White Paper

Cisco UCS with ParAccel Analytic Platform Solution: Deliver Powerful Analytics to Transform Business

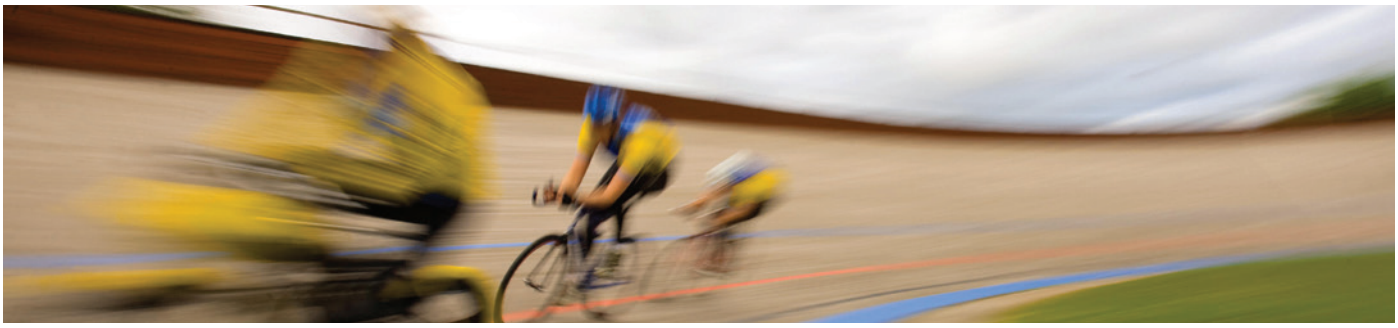
December 2012

In Collaboration With: ParAccel

Contents

Cisco UCS with ParAccel Analytic Platform Solution: Deliver Powerful Analytics to Transform Business

White Paper
December 2012



Introduction

In today's global, highly connected economy, enterprises are faced with a volatile business environment that is full of unexpected risks and opportunities. With windows of opportunity short, companies that effectively use analytic power to take advantage of data can reap rewards that are extremely lucrative. These companies have moved beyond traditional business intelligence that relies on reports, dashboards, and performance indicators that provide a look back at what happened. With the capability to analyze data in near real time using predictive and prescriptive analytics, companies can stay ahead of market conditions, rapidly compare numerous what-if scenarios, and gain a competitive edge.

Staying ahead of market conditions and competitors requires the capability to answer a wide range of questions. The answers are more compelling and accurate when all available data is used. Results can be unexpected and trigger further analysis, along many different analytic vectors. The capability to quickly rerun these analyses supports a process of iterative and interactive data discovery as analysts start to have deep meaningful "conversations" with their data. These conversations enable analysts to uncover insights about the current state of the business and guide strategic decisions so that enterprises can make the choices most likely to benefit the business in the weeks, months, and years to come.

A discovery-based approach requires a platform that supports open, iterative, interactive, and complex analysis of vast and diverse data sets. Results must be available in seconds and minutes rather than hours or days. The combination of the Cisco Unified Computing System™ (Cisco UCS®) and the ParAccel Analytic Platform provides a high-performance big data discovery platform. Deployed as part of a comprehensive data center architecture, the Cisco UCS with ParAccel Analytic Platform solution delivers a powerful and flexible analytics infrastructure that increases business and IT agility, reduces total cost of ownership (TCO), and delivers exceptional return on investment (ROI) at scale, while fundamentally transforming the way that organizations do business.

Cisco UCS with ParAccel Analytic Platform Solution

The Cisco UCS with ParAccel Analytic Platform solution combines ParAccel's powerful, software-based analytics platform with the high-performance, flexible, and efficient unified infrastructure of Cisco UCS. With this tested and validated solution, companies can accelerate analytics projects and give users the capabilities they need to transform their businesses.

The solution helps organizations derive optimal value from all their data, finally delivering on the vision of the analytics-based enterprise. This high-performance, interactive platform for unconstrained analytics allows companies to run the most sophisticated queries against the entire data sphere, including new big data sources.

The ParAccel Analytic Platform is powered by a columnar, massively parallel processing (MPP) analytic database combined with an extensibility framework that enables innovative integration and analytics capabilities. It performs operations, such as aggregation and joins, against large amounts of disparate data. This approach contrasts with traditional online transaction processing (OLTP) databases, in which operations are optimized to perform single-row insertions and deletions.

