

Brochure

Deploy a successful roadmap for Hadoop™

HP Services for Hadoop™



Pioneered by sensors, smart devices, and social collaboration technologies, a new information world has arrived. As these change drivers take hold, a huge mass of transactional data is starting to emerge—structured, semistructured, and unstructured—capturing trillions of bytes of information about customers, suppliers, and operations. In itself, this sheer volume, velocity, and variety of data, also called “big data,” is a global phenomenon—but what does it imply? Though many organizations around the world regard this collection of data and its value with skepticism, the use of big data is becoming a key way for leading organizations to outperform their peers. McKinsey estimates that retailers embracing big data have the potential to increase their operating margin by more than 60 percent.¹

In capitalizing the value of data available to them, organizations must go through a paradigm shift—rethink their management of the sheer scale of this data, their infrastructure, and their current management frameworks and IT processes.

To capture value from big data, organizations need to deploy new technologies. The integration of sophisticated analytics and unstructured data to create new business value is often inhibited by legacy or proprietary systems with incompatible standards. Apache Hadoop, an open-source nonproprietary technology and a new way for enterprises to store and analyze data, opens the door to a world of possibilities. But, the decision to deploy and start developing on a Hadoop cluster is not something you can make overnight. It needs planning, strategic thinking, a vision of where the change could possibly take you, how quickly and how long it will take for the returns on your investment to impact your business growth positively. It’s a journey that can be accomplished better with the HP Roadmap Service for Hadoop.

Built on the underlying principles of scalability, performance, and low cost, Hadoop is a Java-based, open-source framework that enables users to run massive and parallel data processing projects on a low scale-out architecture.

Coping with big data

Big data is a relatively new term in the IT world. The rise of big data, which is much more than just the immense volume of raw data pooling inside of most organizations, has already become a signature challenge in industries where rich data is standard fare.

Ignoring big data is no longer an option. IDC predicts that big data will join mobile and cloud as the next “must have” competency, as the volume of digital content grows to 2.7 ZB in 2012, up by 48 percent from 2011, rocketing toward 8 ZB by 2015.² Organizations unable to manage this data are overwhelmed by it. And this leads to the current predicament for businesses: organizations have more access to potential insight than ever before, yet as this gold mine of data grows, the percentage of data that businesses can process is shrinking—fast.

Paradoxically, big data is more than a challenge; it’s an opportunity to find insight in new and emerging types of data, to make your business more agile, and answer questions that, in the past, were beyond reach. Organizations can apply big data methods to solve existing business problems, implement new business models, and drive growth in innovative, new ways. Sophisticated analytics of big data can help you recognize new patterns and ask better questions with greater confidence, substantially improve decision making, minimize risks, and unearth valuable insights that would otherwise remain hidden.

IDC predicts that big data will join mobile and cloud as the next “must have” competency, as the volume of digital content grows to 2.7 ZB in 2012, up by 48 percent from 2011, rocketing toward 8 ZB by 2015.

¹ “Big Data: The next frontier for innovation, competition, and productivity,” McKinsey & Company, June 2011.

² “IDC Predictions 2012: Competing for 2020,” IDC, December 2011. Note: 1 ZB = 1 billion terabytes

Real-world Hadoop implementation



Enable experimentation to discover needs and improve performance	Support human decision making with automation	Innovate business models, products, and services	Strategize segmentation of customers	Manage inventory based on supply-demand signals
Financial institutes can gather its massive amount of transaction information and process analytics to reveal opportunities.	Oil and gas providers can use predictive analytics with Hadoop, to identify possible oil and gas reserves.	Telecom can understand market trends to better analyze its penetration, coverage, and revenues, and forecast the potential of front-end customer-facing products and services.	Retail chains can use the tremendous amount of customer data it generates, from its loyalty programs, to create promotions targeted to specific customer segments.	By analyzing supply and demand signals, retailers can optimize the supply chain.

To capture value from big data, organizations need to employ new technologies of storing, processing, and analyzing it. Stores of business data are only as good as the methods for extracting and putting them to work. Until now, there were few technologies or approaches that allowed organizations to harvest value from big data. Today, Apache Hadoop, an open-source framework for big data, opens the door to a world of possibilities, providing organizations a cost-effective, innovative approach for advanced analytics. Hadoop gives you a way to economically store and process unstructured data in a reasonable time, making it perfect to run data-intensive applications on cost-effective hardware.

