Cisco Integrated Infrastructure for Big Data and Analytics with Vertica Advanced Analytics Platform

Highlights

Proven enterprise-ready converged data platform

- Cisco UCS® Integrated Infrastructure for Big Data and Analytics offers complete integration of computing, networking, and storage resources with unified management and provides high performance, expandable storage, and scalable, advanced analytics with Vertica.

Built on the Cisco Unified Computing System™ (Cisco UCS)

- The solution uses a fabric-centric architecture, providing business acceleration, a true on-demand infrastructure, and a system that grows gracefully and incrementally.

Designed, tested, and validated for faster time to value

- Prevalidation through industry-standard benchmarks, tighter integration, and performance optimization reduce integration and deployment risk and provide more predictable customer deployments.

Advanced analytics engine

- The Vertica Advanced Analytics Platform is designed from the foundation as a massively parallel, cluster-aware columnar SQL analytical database. It can be installed natively as a database or directly on Apache Hadoop as a query engine. The unique architecture supports simultaneous load and query operations with high availability, compression, and more.

Actionable insights on database layer

- With machine learning built directly into the core of Vertica (in-database machine learning), and with many SQL analytics functions for data preparation, Vertica is excellent for anomaly detection, predictive maintenance, smart manufacturing, and deep analysis of massive amounts of data.

ANSI SQL–compliant structured SQL interface

- Vertica is ANSI SQL compliant and tightly integrates with a wide range of business intelligence and Extract, Transform, and Load (ETL) tools. It supports the Hadoop framework, with optimized support for queries on Parquet, ORC, JSON, and Avro files.

Service-Level Agreement (SLA) based Reliability, Availability, and Scalability (RAS) for analytics

- The solution has no single point of failure. The Vertica analytical database responds appropriately if a node or disk drive fails.
- Vertica includes both a proprietary highly optimized format and support for optimized Hadoop formats. Users can meet even the most demanding SLAs by analyzing the data in place.

Real-time big data analytics platforms in the modern data center

In today's highly competitive business environment, where nearly every industry is experiencing transformation and disruption, change is constant. Maintaining a competitive advantage requires data-driven organizations to think differently and develop new core competencies derived from analytical insight. Success and growth have become intrinsically linked with machine learning for predictive analytics. As companies collect and rely on data more frequently, the velocity, variety, and volume of data sources evolves daily. This continuous change creates greater complexity and scale that traditional systems, and even newer open-source systems, either were not designed to handle or have not yet been proven to handle in real-world, unforgiving environments with stringent Service-Level Agreements (SLAs). While organizations are grappling with these challenges and market demands, a new trend has emerged in which data is the "new oil," and information and insights from this data is the "new gold." As industry markets have demonstrated, generous valuations and financial success are often rewarded to organizations capable of monetizing business data now—not after the opportunity has passed.

Real-time analytics is becoming a fundamental differentiator in all industry markets: for instance, predictive maintenance for the Internet of Things (IoT), risk management for banking, fraud detection for e-retail, and predictive analytics for customer retention and uplift. And modern systems need to support the most demanding analytics initiatives and adapt to big data strategies as business demands change.

Actionable insights provide greater value

Built by design to dramatically improve query performance over traditional relational database systems, Vertica is purpose built with in-database advanced analytics capabilities that are fine-tuned to handle a range of complex big data queries. At the base of Vertica is the insight achieved with a fast SQL analytical database. The Vertica Advanced Analytics Platform supports ANSI SQL standards and augments SQL with geospatial, time-series analytics, data-preparation functions, live aggregation, and custom analytics. In addition, you can perform supervised and unsupervised Machine Learning using popular algorithms such as linear regression, logistic regression, k-means, Naïve Bayes, and support vector machine (SVM). You don't have to move data out of the database to achieve all the insights you need. You can use the power of Vertica's optimizations to move quickly through analytics: both descriptive and predictive.