The platform supports integrated analytics, with more than 500 mathematical, statistical, data-mining, and financial functions that run natively within the database, returning results quickly. A set of on-demand integration (ODI) modules also allows easy integration with Hadoop as well as with data warehouses including Teradata, Oracle, and Microsoft SQL Server. Together, these integration and analytics features allow complex, impromptu big data analytics across multiple data sources, delivering new insight that was not possible with traditional systems.

The platform's superior performance is the result of its innovative architecture, explicitly built for analytics. ParAccel Analytic Platform is a software-only solution that frees customers to take advantage of hardware innovations rather than waiting up to 18 months for them to become available in analytic appliances. This approach also ensures compliance with data center standards for servers and storage ParAccel's software platform uses Cisco® hardware to deliver analytic insight at incredible speeds, enabling faster time to value and greater ROI.

Cisco UCS has redefined data center infrastructure by uniting computing, network, storage access, and virtualization resources in a unified system designed to reduce complexity. Composed of blade and rack servers, redundant switch fabric interconnects, and embedded model-based management, Cisco UCS delivers outstanding flexibility to meet the most rigorous performance and capacity demands. All components are provisioned and managed through a single management interface, simplifying administration and lowering operating costs.

Together, Cisco and ParAccel help organizations gain new insight and get real value from all their data, including big data. The best-in-class solution uses proven technology from industry leaders to deliver innovative analytics capabilities that empower companies to accelerate decision making, innovate, and compete.

About ParAccel

ParAccel is a leading Analytic Platform provider. ParAccel's high-performance, interactive platform creates an environment for unconstrained analytics where companies ask the most sophisticated questions against the entire datasphere, including big data.

ParAccel Analytic Platform

The ParAccel Analytic Platform is an analytics database built with extensibility and integration technology that delivers a foundation for enterprises to run big data analytics at any time and from any location.

ParAccel Analytic Platform Architecture

The ParAccel Analytic Platform relies on an enterprise-class, massively parallel, shared-nothing architecture that is designed explicitly for high-performance, complex analytics. Within the ParAccel cluster topology, Cisco UCS C-Series Rack Servers act as leader, computing, or standby nodes and communicate over a high-speed cluster fabric (Figure 1).

Leader Node

The leader node is responsible for managing communications with client programs and for communicating and coordinating work among computing nodes. It parses and develops query execution plans to carry out analytic operations, particularly the series of steps necessary to obtain results for complex queries and joins. Optimized binary code is distributed to the computing nodes, which perform the steps needed to run queries. After the task is complete, the leader node sends the query results back to the client programs.

Computing Nodes

To increase performance and decrease query run times, computing nodes are designed to use all the performance that the server hardware (CPU, RAM, and disk) can deliver. These nodes store all the data in columnar block format and perform query processing using optimized binary code received from the leader node. Queries are run in parallel simultaneously, both within a server and across all computing nodes.

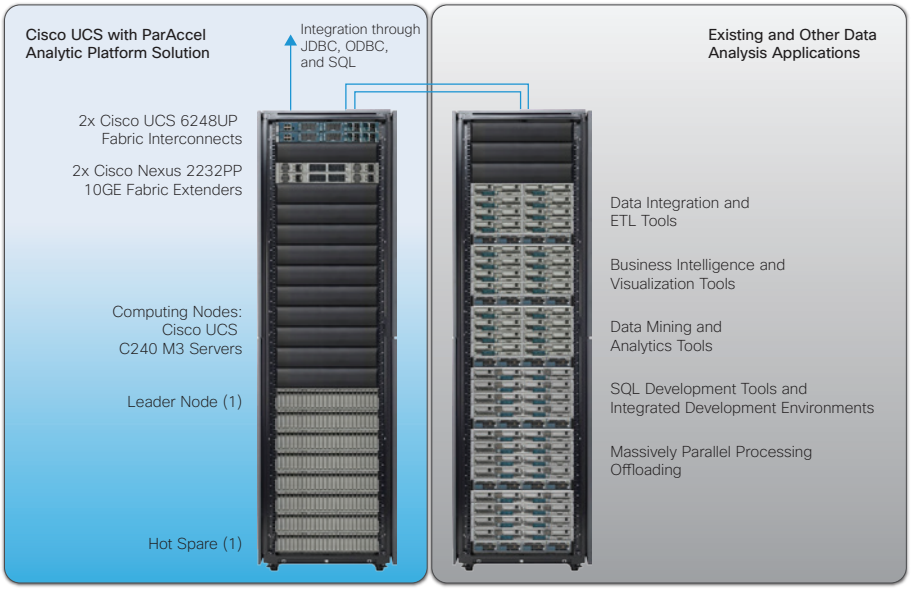


Figure 1. ParAccel Analytic Platform

Hot Standby Nodes

The solution sets aside one or more Cisco UCS servers to act as hot standby nodes. These servers are automatically activated as hot replacements by lights-out management techniques if a leader or computing node fails.

