opentext™

Cloud Without Compromise

Crucial Requirements for the Analytical Data Platform

Contents

Introduction

Climate LLC

At a glance

The future of sustainable agriculture

Maximize yield through advanced analytics

Doing more with less

Gain the freedom of cloud deployment options

ThinkData Works

At a glance

Embracing data-driven innovation

Next-level data delivery

Helping organizations gain new insight

Our Fungible Licensing

The meaning of “no compromise”

McKnight Consulting puts OpenText, Snowflake, and Redshift to the test

Here are highlights of this report

Data preparation

Price per performance

OpenText’s superior concurrency

Conclusion

Introduction

If yours is like most organizations, you're spinning up new analytical workloads and considering moving them to the cloud, whether your workloads are based on data warehouses, data lakes, or both. But, as you may have learned, it can be difficult to find the right platform to fit your organizational realities, including your technology strategy and direction, and your product requirements.

It may seem simple to move everything to a single cloud vendor's resources. But you'll face compromises in choosing a platform that is only available as a cloud service, or one that's available on only one cloud. Few individual cloud vendors support all the tools your analytical teams are familiar with, and want to use, whether they're doing business intelligence, or machine learning and AI. On top of that, not all workloads are suited to every single cloud environment. Some run best on one vendor's offering, some on another's.

This is why a multi-cloud approach may be best for your organization. But you still need to address the question: Should all of your analytical workloads move to a cloud-based platform, or should some of what you've maintained for years in your data center remain there? A combination of cloud and on-premises resources may be right for you—and it may be your only option if your industry's regulatory requirements prohibit storage of sensitive data in the cloud.

This eBook showcases businesses who have wrestled with these complex decisions, and have landed on data analytics configurations that proved successful. These companies have considered the evolution of analytical data lakes, data warehouses, and combined platforms. They've weighed their aspirations for data analytics against the technical realities of their organizations; and they have factored the cost, security, and performance trade-offs of cloud services into their choice of data analytics technology.

Again, if you're like thousands of other technology leaders seeking the business advantages that today's analytical platforms have to offer, I think you'll see some familiar decision-making in these stories. If they help clarify some of the confusion about data analytics, or simply offer a familiar scenario you can relate to, then we've done our job. I look forward to hearing your own stories about your data analytics journey.

Best to you,

Jeff Healey



Jeff Healey
Vice President of
Marketing, OpenText

Climate LLC



Data science and analytics supports an evolution in farm data and farm decision-making.

At a glance

Industry
Agriculture

Customer
Climate LLC

Location
Missouri, USA

Context
Leverage data and data science to help farmers around the world make management decisions that optimize crop yields, field inputs, management efficiency, and farm profits.

Our Response
Vertica Analytics Platform

Impact
- Reduce the impact of hunger around the world
- Minimize environmental impact
- Sustainable and innovative farming with higher yield

Focus Area
Predictive Analytics



Climate LLC is dedicated to creating technologies that transform field data into meaningful insights to help farmers sustainably enhance yield potential, improve efficiency, and manage their risk.



Digital agriculture for actionable insights

Digital agriculture is not only one of the most exciting new frontiers in the advance of technology and science, but serves as a central element supporting one of agriculture's—and humanity's—most pressing concerns: increase crop yields for a growing population, right-size farm inputs to reduce environmental impact, and enable farmers to more purposefully navigate the increasingly complex set of decisions they make for their fields throughout the year.



Just 1.3% of the domestic U.S. workforce1 is responsible for producing the food we eat. This small portion of the population manages massive amounts of land; in 2017, U.S. farmers managed just over 2 million farms compared to nearly 7 million in 19352, representing not only a sharp decline in number of farms but a vast increase in the information being managed by each farmer.

The scale of the data coming off any farm is massive, varied, and complexly interrelated. The need for clean data to drive future mathematical model development and thus actionable insights to farmers is great and urgent. The agriculture of the past, with manual notebooks kept inside tractor cabs, has passed but it's not that far in the past. Digital agriculture has evolved rapidly over the last decade and with it digital solutions.