These unique capabilities enable actionable insight and remediation: the capability to engage and take action based on information while it maintains relevance and value. Vertica customers separate themselves from their competitors by producing value from petabytes of data and beyond.

Bringing Apache Hadoop data alive with analytics

Many companies have chosen to use Apache Hadoop to store massive volumes of unstructured data, but performing analytics processing on Hadoop workloads can be challenging. Because Hadoop is not a database, it can be challenging to handle concurrent queries or gain access to all the insights you need. If you apply the right set of SQL to Hadoop data analytics capabilities, you can dramatically reduce complexity and accelerate time to insight.

At the same time that Hadoop is being used for persistent storage, Enterprise Data Warehouses (EDWs) are growing in size and complexity due to rapidly changing data requirements. The increase in data complexity and heightened demand for access from business stakeholders places a burden on EDWs to support increasing numbers of data consumers. New data sources from IoT sensors, mobile and online applications, and machine-generated sources such as web and application logs place a burden

on both the data warehouse and the Hadoop environments. These requirements create a burden that the solutions were not designed to handle.

Another aspect of the data deluge challenge is economics. Data lakes based on Hadoop offer economical alternatives for data ill-suited for the EDW. Although the data warehouse is well-suited for mixed analytical workloads, the data lake can support many of the functions of the EDW, but with less support for concurrency and fixed SLAs. Information lifecycle management is on the collective minds of many organizations as they seek to keep hot data hot and less frequently queried data in a more economical storage location. To increase adoption and distribution of these projects, data lakes need to support SQL-based tools to provide business stakeholders with direct access.

Vertica recognizes these market conditions. In addition to being deployed as a Linux-based database, Vertica can be deployed as a query engine on Hadoop. Vertica SQL on Apache Hadoop offers an enterprise-ready method for performing SQL queries on Hadoop data. Using years of experience in the big data analytics marketplace and opening Vertica to use the full power of Hadoop, organizations can perform analytics regardless of the data format. Vertica SQL on Apache Hadoop handles mission-critical analytics projects by merging the best of their analytics platforms with the best that Hadoop data analytics can offer. The following principles help to deliver on the promise of:

- A unified analytics platform merging the economics of Hadoop with the power of Vertica
- Complete SQL support to empower business stakeholders to access and use data directly
- The use of ORC and Parquet files on the Hadoop Distributed File System (HDFS) for information lifecycle management
- Workload management to support scheduled and impromptu analytics workloads from across the organization

Next-generation advanced analytics platform

Together, the Cisco Unified Computing System™ (Cisco UCS®) and Vertica's industry-leading performance at petabyte scale, machine learning, and advanced analytics, offer a tightly integrated solution that helps data-driven organizations identify trends and predict behaviors. The power of Cisco UCS and Vertica comes from the capability to deliver advanced, complex analytics to a broad universe of users on any device, promoting the democratization of data and analytics everywhere.

Cisco UCS Integrated Infrastructure for Big Data and Analytics

Organizations today must help ensure that the underlying physical infrastructure can be deployed, scaled, and managed in a way that is agile enough to change as workloads and business requirements change. Cisco UCS Integrated Infrastructure for Big Data and Analytics has redefined the potential of the data center with its revolutionary approach to managing computing, network, and storage resources to successfully address the business needs of IT innovation and acceleration. Cisco UCS Integrated Infrastructure for Big Data and Analytics provides an end-to-end architecture for processing high volumes of structured and unstructured data, for both real-time use and archival storage.

Cisco UCS 6300 Series Fabric Interconnects

Cisco UCS 6300 Series Fabric Interconnects provide high-bandwidth, low-latency connectivity for servers, with Cisco UCS Manager providing integrated, unified management for all connected devices. The Cisco UCS 6300 Series Fabric Interconnects are a core part of Cisco UCS, providing low-latency, lossless 40 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and Fibre Channel functions.

