

MarkLogic® and Cisco: A Next-Generation, Real-Time Solution for Big Data



MarkLogic Enterprise NoSQL Database and Cisco Unified Computing System provide a single, integrated hardware and software infrastructure that can support any big data application or data type in real time.

Meeting Big Data Demands

With growing volumes of both unstructured and structured data flooding into data centers, enterprises are finding that traditional relational databases are too limiting and inflexible. A MarkLogic and Cisco® infrastructure can meet the strenuous demands of today's big data deployments: volume, velocity, variety, and complexity. This combined offering provides enterprises with a seamless data center solution designed to lower operating costs and provide the agility needed to respond to the fast-changing big data environment.

Two Market Leaders Join Forces

MarkLogic is the global leader in Enterprise NoSQL (not only SQL) database technology for high-volume big data implementations. Over the past decade, MarkLogic has deployed more customer-facing, mission-critical databases in large enterprises than all other NoSQL database vendors combined. Cisco, the world leader in networking technology, offers a broad range of options for data centers that use a dynamic network fabric for handling server cluster and other data traffic. Cisco data center options include compute, storage, connectivity, and unified management components that reduce complexity, improve agility, and greatly reduce cost of ownership.

Major Challenges

Business managers want an enterprise-proven, low-TCO system that enables them to optimize the business intelligence they derive from their data, improve productivity and efficiency, and maintain a competitive advantage. With such a system, the business can become more proactive by quickly predicting patterns and spotting trends. Information and communications technology (ICT) managers require a common infrastructure that is scalable, flexible, cost-effective, and fault-tolerant, that supports many different applications and data types, that overcomes the limitations of isolated systems, and that

augments their current data structure rather than replacing it. Users need to have interactive, real-time conversations with a large store of diverse data.

Typical Case Studies

- Content authoring and delivery
- Data unification and virtualization
- Digital asset management
- Metadata cataloging
- Open source intelligence
- Search and discovery applications
- Analysis of large data sets
- Logical data warehouse

Key Benefits

- Single hardware and software solution designed to speed time to implementation of your big data applications.
- Augments open Apache™ Hadoop® technology with real-time capabilities
- Reduces development costs and speeds software evolution with Hadoop.
- Scales easily and linearly as the data grows.
- Simplifies integration of tools and customization of resources.
- Helps ensure enterprise-class availability and offers granular security.
- Disaster recovery protects systems against site-wide disasters.
- Launch a full-featured application quickly with application development tools, and add new features to the application efficiently over time.
- High-availability provides continuity at a single site and protects against component failure.
- The platform supports a spectrum of big data applications on one ICT infrastructure.
- Data and content is loaded as is, without the need for a schema, so you can search the data as soon as it is loaded.

- The solution comes with integrated search features, including full language support and geospatial search.
- Consolidates storage and server clusters on a unified Cisco fabric over Ethernet.
- Unifies management and system maintenance across an entire cluster.
- Uses existing business-intelligence tools and ICT expertise.
- Lowers the administrative burden across processing environments.
- Delivers world-class MarkLogic and Cisco professional support.

An Enterprise-scale Big Data Architecture

The MarkLogic and Cisco next-generation model for managing big data is based on a MarkLogic® Enterprise NoSQL Database running on Cisco Common Platform Architecture (CPA) for Big Data. The solution features a distributed “shared-nothing” architecture. This means that each node is independent; there is no shared memory or storage, and no single point of contention. The model incorporates elements with enough processing power to meet the big data requirements for high volume, velocity, variety, and complexity. Enterprises can determine which elements they need based on the organization’s business models, strategies, and specific data center and network attributes.

MarkLogic Enterprise NoSQL Database captures, reads, and updates unstructured, semi-structured, and structured data at scale, in real-time. It provides a reliable, scalable, and secure big data database platform that enables enterprises to deliver value by taking advantage of existing tools and expertise. MarkLogic is:

- *Powerful and complete.* Enterprises get everything they need to deliver big data value: a schemaless database (no rigid representation of the data it holds); fast, high-quality search; and application services.
- *Accessible and flexible.* ICT staff can code in the language of their choice and deliver applications quickly; the database integrates smoothly with business intelligence tools.
- *Proven and trusted.* The database is enterprise-hardened; there are over 500 deployments in organizations worldwide.

