White Paper

Analytics and Big Data

How Diverse Cloud Deployment Boosts the Value of Analytics and Future Proofs Your Business

VERTICA | MICRO FOCUS

In this white paper, you will learn about three major cloud deployment methods, including in-cloud analytics, hybrid on-premises/on-cloud deployments, and a new method for using cloud object storage with separate compute optimized for dynamic or variable workloads.

Vertica in Eon Mode with Separation of Compute and Storage Architecture Addresses Variable Workloads

Cloud is a buzzword that nebulously denotes the promise and convenience of storing and accessing compute and storage resources—outside of on-premise environments. However, these days, when you use the term "cloud" in a sentence with big data analytics, there could be multiple meanings. The cloud does not represent a single method for deploying analytics. There are many ways to exploit the power of the cloud for better ROI on your analytics.

In this white paper, you will learn about three major cloud deployment methods, including in-cloud analytics, hybrid on-premises/on-cloud deployments, and a new method for using cloud object storage with separate compute optimized for dynamic or variable workloads.

Cloud Deployment Keeps Getting Better

The cloud ecosystem has long been an optimal way to start fast and keep costs low, particularly when analytical challenges don't mandate extraordinary measures. If your total data is just a few terabytes and the workload from your business analysts is manageable, the cloud is a great way to start. Cloud deployments for analytics can help you save costs, too. There are fewer burdens on database administrators and IT when you don't have to provision hardware, licenses, and deployment expertise.

New innovations have pushed cloud deployments beyond this simple story. However, lately, a new breed of cloud deployment strategies, driven by technology innovation in cloud, promises further cloud economic advantages, enabling organizations to pay for only what they need, when they need it, lowering the total cost of ownership for cloud deployments, and making it one of the most attractive ways to perform analytics on data. This white paper will help you choose the appropriate cloud deployment option for your organization, while increasing revenue and cutting costs to achieve the most effective analytics platform.

Deploy Option 1: Simple Deployment of Analytics in the Clouds

Typically, when you think of cloud deployments, you may envision taking your enterprise analytics platform otherwise running in an on premise data center and installing it on the cloud *(see Figure 1 on the following page)* and then loading data into the cloud platform. Such a cloud deployment gets you running fast without provisioning hardware. You can reduce costs with a cloud deployment, particularly server purchases, hardware depreciation, software maintenance, power consumption, data center floor space, backup infrastructure costs, and more. The cloud is a very attractive option for both the costs savings that you realize and the speed at which you can deploy.

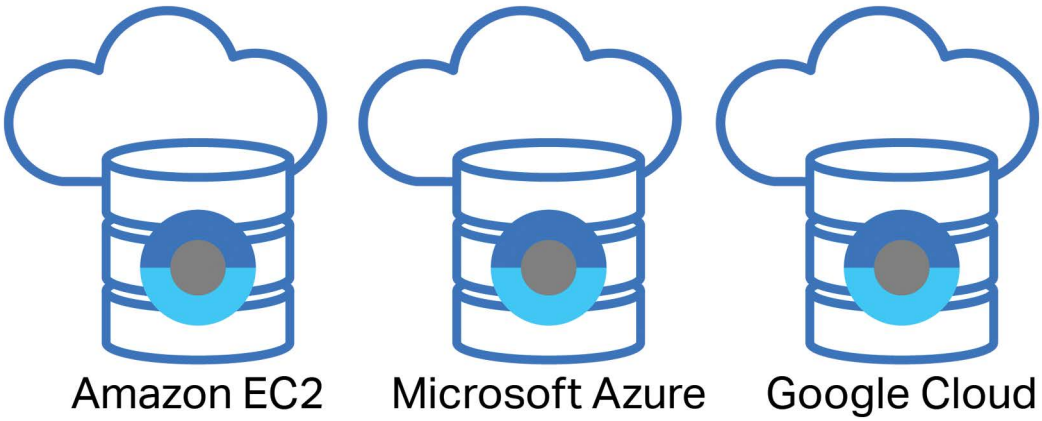

Figure 1. Vertica in the Clouds

When considering these variances for deploying an analytical platform, consider the analytics workloads that you will be asking it to complete. Will your project serve a handful of business analysts working on roughly the same workload, or will you have mixed workloads?

