

WINTERCORP REPORT

W I N T E R C O R P O R A T I O N

A GAME CHANGING PLATFORM FOR CLOUD DATA ANALYTICS

.....

*Teradata VantageCloud Lake
with ClearScape Analytics*

.....

SPONSORED
RESEARCH
PROGRAM

W W W

EXPERTS IN DATA MANAGEMENT AT SCALE

W I N T E R C O R P O R A T I O N

A GAME CHANGING
PLATFORM FOR CLOUD
DATA ANALYTICS



*Teradata VantageCloud Lake
with ClearScape Analytics*

RICHARD WINTER



WinterCorp

www.wintercorp.com

42 DERBY LANE, TYNGSBORO, MA

617-695-1800

EXPERTS IN DATA MANAGEMENT AT SCALE

Summary

MANY DATA WAREHOUSES have now been migrated to or built in the public cloud but analytic data platforms have not previously offered all the capabilities that customers are seeking.

Teradata's latest major announcement, of *VantageCloud Lake with ClearScope Analytics*, delivers an analytic data platform that takes full advantage of the cloud architecture while also providing an open, flexible environment for data exploration and business autonomy. Vantage does this while still meeting the most demanding requirements of enterprise customers. No other cloud analytic data platform service offers comparable flexibility, functionality and performance at scale. And, no other cloud data platform offers comparable cost efficiency and manageability under heavy workloads.