Cisco® Common Platform Architecture (CPA) for Big Data

The Cisco Common Platform Architecture (CPA) for Big Data is designed to meet a broad range of big data platform demands, including NoSQL and Hadoop. The Cisco CPA for Big Data solution stack provides integrated compute, storage, connectivity, and unified management. Components include Cisco Unified Computing System™ (Cisco UCS®) servers designed for diverse compute, I/O, and storage capacity demands, and fabric interconnects and extenders that consolidate IP traffic, storage, and server clusters over Ethernet. Cisco CPA for Big Data meets diverse big data requirements and supports leading software

distributions from Cisco ecosystem partners. Cisco CPA for Big Data offers high-performance and high-capacity options.. See Table 1. The high-performance option provides a balance of compute power with I/O bandwidth optimized for price/performance. The high-capacity option is optimized for low cost per terabyte.

MarkLogic’s Unique Enterprise NoSQL Database

Data analysts today must deal with growing volumes of information that make it difficult and time-consuming to separate useful and actionable information from the rest. They risk drawing incorrect conclusions because they overlook buried information, or they miss a critical deadline or opportunity because they cannot get relevant data fast enough—or get it at all.

For more than a decade, MarkLogic has delivered a powerful and trusted platform that includes an enterprise-grade NoSQL database, search engine and application services that enables organizations to unify their data and turn it into valuable and actionable information. MarkLogic Enterprise NoSQL Database is designed to deal with huge volumes of highly diverse data, allowing users to interact with all the data they need—right now, in real time. The database provides a reliable, scalable, and secure big data platform that gives enterprises an information advantage by making more kinds of data immediately available, and also by using the enterprise’s existing investment in data tools and expertise.

Next-generation big data requires a next-generation database. MarkLogic Enterprise NoSQL Database meets the needs of enterprise big data implementations because it is engineered for volume, velocity, variety, and complexity.

Volume

Volume refers to the massive quantities of data that organizations must harness if they are to improve decision-making. Data volumes are increasing at an unprecedented rate. The MarkLogic database is optimized for today’s advanced scale-out hardware. It supports fast look-up and a shared-nothing architecture that features independent nodes and no contention.

Velocity

The speed at which data is created, processed, and analyzed continues to accelerate. Contributing to higher velocity is the real-time nature of data creation, and the need to incorporate streaming data into business processes. Velocity impacts latency, the lag time between when data is created or captured, and when it is accessible. For time-sensitive processes, certain types of data must be analyzed in real time to be of value. MarkLogic’s database is both fast and agile. It delivers data in real time with performance alerting, and provides multi-version concurrency control so that transactions that are writing and reading do not block each other.

Figure 1: Integrating MarkLogic NoSQL Database and Cisco Common Platform Architecture for Big Data