1 https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/ag-and-food-sectors-and-the-economy.aspx

2 https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/farming-and-farm-income/

Context

The future of sustainable agriculture

Millions of acres of clean data, well curated and organized, across a range of geographical and management conditions, is the future of sustainable agriculture. OpenText™ Vertica™ Analytics Platform plays a key role in the technology landscape at Climate. Erich Hochmuth, Senior Director of Data and Analytics, comments, “Vertica’s (now part of OpenText™) analytical and spatial functions allow Climate to sift through the diverse datasets and get an accurate lay of the land to enable decisions from the direction of new products to the accuracy of our scientific models.”



When it comes to the need to transform the way data is managed on the farm, understanding that data is a key piece. OpenText has enabled critical insights coming out of our trove of FieldView™ data. The team responsible for deployment of the FieldView products use OpenText to better get a sense for how Climate customers are using its apps, and all of the core metrics of success are computed and powered from OpenText.

Maximize yield through advanced analytics

Agronomic recommendations are made available back to the farming clients via a SaaS application to assist them in picking the best seed for next season. OpenText's analytics and spatial capabilities helped enable Climate's data science teams to build and validate the models powering these agronomic recommendations, and drive insights to help its customers successfully use Climate products.

Hochmuth, on the nature of the data Climate has to correlate and analyze, "At Climate we're working with data not just coming from what's being planted, but from the clouds and weather in the sky down to the composition of the soil that holds the crops in place.

> **"Vertica's (now part of OpenText™) analytical and spatial functions allow Climate to sift through the diverse datasets and get an accurate lay of the land to enable decisions from the direction of new products to the accuracy of our scientific models."**
>
> **Erich Hochmuth**
> **Senior Director of Data and Analytics**

Traditional data solutions couldn't make sense of the multitude of data layers a farmer has to care about. OpenText's analytics and reporting functionality greatly reduces the time our teams would need to spend synthesizing data into insights."

Doing more with less

Hochmuth concludes, “Vertica (now part of OpenText™) is a critical technology that helps Climate collect, clean, and organize the big data within FieldView. One of my favorite things about FieldView is how it enables farmers to see all their data in one place. Having access to organized data that paints a clear picture of what’s happening on a field is step one; step two is pushing that data a little further to help enable farmers to make the best on-farm decisions they can, raising their yields, their profits, and the sustainability of their land.”

This is the real impact of FieldView: doing more with less. Over the next several decades, the demand for food will greatly increase while the land available for, and suited to, the cultivation of crops that the world needs will decrease or remain flat. Data and analytics are driving us toward the future of farming in which we’re producing more, making the best decisions possible, and feeding the world.



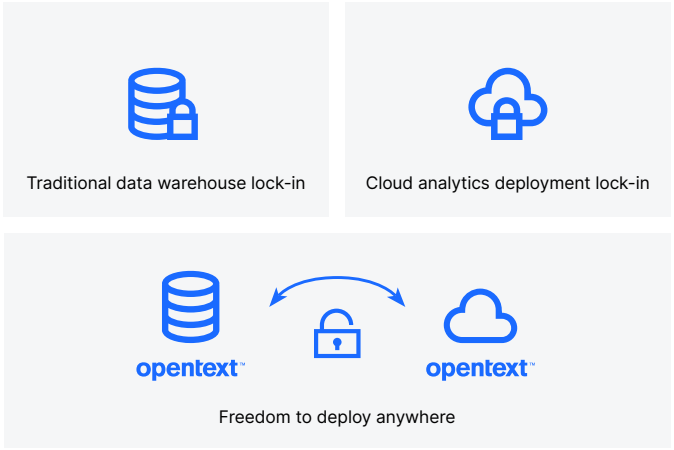


Gain the freedom of cloud deployment options

When you have total freedom over your data analytics deployment options, you put your customers and your business in control. When you can bring your own license to your deployment resources—whether the public cloud, multi-cloud, on-premises hardware, or any combination of those—you avoid getting locked into a single vendor's idea of where you store data and run your analytics.

Multi-tenant cloud implementations often store metadata for multiple customers in a single database, potentially creating a security issue. Performance can also suffer in a shared-infrastructure environment. One customer may use a great deal of resources, reducing the resources available to other clients. This "noisy neighbor effect" means your performance may suffer even when you are not overusing the database.

OpenText™ Vertica™ Accelerator is single-tenant. You own all your own bandwidth, with no noisy neighbors, and all metadata is stored in your own instance—not in a shared system. Your data remains secure in your own AWS account, with your own organization's security rules in place.