Growing adoption of Hadoop

An increasing number of businesses are recognizing the advantages of Hadoop due to its ability to scale and process large amounts of data using low cost components. The reduction in costs associated with storage and processing of data is what makes Hadoop an attractive proposition for delivering value from unstructured data. What's more, it delivers results in a condensed structured summary, suitable for use by existing business intelligence systems.

Hadoop utilizes clusters of machines and coordinates work among them. Clusters can be built and scaled out with inexpensive computers. The Hadoop software package includes:

- **Hadoop Distributed File System (HDFS):** A massively data parallel file system that provides a low-cost redundant file system utilizing distributed storage capability across large numbers of commodity disk drives, (for instance, SATA) within each of the compute data nodes.
- **MapReduce:** A parallel programming framework for extracting, transforming, and exporting information from unstructured data.

The Hadoop infrastructure is termed “elastic” because adding extra data nodes increases compute and bandwidth to network and disk. The framework schedules and manages data processing at the node closest to the data source, to achieve maximum performance. The HDFS is also aware of the physical cluster architecture, replicating data throughout the cluster for added resilience. If a machine fails, Hadoop continues to operate the cluster by shifting work to the remaining machines, while automatically creating an additional copy of the data from one of the replicas it manages. As a result, clusters are self-healing for both storage and computation without requiring frequent intervention by systems administrators.

Low-cost and scalability are key requirements of any unstructured data management platform for it to be attractive to adopters