SAAS VS BYOL

This type of deployment comes in two sub-models for deployment and licensing: 1) You can leverage the software-as-a-service (SaaS) model to automatically spin up nodes and architecture necessary for the task, or; 2) Spin up nodes in the cloud and Bring Your Own License (BYOL) to your cloud deployment.

There are some differences between these two options that are typically SaaS versus those where you BYOL. In SaaS deployments, you typically see analytical platforms that have fewer "knobs" to control things like concurrency, workload management, memory allocation, query optimization, and more. After all, the whole point is that you can easily spin up a cluster and not worry about tuning.

By contrast, when you BYOL, set-up may take you a few more minutes, but you should have more control over how your analytical platform performs queries and allocates its resources. Features like workload management let you have more control over concurrent queries. In most corporate analytics projects, you generally will have queries that need to run fast running alongside queries that can afford to take more time. Resource pools is an important capability to help you to manage mixed workloads as well as database tuning to increase the performance of specific queries.

When considering these variances for deploying an analytical platform, consider the analytics workloads that you will be asking it to complete. Will your project serve a handful of business analysts working on roughly the same workload, or will you have mixed workloads? The old tried-and-true way of handling increased or volatile workloads is less available in many cases in the SaaS model. So, if you need an enterprise architect to look at the workload and design an optimal strategy for delivering analytics, there may be less that they can do. The architect may want to leverage a system that can sense long-running queries and assign different resources (memory, CPU, etc.) so that long-running queries don't appreciably impact daily analytics.

The fact is, it is not always convenient to migrate all systems to the cloud.

The newer auto-scaling databases that are part of the SaaS ecosystem don't offer much in tuning. After all, you can just auto-scale when queries get slow. However, this comes at an incremental cost! Make sure you can control the auto-scaling and consider the benefits of optimizing a smaller cluster rather than simply adding hardware. Additionally, many of the license agreements for the SaaS-bases analytical systems penalize you and charge for additional nodes for at least 30 days. Even if you realize that an analyst has scaled the cluster with a big query, the resources are now locked in for a committed time period.

Deploy Option 2: Hybrid Architectures

Another way that companies use the cloud is to augment on-premises systems with cloud-based systems. In this scenario, it's a matter of analyzing data in the right place for the right price.

The fact is, it is not always convenient to migrate all systems to the cloud. When companies grow, either naturally or through acquisitions, the number of systems where data resides can also grow, and become challenging to manage. Companies are locked into certain systems until it makes economic sense to move to a more modern platform like the cloud. Some companies are saddled with multiple CRM systems or ERP systems. Even mainframe applications may be left over in more mature companies.

Meanwhile, new big data is stored in growing data lakes in Apache Hadoop (HDFS) and Cloud data stores like Amazon S3. Although it makes economic sense for new analytical workloads to run in the cloud, you may need to feed the analytics with legacy systems data, too.

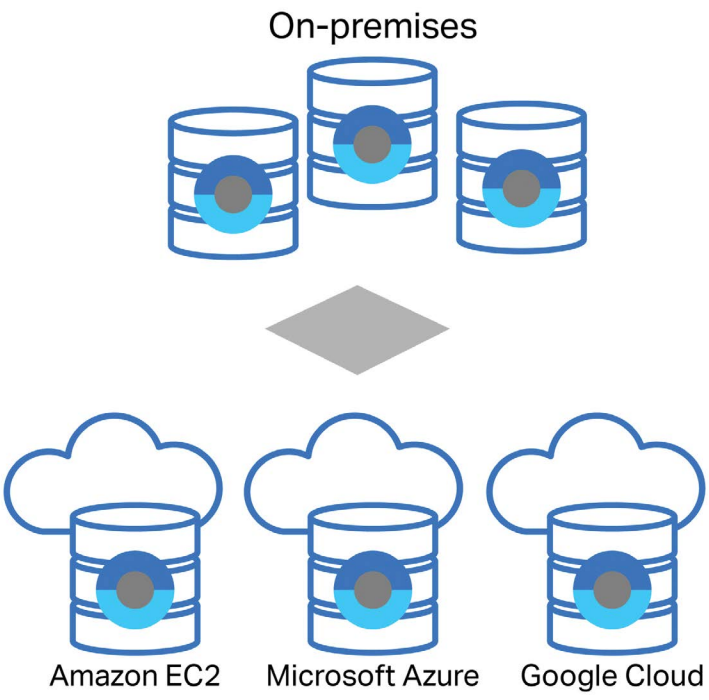

Figure 2. Hybrid: Hot Backup and Colocation

There are four key reasons why a hybrid architecture *(see Figure 2 on the previous page)* may make sense to the entire analytics community within your company:

- **External Data**—You can gain additional insight by creating JOINs between your data warehouse and data that is sitting untapped in Amazon S3 or HDFS. For example, users can leverage Web logs to gain additional customer insight, use sensor and IOT data that is sitting externally for pre-emptive service, or track the success of marketing programs by joining data with your data warehouse.
- **Data Science**—Now you can use your analytics platform to explore big data. You should be able to run your SQL queries and use your visualization tools, SQL, R, and Python on cloud-based data.
- **Training the Machine Learning Models**—A common practice is to move data, or subsets of data, from the data warehouse into Apache Spark to train your machine learning models. However, analytical platforms like Vertica can leverage cloud-based data to train your models without moving the data. You can use higher volumes of data and get more accurate models.
- **Information Lifecycle Management**—After your hot data cools off, a common practice is to move your data off to lower-cost storage in support of information lifecycle management. Cloud storage is perfectly suited for this. With Vertica, you can move cooler data out of the data warehouse into Hadoop or the clouds and still analyze it. Hadoop and cloud storage help companies store massive volumes of data at a low cost and make for operational efficiencies.

COLLOCATION AND HOT BACKUP

To protect your database from site failures caused by catastrophic disasters, many users can use the cloud to maintain an off-site replica of your database to provide a standby. In case of disaster, you can switch database users over to the standby database. This is approach is generally straightforward. During each data load process, the feed is forked to simultaneously load a second cloud-based database while loading the primary database on-premise. You can achieve this easily with off-the-shelf ETL software or writing your own code.

To protect your database from site failures caused by catastrophic disasters, many users can use the cloud to maintain an off-site replica of your database to provide a standby. In case of disaster, you can switch database users over to the standby database.

Keeping multiple copies of the database might be just the solution for disaster recovery, but it can also serve the purpose of handling data locality. Global companies seeking to store data closer to home can leverage cloud servers without setting up branch offices with server farms. Under some industry regulations, corporations must store certain data locally, which a local cloud deployment can help.

Deploy Option 3: Separation of Compute and Storage Architecture

Each of the above options work ideally with Vertica in Enterprise mode when you have a fairly steady and unchanging pace of analytics at your company. However, some organizations have dynamic workloads. In many analytics scenarios, there is an end-of-day/month/quarter calculation that requires additional computing power. You don't want to have to provision for the peak work load as it will likely cause you to incur unnecessary costs during your low-demand periods. On the other hand, provisioning computing resources below peak may cause queries to fail and will certainly take longer to complete.

For this scenario, Vertica introduces Eon Mode.

Vertica in Eon Mode delivers on the promise of cloud economics so that data engineering teams only have to provision the compute and storage resources they need from month to month, day to day, or hour to hour, while supporting efficient elasticity as new workloads emerge.

WHAT IS VERTICA IN EON MODE?

Vertica in Eon Mode *(see Figure 3)* is the same high-performance Vertica SQL engine with a new underlying data architecture that separates compute from storage. This architecture enables rapid scaling of your compute cluster to meet dynamic workloads, independent from scaling your storage to meet the growing data volumes. When activity is low, you can scale down your cluster to reduce costs. When activity peaks, you can scale up your cluster to keep pace.

