process:

- **Software Limitations** – Whenever software is built to be offered as a service, there are always capability sacrifices made to make the software easier to use. Database functionality is sacrificed for simplicity. Be certain to understand the sacrifices.

- **Hardware Limitations** – Some DWaaS offerings limit the number of hardware configurations that are available for each deployment. For example, there may be a requirement of jumping from 16 nodes to 32 nodes. Hardware functionality and configuration flexibility are sacrificed for the sake of easy mass deployment to the cloud.

- **Hidden Costs** – Most DWaaS offerings have hidden costs that don't surface until you want enterprise strength in terms of backup, recovery, high availability, or analytic functionality. DWaaS offerings almost universally charge extra for these things. Security features can also be an upcharge in some cases. There are other hidden costs for both unexpected growth in data or compute power required by growing analytics programs.

- **Vendor Lock-In** – It is often assumed that it is easy to move from one DWaaS vendor to another, but this is not the case. Every platform is different and therefore requires a significant amount of migration effort, especially if moving from one public cloud, with one type of storage, to another. Also, DWaaS vendors seek to sell a broad variety of adjacent services to lock customers into complex deployments, which are difficult to replicate in other environments. Generally, the only adjacent services available on that cloud are the ones offered by the cloud provider.

- **Hybrid Limitations** – Since on-premises analytics are not going away, hybrid capabilities are vital to successful analytics programs. Some DWaaS offerings do not offer any on-premises equivalent, making it necessary for companies to have multiple types of data warehouses. This complexity can vastly increase the cost of operating and orchestrating overall corporate analytics programs. Data warehouse vendors with both on-premises and BYOL offerings have an advantage in hybrid environments over those who offer their software only as a service.

- **Pricing Limitations** – DWaaS offerings are well-known for their pay-as-you-go pricing models. While these pricing models offer low points of entry, they are tied to usage and performance. Because the bill comes to the consumer after a period of usage, it is often not as predictable as set-price subscription software licensing, enterprise licensing, or all-in-one licensing.

BRING YOUR OWN LICENSE TO THE CLOUD

While DWaaS gets a lot of attention these days, there remains a viable alternative: BYOL. There are several analytic and data warehouse platform vendors that have invested years and millions of dollars to mature their platforms for high-performance, multi-workload analytics via software-only solutions.

Organizations moving to the cloud should carefully weigh the advantages of BYOL against the DWaaS option. Many of the same cloud migration goals can be achieved more economically and efficiently by running your licensed software on cloud infrastructure.

There are several important factors to consider when making a CDW decision:

- **Hybrid Focus –** In recent EMA research, a large percentage of business and IT executives indicated that they prefer to keep existing analytics projects on-premises and move new analytics projects to the cloud. Based on this data point, it is clear that hybrid analytic solutions are an imperative. If your organization is committed to both cloud and on-premises for the near future, then your decision should strongly support a hybrid environment with interoperability among cloud and on-premises analytics.

- **Analytic Focus –** Organizations striving to become insight-driven require analytics that are both deep and wide. Depth consists of the complexity of analytics, machine learning, and artificial intelligence and creates a competitive advantage. Width consists of the broad number of business areas using analytics to drive decisions. Width drives the culture of data-driven decisions. Analytics are more deeply embedded and widely supported in some of the licensed software products, compared to DWaaS.

- **Workload Focus –** When making a move to the cloud, it is important to understand your analytic workload requirements from two different perspectives. First, determine the need for compute intensity versus data intensity. Second, weigh the need for workloads that vary over time versus set workloads. For those organizations that have a broad set of workload requirements and changing dynamics, the flexibility of a licensed analytic platform that can work in both environments may be better suited than the DWaaS model, which is restricted to the cloud environment.

- **Performance Focus –** Pure analytic platform software vendors implicitly have performance as a core design element. Higher performance results in greater customer satisfaction and, consequently, greater success for the software vendor. In situations in which the DWaaS vendor also owns the infrastructure, there is not as much incentive to drive performance. Because of this conflict of interest for DWaaS vendors, pure software vendors tend to provide better performance, especially across diverse workloads.

- **Control Focus –** By virtue of an "as a service" mandate, DWaaS vendors remove some of the flexibility of the platform for the sake of usability, often handling database tuning behind the scenes. User control is sacrificed in lieu of simplicity. With license-based software vendors, customers can fine-tune software and data structures for their own very specific workloads. Tuning does require administrative resources, but for real-time, digital requirements, the additional cost may be worth the increase in speed, control, and flexibility.

- **Pricing Focus –** A primary reason for a move to the cloud is the "pay per use" and subscription-based pricing models offered by DWaaS vendors. In today's competitive market, many software vendors offer this same kind of pricing both on-premises and in cloud offerings. It may be advantageous to utilize subscription pricing with a pure software vendor or simply leverage enterprise pricing across all on-premises and cloud implementations. Pure software vendors may offer the widest variety of pricing models, including perpetual, subscription, pay per use, or all-in-one—the ability to apply the same license across workloads, regardless of the deployment mode.

- **Data Focus –** Recent advancements in cloud technology enable cost-effective solutions for diverse data types. Customers can reduce the cost of data storage by moving some data to file systems or object storage. Organizations that go with DWaaS vendors who run across file systems and object storage lose their chance to negotiate with cloud vendors for large-scale pricing on those systems. The advantage here is with pure software vendors or DWaaS vendors who allow separate negotiations for tiered storage. In addition, most pure software vendors have allowed access to data lakes and object storage for some time now, and have mature solutions for multi-tier access.