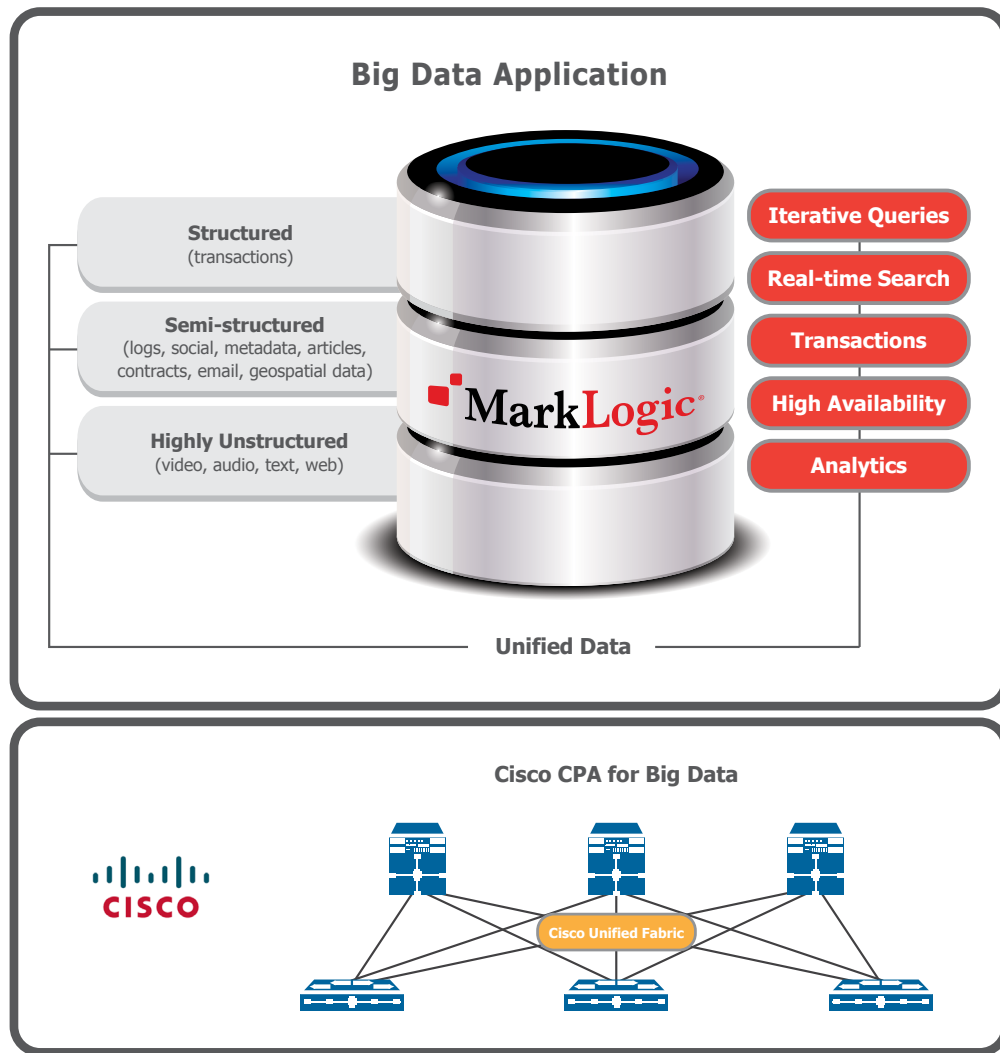


Table 1: Cisco CPA for Big Data: High Performance and High Capacity Racks

	High Performance Base Rack	High Capacity Base Rack
Connectivity and Management	10-Gbps unified fabric supported by two Cisco UCS 6296UP 96-Port Fabric Interconnects (supports up to 160 servers) and 2 Cisco Nexus 2232PP 10GE Fabric Extenders	10-Gbps unified fabric supported by two Cisco UCS 6296UP 96-Port Fabric Interconnects (supports up to 160 servers) and 2 Cisco Nexus 2232PP 10GE Fabric Extenders
Management	UCSM Manger	UCSM Manger
Servers	16 Cisco UCS C240 M3 Rack Servers, each with: 2 Intel Xeon processors E5- 2665 at 2.4 GHz, 256 GB of memory, Cisco UCS VIC 1225, 12 SFF 1-TB 7.2K 2.5-inch SATA HDDs, LSI MegaRAID 9266-CV 8i card	16 Cisco UCS C240 M3 Rack Servers, each with: 2 Intel Xeon processors E5- 2640 at 2.5 GHz, 128 GB of memory, Cisco UCS VIC 1225, 12 LFF 3-TB 7.2K 3.5-inch SAS HDDs, LSI MegaRAID 9266-CV 8i card
Processors/Cores	32/256	32/192
Memory	4TB	2TB
Storage Capacity	384TB	576TB
IO Bandwidth	32Gbytes/sec	16Gbytes/sec

The solution is offered as reference architectures and as [Cisco UCS SmartPlay solutions](#) that can be purchased by ordering a single part number.