ThinkData Works


Many businesses lack the time, resources, and skills to prepare valuable external data sources for rapid analysis. Building on OpenText, ThinkData Works is eliminating barriers to the analysis of data—making rapid, reliable, and cost-effective third-party data analysis available to businesses around the world.

At a glance

Industry
Technology

Customer
ThinkData Works

Location
Toronto, Canada

Context
How could ThinkData Works help businesses unlock the power of external data rapidly and cost-effectively?

Our Response
Vertica Analytics Platform

Impact
- 70% increase in revenues typically realized by ThinkData Works' customers
- 2400% efficiency saving achieved by one bank using ThinkData Works
- 51% reduction in time to create client PoC's achieved by a leading consultancy firm

THINKDATA WORKS

ThinkData Works is a Toronto-based tech startup that specializes in creating platforms for the aggregation, standardization, and distribution of high-value public and third-party data.

Embracing data-driven innovation

Organizations of all kinds depend on data from a variety of sources to improve the speed and quality of decision-making, to better understand customer needs, and to mitigate future risks. The ability to make sense of data quickly and cost-effectively is essential to maintaining competitiveness.

Sourcing and preparing data for analysis is a sophisticated task, requiring significant investment and specialist skills. For many companies, especially small- and medium-sized enterprises, the costs are simply too high to be able to carry out these activities in-house. For other businesses that can afford to hire specialist data-science specialists, the time-intensive nature of data preparation tasks means that many analysts spend up to 40 percent of their time searching for, and cleansing, high-quality data. That effort reduces the time they might spend discovering vital insights.

And this is just for internal sources of data. When it comes to tapping into high-quality public and third-party data, the challenges are greater still. So are the potential rewards: Capgemini research suggests that organizations able to take advantage of external data ecosystems on average improve customer satisfaction by 15%, boost productivity by 14%, and reduce costs by 11%.

To make it easier for businesses to analyze high-quality external data sources, ThinkData Works developed Namara Marketplace: an online data repository comprising 250,000 data sets from over 3,500 sources, powered by OpenText. With Namara, businesses can browse, access, and establish connections with deployment-ready data, quickly and easily.



Next-level data delivery

To meet these emerging needs and help businesses ingest and extract value from new data sources, ThinkData Works set out to develop its existing solution into a fully fledged platform for fast, simple, and efficient external data analytics.

“Vertica (now part of OpenText™) has been at the heart of Namara Marketplace for over eight years,” says Tim Lysecki, Manager, Product Marketing at ThinkData Works. “Throughout that time, we’ve seen just how reliable, cost-effective, and high performing the solution is, so continuing to innovate with Vertica (now part of OpenText™) was the obvious choice for us. And the fact that Vertica (now part of OpenText™) supports on-premises, hybrid, and public cloud deployments was a huge advantage, because it provides all the flexibility that our customers could need.”

By combining the power of OpenText with its proprietary data ingestion engine, ThinkData Works has built a platform that effectively gives businesses their own private version of Namara Marketplace. Here, they can easily maintain links to all ThinkData Work’s public datasets without the need for extensive integration work, but also bring in and manage their own private data sets—whether internal or external—in the same way. In addition to providing fast, easy access to high-quality external data, the solution automatically handles changes in data types and data sources, enabling users to focus on data analysis rather than data administration. To enable customers to view their data in their preferred format, the ThinkData Works platform integrates fully with popular visualization tools such as Tableau, Microsoft PowerBI, and Diablo Data Systems.

> **“Namara Marketplace proved hugely popular, so much so that our customers approached us to see if we could evolve the solution to support their other data sources, both external and internal.”**
>
> Brendan Stennett
> Co-Founder & CTO at ThinkData Works

“To help customers spend more time with their data and less time cleansing it, setting up and monitoring pipelines, we work individually with them to configure the solution to their preferences,” says Brendan Stennett. “Once we’ve set up the solution according to our customers’ storage preferences and connected it, both physically and virtually, to their data warehouse, we provide a set of self-service controls. Users can then start importing their preferred data from external sources and schedule automatic data set refresh intervals as they wish.”