Cluster Fabric

The cluster fabric provides high-speed internode communication for cluster messaging and data movement. A purpose-built, high-speed, scalable, low-latency protocol is used to transmit data between the leader and computing nodes to optimize performance.

Main Capabilities

The Cisco UCS with ParAccel Analytic Platform solution is designed to handle the most complex SQL queries on any data at industry-leading speed. This capability includes queries with a large number of complex table joins, many levels of nested subqueries (correlated, uncorrelated, or skip correlated), and complex aggregations and analytic functions such as window functions. In addition, extensive conditional SQL (CASE statements), full-outer joins, and uniqueness filtering over large data sets (DISTINCT) are supported. As a result, enterprises do not need to invest precious time or budget in rewriting complex, business-based SQL queries to help ensure that they run at high performance, something many organizations must do with solutions from other vendors. This feature is particularly valuable because complex SQL statements often originate from business intelligence or visualization tools that limit the capability to rewrite SQL code, or from advanced SQL users who continue to evolve SQL code as they discover new relationships in the data.

Designed with the need for in-database advanced analytics in mind, the ParAccel Analytic Platform enables access to over 500 analytic functions that are commonly used in complex mathematical, statistical, probability, and data-mining analyses. These in-database analytics enable computations to be performed in the place where the data resides. By eliminating the need to move a subset of the data to an external software platform, these modules can exploit the parallel architecture of the ParAccel Analytic Platform, allowing analytics to be run over the entire data set rather than just a representative sampling, enabling greater accuracy.

The ParAccel Extensibility Framework takes an already impressive analytic database and turns it into a platform. Providing open access, the framework makes it possible to easily bring in data from many different sources, leverage external analytics, and quickly assimilate new functionality that is critical to handling different kinds of analytic workloads. ParAccel On Demand Integration (ODI) modules change the way analysts work by giving them inline access to other data sources, with parallel, bi-directional integration. ParAccel offers a variety of pre-built ODI modules, and custom integration with other source systems can be developed. In addition, the framework gives developers the tools they need to build custom user-defined functions (UDFs) to extend the data processing capabilities of the platform.

Integration with Leading Analysis Tools

The Cisco UCS with ParAccel Analytic Platform solution works with tools from other vendors to deliver comprehensive analytics and greater insight to users.

- **Data integration and extract, transform, and load (ETL) tools:** The ParAccel Analytic Platform supports a wide range of data integration and ETL tools. These tools are used to pull, transform, and cleanse data that is loaded into the ParAccel analytic database.
- **Business intelligence and visualization tools:** A wide variety of vendors and solutions provide business intelligence and visualization tools that are supported by the Cisco and ParAccel solution. In addition, support for industry-standard connectivity enables integration with emerging data visualization vendors to connect to a standard Open Database Connectivity (ODBC) or Java Database Connectivity (JDBC) data source.
- **Data mining and analytic tools:** ParAccel works with traditional advanced analytics workflows that use very large sample data sets. Organizations can run in-database analytics, such as advanced mathematical, statistical, and data-mining functions, on the Cisco and ParAccel solution and use all available data for greater insight and precision.

Interface and Standards Compliance

The ParAccel Analytic Platform is ANSI SQL compliant, enabling companies to standardize on a single analytics platform while using a wide variety of business intelligence and visualization tools. In addition, analysts who prefer to work directly in SQL can submit queries to the platform without the need for a front-end tool. In addition, the platform supports the universally available ODBC standard, making it easy for applications to run queries. Integration with Java technology-based applications is available through industry-standard JDBC interfaces.

Deployment Options for the Cisco UCS with ParAccel Analytic Platform Solution

Organizations can deploy the solution in two ways:

- **Integrated software and hardware bundle:** For the simplest sourcing and deployment, organizations can choose a bundled solution. Some prefer this method of acquisition, with only one Cisco part number to order. A Cisco Authorized Data Center Partner integrates Cisco hardware with pre-configured ParAccel software for a complete analytics solution that can be rapidly deployed.
- **Reference configurations:** Alternatively, organizations can base the solution on one of the predefined Cisco reference configurations for ParAccel. This approach offers more flexibility to optimize the platform for specific use cases. Cisco reference configurations for the ParAccel Analytic Platform are designed and validated by Cisco and ParAccel to enable the rapid deployment of enterprise analytics platforms.

