EXECUTIVE SERIES

THE GORILLA GUIDE TO...®
EXPRESS EDITION

Modernizing Data Analytics

Joey D'Antoni

INSIDE THE GUIDE:

- Get More from Your Analytics
- The Power of Hadoop and Vertica
- Dive Deep into HPE's Turnkey AI Solutions Platform

Compliments of

MICRO FOCUS

VERTICA

Hewlett Packard Enterprise

ActualTech Media

actualtechmedia.com

THE GORILLA GUIDE TO...

Modernizing Data Analytics

Express Edition

AUTHOR
Joey D'Antoni

EDITOR
Keith Ward, ActualTech Media

PROJECT MANAGER
Wendy Hernandez, ActualTech Media

EXECUTIVE EDITOR
James Green, ActualTech Media

LAYOUT & DESIGN
Olivia Thomson, ActualTech Media

ACTUALTECH MEDIA

6650 Rivers Ave Ste 105 #22489

North Charleston, SC 29406-4829

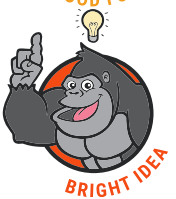
www.actualtechmedia.com

TABLE OF CONTENTS

CALLOUTS USED IN THIS BOOK

The Gorilla is the professorial sort that enjoys helping people learn. In the School House callout, you'll gain insight into topics that may be outside the main subject but are still important.

This is a special place where you can learn a bit more about ancillary topics presented in the book.



When we have a great thought, we express them through a series of grunts in the Bright Idea section.

Takes you into the deep, dark depths of a particular topic.

Discusses items of strategic interest to business leaders.

ICONS USED IN THIS BOOK

DEFINITION

Defines a word, phrase, or concept.

KNOWLEDGE CHECK

Tests your knowledge of what you've read.

PAY ATTENTION

We want to make sure you see this!

GPS

We'll help you navigate your knowledge to the right place.

WATCH OUT!

Make sure you read this so you don't make a critical error!

It's Science, Not Witchcraft

Welcome to The Gorilla Guide To…® Modernizing Data Analytics. If you're looking to get more out of your data, and find ways to move that often-dormant information in more profitable directions, this short book is for you.

Data is the lifeblood of every organization, and there's more of it coming in than ever. But the reality is that doing more with that data than the things you've always done is scary. In a way, ideas like artificial intelligence (AI) and machine learning (ML) almost seem like dark arts—collecting and making use of your existing data for these purposes is something done with wands and smoky cauldrons, with magicians waving their hands over the fumes.

But it doesn't have to be that way. There are real, valid ways to extract your data and turn it into actionable bits that work for you and your bottom line. That's what this Guide is about.

It provides an overview of the current state of data analytics, and goes into detail on ways to level up your operations in that area. In a shorter amount of time than you think, you can be doing things with your data you may

have thought were only possible with hyperscaler-sized budgets and armies of data scientists.

That's not the case, though—what you need are the right tools, and partners who really *get* this stuff. There's help available to get you started, or to ramp up your existing efforts to unlock more value. Come along and find out how to get and do more with less risk, and outstrip your competition.

It starts with a discussion of where much of the new data is being created: the edge.

CHAPTER 1

IoT: From Edge to Core to Cloud

Whether gathering user data from mobile devices or sensor data from manufacturing lines, much of modern computing has shifted to the "edge." The edge is a computing term, but it's also the location where transactions take place—a retail sales floor, your home, a power plant, an oil rig, or an airplane, but not in a centralized data center as in the past.

The technology that brings intelligence to the edge encompasses three "C's"—connect, compute, and control: The devices involved are nearly always connected to the Internet, allowing them to transmit data to a central analytics system; they all have microprocessors, so they can provide computing power to enable access to applications, make decisions, and filter data; and that computing power lets these devices control—change settings to manage specific scenarios or to orchestrate certain actions. Such devices are connected to sensors and actuators to control things—from smart lightbulbs in your home to a turbine in an electric plant.