IT AND DATA MANAGEMENT
RESEARCH | INDUSTRY ANALYSIS | CONSULTING

MAKING THE FINAL DECISION

A combination of business and technical requirements must drive the final cloud data warehousing decision. On the business side, it is important to consider the speed, volume, change, and expansion of your analytics programs. On the technical side, data volume, data variety, analytic complexity, and interoperability will impact your cloud data warehouse decision. EMA recommends answering both business and technical questions in advance, before beginning your search for the right solution.

Business Guidelines

Speed of Business – What parts of your business require real-time or near-real-time engagement with customers, partners, or internal stakeholders? At what speed does that engagement need to take place? How will this requirement change over the next 12-24 months?

Volume of Analytics – What is the current volume of analytics required to run your business? What is the current pipeline of analytics requests coming from the business? What will the volume of analytics requirements be in 12 months or 24 months?

Changes in Analytic Needs – What portion of your analytics are stable and unchanging? What portion of your analytics change on a seasonal or annual basis? What portion of your analytics change monthly, daily, or hourly? Is the change required minimal, moderate, or major?

Expansion of Analytics Program – To what new business areas do you plan to expand your analytics program in the next 6 months, 12 months, and 24 months? For each of the new business areas, what is the expected volume of analytics needed by the business? To what new applications, machines, or services do you plan to expand your analytics programs? For each of the new business areas, what is the expected volume of analytics needed by the business?

Technical Guidelines

Data Volume – What is the current volume of data required for your analytics programs? How do you expect data volumes to grow in the next 6 months, 12 months, and 24 months?

Data Variety – What are the current data sources and data types included in your analytics program? What new data sources and data types will you add to your analytics program in the next 6 months, 12 months, and 24 months?

Analytic Complexity – What is the current volume of complex analytics (multi-source, multi-attribute, predictive, prescriptive, machine learning, artificial intelligence) run by your organization? What is the expected volume of complex analytics run by your organization in the next 6 months, 12 months, and 24 months?

Interoperability – What level of interoperability do you require between your cloud and on-premises analytical systems? What level of interoperability do you require between your cloud analytical systems and other on-premises systems? What level of interoperability do you require between your cloud analytical systems and other cloud systems? With what other applications (visualization, ETL, etc.) will the data warehouse need to integrate to make a complete solution?

VERTICA ANALYTICS PLATFORM

EMA completed a product review of the Vertica Analytics Platform and found it worthy of consideration for on-premises, cloud, and hybrid data warehousing solutions. EMA recommends looking at the four strengths of the Vertica offering:

1. Deployment Flexibility – Because Vertica is truly a software-only, infrastructure-independent product, it provides a consistent experience no matter where it runs. It is the same database on-premises or on any cloud infrastructure. Also, Vertica partners with Pure Storage to deliver the industry's only analytical database solution with a separation of compute and storage architecture on-premises by running on Pure Storage FlashBlades, enabling organizations to adopt all of the benefits of a cloud-optimized architecture from within their own data centers. All of this provides an advantage for cost containment, resource allocation, and interoperability.

2. License Pricing – Vertica's customer focus led them to develop pricing for a broad range of customer needs. Vertica pricing options include perpetual licensing for capital expenditure, subscription pricing for operational expenditure, terabyte-based pricing, per-node pricing, per-usage pricing on AWS, or all-in-one licensing. High availability and development and test environments are all included at no additional charge.

3. Advanced Analytics – Vertica has a rich set of optimized advanced analytic and machine learning functions and algorithms embedded in their database, available as part of their general licensing. There is no extra charge. This significantly reduces the need to utilize third-party tools and eliminates hidden analytics fees charged by some platform vendors.

4. Performance – With decades of investment in query optimization and embedded operations speed, Vertica is well-suited to handle diverse workloads including complex queries, concurrent queries, and complex analytics. Vertica performance tuning and flexibility are optimal for most analytic workloads.

ABOUT VERTICA

Vertica is trusted by thousands of leading data-driven enterprises around the world, including Etsy, Cerner, Intuit, Uber, and more to deliver speed, scale, and reliability on mission-critical analytics. The Vertica Analytics Platform is designed for use in cloud data warehouses and other big data workloads in which speed, scalability, simplicity, and openness are crucial to the success of analytics. Vertica in Eon Mode provides a cloud-optimized architecture that delivers operational simplicity and rapid elasticity by scaling AWS infrastructure to meet peak demand when needed, and scaling back when it's not needed. Vertica in Eon Mode enables this through the separation of compute and storage, which capitalizes on cloud economics and dynamic workloads by scaling EC2 compute resources independent of shared S3 storage. Providing even greater choice, Vertica now offers all of the capabilities of cloud-optimized architecture to on-premises environments with Vertica in Eon Mode for Pure Storage.

IT AND DATA MANAGEMENT
RESEARCH | INDUSTRY ANALYSIS | CONSULTING

Corporate Headquarters:
1995 North 57th Court, Suite 120
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com
3901.102419

IT AND DATA MANAGEMENT
RESEARCH | INDUSTRY ANALYSIS | CONSULTING