Cisco UCS Support for the ParAccel Analytic Platform

The Cisco UCS with ParAccel Analytic Platform solution is based on the Cisco big data common platform architecture (CPA), a highly scalable architecture designed to meet a variety of scale-out application demands with transparent data and management integration capabilities built using the following components:

- **Cisco UCS 6200 Series Fabric Interconnects:** The Cisco UCS 6200 Series Fabric Interconnects provide a single point of management and connectivity for all the solution's servers. Typically deployed in redundant pairs, the fabric Interconnects provide line-rate, low-latency, lossless 10 Gigabit Ethernet connectivity and unified management in a highly available management domain.
- **Cisco UCS 2200 Series Fabric Extenders:** Cisco UCS 2200 Series Fabric Extenders bring the system's unified fabric to rack servers and provide a highly scalable and cost-effective unified server-access platform.
- **Cisco UCS C240 M3 Rack Server:** The solution's servers are 2-socket servers based on the Intel® Xeon® processor E5-2600 series, with up to 768 GB of main memory. Up to 24 Small Form-Factor (SFF; high-performance option) or 12 Large Form-Factor (LFF; high-capacity option) disk drives are supported, along with four built-in Gigabit Ethernet ports.
- **Cisco UCS virtual interface cards (VICs):** Available only from Cisco, these interface cards allow the number and type of I/O devices to be configured on demand, with options supporting up to 20 Gbps of I/O bandwidth in Cisco UCS rack servers. Cisco VICs allow up to 256 Ethernet or Fibre Channel devices to be configured on demand.
- **Cisco UCS Manager:** Cisco UCS Manager resides within the Cisco UCS 6200 Series Fabric Interconnects. It makes the system self-aware and self-integrating, managing all the system components as a single logical entity. Cisco UCS Manager can be accessed through an intuitive GUI, a command-line interface (CLI), or an XML API. Cisco UCS Manager uses Cisco UCS service profiles to define the personality, configuration, and connectivity of all resources in the system, radically simplifying provisioning of resources so that the process takes minutes instead of days. This simplification allows IT departments to shift their focus from constant maintenance to strategic business initiatives.

With Cisco UCS, every detail of a server's configuration and its network connectivity are encapsulated in a Cisco UCS service profile. Applying a Cisco UCS service profile to a server configures it to a known state that complies with predefined organizational standards. Cisco UCS service profiles make migration of workloads between servers with different capacities straightforward. Spare capacity can be maintained in a pool shared by ParAccel and third-party applications and tools and allocated on demand, reducing the cost of burst capacity and disaster-recovery resources, making the data center more agile.

Integrated Solution Configurations

The solution is offered in half-rack, full-rack, and multirack configurations. The half-rack and full-rack configurations use the compression capabilities offered by the ParAccel analytic database to accommodate 127 terabytes (TB; half rack) or 254 TB (full rack) of data. Total capacity varies depending on the data types and workloads. Details about the reference configurations, including the number of nodes, processor cores, storage data capacity, and I/O bandwidth, are provided in Table 1. These reference configurations can be deployed as defined or used as a starting point for the design of large-scale multirack systems.

Table 1. Cisco Reference Configurations for ParAccel Analytic Database

	Half rack	Full rack	Multirack
Servers	8 Cisco UCS C240 M3 Rack Servers with 128 Cores	16 Cisco UCS C240 M3 Rack Servers with 256 Cores	Full rack and 16 additional computing nodes for each additional rack
Internal Storage Capacity	113 TB using 600-GB 10K SAS drives	225 TB using 600-GB 10K SAS drives The total available storage will vary based on data type, workload and how much raw data size is reduced with ParAccel's compression features.	
Memory Capacity	2.0 TB typical (up to 6 TB supported)	4 TB typical (up to 12 TB supported)	
I/O Bandwidth	22 Gbps	44 Gbps	

Each configuration consists of two fully redundant Cisco UCS 6248UP 48-Port Fabric Interconnects (for up to five racks) or Cisco UCS 6296UP 96-Port Fabric Interconnects (up to 10 racks) along with two Cisco Nexus® 2232PP 10GE Fabric Extenders per rack and the number of Cisco UCS C240 M3 servers defined in Table 1. Each server in the configuration connects to the unified fabric through two active-active 10-Gbps links using a Cisco UCS VIC for data traffic and Cisco Integrated Management Controller (IMC) connection for management traffic, as shown in Figure 2. Multirack configurations include two Cisco Nexus 2232PP fabric extenders and 16 Cisco UCS C240 M3 Rack Servers for each additional rack.