Helping organizations gain new insight

By offering a flexible platform for ingesting and managing multiple external and internal data sources, built on the market-leading OpenText platform, ThinkData Works is helping businesses to unlock the full business value of their data.

“Capgemini research suggests that “data masters” enjoy up to 70 percent higher revenues per employee and achieve 22 percent higher profitability,” says Tim Lysecki. “What’s more, companies that use 7 or more external sources of data generate 37% more revenue per employee than their peers. Our platform makes it easy to plug in those external data sources without all the administrative overhead that typically implies. Our platform frees our customers from non-productive, expensive, and time-consuming data processing tasks. With more time spent on analysis, customers gain richer insights and are better placed to drive strategic development.”

ThinkData Works is working with organizations in a variety of sectors to help streamline their analytics capabilities. One leading bank had six different departments connected to the same paid third-party data source without realizing it—creating unnecessary costs and increasing management workloads for data analysts. By deploying the OpenText-based solution, the bank achieved an efficiency saving of 2,400 percent.

Similarly, a consulting firm enlisted the help of ThinkData Works to streamline the creation of proof-of-concept (PoC) exercises. Previously, skilled data analysts at the firm would spend significant amounts of time preparing and aggregating data from over 250 external datasets, limiting their ability to focus on other value-added activities. By deploying the OpenText-powered solution, the firm automated key aspects of its data preparation activities, reducing the time required to carry out PoCs by 51 percent. That drives more responsive client service, accelerating time-to-market, and enabling significant efficiency savings.

In its most recent project, ThinkData Works partnered with Palantir Technologies Canada to build a cutting-edge supply chain resiliency platform.

Brendan Stennett outlines the initiative: “In recent years, global supply chains have experienced unprecedented disruption, creating knock-on negative effects for businesses across multiple sectors. To help businesses better prepare, we’ve combined our data platform with Palantir AI and analytics tools to develop a solution that automatically predicts the likelihood of disruptive global events and warns users before they occur. Martinrea International, one of the world’s leading producers of quality metals, will be the first adopter of the solution, and the company already anticipates savings of $40 million each year by avoiding disruption.”

> **“Vertica (now part of OpenText™) has been at the heart of Namara Marketplace for over eight years. Throughout that time, we’ve seen just how reliable, cost-effective, and high performing the solution is, so continuing to innovate with Vertica (now part of OpenText™) was the obvious choice for us.”**
>
> **Tim Lysecki, Manager, Product Marketing, ThinkData Works**



Our Fungible Licensing

A few years ago, I worked as an evangelist for a midsize software company, and sometimes had to ask my boss to fund a project that hadn't been anticipated in the quarterly plan. As long as my reasons for the change were sound, he was usually accommodating, telling me, "Steve, money is fungible."

He simply meant that funding could be moved from one department to the next, or a project here could be put on hold to help a more urgent project there. At a simpler level, "fungible money" means a dollar could be divided into four quarters, ten dimes, etc. Dollars can't be produced out of thin air, but they can be re-allocated.

This is similar to the way OpenText handles licensing. OpenText's "fungible licensing" gives customers a way to choose how much of their total analytic resources to fund via one cloud vendor, how much for another, and if needed how much to fund on-premises in the data center. You purchase a license for the total size you need in terabytes, or in nodes—and how you spread that licensing across your available resources is totally up to you.

We believe this flexible approach to your overall OpenText licensing is how it should be when it comes to using data analytics software. Our competitors tend to disagree, forcing you to buy a separate license for different deployments. Often there's a separate installable for separate projects. That makes no sense to us.

Fungible licensing simplifies how you allocate, and re-allocate, your overall data analytics capability as requirements evolve. Need to shift a workload from the cloud to on-premises? No problem. Move some of your storage from Azure to AWS? It's simple.

This is just one way OpenText is taking the administrative hassles out of data analytics projects, so you can spend less time on licensing arrangements, and more time delivering data-driven insights.

Steve Sarsfield
Director Product Marketing, OpenText

The meaning of “no compromise”

McKnight Consulting puts OpenText, Snowflake, and Redshift to the test

There are many popular data analytic platforms available for the cloud today, along with many claims about performance, affordability, and ability to support multiple users simultaneously. So it can be helpful when an objective third party performs a benchmark analysis regarding those claims.