Figure 1. Why is Hadoop an attractive proposition

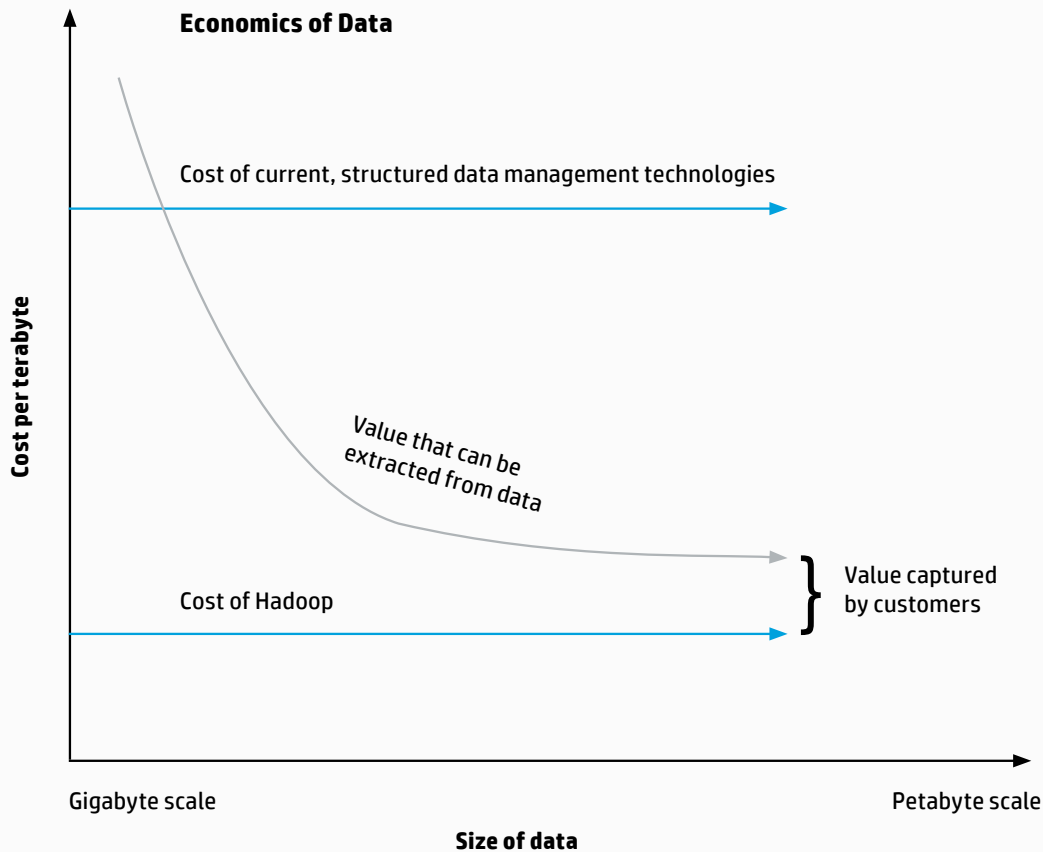
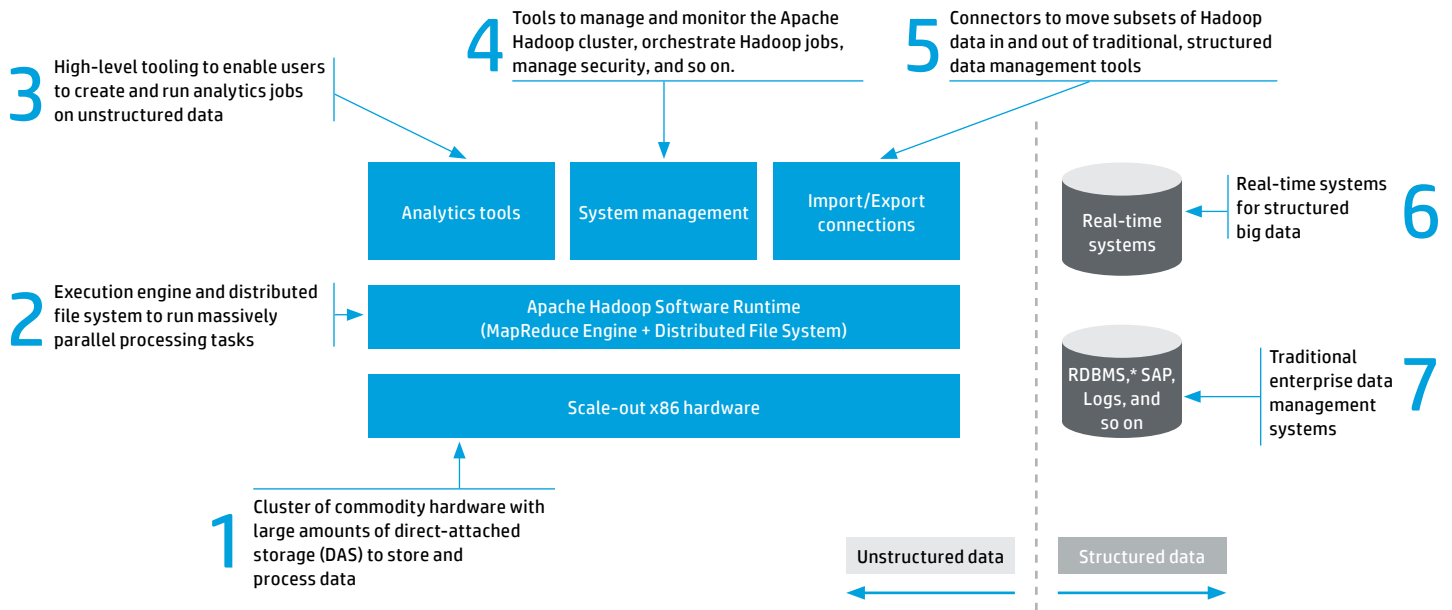


Figure 2. Components of the Hadoop ecosystem

The Hadoop ecosystem consists of a stack of new big data software tools to run massively parallel, batch-based processing jobs on unstructured data, and interfaces with traditional, structured data management tools