Edge computing led to a design pattern called Lambda architecture, which manages large volumes of data by leveraging both stream and batch processing methods. The streaming layer is used to process data in real time and is useful for identifying anomalous values, like a temperature being out of range, and then either taking an action or firing an alert. The batch layer stores the large volumes of data coming from devices and provides the computing power needed to drive analytics.

Whether for data science or more traditional data warehouse queries, the batch layer needs to be able to rapidly load and query immense amounts of data. In addition, the edge allows organizations to capture insights from new sources located closest to customers and operations. The need to quickly turn data into insights and actions necessitates a shift in how data is captured and processed locally. This in turn creates the need within applications to drive increased intelligence and automation with artificial intelligence/machine learning capabilities.

Luckily, there's a platform that can meet all of these challenges. It's called Vertica and, as a scale-out columnar database, it can both ingest large volumes of data and serve massive queries very quickly while also providing faster predictive analytics through in-database machine learning.

The Vertica Platform

Legacy databases weren't designed to deal with the volumes of data modern analytical systems have to handle, and they require extensive tuning, optimization, and proprietary hardware in order to meet the pace of business.

Vertica takes a different approach. It's what's known as a columnstore database, which offers a number of benefits for large-scale analytic workloads. The data is stored in columns, rather than in the traditional mixed data pages of a legacy relational database. This allows for much higher degrees of compression, which greatly reduces the storage space required for the data, and that reduced space in turn allows the data to be processed much more quickly.

The massively parallel engine enables scaling out the workload across multiple servers, and it means data can be stored in an ordered fashion that improves performance on frequently run queries. Some public cloud platforms offer similar architectures running in a Platform-as-a-Service (PaaS) model, but the costs of the service can grow exponentially as consumption grows.

Vertica also supports connections to a wide variety of "Big Data" platforms, such as MapR (now part of HPE),

Which Vertica Is Right for You?

Vertica comes in two flavors: premium and express. Premium Edition has no limits on features, nodes or data size.

Vertica Express Edition includes a subset of Premium Edition features, at a reduced price. For example, it includes all the basic features such as Workload Management, but doesn't have more advanced features like Live Aggregate Projections.

Vertica for SQL on Apache Hadoop is a separate product with its own license.

Vertica Community Edition (CE) is free and allows customers the following:

Store and analyze up to 1TB of data for free with no time limit. Install Vertica CE on-premises, as a VM, on Apache Hadoop, or in your choice of clouds (AWS, Azure, Google).

- Install Vertica CE on up to 3 nodes
- Store and analyze up to 1TB of structured and semi-structured data

- Use Vertica for SQL on Apache Hadoop for data exploration as part of Vertica CE free trial
- Enjoy no time limits or license requirements
- https://www.vertica.com/register/

Cloudera, and Apache Spark. This allows Vertica to be an analytics hub, modernizing data architectures.

Vertica ships with R and Python, as well as with prebuilt machine learning options within the database, allowing your data science to take place next to your data, which reduces the complexity of data movement and improves overall system performance. Vertica also ships with time-series and geospatial data tools, allowing you to quickly build reports on advanced data structures.

This architecture means Vertica can ingest data rapidly, with enough performance to let analysts explore data interactively and enable everyone to make data-driven decisions. Vertica customers report that compared to legacy database platforms, they see as much as a 50x boost in performance, so a report that used to take an hour now runs in less than two minutes.

Vertica also applies deeper analytics capabilities that support the scale and performance of today's speed of

analytics, thus creating insights and driving actions at the edge. And, best of all, Vertica is an enterprise-ready ANSI-SQL analytics solution that's easily deployed by existing IT staff and doesn't require a highly skilled team of data scientists.

Implementing the Intelligent Edge

Edge implementations typically consist of a variety of hardware devices, leading to difficulties in monitoring, connectivity, and standardization. HPE Edgeline systems meet that challenge with three points of convergence of IT/OT systems that enable building an intelligent edge: deep analytics/compute, data acquisition/control, and enterprise management.