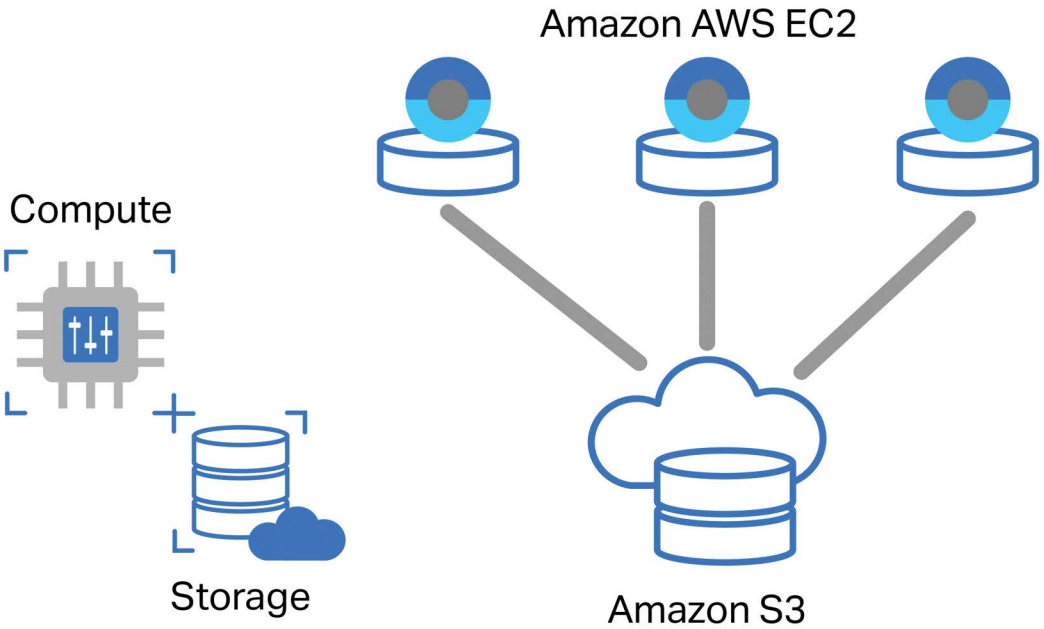

Figure 3. Vertica in Eon Mode—Separation of Compute and Storage

The Eon Mode architecture has a communal storage that stores data on a single, durable communal storage location for all data like data stored in Amazon S3. Nodes cache copies of the data for performance in ephemeral storage called a "depot." Together the communal storage and depot storage provide the durability and blazingly fast performance you expect from Vertica while at the same time enabling rapid elasticity to reduce your infrastructure spending.

Uses Cases for Vertica in Eon Mode

Vertica in Eon Mode delivers on the promise of cloud economics so that data engineering teams only have to provision the compute and storage resources they need from month to month, day to day, or hour to hour, while supporting efficient elasticity as new workloads emerge. The following use cases represent just a sample of how many industries and organizations can benefit from Vertica in Eon Mode:

Use Case	Description	Separation of Compute and Storage Benefit
Intelligent Device Early Release Testing	Medical equipment manufacturer is going to market with medical devices embedded with many sensors. It needs to run a trial at key accounts before formally launching the product.	Provides optimal resources to analyze the trial data, ensuring the device manufacturer maximum reliability upon the generally available mass release.
Experimentation and Discovery for Optimized Device Setting	HVAC manufacturers need to run periodic maintenance tests by pushing an update to edge devices, collecting a large amount of sensor data, and quickly identifying micro patterns before pushing out optimized settings.	Assists HVAC manufacturers by increasing the compute and storage resources seamlessly to accommodate for increased activity.
Engagement analysis for new product introduction	Gaming companies must increase analytical capacity during major launches or tournaments to evaluate the success of a new game or provide customer insights during a big tournament or special event.	Delivers just-in-time analytics on critical feedback by event—in time to make modifications that appeal to their community—as opposed to the ultimate launch, when it may be too late to understand the level of popularity of the game.
Analytics for seasonal sales patterns	Retailers have a variable analytic workload due to high-volume sales seasons or to close the books at the end of the month. They need maximum dashboard performance regardless of the number of concurrent users, so that key members gain access to important metrics.	Ensures maximum dashboard performance regardless of the number of concurrent users, so that key members gain access to important metrics irrespective of the load on the analytics platform and the seasonality.
Project-Based Data Science	Contract data scientists are increasingly used for project-based work, where they ramp up and ramp down analytics projects to meet specific client deliverables and schedules.	Empowers contractors with more flexibility to address client needs faster without the fear of costly overprovisioning.
Data Science and Engineering Team Collaboration	Machine Learning models are often built in the cloud using down-sampled data before the model is ultimately deployed onto a much larger, production, on-premises analytics platform.	Provides a "hybrid" bridge that allows data science and data engineering teams to collaborate on a best-of-both world's approach to machine learning model development, testing, and deployment on the full data set.

From an intelligent device equipment manufacturer to a gaming company to a retailer, all data-insight organizations can derive immense cost savings with Vertica in Eon Mode.