Cisco fabric interconnects offer the full active-active redundancy, performance, and exceptional scalability needed to support the large number of nodes that are typical in clusters serving big data applications. Cisco UCS Manager enables rapid and consistent server configuration using service profiles and automates ongoing system maintenance activities such as firmware updates across the entire cluster as a single operation. Cisco UCS Manager also offers advanced monitoring with options to raise alarms and send notifications about the health of the entire cluster.

Cisco UCS C240 and C220 M5 Rack Servers

Cisco UCS M5 Rack Servers are dual-socket, 2-Rack-Unit (2RU) servers offering industry-leading performance and expandability for a wide range of storage and I/O-intensive infrastructure workloads, such as big data, analytics, and collaboration. These servers use the new Intel® Xeon® Scalable processors with up to 28 cores per socket. They support up to 24 Double-Data-Rate 4 (DDR4) Dual In-Line Memory Modules (DIMMs) for improved performance and lower power consumption. The DIMM slots are 3D XPoint ready, supporting next-generation nonvolatile memory technology.

Depending on the server type, Cisco UCS rack servers have a range of storage options. The Cisco UCS C240 M5 Rack Server supports up to 24 Small Form-Factor (SFF) 2.5-inch drives (with support for up to 10 Non-Volatile Memory Express [NVMe] PCIe solid-state disks [SSDs] on the NVMe-optimized chassis version) or 12 Large-Form-Factor (LFF) 3.5-inch drives plus 2 rear hot-swappable SFF drives with a Cisco 12-Gbps SAS modular RAID controller. The Cisco UCS C220 M5 Rack Server supports up to 10 SFF 2.5-inch drives (with support for up to 10 NVMe PCIe SSDs on the NVMe-optimized chassis version). In addition, all the servers have two modular M.2 cards that are used for boot. A modular LAN-On-Motherboard (mLOM) slot supports dual 40 Gigabit Ethernet network connectivity with the Cisco UCS Virtual Interface Card (VIC) 1387.

Vertica Advanced Analytics Platform

The Vertica Advanced Analytics Platform is purpose built from the first line of code for big data and analytics workloads. It is designed for use in data warehouses and for other big data workloads in which speed, scalability, simplicity, and openness are crucial to the success of analytics. Vertica relies on a tested, reliable distributed architecture and columnar compression to deliver exceptionally fast speed. A simplified license that is based solely on data volume and the capability to deploy the solution anywhere delivers on the promise of big data analytics.

Database schemas often need refactoring, but Vertica offers a flattening table feature that can mitigate the need for refactoring. Flattened Tables optimize the performance of query join operations by creating a big virtual flat fact table for your SQL query writing convenience. Flattened Tables speed up queries significantly by denormalizing the schema. Queries are easier to write and faster to process with Vertica Flattened Tables.

Vertica Enterprise is the core "shared nothing," distributed analytical database designed to work on clusters of cost-effective, off-the-shelf servers in your data center with outstanding performance and scale. Vertica provides an advanced SQL database analytics portfolio – all based on the same unified architecture -- built to address the most demanding big data analytics initiatives.

- Fast loading and querying: Vertica helps ensure extremely high query concurrency, while simultaneously loading new data into the system. Administrators can designate resource pools to specify the amount of memory and CPU resources allocated to critical tasks versus more routine tasks.
- Support for multiple file formats and structures: Vertica uses its own proprietary format for the best optimization and speed. However, other formats, such as Parquet, ORC, and Avro, are supported as well. Vertica's Flex Tables can be used to query unstructured data, such as JSON data.
- Advanced in-database analytics: You can conduct the analytics computations closer to the data and get immediate answers from a single place without the need to extract information to a separate environment for processing, allowing machine learning at scale as well as SQL, Python, and R processing.
- Administration and management tools: You can tune and control queries with little administration required using Vertica's Database Designer with Administration Tools. Spend less time identifying performance problems and optimizing a database physical design.
- Columnar design and data compression: You can use columnar data storage for significant gains in performance, I/O, storage footprint, and efficiency. Compression works better on columnar data than on row-based data.
- Supported in current ecosystem: With support for all leading business intelligence and visualization tools and open-source technologies such as Apache Hadoop, Kafka, and Spark, you can streamline the transition to Vertica to modernize your analytics ecosystem and accelerate your data pipelines.