Variety

Organizations need to integrate and analyze data from an array of both traditional and non-traditional information sources, from within and outside the enterprise. Data can be generated in countless forms, including: text, web, tweets, audio, video, click streams, log files, application-specific documents, and other formats. MarkLogic can handle all these types of data, with metadata being extracted, indexed, and stored to aid in search.

Complexity

MarkLogic uses XML to handle variable-length elements, hierarchical relationships, sparse data, and schema-independent data. Information that should not, or cannot, be disassembled into rows and columns—such as contracts, manuals, books, emails, tweets, and metadata—is ideally suited to the XML-based, document-centric model used in the MarkLogic database. Universal indexes allow loading information as is, thus avoiding rigid, predefined schemas. This is especially efficient for indexing and querying information with poorly defined, poorly followed, changing, or unknowable schemas. MarkLogic also readily supports information that adheres to a schema.

Use Case #1: A Large Information Provider

Organization

This major information provider for the petrochemical and energy market gathers a staggering amount of market data from reporters and researchers around the world and packages it as statistics, reports, and website content.

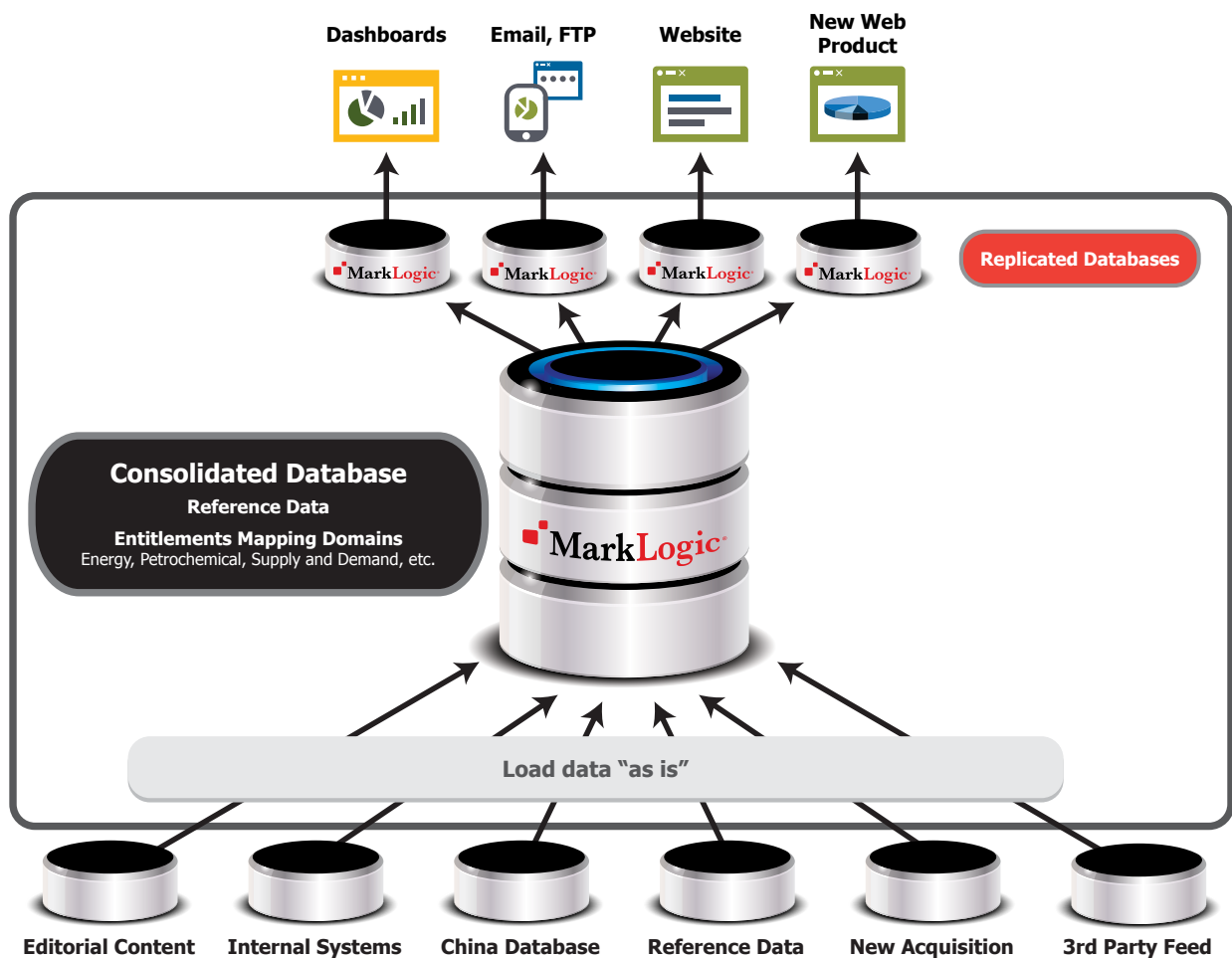
Challenges

- Large data volumes made logistics difficult.
- New acquisitions have complicated the data mix.
- Islands of information and disparate data have hindered integration and innovation.
- Timely data analysis is of paramount importance.

Solution

The company chose MarkLogic to meet their data management and distribution needs. The most important advantage was its schema-less database. They needed to add and consolidate new data structures from multiple sources without making code changes, and MarkLogic gives them that flexibility.

Figure 2: An Infrastructure for Global Information Distribution



They also appreciate the solution's programmable structure, which simplifies data publishing and retrieval. The database's robust search capabilities provide another important advantage. Yet another benefit is asynchronous replication. As shown in Figure 2, the company is now able to set up database replicas for their global customers in different locations.

The firm has sharpened its competitive edge. By centralizing key information and simplifying integration from outside sources, they are able to deliver data to global customers with greater efficiency and precision. Perhaps most importantly, they have accomplished this while continuing to maintain their existing services.

Use Case #2: A Multimedia Content Provider