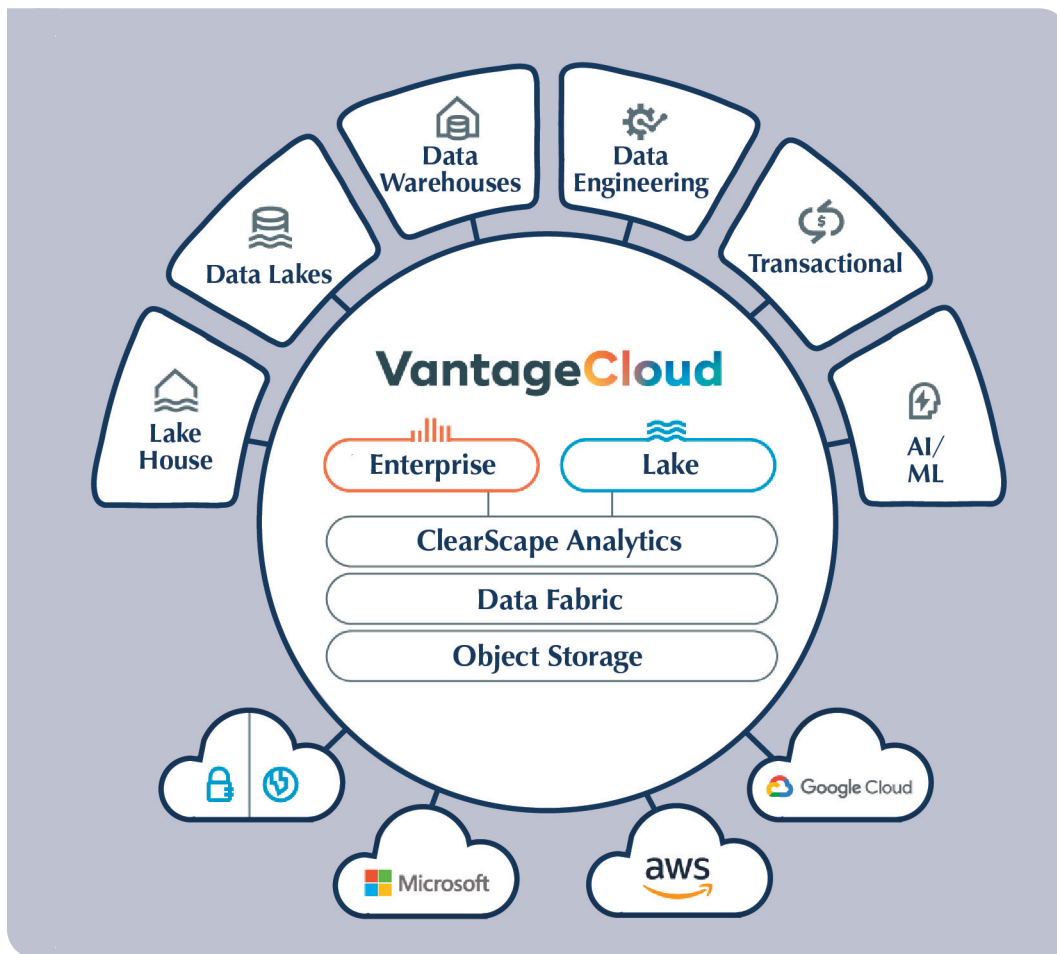


Figure 1: VantageCloud Lake Extends the Teradata Ecosystem

While Teradata *Vantage* has previously been identified primarily with high-end enterprise requirements, it has not been widely seen as a desirable solution for cloud-based line of business projects. In these projects, the customer is typically seeking autonomy, agility and independence from the enterprise IT governance while often still wanting to use the enterprise data.

Now Teradata provides an answer for such line of business customers, without sacrificing the benefits of its enterprise approach.



Figure 2: Key New Capabilities of *VantageCloud Lake*

Key new capabilities of *VantageCloud Lake* are:

- **Shared low-cost object storage optimized for analytic workload processing, with complete separation from compute;**
- Multi-cluster compute with elasticity, where **clusters are easily spun up and spun down for business user autonomy;**
- Optimized pricing for the cloud with central administration and financial governance;
- Open connections to external data, tools, libraries and data sources;
- A powerful array of new in-database functions for machine learning, data science and advance analytics, named *ClearScape Analytics*.

These advanced, cloud-based, capabilities are now available in Teradata’s unique *Vantage* platform that has long led the industry in its ability to deliver advanced analytics and machine learning in production at scale; operational (aka “tactical”) query on live data at very high volumes; continuous operation; exceptional performance; and, exceptionally low cost per query at high volume.

This report describes the new cloud capabilities of *VantageCloud Lake* and discusses the implications for customers of this newly available, remarkably complete capability for enterprise analytics in the cloud.

On the basis of its independent research, WinterCorp recommends that companies with a need for enterprise class data warehouse capabilities or advanced analytics in the cloud look closely at *VantageCloud Lake*, with an eye to near term implementation. ●

1 Issues in Cloud Data Warehousing

AS CUSTOMERS MOVE THEIR ENTERPRISE CLASS DATA WAREHOUSES to the cloud, they want to take full advantage of the innovative services and capabilities in the cloud, such as the remarkably low-cost object storage that is such a good option for historical data, and the ability to increase and decrease capacity on short notice, as needs fluctuate. Further, customers want the ease of operation and management that the cloud promises.

Full separation of storage and compute. To many customers, the flexibility of the cloud is one of its greatest attractions. In some situations, they want a very large amount of data available for use in a very small compute cluster. In other situations, it may be the opposite. This flexibility is possible in the cloud, and it can translate into large cost savings. For many customers, this complete separation of storage and compute is a key requirement.

Real Elasticity. The cloud promises elasticity: the ability to increase or decrease capacity as a system is running to respond to changing customer workloads. Some vendors promote elasticity but don't deliver the real thing: either the capacity changes are unreasonably large; or they don't happen immediately; or, even worse, the system must be taken down and restarted for the capacity change to take effect. What customers actually need is rapid, granular changes in capacity in response to moment-to-moment events — providing them with both the throughput and the economics that their business demands.

Large, high priority jobs. In data warehousing, it is not unusual for a large job, requiring a lot of system resources, to run at the same time as other critical processes. Examples, often requiring large amounts of computation, include:

- An ETL job doing complex transformations;
- Machine learning on IoT data from a large factory to predict machine failures;
- Analyzing the journey of a large population of customers to identify the behaviors that lead to high lifetime value.

In the cloud, customers want simple ways to initiate such large processes without impact on existing data warehouse operations.

Business Autonomy. Project teams and business units often want their own, dedicated, analytic compute resources to deploy as they think best, while continuing to leverage corporate data. Similar needs often exist for application developers and quality assurance teams, whose needs may exist over a few weeks or months, but at times may be resource intensive. From an enterprise perspective, it may be advantageous to provide these dedicated compute resources, especially within a framework in which the users can be charged for the resources they use.

Optimized low-cost data storage. Cloud object storage, typically a factor of ten lower in cost than previously available options, is often suitable for most data in a data warehouse. Customers want the benefit of this low-cost storage as a way to store all the corporate data including files that are optimized for data warehouse use.

