Data Sheet

Cisco Big Data Warehouse Expansion Solution

Featuring MapR Distribution including Apache Hadoop

Driven by ever-growing enterprise data, companies are facing huge expenses adding capacity to their existing Data Warehouses (DW). To decrease this spend, companies are looking at lower-cost alternatives that preserve their existing reporting and analytics.

Cisco Big Data Warehouse Expansion (BDWE) reduces warehouse management costs by enabling organizations to collect and retain data that was previously too expensive to store---data can now be made available for analysis and improved business insights at 1/10th to 1/50th the cost on a per terabyte basis. In addition, identifying and offloading infrequently used data from the existing data warehouse to low-cost Big Data stores yields immediate performance and cost benefits. The Cisco BDWE solution optimizes Cisco UCS hardware for running the award-winning MapR Distribution including Hadoop, and includes software for federating multiple data sources, and a comprehensive services methodology for assessing, migrating, virtualizing, and operating a logically expanded warehouse. It uniquely blends a best-in-class data, computing, and network infrastructure that reduces risk and delivers accelerated performance and scalability.

The BDWE with MapR solution benefits include the ability to:

- **Enhance Analytics** – Access not just current and recent history, but also extended historical data that is typically archived and not easily accessible.
- **Control Costs** – Economically locate hot/cold data in its proper place to fully take advantage of technology investments.
- **Improve Performance** – Benefit from an optimized enterprise-class, high-performance Hadoop and network infrastructure solution that's uniquely bundled for the job.
- **Enable Hadoop Production Success** – Serve business-critical needs for Big Data applications that cannot afford to lose data, must run on a 24x7 basis, and require immediate recovery from node and site failures—all with a smaller data center footprint.
- **Gain Competitive Advantage** – Cost-effectively collect, retain, and utilize all of your company's data assets with enhanced analytics for higher productivity that address business change.

- **One Platform for Big Data Applications** – MapR provides an enterprise data hub for Big Data with Hadoop at its center. Hadoop provides a general purpose platform for a variety of workloads including data storage, integration from multiple sources, database operations, analytics, search, and real-time stream processing.
- **Reduce Risk** – Use proven software, network and computing infrastructures to adopt Big Data and logical data warehousing.

Implementation Methodology

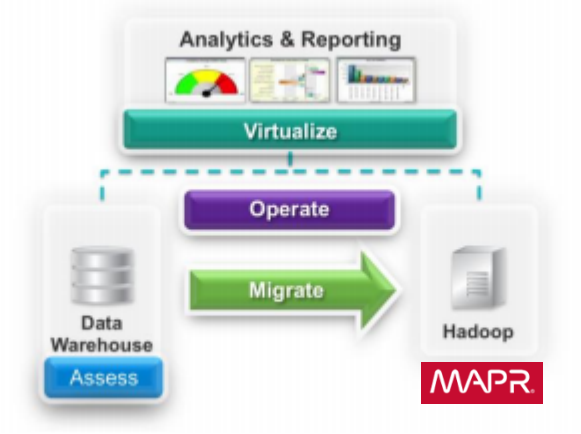
At the foundation of the solution is the methodology, which provides:

- **Proven Formula for Success** – Confidently achieve your goal, while saving time and money, using documented best practices.
- **World-class Experts and Technology** – Experienced consultants working with best-of-class technology deliver great results.
- **Reduce Risk While Advancing Your Data Strategy** – With an end-to-end solution from a trusted vendor, you achieve your data and business goals at minimal risk.

Table 1. The methodology consists of four major phases.

Feature	Description
Assess	Collect data usage statistics, analyze data, prepare ROI statement, and propose recommendations.
Virtualize	Review existing usage profiles, determine sources to be migrated, connect and virtualize DW data, test and tune applications.
Migrate	Determine data migration approach, migrate identified DW data to Hadoop target, connect and virtualize Hadoop data, test and tune applications.
Operate	Manage and optimize queries, and periodically assess DW data load.