When Vertica (now part of OpenText™) announced a fully managed SaaS option for its unified analytics platform in 2021, called Vertica Accelerator, McKnight Consulting Group did a head-to-head comparison of this new offering against two other data analytic SaaS vendors: Google Redshift and Snowflake.

Here are highlights of this report

Three analytical databases compared. As McKnight describes in this report, Vertica Accelerator, Snowflake, and Amazon Redshift are "three relational analytical databases based on massively parallel processing (MPP) and columnar-based database architectures that scale. They all three provide high-speed analytics and managed storage that uses high-performance SSDs for fast local storage and AWS S3 object storage for longer-term durable, persistent storage."

The following table in McKnight's report shows basic characteristics of the three databases (note that only OpenText offers ANSI SQL compliance):

	Vertica Accelerator	Redshift	Snowflake
Company	Micro Focus*	Amazon	Snowflake
First Released	2005	2014	2014
Version Tested	11.0.0	1.0.34355	6.0.1
Storage	Amazon S3	Managed Storage	Managed Storage
SQL	ANSI SQL-compliant	PostgreSQL 8	Snowflake SQL
Massive Parallel Processing (MPP)	✓	✓	✓
Columnar	✓	✓	✓

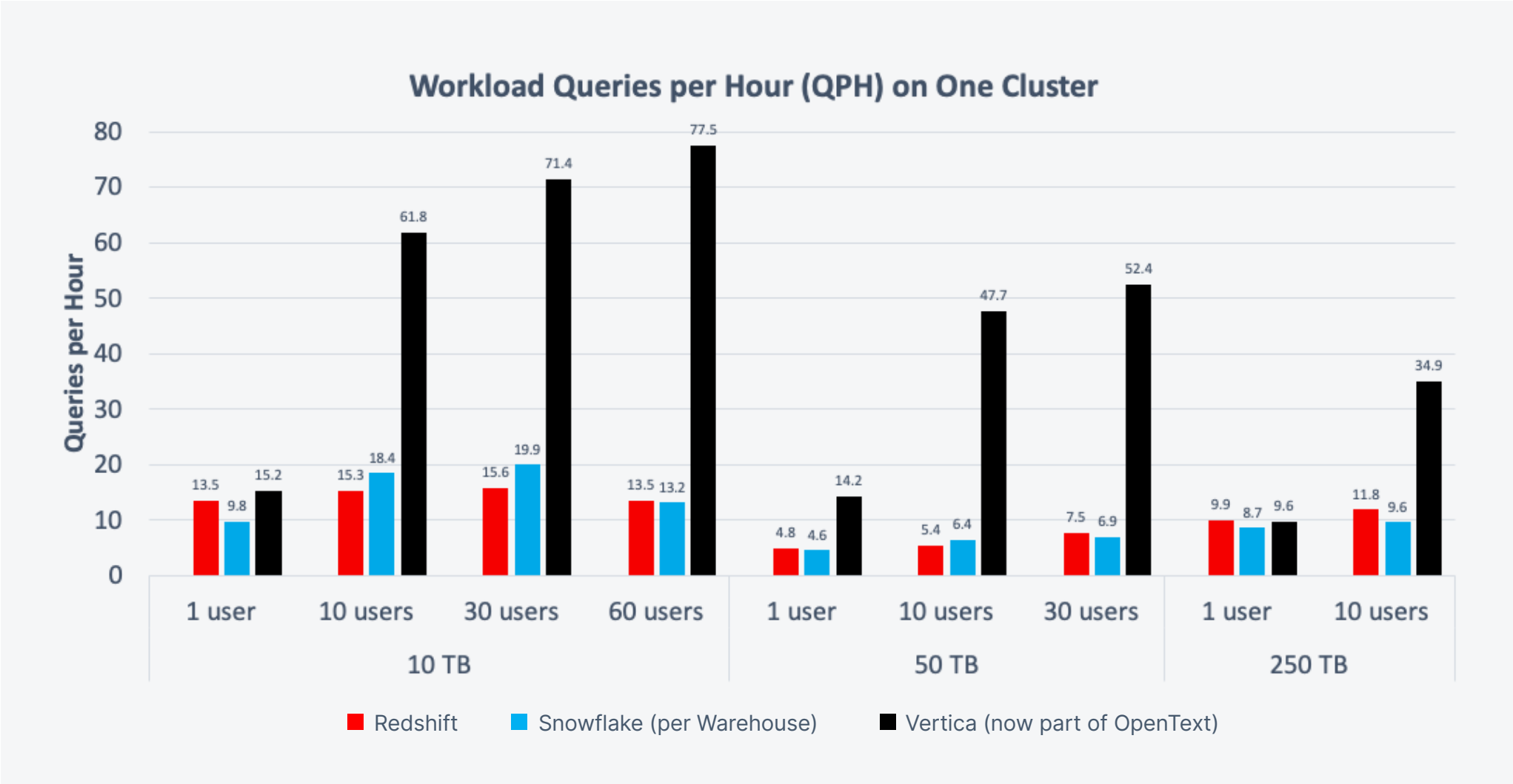
*Micro Focus is now part of OpenText

McKnight describes how essential elements of the test were prepared.