The half-rack configuration includes one leader and one failover node and six computing nodes. The full-rack configuration includes one leader, one leader failover, and 14 computing nodes. Multirack configurations use the full-rack configuration along with two Cisco Nexus 2232 fabric extenders and 16 servers configured as ParAccel computing nodes per rack.

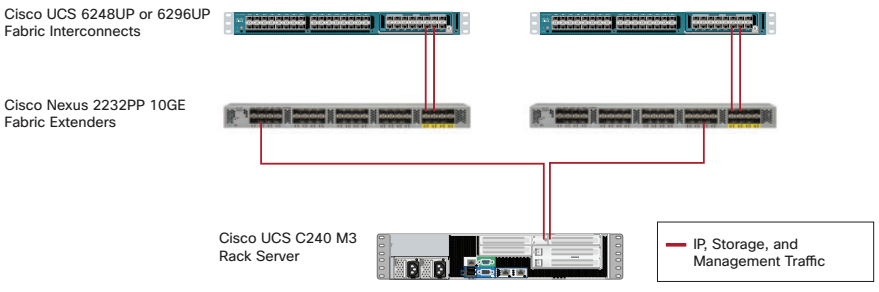


Figure 2. Cisco UCS Fabric Architecture Uses a Single Set of Wires for IP, Storage, and Management Traffic

High Performance, Exceptional Scalability, and Simplified Management

The Cisco Unified Fabric provides fully redundant, highly scalable, lossless 10 Gigabit Ethernet and Fibre Channel over Ethernet (FCoE) connectivity. Powered by the latest Intel Xeon processors, the solution delivers best-in-class performance and internal storage capacity. These configurations can easily scale to hundreds of nodes when required by business demands. The advanced management capabilities of Cisco UCS radically simplify this process with a single point of management that spans all nodes in the cluster. In traditional environments, it can be challenging to manage these large numbers of servers effectively. Cisco UCS Manager delivers unified, model-based management that applies personality and configures server, network, and storage connectivity resources, making it as easy to deploy hundreds of servers as it is to deploy a single server. Additionally, Cisco UCS Manager can perform system maintenance activities such as firmware updates across the entire cluster as a single operation.

Cisco Solution Bundle

The hardware components for the Cisco UCS with ParAccel Analytic Platform solution are available through the Cisco Smart Play program (Table 2). With only a single part number to order, the program makes it easy to quickly deploy a powerful, secure big data environment in your enterprise without the expense or risk entailed in designing and building a custom solution.

Table 2. Cisco Big Data High Performance SAS Half Rack Smart Play Bundle: All Hardware Components Can be Ordered Using a Single Part Number

Big Data High-Performance SAS Half Rack
Computing 8 Cisco UCS C240 M3 Rack Servers, each with: • 2 Intel Xeon processors E5-2640 at 2.9 GHz (2 sockets and 12 cores) • 128 GB of memory • Redundant power supplies • Cisco UCS P81E Virtual Interface Card (VIC)
Networking 10-Gbps unified fabric with: • 2 Cisco UCS 6248UP • 2 Cisco UCS 2232PP • Low-latency Twinax cabling
Storage 24 SFF SAS Hard Disk Drives per server: • 600-GB 10K RPM SFF SAS • LSI MegaRAID 9266-CV 8i Card

Conclusion

When companies need to run complex analytics on big data across their enterprise, they turn to Cisco and ParAccel. With the Cisco UCS with ParAccel Analytic Platform solution, organizations can provision new, powerful analytics applications and extend their analytics reach across customers, partners, and an ever-changing marketplace. This solution from Cisco and ParAccel brings these powerful capabilities together on Cisco UCS, the industry's first unified data center platform. By running the ParAccel Analytic Platform on the intelligent, programmable infrastructure of Cisco UCS, companies can analyze more information faster and make better decisions to respond to a rapidly changing business environment.

For More Information

- To learn more about Cisco UCS, visit http://www.cisco.com/go/ucs.
- To learn more about the ParAccel Analytic Platform, visit http://www.paraccel.com/UCS.
- To learn more about the Cisco Smart Play program, visit http://www.cisco.com/go/smartplay.
- To learn more about Cisco's Common Platform Architecture (CPA) for Big Data, visit http://blogs.cisco.com/datacenter/cpa/.

Acknowledgments

This document was made possible through the efforts of Raghunath Nambiar, Girish Kulkarni, and Wendy Heaton.

Cisco UCS with ParAccel Analytic Platform Solution: Deliver Powerful Analytics to Transform Business
December 2012

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices**.

LE-36901-00 12/12