Vertica for SQL on Apache Hadoop lets you run the industry's most comprehensive Vertica SQL analytics engine directly on Hadoop clusters and tap into advanced SQL on Hadoop capabilities to achieve greater concurrency. It runs on any Hadoop distribution.

- Advanced analytics: Gain full ANSI SQL capabilities, not a subset of commands. Run 100 percent of TPC-DS benchmark queries without modification.
- Platform independent: Install Vertica directly on all major distributions of Hadoop, including Hortonworks, Cloudera, and MapR, with no helper nodes or single point of failure.
- Data access: Query data using Avro, Parquet, ORC, (JSON), and many other formats.

Use cases

The Vertica Massively Parallel Processing (MPP) SQL analytical database is prominently deployed in the following use cases for financial services, healthcare, IoT, IT infrastructure, marketing and ad serving, retail, telco, and other markets.

- Assurance: Secure, manage, and enhance networks, revenue, and devices, protecting against unexpected events.
- Customer experience management: Gain better knowledge of customers and devices and infrastructure supporting them to improve service quality, reduce churn, and deliver incremental revenue growth.
- Operational analytics: Infuse the supply chain with just-in-time, dynamic actions and remediation based on Vertica's ability to analyze the entire data set upon ingest with advanced analytical functions and machine-learning algorithms.

Real-time analytics platform reference architecture with Cisco UCS and Vertica

Cisco UCS Integrated Infrastructure for Big Data and Analytics for Vertica includes eight or more Cisco UCS C240 M5 (or C220 M5) Rack Servers, each with dual Intel Xeon Scalable 6132 CPUs (2 x 14 cores and 2.6 GHz), 384 GB of RAM, dual 40-Gbps network connectivity, and 8 (or 16) SSDs. These servers are connected to Cisco UCS 6332 Fabric Interconnects (Figure 1).

Figure 1. Real-time analytics platform reference architecture with Cisco UCS and Vertica.