Advanced analytics and machine learning *in database*. Innovation and competitive advantage increasingly result from data science, machine learning and other forms of advanced analytics. In the context of a data-intensive enterprise, it is critical for these functions to be performed on data in place in the database; in parallel; and, in a highly efficient fashion. Otherwise, the volumes of data involved make these processes cost too much or take too long. In addition, if the data is sensitive, there is a significant security advantage to keeping it in one place, where it is carefully governed, rather than transferring it out to be analyzed on other systems.

Mission critical operations. In many companies, the days in which the data warehouse was used only for strategic analysis and reporting are long gone. It is increasingly common for the enterprise data

A W I N T E R C O R P R E P O R T

warehouse to support the moment-to-moment operation of the enterprise, handling tactical queries on near real time data and delivering services which can never be down.

Enterprise levels of scale and complexity. Larger data warehouses manage petabytes of data, service enormous populations of users and support data models in which there is a complex web of relationships among many large tables. Workloads are numerous, varied and dynamic. In these data warehouses, the platform must provide efficient, scalable solutions to a demanding set of integrated requirements, not subject to shortcuts and glib simplifications. These requirements are fundamental to the operation of a large, complex data warehouse and are not diminished by moving it to the cloud. Customers with such requirements need a cloud platform that can meet them head on, with mature and comprehensive capabilities.

2 Introducing *VantageCloud Lake*

TERADATA HAS INTRODUCED **VANTAGECLOUD LAKE** to address the full set of requirements for cloud data warehousing and AI/ML at enterprise levels of scale and complexity. While the resulting service is uniquely capable of meeting high end needs, it also provides exceptional analytic capabilities for customers operating on a more modest scale and also for customers that are just starting out. Key new capabilities are described in the following paragraphs.