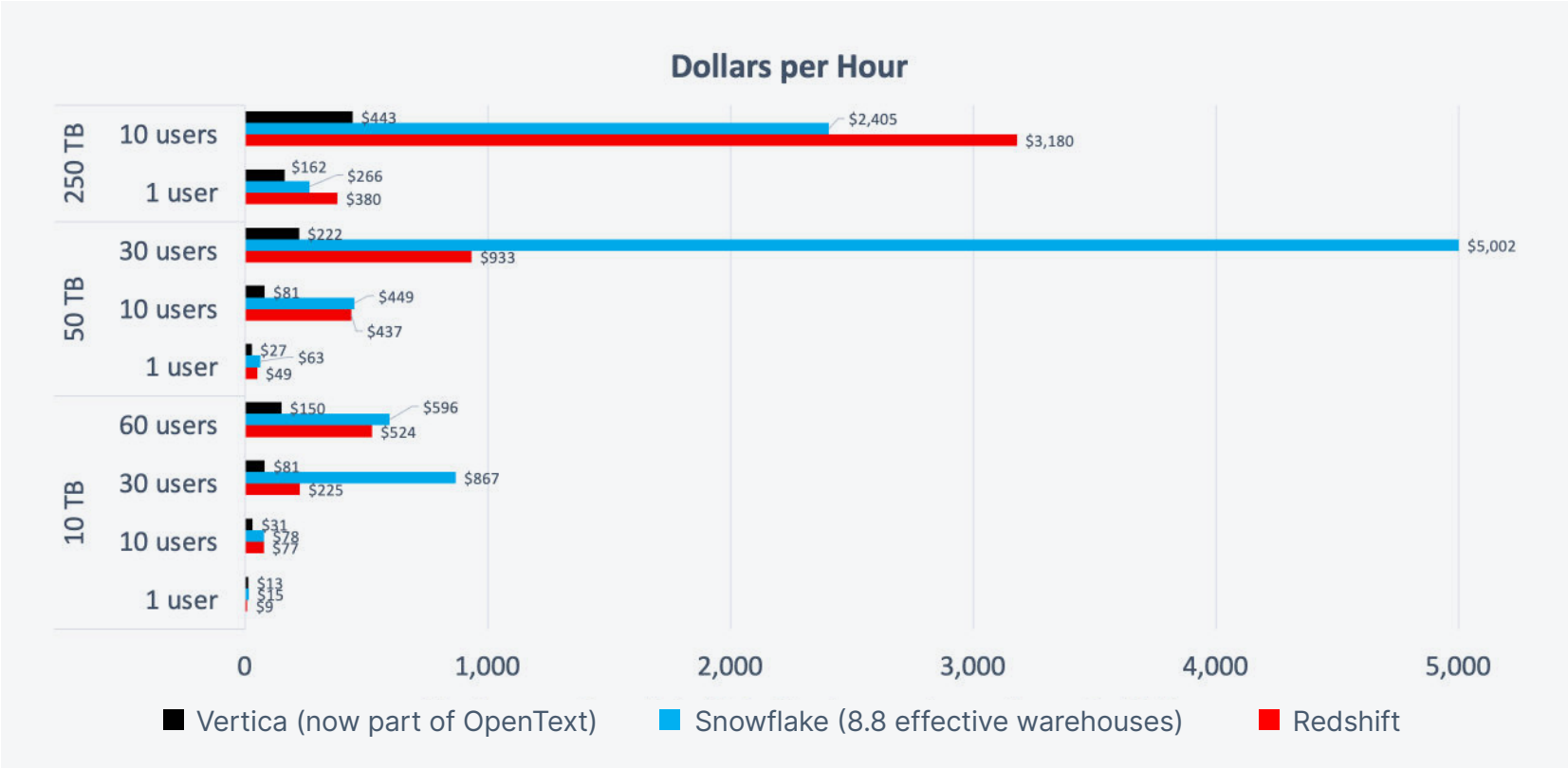
Data preparation

The data sets used in the test were an extension of the original UC Berkeley AMPLab BDB dataset. The pre-existing Big Data Benchmark (BDB) that we modeled our datasets after was provided by the UC Berkeley AMPLab. Sample AMPLab BDB datasets are publicly available in an S3 bucket at s3n://big-data-benchmark/pavlo/. For more on the AMPLab BDB Data Set, please see https://amplab.cs.berkeley.edu/benchmark."

The following graph from the McKnight report illustrate Vertica Accelerator's superior ability to complete far more queries per hour, especially for multi-users.



McKnight explains: “Vertica (now part of OpenText™) has the highest QPH for 8 of the 9 scenarios, quite often by far, especially in the concurrent profiles while Redshift and Snowflake (8.8 effective warehouses) produced similar results at 30 users of 10 TB, and all 50 TB and 250 TB levels.”



McKnight also found that Vertica Accelerator was much more cost-effective, with greatly reduced price-per-performance compared to the two other vendors:

Price per performance

"For the 10 TB workload, with 10 concurrent users, Vertica Accelerator was 60% less expensive than Redshift. At 30 concurrent users, Vertica Accelerator was approximately 63% less expensive than Redshift.

"Snowflake consistently had the highest price-performance through 50 TB. At 250 TB, Redshift showed the highest price-performance.

"For the 50 TB workload, with 10 concurrent users, Vertica Accelerator was 81% less expensive than Redshift. At 30 concurrent users, Vertica Accelerator was 76% less expensive than Redshift while Vertica Accelerator and 95% less expensive than Snowflake."

OpenText's superior concurrency

Finally, OpenText exhibited far superior concurrency (ability to support multiple user simultaneously) compared to Redshift and Snowflake. In fact, a test of Snowflake at high scale and high concurrency could not be completed, because the system "timed out and failed after two hours" as explained below:

Connect with Us
OpenText CEO
Mark Barrenechea's blog

Like what you read? Share it.

"Even with 6+ clusters, [compared to Vertica's (now part of OpenText™) 1 cluster] Snowflake performance was very similar to Vertica (now part of OpenText™) in elapsed time for workload completion at 50 TB and 250 TB with concurrency. The exception is that at high data scale and high workload concurrency—i.e., 60 concurrent users with 50TB of data, and 30 and 60 concurrent users with 250TB of data—Snowflake timed out and failed after two hours, so valid comparison numbers were not possible.

"Vertica (now part of OpenText™) has the highest Queries per Hour (QPH) for 8 of the 9 scenarios, quite often by far, especially in the concurrent profiles while Redshift and Snowflake produced similar results at 30 users of 10 TB, and all 50 TB and 250 TB levels. Higher concurrency tests at 50TB and 250TB levels could not be completed, due to Snowflake concurrency limits at that scale."

Conclusion

McKnight's conclusion leaves no doubt that, in an objective evaluation of OpenText against Snowflake and Redshift, OpenText is the clear winner:

"For our largest test on the 250 TB workload with 10 concurrent users, Redshift was over 700% the cost/hour of Vertica Accelerator. Snowflake was over 540% the cost/hour of Vertica Accelerator. Vertica's (now part of OpenText™) price-performance increased predictably with increased users while increased users generated much higher and unpredictable results with both Snowflake and Redshift.

Overall, Vertica Accelerator for AWS is an excellent choice for companies needing a high-performance and scalable analytical database in the cloud or to augment the current, on-premises offering with a hybrid architecture—at a reasonable cost."

Learn more at microfocus.com/opentext

© 2023 Open Text

opentext™