The deep compute engine handles data aggregation and preparation and provides artificial intelligence by running machine learning models at the edge, in a hardware package designed to work in hardened environments with extreme temperatures and conditions. This package can be combined with systems likes Spark or Kafka to do stream processing, and you can use Vertica on the Edgeline to handle edge use cases like meter readings.

While these tools are very powerful, not everyone can take advantage of them: many analysts experienced in working with business intelligence platforms and data warehouses lack similar expertise with open source tools like Spark or Kafka, which require a different set of development skills. While most data scientists possess these skills, many organizations are still growing that skillset. Using Vertica as an edge tool allows for the BI analysts to utilize their existing skill set on edge data.

Bringing intelligence to the edge can reduce both network and decision-making latency, shrinking the amount of network bandwidth required. It can also protect data against network threats and data corruption from intercepted traffic or broken connections.

Vertica and Hadoop: Better Together

Big data workloads are defined by the five "V's": velocity, volume, value, variety, and veracity (see **Figure 1**). "Velocity" and "volume" are implied with edge workloads—there's a lot of data, and that data is typically "as it happens" data, so it can flood the target system in no time. The "value" proposition is specific to the data as it relates to your business, and "variety" can refer either to the type of data (video, audio, documents, meter

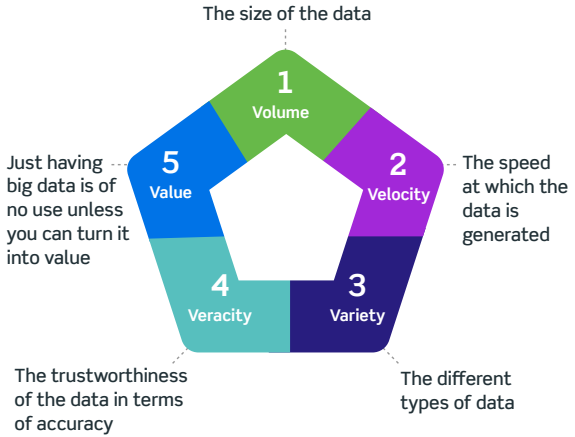


Figure 1: The five "V's" that define big data workloads
Source: Science and Education publishing (SciEP)

readings) or the business function of the data (supply chain versus sales, for example).

Hadoop is the most broadly adopted big data system, but there are challenges: Many organizations that made investments in Hadoop failed to see any business benefit, because performing analytics on massive volumes of unstructured data isn't easy. The tools used to analyze data in Hadoop were difficult for developers and analysts, and the methods of data extraction differed from normal SQL queries.

But there's enormous business value in capturing large volumes of data and running fast analytics for deeper insights at the edge. Most modern implementations use a data lake methodology to capture raw data in its original format.

Vertica can help accelerate your time to value with this data by providing an interface that enables access to scalable ANSI SQL. Combining Vertica with Hadoop ecosystems like HPE MapR Data Platform builds a complete, integrated analytics solution that offers fast, native SQL on Hadoop without the hassle of connectors. And with a single data set from Hadoop and Vertica, analysts avoid writing map reduce jobs and using unfamiliar tools, and simply write queries or use a reporting tool to analyze data.

Vertica also supports business intelligence solutions such as Tableau, QlikView, and MicroStrategy.

Securing Your Data with Voltage SecureData

Given the volume of data that can be stored in a system like Vertica, it can become a target for hackers. The common solution in databases has been to simply encrypt the data in the database, but that means the data has to be decrypted for use in other systems like business intelligence tools.

The only way to protect data end to end is to put the encryption controls within the data itself using format-preserving encryption (FPE), which persists the protection as the data moves from database to application. Voltage SecureData offers datacentric security that includes FPE and works with big data platforms such as Vertica and Hadoop.

Let's look at a quick example that shows how FPE—and Voltage SecureData—can help. Suppose you want to encrypt U.S. Social Security numbers (SSNs), which are stored in the format nnn-nn-nnnn. If you send SecureData a value, such as 123-45-6789, to encrypt using the SSN format, SecureData replies with, for example, 452-32-2323, which means the encrypted values don't require any changes to the underlying data structure. Not only is this much easier to implement, it means you can expose "real" values for a portion of the data, such as showing the last four digits of the SSN to a customer service representative.