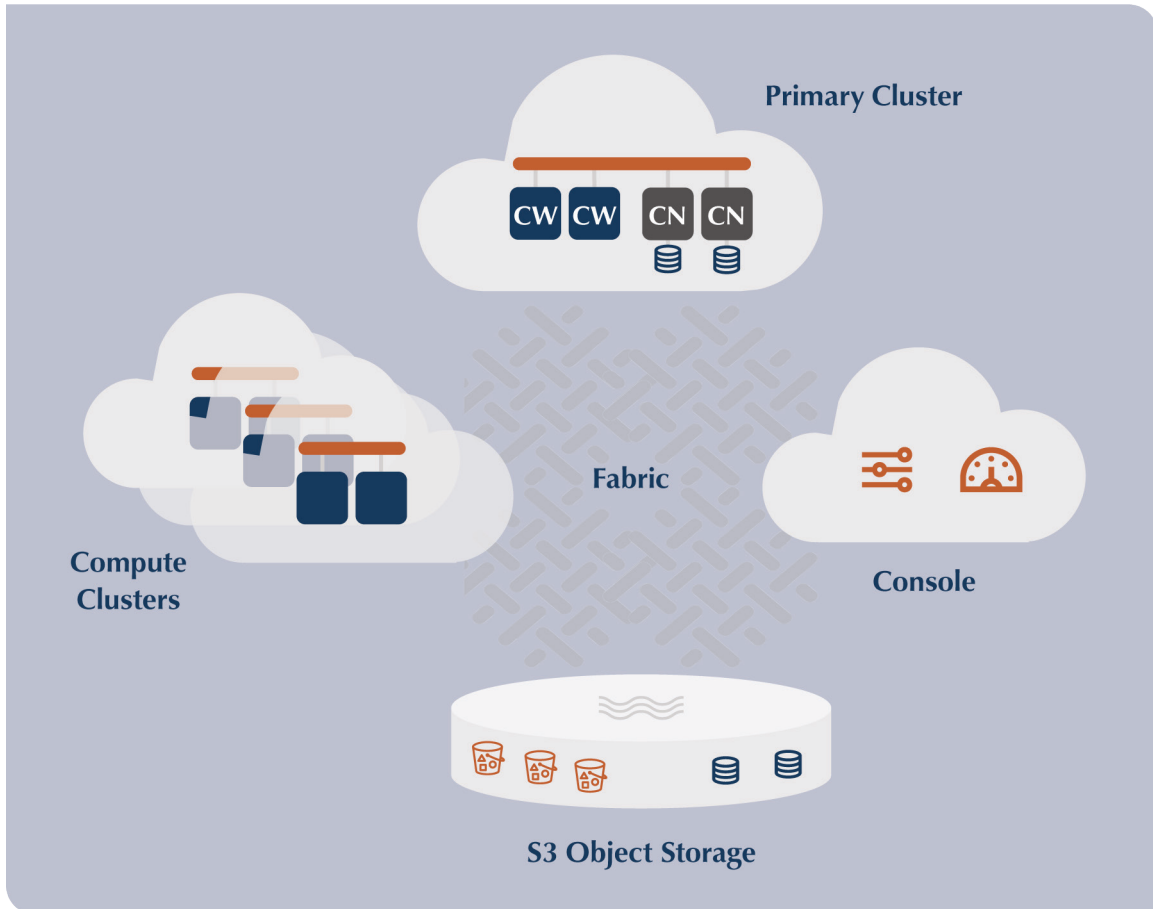


Figure 3: Elastic Compute Clusters Can Be Spun Up and Down for Business Autonomy

Separate compute clusters. *VantageCloud Lake* features a new cloud native architecture that makes it easy for the customer to bring up and bring down elastic compute clusters to handle specific capacity demands and to provide dedicated resources for user projects and departments. Users of these separate clusters can access the same data, while not contending with any other users for resources.

The architecture features two types of clusters: primary and compute. The primary cluster provisions resources for query planning and work distribution amongst compute clusters. Compute clusters execute query workload; they can be scaled in or out — that is, nodes can be added or removed — manually or automatically, according to defined policies and they can be hibernated when not in use.

VantageCloud Lake compute clusters have smart scaling, a technique that uniquely monitors resource utilization and provides automated scaling as compute resource usage thresholds are met. Smart scaling

makes the most efficient use of resources and as such, results in lower overall costs. In contrast, other offers rely on static metrics like query counts which results in unnecessary and more frequent scaling at a greater cost.

All clusters can share all the data — both data maintained in block storage for high performance and data maintained in object storage for lower cost. This capability provides the business autonomy that many users are seeking in the cloud, while still providing a shared core of enterprise data that can be leveraged across the organization.



Figure 4: Two Supported Options for Object Storage

VantageCloud Lake supports both Optimized Lake File System and open data format. Lake uses object storage in two key ways: it uses the *Vantage* Object File System to optimize how the data is organized (stored) in a modified Parquet format. The Optimized Lake File System provides ACID-compliant operations on data formatted and governed for scale and optimized for performance and cost efficiency. In addition, it uses *Vantage* Native Object Store to read and write in open data format (e.g., Parquet, CSV, JSON) optimized for data sharing.

Central administration. In the unique architecture of *VantageCloud* Lake, the entire ecosystem of clusters is managed from a central point as one system. An admin cluster enables the central management, governance and operation of all operations. The primary cluster is the initiation point for all queries and provides all the benefits of the previous versions of *Vantage*, including its distinctive, advanced workload management. And now the compute clusters enable dedicated resources for high priority batch jobs and for projects and departments wanting the benefits of their own, separate resources— all on one centrally governed copy of the shared enterprise data.

Compute nodes in the primary cluster. In another key innovation, *VantageCloud* Lake can be operated with additional compute nodes in the primary cluster. These compute nodes can be dedicated to specific tasks, such as high-performance tactical queries with sub-second response time requirements; or particularly compute intensive analytical tasks. This provides an additional dimension of flexibility in operating the environment to meet specific needs.

Full separation of storage and compute. In an earlier era of parallel data warehouse architectures, the only way to get sufficient storage bandwidth was to dedicate storage devices to compute nodes. This provided high performance but could limit the customer's options for expanding capacity. To get increased storage capacity, the customer could be required to also increase compute capacity. While *Teradata Vantage* offered a high degree of flexibility in this regard, it did not previously offer complete separation of compute and storage.

VantageCloud Lake now offers optimized object storage of data with complete separation from the compute nodes. Customers have complete flexibility on the ratio of object storage capacity to compute capacity. In fact, this ratio can now be configured differently in different in both the primary and compute clusters.

