Solution Overview

LexisNexis High-Performance Computing Cluster Systems Platform with Cisco Unified Computing System

Get More Scalability and Flexibility for Big Data

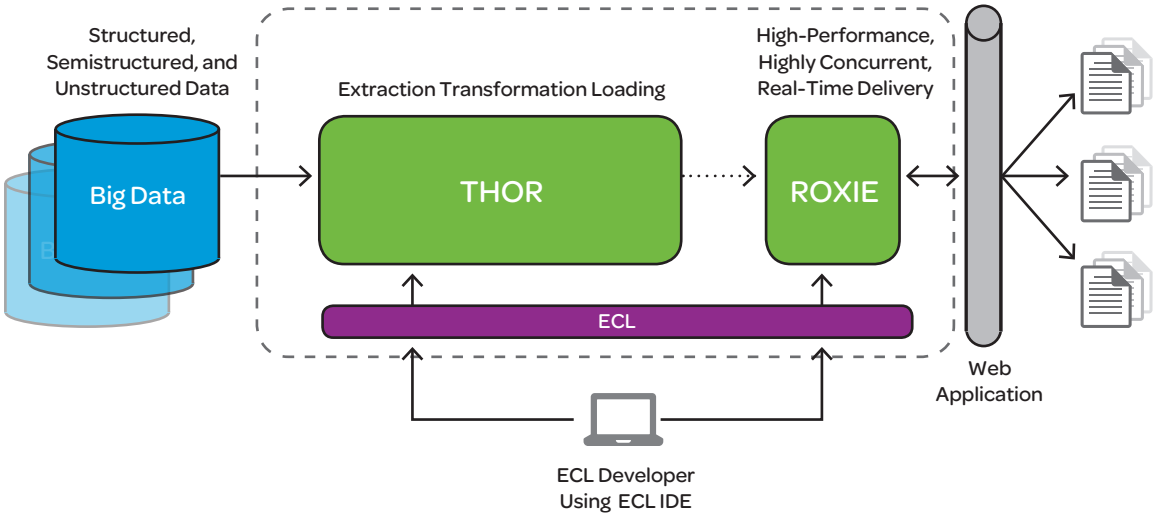
LexisNexis®

Risk Solutions
LexisNexis HPCC Systems Platform
with Cisco Unified Computing System

What You Will Learn

Modern enterprises are challenged with the need to store and analyze massive amounts of structured and unstructured data. This is essential to derive crucial insights that may give the company the necessary advantage over its competition. The amount of this data grows at a faster pace than ever before. Timely and accurate analysis of this data is of utmost importance. High-Performance Computing Cluster (HPCC) Systems is an open source big data processing platform from LexisNexis. LexisNexis has collaborated with Cisco to offer a high-performance analytics platform that is scalable, flexible, and cost effective (Figure 1).

Figure 1 LexisNexis HPCC Systems



The solution is based on the Cisco Unified Computing System™ (Cisco UCS®) Common Platform Architecture (CPA) for Big Data. The software includes a Thor cluster that acts as the extract, transform, and load (ETL) engine; a Roxie cluster for that acts as a high-performing concurrent data delivery platform; and an Enterprise Control Language (ECL) Machine Learning (ECL-ML) module. HPCC Systems has validated and certified this solution on Cisco UCS CPA for Big Data.

Challenges of Big Data and High Performance Computing

"Big data" refers to the data in various structured and unstructured forms that originates from a multitude of devices and applications connected to the Internet. This data typically doesn't fit into traditional relational models because of its size, velocity, and composition: the amount of big data increases rapidly, and big data is a mix of structured and unstructured data. As a result of the continuing dramatic increase in data, many organizations are overwhelmed by the data in their systems, with the result that they are unable to process the information and use it effectively. The answer to these challenges is a scalable, integrated computer hardware and software architecture designed for parallel processing of data- intensive computing applications.

HPCC Systems

HPCC Systems, from LexisNexis, is an emerging enterprise platform with massive parallel-processing performance and high levels of reliability, security, and uptime. To manage, sort, link, and analyze billions of records in less than a second, LexisNexis Risk Solutions offers a data-intensive platform that has been proven for the past 10 years with customers who need to sort through billons of records. Customers such as leading banks and insurance companies, utilities, law enforcement, and the federal government depend on LexisNexis HPCC Systems technology and information solutions to help them make better decisions faster. HPCC Systems comes with optional subscription-based support services. It includes software, services, and support to help organizations discover the insights of new and emerging types of information, a capability that did not exist before. With the extensible set of fully parallel HPCC Systems Machine Learning and Matrix processing modules, enterprise customers are fully equipped to meet the challenges of big data.

Cisco UCS CPA for Big Data

The Cisco® solution for HPCC Systems is based on the Cisco UCS CPA for Big Data, a highly scalable architecture designed to meet a variety of scale-out application demands with transparent data integration and management integration capabilities.

The new Cisco UCS CPA Version 2 (CPAv2) for Big Data improves both performance and capacity. With complete, easy-to-order packages that include computing, storage, connectivity, and unified management features, Cisco UCS CPAv2 for Big Data supports rapid deployment, delivers predictable performance, and reduces total cost of ownership (TCO).

Cisco UCS CPA v2 for Big Data is built using the following components:

- Cisco UCS 6200 Series Fabric Interconnects establish a single point of connectivity and management for the entire system. The fabric interconnects provide high-bandwidth, low-latency connectivity for servers, with integrated, unified management for all connected devices provided by Cisco UCS Manager. Deployed in redundant pairs, Cisco fabric interconnects offer the full active-active redundancy, performance, and exceptional scalability needed to support the large number of nodes typical in clusters that serve big data applications. Cisco UCS Manager enables rapid and consistent server configuration using service profiles, automating 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 activate alarms and send notifications about the health of the entire cluster.