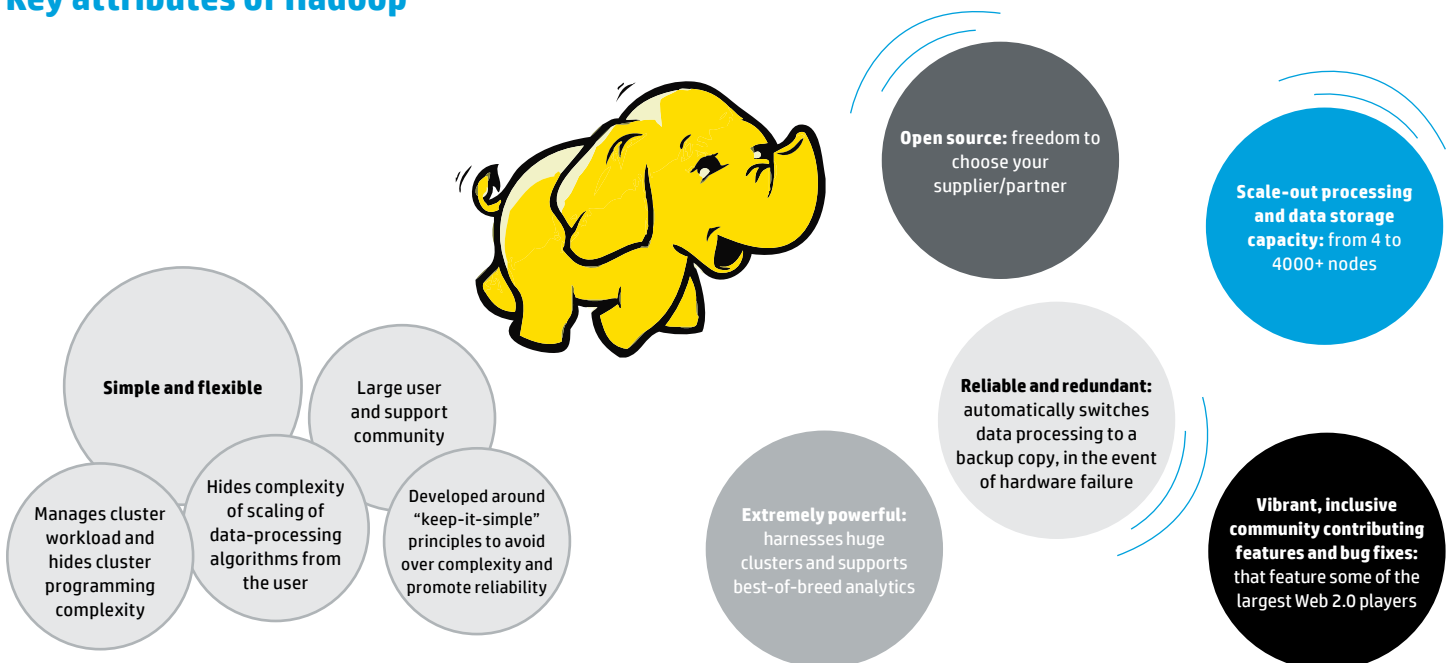
*(Relational Database Management System)

Building a production quality cluster and deploying the Hadoop framework to enterprise standards is not a simple task. An initial lack of experience and desire for experimentation frequently leads to delays in successfully implementing big data initiatives, sabotaging future efforts where big data could have significant impact. The unfamiliarity, complexity, and rapidly evolving nature of this open-source framework add to the associated risk of approaching its implementation with a wrong mindset—applying traditional data warehousing to Hadoop.

Impediments, such as, misleading information from vendors to protect margins and revenues, internal conflicts on the meaning and significance of big data, and lack of awareness about risks and issues associated with Hadoop implementation, can hinder the success of Hadoop deployments. What do you do?

A targeted roadmap service can assess the Hadoop fitment, implementation, and optimization for your organization.

Key attributes of Hadoop



HP Roadmap Service for Hadoop: Tailored to your specific needs

HP Roadmap Service for Hadoop can help you develop a common vision, understand suitable sources of data and its sensitivity, identify organizational inhibitors, understand risks and mitigation strategies, and develop a roadmap that drives the successful planning, deployment, and support for Hadoop within your organization. We leverage our experience, best practices, and proven and tested HP reference architectures—all tailored to the specific needs and drivers of your roadmap.

HP Roadmap Service for Hadoop adopts a four-step approach to successfully deploy Hadoop within your organization, without investing in significant IT muscle.

Conduct assessments: Identify possible use of Hadoop within the business, sources and volumes of input data, and potential data transformations

Leverage best practices: Walk through current best practices for managing, operating, and maintaining Hadoop clusters and identifying potential opportunities, risks, and mitigations

Suggest design and configuration recommendations: Produce configuration and process recommendations for the implementation of Hadoop based on HP reference architecture, assessments, and best practices

Identify risks, conduct gap analysis, and propose a roadmap: Propose a roadmap and plan for a successful Hadoop implementation based on specific customer requirements and HP recommendations