High performance block storage still available. In addition, *VantageCloud Lake* still offers block storage, more closely tied to the Primary compute node, for high performance operations. Several other cloud data warehouse platforms do not offer this option. Thus, the *Vantage* customer now has the best of both worlds: ***block storage for low latency tactical workloads and object storage for higher latency but high throughput.***

While these cloud capabilities have been available elsewhere, others haven't been able to meet the enterprise class requirements to which *Vantage* addresses itself. **We have thus entered a new era in cloud data warehousing, in which enterprise class data warehousing is delivered on a platform with an architecture that takes full advantage of the cloud.** ●

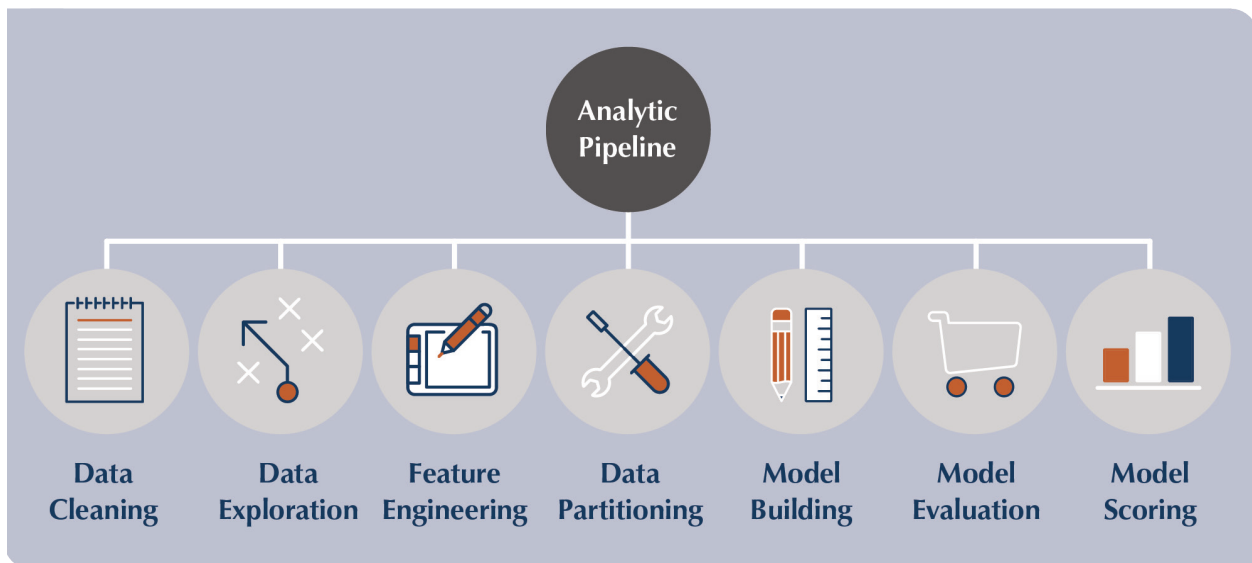


Figure 5: In-Database Analytic Pipelines

ClearScape Analytics. *ClearScape Analytics* is a unique, extraordinary new capability for in-database AI/ML analytics offered with *Vantage Data Lake* edition. It includes a long list of built in in-database functions to help customers solve complex analytic challenges on data in place in the database, with parallel execution. It supports end-to-end data pipelines — including streaming data — and strengthened integrations with partners such as Dataiku, H2O.ai and AWS services for AutoML and data science.

Financial Governance and Pricing. *VantageCloud Lake* is offered with a simple unit-based compute consumption model in which the customer pays only for compute when on or provisioned. Data is priced comparably to cloud object storage. Teradata offers low minimum requirements for utilization in connection with its consumption pricing. And, *VantageCloud Lake* provides capabilities to manage and allocate costs to groups of users via an intuitive self-service console with transparent usage and pricing data.

Highly Optimized In-Database Analytic Functions

-  Descriptive Statistics
-  Data Cleansing/Transformation
-  Feature Engineering
-  Hypothesis Testing
-  Multivariate Statistics
-  Machine Learning
-  Time-Series Forecasting
-  Digital Signal Processing
-  Pathing Analytics
-  Geospatial/Temporal