Understanding the Financial Benefit of Vertica in Eon Mode

From an intelligent device equipment manufacturer to a gaming company to a retailer, all data-insight organizations can derive immense cost savings with Vertica in Eon Mode. Here is an example of a typical retailer to illustrate this point.

Retailers struggle to accommodate the busiest shopping season of the year in the U.S., which begins late on Thanksgiving Thursday, continues through Black Friday and into the weekend, and concludes on Cyber Monday. The retailer must increase their compute capacity just for this seasonal peak, maybe a partial peak after Cyber Monday until about 10 days after the New Year once the returns are all in. For the rest of the year, the retailer is operating within a nominal compute capacity. Let's break it down.

Scaling capacity to workload offer significant cost saving. A traditional database (provisioned for peak workload all year long) vs. Vertica in Eon Mode will deliver savings of approximately 66%, just in compute costs alone with Vertica in Eon Mode.

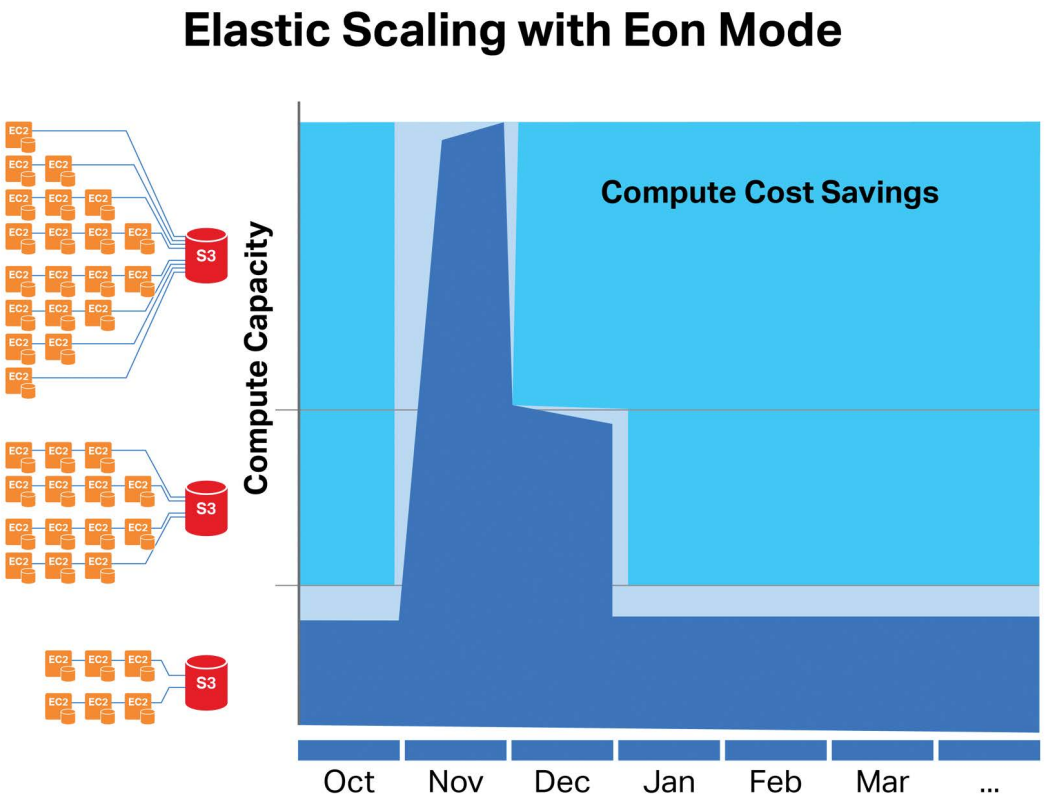

Figure 4. Vertica in Eon Mode—Cloud Economics

Assuming that the retailer needs:

- 20-node Vertica cluster for the Thanksgiving weekend and Cyber Monday
- 12-node cluster until after New Year, and
- 6 nodes for the other months of the year

Scaling capacity to workload *(see Figure 4)* offer significant cost saving. A traditional database (provisioned for peak workload all year long) vs. Vertica in Eon Mode will deliver savings of approximately 66%, just in compute costs alone with Vertica in Eon Mode. Bear in mind that usage may vary based on the analytics need and workload variability from retailer to retailer.