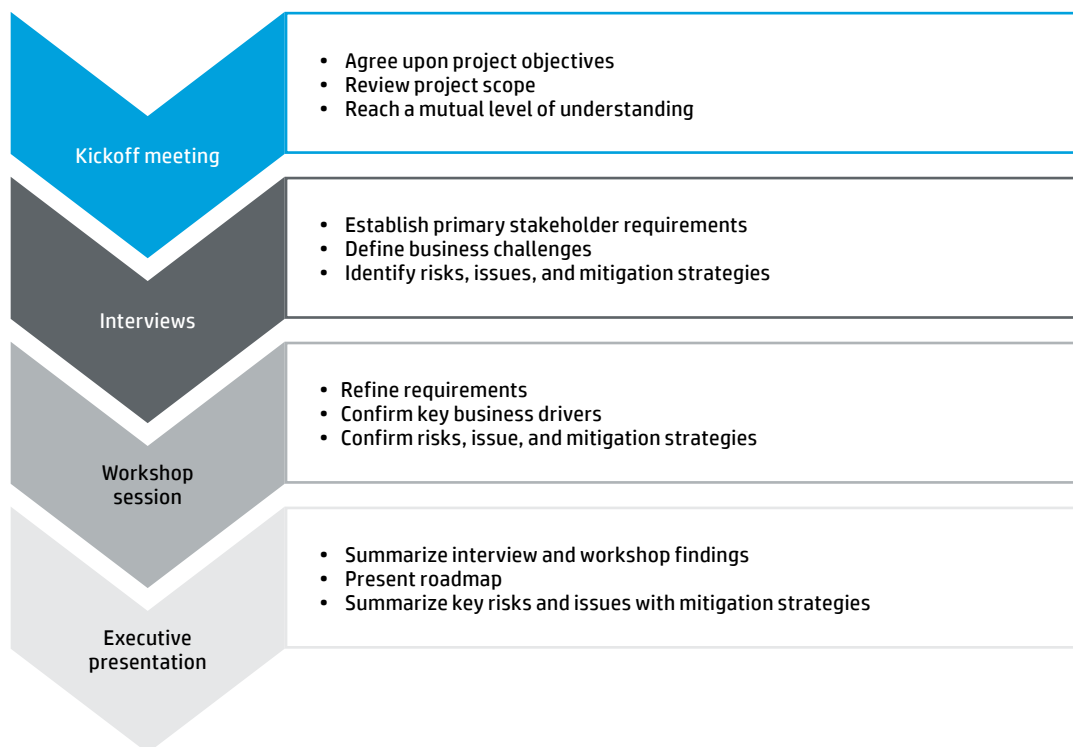
As part of implementing Hadoop, HP can provide additional services that help ensure successful Hadoop deployment and integration into your environment. These services range from the HP Start Service for Hadoop—installs and configures the Hadoop framework and infrastructure, according to the design—to support, education, cluster management, and data management services.

Delivering business value

Establish an implementation roadmap

Deploying Hadoop requires change on many levels—from establishing a starting point to determining the right plan, to defining an appropriate architecture, to instituting new operating models, and creating a workable execution plan. The HP roadmap service offers a collaborative approach to define a focused, prescriptive, and actionable roadmap for deploying Hadoop. It eases the transition and accelerates your journey in order to achieve benefits more quickly. We enable you to build a prioritized list of activities that increase the value and business outcomes of your Hadoop cluster.

Figure 3. HP Roadmap Service for Hadoop: Key highlights



Identify business drivers

HP Roadmap Service for Hadoop offers recommendations for moving your Hadoop project from the consideration phase to production. At each step, we identify key business drivers for your business to help you plan an effective Hadoop design for your organization. With HP, discover ways to align your Hadoop implementation with best practices and determine which next steps to take to prepare for deployment.

Categorize risk and mitigation strategies

Your organization needs to plan for Hadoop deployment to avoid a constant state of reactive crisis management. What should be the scale of your Hadoop deployment? How do you manage it? What are the key considerations in interfacing, operating, implementing, and migration of a Hadoop design? HP Roadmap Service for Hadoop is designed to help you understand, establish, and outline the strategies for implementing, managing, operating, and interfacing your Hadoop deployment. Our consultants guide you through onsite work sessions and discussions that help determine your Hadoop initiatives and goals, identify the scale, and provide a high-level implementation strategy roadmap for successfully achieving your Hadoop requirements.