The Best Design for Your Architecture

The number of edge computing devices continues to grow exponentially year over year, and designing IT architectures to capture, process, analyze, and derive deeper insights from that data will be one of

the defining problems of the next decade. Using HPE Edgeline Systems for your edge analytics, in conjunction with Vertica, lets you process massive amounts of data at the edge.

The advanced analytics offered by Vertica enables deep exploration of the data, and when combined with HPE MapR Data Platform, allows you to derive insights from petabytes of data.

And you can keep all that data secure using Voltage SecureData, which can directly integrate into your Vertica environment and encrypt your data with virtually no changes to your data structure or reports.

Once you've captured all that edge data, where do you go next? That's what we explore in Chapter 2, so read on.

CHAPTER 2

Modernizing the Enterprise Data Warehouse: How to Level-Up Hadoop Environments with Vertica

The journey to modernizing your data analytics strategy isn't always a straight line, and it depends on your starting place. Organizations typically fall into one of these categories:

- Those with little data analytics capacity but a great deal of meaningful data to analyze
- Those who have data analytics capacity but fall short on delivering the expected results and value. These organizations frequently have data in silos throughout the organization.
- Those who have moved to a cloud data warehouse or Data Warehouse as a Service model, and are often challenged with ever-growing infrastructure costs

Each type of organization confronts a different set of challenges to reach its data analytics goals. There are myriad options in the business intelligence space, and,

depending on their maturity level and skills, organizations face a series of decisions around cost and technology to best meet their needs.

Whatever option is chosen, the foundation of any solution can be Vertica and MapR Hadoop, which work together to meet the data needs of any company. Vertica offers a scale-out, massively parallel processing data warehouse with in-memory query execution, and native ANSI SQL with in-database machine learning functionality.

Hadoop provides a converged data platform that can ingest petabytes of data and support semi-structured and unstructured data types. Vertica integrates with Hadoop to provide a best-in-class SQL-on-Hadoop layer, with a rich collection of analytic functions that can perform queries directly on both Hadoop and the data warehouse without moving the data.

Together they give companies a powerful, unified platform to perform advanced analytics on massive volumes of data, wherever it resides, with the SQL skills they already have in their organization.

With that in mind, let's look at how each type of organization can build the right data analytics solution for their business.

Greenfield Analytics

A firm that's new to this journey has multiple options and minimal technical debt, but its decisions should be driven by its source data and its business goals. An example of such a firm might be a startup Internet-based company with a mobile app that captures data in JSON format. It might also look at data from the web, and application interactions might be processed in its analytics environment.

This is a firm that likely has decent software development skills but limited experience with building a data model or executing analytical tasks. Depending on the volume of data, it could deploy a Vertica data warehouse solution and, as its data volume grows over time, add Hadoop nodes to achieve a lower cost for long-term storage of its data.

Typically, such a firm would leverage a business intelligence tool like Tableau or QlikView to provide a presentation layer for reports. From a skillset perspective, the firm would want to ensure that it had SQL development expertise and, as it moved into deeper analysis of its data, it should evaluate adding data science skills to move from real-time to predictive and eventually prescriptive analytics.

Siloed Data—All the Hadoop Clusters

Within large organizations, this challenge typically results when funding for such projects comes from the business side of the organization rather than a central IT function. Each business function declares its Hadoop cluster its own sandbox, and doesn't want to let other parts of the organization share its cluster.

A "data silo" is a data repository under the control of one part of an organization. As such, the data is cut off and isolated from the rest of the company, making it much more inaccessible.

Data silos tend to grow organically within enterprises, since each department has different priorities, mandates, and goals. Silos typically stifle innovation, as well, since the data can't be used by other departments for different activities.

This leads to two major problems—proliferation of Hadoop clusters, and data silos throughout the organization. Or, even worse, the data doesn't make it into

a Hadoop cluster and lives on file shares or in Excel spreadsheets. In any case, the company isn't maximizing its data insights by centralizing all its data.