Organization

Together with its subsidiary companies, this enterprise provides a range of multimedia content and editorial services, including sports data, entertainment news, weather forecasts, syndicated photos, and more. Their diverse customer base ranges from small businesses to global corporations and government agencies.

Challenges

- Different data types were isolated in different systems and silos.
- Traditional databases could not align all the diverse and complex content.
- The company was spending too much time managing and manipulating data.
- Data security is important

Solution

According to the organization's managers, the new MarkLogic platform aligns perfectly with the company's business strategy and development goals. MarkLogic delivers a central, common database—one system to accommodate all types of data. The new implementation has the flexibility and scalability needed to deal with varied content. It can accommodate future growth and also offers a high degree of security. The company is able to complete projects much more quickly. As an example, one

project would have taken up to 100 workdays to complete with the old platform, but with the MarkLogic solution it took only 34 days—a 66 percent boost in efficiency.

The company can now store and manage content in a way that aligns with how it wants to present data to customers. It can deploy new solutions that were previously too difficult, time-consuming, or expensive to build. The organization also finds that even the most traditional developers become invigorated by the clever and unique ways of handling structured and unstructured data that are enabled by the MarkLogic solution.

For any enterprise looking to enhance its business operations by building big data applications, the MarkLogic and Cisco partnership offers a powerful next-generation solution.

More Information

For more information about MarkLogic, visit www.marklogic.com or www.cisco.com/go/bigdata, or contact:

David Ponzini

SVP, Corporate Development
david.ponzini@marklogic.com
650.655.2328 (Office)
925.997.9872 (Mobile)

Jeff Faraday

Director, Alliances
jeff.faraday@marklogic.com
650.655.2372 (Office)
925.872.6545 (Mobile)

© 2013 MarkLogic Corporation. All rights reserved. This technology is protected by U.S. Patent No. 7,127,469B2, U.S. Patent No. 7,171,404B2, U.S. Patent No. 7,756,858 B2, and U.S. Patent No 7,962,474 B2. MarkLogic is a trademark or registered trademark of MarkLogic Corporation in the United States and/or other countries. All other trademarks mentioned are the property of their respective owners. [SS-MLIH-13-02]

MarkLogic Corporation
www.marklogic.com
sales@marklogic.com
+1 877 992 8885

Headquarters
999 Skyway Road, Suite 200
San Carlos, CA 94070
+1 650 655 2300