Identify critical success factors

By identifying your critical success factors, you can create a common point of reference to help you direct and measure the success of your Hadoop project. HP Roadmap Service for Hadoop helps you identify and categorize critical success factors for your Hadoop cluster while enabling better planning and designing of your Hadoop initiative.

Establish a consistent view across stakeholders

HP Roadmap Service for Hadoop helps establish an aligned view between stakeholders, and an acceptable and durable decision. This service aims to build trust and consensus across stakeholders and provides a structured plan for your Hadoop deployment.

Address your organization's unique needs using our comprehensive portfolio

HP has been helping customers implement high-performance computing (HPC) for many years and will leverage its experience and knowledge to help you plan for and ease the risk of implementing Hadoop within your organization.



Numbers speak ³

8 out of 10 of the world's most trafficked websites

4 out of 5 of the world's largest search engines

7 out of 10 of the world's largest cloud service providers

3 most popular social media properties in the U.S.

15,000

Issued more than 15,000 private cloud licenses

500

Many of the top 500 HPC Clusters

³ Tabulations based on <http://www.alexa.com/topsites> and <http://www.top500.org/> estimates.

Customer intelligence at HP: Closing the information asymmetry gap using unstructured data

A pilot project, called project Fusion, completed by HP Labs and HP Global Customer Intelligence shows that data from social media can be merged with an organization's data to predict customer behavior, with up to 90 percent accuracy.

The pilot demonstrates that the collaboration between marketing and IT departments is critical in unlocking the business value of social media. The system used can correlate social-media conversations about specific product features to actual customer transactions in real time.

How does the HP system work?

Like many companies, HP has already wrung most of the useful information out of its existing customer database. To tap into the richness of social media, HP had to invent a new technology.

The text analytics HP Labs created are part of a broader live customer intelligence project that includes data visualization, parallel processing, and other key components.

The software is also designed to work in real time and works with database technology such as that of Vertica.

Project Fusion: Results

Project Fusion combines two different kinds of data:

- Unstructured data (**Amazon.com** reviews, customer surveys, customer support logs, and other natural-language text)
- Structured data (customer support tickets, sales transactions, customer demographics, and so on)

Using the new HP Lab text analytics technology, the team first converted unstructured data into a structured format. Analysts could then use standard data-mining and statistical tools to analyze the two data sets together. In one case, social signals predicted support tickets with 90 percent accuracy, while in the other, different social signals were highly correlated to sales.

Bringing the solution to HP customers

Complementing the capabilities created in the lab, HP further expanded its portfolio of unstructured data analytics by acquiring Autonomy Corporation plc—a leader in enterprise information management and meaning-based computing.

“We are committed to helping our customers solve their toughest IT challenges. The exploding growth of unstructured and structured data and unlocking its value is the single largest opportunity for consumers, businesses and governments,” says *Meg Whitman, President and Chief Executive Officer, HP.*

“Autonomy significantly increases our capabilities to manage and extract meaning from that data to drive insight, foresight, and better decision making,” *she adds.*⁴

Tap into potential, tap into limitless growth

No matter how you slice it, the case for Hadoop as a framework for managing, coordinating, and analyzing large sets of data is very powerful. Chances are, because the potential business value is so compelling, you'll deploy it sooner than expected. HP Roadmap Service for Hadoop will help you create a unified roadmap to effectively integrate Hadoop into your data management environment, allowing you to transform your big data challenges into big business opportunities.

Global citizenship at HP

At HP, global citizenship is our commitment to hold ourselves to high standards of integrity, contribution, and accountability in balancing our business goals with our impact on society and the planet. To learn more, visit hp.com/hpinfo/globalcitizenship, and for information about the HP Eco Solutions program, go to hp.com/ecosolutions.

Work with our experts to plan your roadmap

HP Services for Hadoop can help you plan, strategize, and manage Apache Hadoop for large-scale data processing and analysis. To know more about the services, visit hp.com/us/en/business-services/it-services.html?compURI=1079292.

⁴ For more information, refer to the HP press release, visit hp.com/hpinfo/newsroom/press/2011/111003xb.html.

Get connected

hp.com/go/getconnected

Get the insider view on tech trends, support alerts, and HP solutions

© Copyright 2012 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

4AA3-7986ENW, Created February 2012; Updated May 2012, Rev. 1