By moving to a converged data platform, organizations can lower their total cost of technology ownership, offer a common analytics platform, and gain deeper insights by treating organizational data as a holistic data set, allowing data scientists to see a broad picture across the firm.

Many large enterprises have moved to a model of storing all data in a data lake using Hadoop, then adding other analytic tools on top. This gives the advantage of a central data store and allows for granular security. Vertica for SQL on Hadoop provides the performance and functionality that organizations have often lacked when trying to perform analytics on their Hadoop environment.

Overloaded with Data Capacity—My Cloud Bill Is Too High

Cloud computing provides organizations many benefits, including workload flexibility, operational efficiency, and ease of use. A major challenge, though, is predicting monthly costs, especially for Platform-as-a-Service (PaaS) offerings. It can be difficult because you can't simulate your environment in either virtual machines or on-premises hardware.

This may not always be a problem, but in the case of rapid growth it can lead to unpredictable costs, which no business wants. In some cases, the platform may simply not be able to meet the performance requirements of the customer without drastically increasing compute spend.

In an ideal cloud world, customers would increase their workloads in a linear fashion. As the workload grows, they'd add nodes and possibly resize existing nodes. Both Vertica and MapR allow you to scale your cluster as your data volumes and workload grow.

Rather than constantly adding expensive compute resources as data volumes grow, Vertica gives customers the control and freedom to tune queries and optimize performance. Vertica can typically run on smaller amounts of hardware than other MPP data warehouse platforms. Whether you're on-premises or in the cloud, this lets you predictably manage your costs. If your workload is seasonal, for example, you can easily scale your cluster up or out to meet peak demand.

Changing Analytics Requirements

A challenge many organizations face is moving from traditional, historical analytics to a more modern, near-real-time analytics system. The major difficulty in making this shift is that it requires changing the

extract, transform, and load (ETL) process for the data warehouse from something that runs overnight or every four hours into a process that manages streaming data. This challenge is twofold—the ETL process needs to be rebuilt from the ground up, and the data warehouse needs to able to manage streaming data.

Traditional data warehouses are designed and optimized for batch loading, not for streaming. Vertica provides high performance for ingesting data streams at

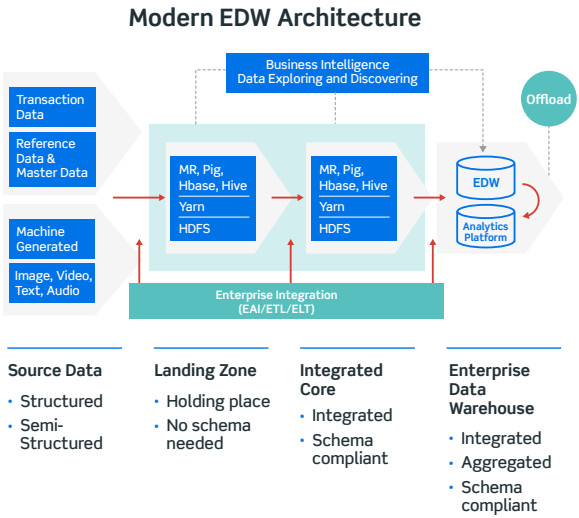


Figure 2: A modern enterprise data warehouse architecture

low latency, whether it's sentiment analysis data from a new social media campaign or sensor data coming from Internet of Things (IoT) devices, using its streaming message bus technology. It also allows for data to be streamed to other targets (see **Figure 2**).

Vertica and Hadoop—Better Together

While Hadoop is a powerful platform, its ecosystem is challenging, especially to non-developers. Many of the data interfaces are problematic for business analyst users. And while there are some SQL tools that work with Hadoop, they can be limited and may not run a full set of queries to perform the needed analysis.

Such users are able to query Hadoop tables through Vertica, which allows external tables to be created that can be queried using standard ANSI SQL functionality. This lets you leverage your existing IT and analyst staff to modernize your analytics practice.

Procurement Options

Modern IT offers an abundance of hardware and software options. Depending on the nature of your business, you may want to acquire hardware and software through Hewlett Packard Enterprise as a capital expense. This

gives you the tax advantage of deprecation, as well as full ownership of your solution.