Leverage Languages and Tools of Choice

Languages

Use preferred languages R, SQL, Python, SAS

SQL  python  R  sas

Partner Integration

Best ML/AI tools with tight integration and execution with Vantage



Bring Your Own Analytics

Open analytics framework containers and model sharing

Deploy at Scale to Drive Transformative Results

Robust APIs

Real-time or Batch scoring
Enable open ecosystem access to features or data

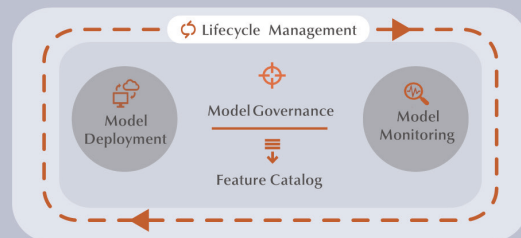
 **Access**
REST, SQL, Python, R, Java

 **Orchestration**
QueryGrid, NOS

 **Replication**
TPT, DSA, SFTP, Native API's

Integrated ModelOps

Models Governance, Lifecycle Management and Monitoring Extensions



Tightly Integrated with Vantage Database and Workload Management

Figure 6: ClearScape Analytics

3 Distinctive Strengths of *VantageCloud Lake*

UNLIKE MOST OTHER CLOUD DATA WAREHOUSE SERVICES, the *Vantage* database has a long history of technical leadership in data warehousing and ML/AI analytics. As a result of focused engineering and R&D over decades, it has developed unmatched capabilities for handling the demanding, real-world requirements of large enterprises with business critical needs to solve large scale, complex analytic problems.

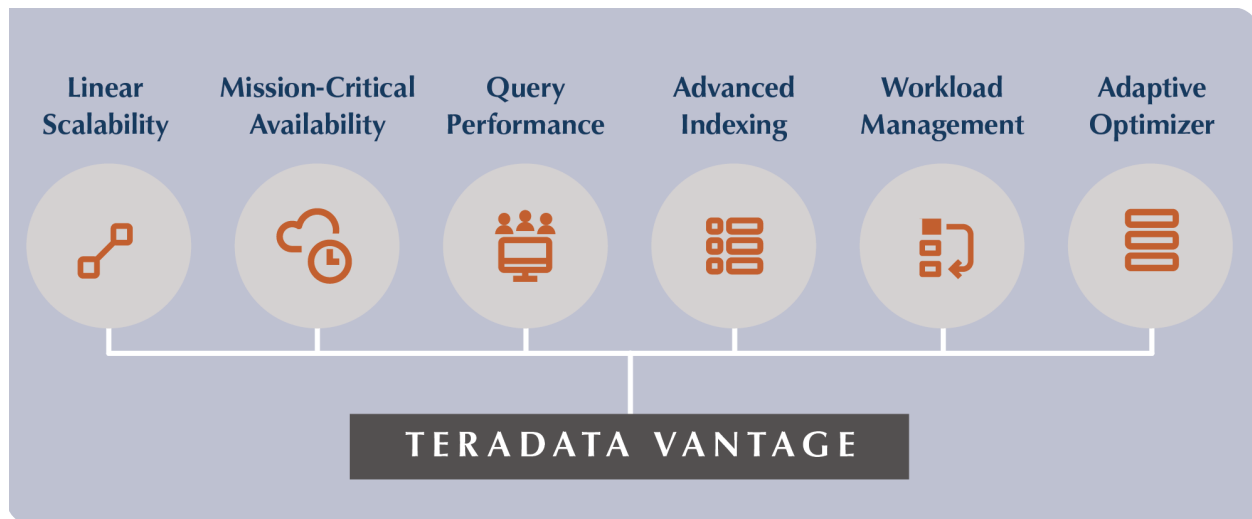


Figure 7: Distinctive Strengths of Teradata *Vantage*

While the *VantageCloud Lake* Architecture moves the product ahead in the cloud dimension, it is critical to appreciate that nothing has been lost in the fundamentals of data warehousing and AI/ML analytics, where the unique capabilities of Teradata *Vantage* database are more important than ever.

Here is a brief review of a few of these capabilities that the author has seen in action first hand, under the most demanding customer requirements. All of these capabilities are supported at small, moderate and very large scale:

- **Multiple dimensions of scalability**
 - » Complexity of queries from 2-way to 64+ way joins
 - » Many concurrent users
 - » Large data volumes
 - » Freedom to run complex and unpredicted queries
 - » Complex data structures
 - » Large, dynamic workloads
- **Vantage can handle requirements that are demanding, and growing, in all of these dimensions at once. This is the true meaning of data warehouse scalability;**
- Efficient management of complex workloads. The customer can have many workloads, each with its own service level objectives and manage them from a central console, with the benefit of significant automation;
- Support of a denormalized data architecture. The database can include many large and small tables with a complex web of data relationships; the *Vantage* database optimizer is capable of efficiently

joining such tables along multiple different paths and in complex combinations. Most other platforms cannot match this capability and force the customer to simplify the data model in ways that will undermine the business value of the data warehouse;