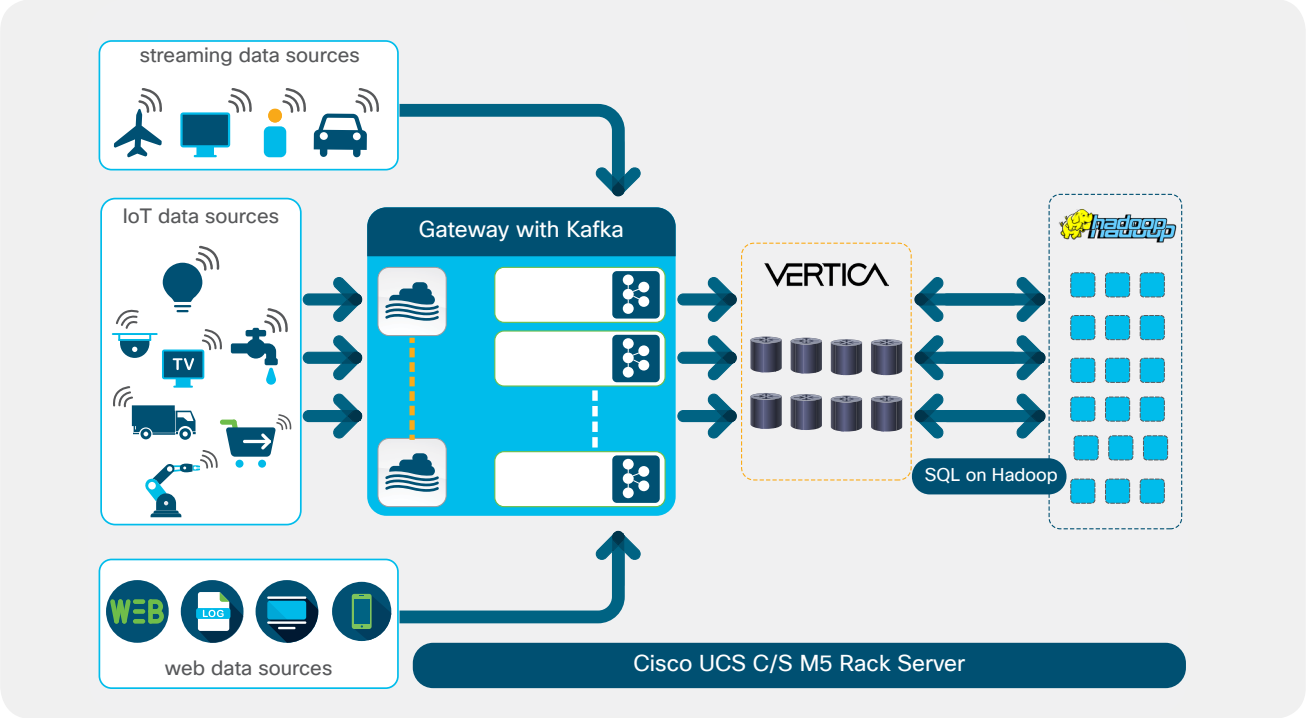
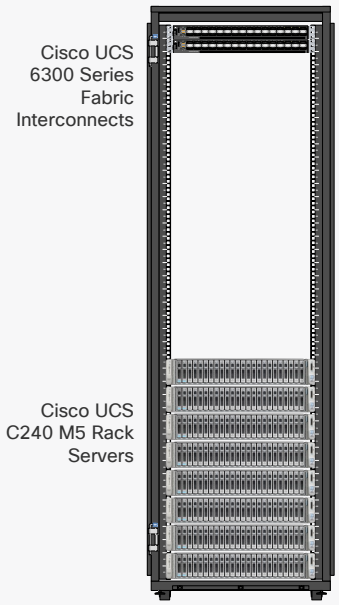


Figure 2. Shows Cisco UCS reference architecture bundles for Vertica.



Cisco UCS reference architecture for Vertica

Network connectivity:

Cisco UCS 6332 Fabric Interconnect (40 Gbps)

Starter	Advanced
Servers: 8 Cisco UCS C220 M5 Rack Servers, each with: • Dual Intel Xeon Scalable 6132 CPUs (2 x14 cores and 2.6 GHz) • 6 x 32 GB 2666-MHz (192 GB) DDR4 RAM • 8 x 960-GB (or 1.6-TB) SSDs • Cisco UCS VIC 1387 with dual 40-Gbps port connectivity • Cisco 12-Gbps SAS modular RAID controller with 2-GB flash-based write cache (FBWC) • M.2 with 2 x 480-GB SSDs	**Servers:** 8 Cisco UCS C240 M5 Rack Servers, each with: • Dual Intel Xeon Scalable 6132 CPUs (2 x 14 cores and 2.6 GHz) • 12 x 32-GB 2666-MHz (384 GB) DDR4 RAM • 16 x 960-GB (or 1.6-TB) SSDs • Cisco UCS VIC 1387 with dual 40-Gbps port connectivity • Cisco 12-Gbps SAS modular RAID controller with 4-GB FBWC • M.2 with 2 x 480-GB SSDs

Conclusion

The fifth generation of Cisco UCS Integrated Infrastructure for Big Data and Analytics builds on the previous generation of platforms with new processors, faster memory, and more storage options. It is designed, tested, and validated for enterprises to lower the cost of ownership and to scale from small to very large as applications demand for on-premises data.

For more information

- Cisco UCS Integrated Infrastructure for Big Data and Analytics: http://blogs.cisco.com/datacenter/cpav5
- Cisco Validated Designs for big data: www.cisco.com/go/bigdata_design
- Cisco UCS big data solutions: www.cisco.com/go/bigdata.
- Cisco UCS: www.cisco.com/go/ucs
- Vertica: www.vertica.com/
- Vertica: www.vertica.com/try