Consider HPE Elastic Platform for Analytics (EPA) that enables independent scaling of compute and storage through infrastructure building blocks that are optimized for density and disparate workloads. Combine it with HPE MapR converged data platform that offers data services for ingesting, storing, and managing data, and you have a powerful data analytics solution.

If you prefer a more cloud-based model for meeting your IT needs, choose HPE GreenLake, which aligns to your capacity usage for easy scale up. This allows you to purchase hardware, software, and support on an ongoing basis in a consumption-based model, offering the ultimate flexibility as you can light up more hardware as your data needs grow.

Vertica is a software solution, completely free from underlying infrastructure, and can be installed and run on the right type of HPE hardware for your organizational and budgetary needs.

Those tools are great, but what if you need an entire platform, either on-premises, as a hybrid, or delivered entirely as a service? A platform that incorporates AI/ML seamlessly? There are solutions for whatever need you have, and that's the focus of Chapter 3.

CHAPTER 3

Future State of AI Architecture: The Turnkey AI Solutions Platform

Data science, ML, and AI are all the buzz in business technology. Various vendors have made efforts to reduce barriers to entry for using these technologies or integrated them into existing projects. While this can be beneficial for helping users organize their calendars or help your customers via a chatbot on your website, the types of ML/AI projects that really drive business value are much harder to achieve.

There are several reasons for this, but the first is that the toolset in this space is massive—there are open source tools, SDKs from vendors like Microsoft and Google, independent models, and curated ISV solutions. Data wrangling from a variety of internal and external data sources often requires a combination of system administration, development, data engineering, and mathematical skills that are a challenge to find in a team of people, much less one or two hires. Additionally, to hire these rare individuals, you're likely competing against

large software firms, financial services companies, and private consulting firms—all who can afford to pay them *a lot*.

Another option is to try to elevate the skills of your business intelligence team from retrospective analysis to a predictive model. This approach is particularly effective when the questions you're asking of your data are straightforward, like changing production volumes based on historical trends. But those teams may lack the skills needed to integrate more advanced external datasets and build ML pipelines to improve the accuracy of models.

Vertica for Data Science

Gartner Inc. reported in January 2019 that 80% of analytics insights projects won't deliver business outcomes through 2022, and that 80% of AI projects will "remain alchemy, run by wizards" through 2020. Key to building a successful data analytics project is getting the right datasets to the project, and being able to quickly perform analysis on the data.

Even a traditional data warehouse project involves pulling together data from a wide variety of transactional systems. In most warehousing projects, 70% to 80% of project effort is spent on building extract, transform,

Biggest AI Hurdles for Organizations

The top challenges to adopting AI for respondents were a lack of skills (56%), understanding AI use cases (42%), and concerns with data scope or quality (34%). "Finding the right staff skills is a major concern whenever advanced technologies are involved," said Jim Hare, research vice president, Gartner.

and load (ETL) processes, which convert the data into a suitable format for analysis.

In a modern analytics project that combines traditional business intelligence (sales, inventory, production, and so on) with mobile app, clickstream, and social media, data is far more complex, especially when trying to integrate with modern data storage technologies such as object stores or Hadoop. While open source systems are robust for storage and data analysis, integrating them into an enterprise data landscape is challenging because of varying connectivity and programming language options.

One answer to overcoming these challenges, as you've already seen, is Vertica. Vertica easily integrates with Hadoop and provides a full ANSI-SQL interface across Parquet, ORC, JSON, and many other data formats. This means your analysts can quickly map and begin querying other data sources or your data lake, using Vertica's SQL-on-Hadoop engine. You can load structured data directly into Vertica, or take advantage of data virtualization using external tables on HDFS.

Built-in Machine Learning

One of the benefits of using Vertica for ML projects is the in-database framework for advanced analytics and ML. This allows your analysts to quickly analyze your data using familiar SQL algorithms and combine them with ML toolsets like R and Python.