- Fast, efficient tactical query. *Vantage* can perform short queries of simple to moderate complexity with great efficiency — often 100 times or more faster and at 100 times or more lower cost than other data warehouse platforms; some customers have tens of millions of such queries a day;
- Lowest cost per query. Looking across a large, diverse data warehouse workload, *Vantage* often exhibits a much lower cost per query than other platforms;
- Four dimensional analytics. Three dimensions of space plus one of time. *Vantage* has built in capabilities to define, store and maintain geo and spatial data; and temporal data; and to use these in complex queries at scale;
- Temporal data support. When a column is defined as temporal, *Vantage* will automatically and transparently maintain previous values or rows and support “as of” queries that leverage the values that existed at a previous point in time; especially valuable for slowly changing dimensions;
- Integrated path analytics — *Vantage nPath* supports **path** analytics. One example is automated analysis of the journey a customer follows in the course of a relationship, which can be used to identify the path to the first sale or to a high lifetime customer value. This is extremely difficult analysis to do in SQL on other data warehouse platforms. But, built in constructs make it straightforward and efficient with *Vantage nPath*, where such analysis is supported with Sankey visualizations;
- Fast, efficient, large, complex joins- big table, big table joins are rarely implemented efficiently at scale; *Vantage* does this but also optimizes complex joins;
- Join indexes and aggregate join indexes — for frequent joins, these are system-maintained pre-joins and aggregates, invisible to the user, automatically used by the optimizer to reduce query time and effort when possible. Note that most other systems rely instead on user-visible pre-joins, a mechanism that becomes unworkable as database complexity increases;
- Continuous operation — *Vantage* has all needed capabilities to operate with no scheduled or unscheduled downtime; it can fail over to a hot standby across site and region boundaries.

This is the combination of capabilities that has drawn companies in every major industry to *Vantage* when they want to pursue data driven strategies without limitation and without concern that they might outgrow their data warehouse platform.

4 Conclusions and Recommendations

For customers who use analytics strategically in their business — to support critical business operations and to leverage otherwise unavailable business opportunities — *Vantage* has long provided a distinctive, uniquely capable data warehouse platform. It is unmatched in its ability to handle the demanding requirements that emerge at enterprise levels of scale and complexity.

Until now, it has not been able to deliver the agile and flexible cloud capabilities that many customers are looking for.

That has changed. *VantageCloud Lake* is now in the premier tier of cloud data warehouse services with full separation of compute and storage; dedicated compute clusters that are easily spun up and spun down; dedicated compute nodes for special requirements; optimized object storage as well as open object storage; open, connected in-database analytic and machine learning; and a raft of other appealing cloud features.

At the same time, *Vantage* delivers its industry leading analytics and machine learning all with *ClearScape Analytics* in production at scale, supported by many other uniquely valuable features.

On the basis of its independent research, WinterCorp recommends that companies look closely at *VantageCloud Lake Edition*. This is a remarkable new capability for data warehousing in the cloud, uniquely able to satisfy needs for line-of-business autonomy and ease of use, while at the same time meeting the most demanding enterprise requirements at scale. The resulting combination is one that no company should overlook.

At the same time, it is important to remember that no one platform or service is best for every customer. WinterCorp recommends that customers adopt a cloud data warehouse platform only after a careful, fact- and measurement-based evaluation focused on their particular needs. ●

WinterCorp is an independent consulting firm expert in the architecture and strategy of the modern database management ecosystem.

Since our founding in 1992, we have architected and engineered solutions to some of the toughest and most demanding database management challenges, worldwide.

We help customers define their data-related business interests; develop their data strategies and architectures; select their data platforms; and, engineer their solutions to optimize business value.

Our customers, with our help, create and implement cloud, multi-cloud and hybrid cloud architectures; they create the data foundation needed for data science, artificial intelligence and machine learning.

Our customers get business results with analytics in which their return is often ten or more times their investment.

When needed, we create and conduct benchmarks, proofs-of-concept, pilot programs and system engineering studies that help our clients manage profound technical risks, control costs and reach business goals.

We're expert with structured data, unstructured data, and semi-structured data — with the products, tools and technologies of data management for data analytics in all its major forms.

With our in-depth knowledge and experience, we deliver unmatched insight into the issues that impede scalability and into the technologies and practices that enable business success.



WinterCorp

www.wintercorp.com

42 DERBY LANE, TYNGSBORO, MA

617-695-1800

©2022 Winter Corporation, Tyngsboro, MA. All rights reserved.