Perhaps your workload is uniform, just periodic. For short-term projects, Vertica in Eon Mode enables "Hibernating." Hibernating lets you shut down all compute nodes until you need them again. Come out of hibernation when your project starts up again by simply creating a new cluster and reviving your database. It is that simple.

How Do You Choose the Right Deployment Option?

When choosing between these modes, the most important thing is to consider the workload you have today and how analytics will be increased or used in the immediate future.

1. Does your organization need to simply deploy analytics in the cloud with a fairly consistent workload?
2. Will you use the cloud with your on-premises analytics to solve locality and disaster recovery concerns?
3. Can you use cloud-based data to train machine learning models?
4. What about moving colder data to a lower-cost cloud storage platform in support of information lifecycle management?

This is all possible with a modern cloud-based analytics platform.

When choosing between these modes, the most important thing is to consider the workload you have today and how analytics will be increased or used in the immediate future.

CLOUD LOCK-IN

Beware of vendor lock-in. Make sure that if you choose any cloud platform for analytics that you can easily shift gears when needed. Many factors can necessitate the need to switch between on-premises and cloud. When the cloud vendor raises prices or if better pricing exists in other clouds can you easily move your platform? If your company is acquired, can you merge your analytics platform with the new systems?

You may be familiar with the lock-in phenomenon from dealing with traditional enterprise database companies. You begin to use a vendor only to realize that not only is it difficult to move off of the platform, but hidden cost and add-ons creep into the deal. You're locked in. The same is true with selecting a cloud vendor. It is very important to set up your analytics platform so that you maintain enough freedom and negotiating power to settle on a solid agreement for your business. Make sure the system you choose is flexible and provided you with freedom from underlying infrastructure.

See the matrix below how Vertica provides freedom and flexibility to organizations for all the three deployment options:

Criteria	Option I—Simple Deployment of Analytics in the Clouds	Option 2—Hybrid Architectures	Option 3: Separation of Compute and Storage
Vertica Modes	Enterprise Mode		Eon Mode
On-premise, VM, Google, Azure	Yes		No
AWS Cloud Deployment	Yes		Yes
Operational Characteristics			
Workload Suitability	Fixed (Fixed—Consistent number of Nodes)		Dynamic (Seasonal, Hour to hour, etc.)
Workload Isolation SLA Objectives	No		Yes
Grow data indefinitely while paying only for compute	No		Yes—,via AWS Paid Listing
Storage Type	Local Attached Disk (Raid 8 for Durability and Performance)		Shared S3 Object Store—Durability (Ephemeral Cache disk for Performance)
Pause all compute nodes	No		Yes
Disaster Recovery—Cluster Replication	Backup/Recovery		S3 Snapshots

Continued on next page

Vertica provides your business with a number of deployment choices—the option to deploy analytics in an on-premise data center, in the cloud or in a hybrid environment. Vertica also enables you to buy only the compute power you need and reduce storage costs based on use case requirements.

Criteria	Option I—Simple Deployment of Analytics in the Clouds	Option 2—Hybrid Architectures	Option 3: Separation of Compute and Storage
Features			
Advanced Analytics	Yes		Yes
Machine Learning	Yes		Yes
External Data	Yes		Yes
Community Ed | BYOL | AWS Marketplace Hourly Pricing*	Yes		Yes

** Vertica is also available on the AWS Marketplace as an hourly paid subscription for organizations who prefer usage-based consumption and whose financial structure prefers OPEX vs. CAPEX .*

Summary—Demand the Freedom to Choose Your Deployment Model as Needs Evolve

Most organizations have analytic use cases that require consistent compute requirements as well as use cases that have variable compute needs and the ability to query data lakes. Vertica provides your business with a number of deployment choices—the option to deploy analytics in an on-premise data center, in the cloud or in a hybrid environment. Vertica also enables you to buy only the compute power you need and reduce storage costs based on use case requirements. Organizations need to support the right use cases today, knowing with certainty that they may change in the future. To achieve that end, Vertica is capitalizing on cloud infrastructure to provide organizations that choice and flexibility. However, deployments in the cloud are always likely part of a larger, more hybrid strategy, and that fact must be kept in mind as analytics initiatives evolve.

Learn More At

www.vertica.com

Additional contact information and office locations:
www.vertica.com

www.vertica.com

VERTICA | MICRO FOCUS®