Vertica's built-in ML functions include linear and logical regression, k-means clustering, and Bayesian analysis, among others. This enables you to prepare your data for normalization, outlier detection, and sampling, and to create, train, and score those models using SQL skills commonly found in enterprise organizations.

The power of Vertica's column-oriented Massively Parallel Processing (MPP) architecture means you can execute queries over hundreds of terabytes of data very

quickly, on a single platform. Developers also have the option of creating functions in R, Python, Java, and C++ for custom applications. Unlike many open source tools, Vertica has full commercial support and provides regular upgrades and updates to its functionality. This delivers low management overhead for your data analytics platform.

HPE Container Platform, HPE MapR, and Vertica

Hadoop is an established platform, but it lacks many enterprise management resource controls, and generally requires a team of skilled administrators to secure and manage your clusters. Historically, multi-tenancy has also been a challenge in the environment. The HPE MapR Data Platform offering has always been at the forefront of reducing these challenges to enterprises, by minimizing the management effort through enhanced tooling and easier enterprise security integrations.

The HPE Container Platform is an enterprise-grade container platform that supports both cloud-native and non-cloud-native monolithic applications with persistent data. It includes innovations from HPE's recent acquisitions of BlueData and MapR, together with open source Kubernetes for orchestration.

HPE MapR is used by companies for storing and processing large amounts of data. The HPE MapR platform provides a number of capabilities for running distributed applications. The software exposes storage APIs for the Amazon S3 API, to go along with APIs for HDFS, POISX, NFS, and Kafka.

MapR was acquired by HPE in August 2019.

BlueData has a proven track record of deploying non-cloud-native AI and analytics applications in containers, and HPE MapR brings a state-of-the-art file system and data fabric for persistent container storage. Now enterprises can extend the agility and efficiency benefits of containers to more of their enterprise applications—running on either bare-metal or virtualized infrastructure, either on-premises, in multiple public clouds, or at the edge.

Modern Data Flow

Vertica is typically used as a big data analytics engine because of its high performance. However, with the massive data volumes coming from Internet of Things (IoT) devices and other streaming data sources, many

organizations dump the raw data into a data lake, which represents a lower-cost storage platform for the undistilled data.

Kafka, a streaming query engine, is also commonly used in this architecture to provide alerts on out-of-bound values in real time. A good example of this is data coming from an airplane engine—most of the data is uninteresting, except for longer-term analysis. High temperatures, on the other hand, might require action as soon as the plane touches the ground.

The full data set will typically live in HDFS on Hadoop. The HPE MapR Data Platform provides some major enhancements around security and performance. HPE MapR's approach to security is that the product is secure out of the box, offers ends-to-end encryption, and provides a unique data governance and lineage solution, allowing you to track the data in your environment. HPE MapR also offers enhanced performance over open source Hadoop solutions by using the HPE MapR XD file system that accesses storage directly.

BlueData EPIC—a Component of HPE Container Platform

The container-based BlueData software platform is the foundation of the HPE Container Platform. With BlueData EPIC software, you can create distributed environments for ML, data science, and analytics in minutes rather than months. You can offer a self-service experience with the data and tools that your data science teams need, while providing enterprise-grade security and reducing costs.

The additional value BlueData EPIC brings to the table includes:

- IOBoost™, an application-aware caching service to reduce the cost of IOs against a large data set
- DataTap™, which allows in-place access to data, reducing costly data movement inside of a cluster
- ElasticPlane™, a management control plane that allows for multi-tenancy, and even multi-cloud deployment and management.

ElasticPlane allows for automated integration with your Active Directory or LDAP solution, and management of your containerized AI and big data solutions can be easily secured and managed (see **Figure 3**).