- Cisco Nexus® 2200 platform fabric extenders bring the system's unified fabric to each rack, establishing physically distributed but logically centralized network infrastructure. These low-cost, low-power-consuming devices act as remote line cards for the fabric interconnects, providing connectivity without adding the cost and management complexity required by top-of-rack switches. The result is highly scalable and cost-effective connectivity for many nodes.

- Cisco UCS C240 M3 Rack Servers support a wide range of computing, I/O, and storage-capacity demands in a compact two-rack-unit (2RU) design. The servers use dual Intel Xeon processor E5-2600 v2 series CPUs and support up to 768 GB of main memory (128 or 256 GB is typical for big data applications) and a range of disk drive options. Cisco UCS virtual interface cards (VICs) are optimized for high-bandwidth and low-latency cluster connectivity, with support for up to 256 virtual devices that are configured on demand through Cisco UCS Manager.

Cisco CPAv2 for Big Data supports the massive scalability demanded by big data environments. Up to 160 servers are supported in a single management domain with a pair of Cisco fabric interconnects. Additional scaling can be accomplished by interconnecting multiple domains using Cisco Nexus 6000 or 7000 Series Switches. With Cisco UCS Central Software, thousands of servers and hundreds of petabytes (PB) of storage can be managed through a single interface with the same automation provided by Cisco UCS Manager.

The HPCC System and Cisco solution is offered as a Cisco UCS CPAv2 reference architecture and as a single SKU bundle (Table 1). The bundle can be used as it is or as a base for larger configurations.

Table 1 Cisco CPAv2 for Big Data: Performance- and Capacity-Balanced Configuration

Network Fabric	Management	Servers per Rack
• 2 Cisco UCS 6296UP 96- Port Fabric Interconnects (per domain) • 2 Cisco Nexus 2232PP 10 GE Fabric Extenders (per rack)	• Cisco UCS Manager	• 16 Cisco UCS C240 M3 Rack Servers, each with: - *2 Intel Xeon processors E5-2665 v2* - *256 GB of memory* - *LSI MegaRaid 9271CV-8i storage controller card* - *24 x 1-TB 7200-rpm SFF SAS drives (384 TB total)*

Benefits of Cisco and HPCC Systems

- **High performance and exceptional scalability:** Cisco UCS unified fabric architecture provides fully redundant, highly scalable, lossless 10-Gbps unified fabric connectivity for big data traffic. Powered by the latest Intel Xeon processor, the joint Cisco and HPCC Systems solution delivers best-in-class performance and internal storage capacity. The Cisco and HPCC Systems reference configurations can easily scale to support a large number 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.

- **Simplified management:** Big data analytics implementations tend to involve very large numbers of servers. In traditional environments, managing these large numbers of servers effectively can be challenging. 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.

- **Coexistence with enterprise applications:** In building big data solutions that involve unstructured or NoSQL data, organizations need ways to transfer data transparently between their enterprise applications and big data platforms. This solution can connect, across the same management plane, to other Cisco UCS deployments running enterprise applications, thereby radically simplifying data center management and connectivity. HPCC Systems provides a comprehensive platform for designing and managing solutions that cross the boundaries of traditional and big data platforms. By providing easy-to-use tools and familiar design concepts for both traditional and big data platforms, HPCC Systems empowers organizations to use existing IT skill sets to build big data solutions.

- **Rapid deployment:** Deployment of large numbers of servers can take time. Systems need to be racked, networked, configured, and provisioned before they can be put into use. Cisco UCS Manager uses a model-based approach to provision servers, applying the desired configuration to physical infrastructure quickly, accurately, and automatically. The capability to create consistent configurations improves business agility and eliminates a major source of errors that can cause downtime. The tightly integrated HPCC Systems platform addresses the challenges of building end-to-end solutions that encompass the workflow from data acquisition and processing to comprehensive analytics.

- **Enterprise service and support:** Enterprises want to know that the vendors providing a solution have the expertise to help them quickly proceed through the design, deployment, and testing processes for strategic big data initiatives. Businesses also need to have confidence that if a critical system fails, they will be able to get timely and competent support. The Cisco reference configurations bring together world-class service and support from Cisco and HPCC Systems.

For More Information

For more information about HPCC Systems, please visit http://hpccsystems.com/products-and-services/products.

For more information about Cisco UCS big data solutions, please visit http://www.cisco.com/go/bigdata.

For more information about the Cisco UCS CPAv2 for Big Data, please visit http://blogs.cisco.com/datacenter/cpav2.

For more information about the Cisco Smart Play program, please visit http://www.cisco.com/go/smartplay.

About LexisNexis® Risk Solutions
LexisNexis® Risk Solutions (www.lexisnexis.com/risk/)is a leader in providing essential information that helps customers across all industries and government predict, assess and manage risk. Combining cutting-edge technology, unique data and advanced scoring analytics, Risk Solutions provides products and services that address evolving client needs in the risk sector while upholding the highest standards of security and privacy. LexisNexis Risk Solutions is part of Reed Elsevier, a leading publisher and information provider that serves customers in more than 100 countries with more than 30,000 employees worldwide.

About Cisco
Cisco is the worldwide IT leader that enables businesses to seize the opportunities of tomorrow by proving amazing things happen when you connect the previously unconnected. Cisco brings together people, process, data and things to transform how organizations meet next-generation demands. For more information: www.cisco.com.