Figure 1: Big Data Warehouse Expansion Reference Architecture



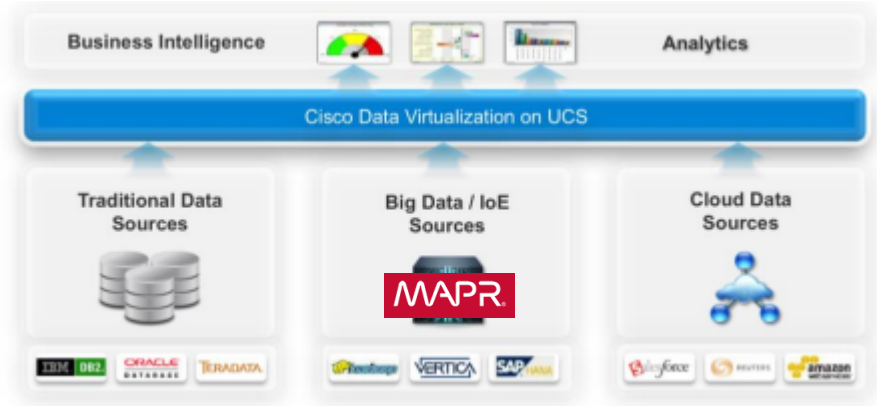
Data Virtualization

After selected warehouse data has been offloaded to Hadoop, the Cisco Information Server is used to federate both data sources and offer a “single view” of data. Analytic and business intelligence reports are now enriched because they now have access to *more* data, from the warehouse as well as the Hadoop Big Data store.

Table 2. Data virtualization logically makes all data accessible.

Feature	Description
Data Access	Connect and expose data from diverse sources.
Data Federation	Execute and optimize queries across the data warehouse, Hadoop Big Data stores, and more. Optimized algorithms speed queries across disparate data sources.
Data Delivery	Deliver data to diverse consuming applications including analytics and BI tools.

Figure 2: Data Virtualization: Logically Access Data Warehouse with Offloaded Data and More



Unified Computing System (UCS)

For implementing a MapR Big Data cluster, Cisco offers a comprehensive solution stack. The Cisco UCS Common Platform Architecture (CPA) for Big Data with MapR includes computing, storage, connectivity, and unified management capabilities. Unique to this architecture are transparent, simplified data and management integration features with an enterprise application ecosystem.

Table 3. UCS servers are optimized for Hadoop deployments.

Feature	Description
High performance and scaling	UCS C240 M3 ideal for Big Data deployments.
Ease of deployment	Rapid deployment of server using "service profiles."
Comprehensive manageability	Easy to manage and maintain entire cluster.
Coexistence with enterprise applications	Transparent, simplified management and data integration.
Enterprise-class service and support	Leading industry support from Cisco and its partners.

For more information about Cisco CPA for Big Data, please visit: http://blogs.cisco.com/datacenter/cpa/

MapR: A Complete Hadoop Platform

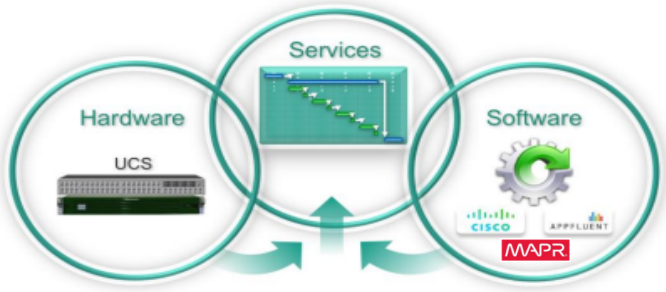
As the technology leader in Hadoop, MapR provides an enterprise-class, high-performance solution that is fast to develop and easy to administer. With significant investment in architectural innovations, MapR delivers more than a dozen tested and validated Hadoop software modules over a fortified data platform, offering exceptional ease of use, reliability, and performance for Hadoop solutions.

Table 4. MapR Distribution including Apache Hadoop

Feature	Description
Ease of Use	As the technology leader in Hadoop, MapR provides an enterprise-class, high-performance solution that is fast to develop and easy to administer. With significant investment in architectural innovations, MapR delivers more than a dozen tested and validated Hadoop software modules over a fortified data platform, offering exceptional ease of use, reliability, and performance for Hadoop solutions.
Reliability	MapR provides lights-out data center capabilities for Hadoop. Features include self-healing of the critical services that maintain the cluster nodes and jobs, snapshots that allow consistent point-in-time recovery of data, mirroring that allows wide-area inter-cluster replication, and rolling upgrades that prevent service disruption during software upgrades.
Performance	MapR is twice as fast as any other Hadoop distribution. To provide superior and exceptional performance over other Hadoop distributions, MapR uses an optimized shuffle algorithm, direct access to the disk, built-in compression, and code written in advanced C++ rather than Java. As a result, MapR provides the best hardware usage when compared to any other distribution.

For more information about Cisco UCS CPA with MapR, please visit: http://www.mapr.com/cisco

Figure 3. Cisco & MapR Delivers a Complete Big Data Warehouse Expansion Solution



For more information about Cisco's Big Data Warehouse Expansion solution, visit: http://www.cisco.com/go/datavirtualization or email dv-sales@cisco.com.

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.**

Printed in USA