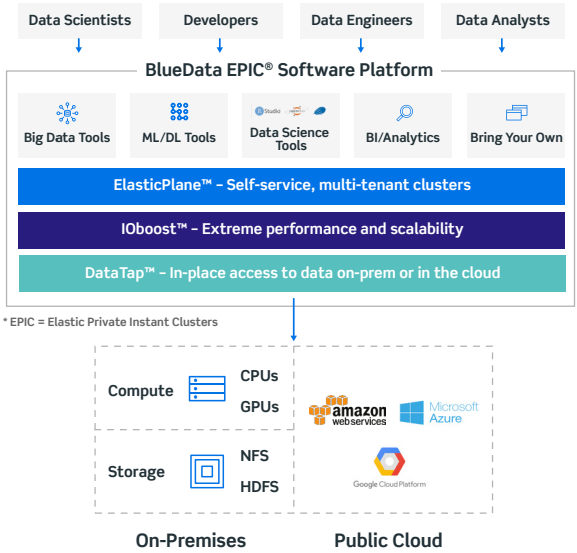


Figure 3: BlueData EPIC—a component of HPE Container Platform.

HPE GreenLake

Choosing the right amount of computing resources and software for your growing data environment can be a challenge. Frequently, firms will underestimate the amount of resources they need as new data sources come into projects and require more software, storage, and compute to meet the demands of the business. In

many cases, the business moves data sources into new computing paradigms over time and the IT organization can remain over-provisioned for years.

The HPE GreenLake model treats software and infrastructure, both compute and storage, like public cloud resources. Software licenses are scaled on an as-needed model. GreenLake right-sizes the solution from Day 1 and adds a buffer layer of hardware that isn't paid for until it's consumed, and grows as the consumptions increases. This allows for nearly on-demand scaling of software, compute, and storage resources.

Derive Insight from Large Volumes of Data

While big data analytics is a rapidly evolving space, the technology stack has settled down in recent years. Most organizations are using some combination of Hadoop and an in-memory database engine like Vertica to derive insight from large volumes of data.

This gives a couple of benefits—the first is that developers and business analysts who have vast experience with the SQL programming language can quickly get up to speed. The second is that the reduced cost for storage in a platform like Hadoop can be realized for a significant portion of the data store.

This architecture makes data and ML options available outside the elusive realm of a few data scientists or statisticians that you have on staff. It also allows for you to use your favorite business intelligence visualization tools like Tableau or Qlik to get your data answers in front of business users.

The Vertica/HPE MapR/HPE Container Platform combination allows businesses to have a production, Internet-scale big data solution very quickly by reducing the integration challenges that many open source projects face. When combined with the ease of deployment of HPE GreenLake, you can derive new business insights in minutes instead of months.

Things to Think About

Where is your enterprise in terms of analytics maturity? Are you just getting started? Or do you have a robust data warehouse that you'd like to augment with outside data sources, or even social or clickstream data?

Wherever you are in your analytics journey, there are ways to get help. Throughout this Gorilla Guide, you've seen solutions that run the gamut. You've also learned that it's not nearly as hard as you probably think to harness the power of AI and ML, and get it working for you. In fact, you can get started today with HPE and Vertica!

Visit the HPE big data analytics solutions website[1] and the Vertica website[2] to discover how you can transform your business from edge to cloud and put your data to work.

[1] https://www.hpe.com/us/en/solutions/data-analytics.html

[2] https://www.vertica.com/

ABOUT MICRO FOCUS & HPE

Micro Focus helps organizations run their business and transform it. Our software provides the critical tools they need to build, operate, secure, and analyze their enterprise. By design, these tools bridge the gap between existing and emerging technologies—enabling faster innovation, with less risk, in the race to digital transformation.

Hewlett Packard
Enterprise

Hewlett Packard Enterprise is the global edge-to-cloud Platform-as-a-Service company that helps organizations accelerate outcomes by unlocking value from all of their data, everywhere. We're built on decades of reimagining the future through innovation.

ABOUT ACTUALTECH MEDIA

ActualTech Media is a B2B tech marketing company that connects enterprise IT vendors with IT buyers through innovative lead generation programs and compelling custom content services.

ActualTech Media's team speaks to the enterprise IT audience because we've been the enterprise IT audience.

Our leadership team is stacked with former CIOs, IT managers, architects, subject natter experts and marketing professionals that help our clients spend less time explaining what their technology does and more time creating strategies that drive results.

For more information, visit
www.actualtechmedia.com