



Preview: IBM PureData System for Hadoop H1001 helps simplify big data for the enterprise

Table of contents

1 Overview	3 Statement of general direction
2 Planned availability date	4 Announcement countries
2 Description	

At a glance

IBM® PureData™ System for Hadoop H1001 is the newest member of the IBM PureSystem family. It is a purpose-built, standards-based expert integrated system that architecturally integrates IBM InfoSphere® BigInsights' Apache Hadoop-based software, server, and storage into a single, easy-to-manage system.

Highlights of IBM PureData System for Hadoop H1001:

- Built-in expertise to help accelerate big data time to value that offers:
 - Fast deployment, ready to load data in hours
 - Built-in visualization
 - Exceptional social data, machine data, and text analytics application accelerators
- A simplified experience to help simplify big data adoption and consumption
 - Single console for full system administration
 - Rapid software updates with automation
 - Single vendor support
- Integration by design for implementation of enterprise-class Hadoop
- Integrated bidirectional archive and restore
- Built-in analytics development environment

Overview

IBM PureData System for Hadoop H1001 is built to optimize Apache Hadoop data services for big data analytics and online archive with appliance simplicity.

IBM PureData System for Hadoop combines IBM InfoSphere BigInsights™ and IBM System x® hardware for an integrated Hadoop system. It delivers enterprise Hadoop capabilities with easy-to-use analytic tools and visualization for business analysts and data scientists. It includes rich developer tools, powerful analytic functions, and exceptional administration and management capabilities, as well as the latest versions of Hadoop and associated projects. In addition, IBM PureData System for Hadoop provides extensive capabilities with enhanced big data tools for monitoring, development, and integration with many more enterprise systems.

IBM PureData System for Hadoop:

- Provides an exploratory environment for data analysts to help understand new and emerging data sources

- Provides a landing area for data from other sources, enabling aggregation, initial integration, visualization, and exploration before data moves to other parts of the analytic ecosystem
- Provides an integrated management console for the entire system
- Provides built-in analytics and enterprise functionality, on top of Hadoop technology, to help meet big data enterprise requirements
- Integrates advanced hardware cluster management capabilities with IBM InfoSphere BigInsights
- Incorporates integrated data archiving and transfer capabilities with connectivity to enterprise data warehouse systems
- Integrates with IBM DB2® , IBM Netezza® , IBM PureData System for Analytics, and IBM InfoSphere Guardium®

Planned availability date

Second half, 2013

Previews provide insight into IBM plans and direction. Availability, prices, ordering information, and terms and conditions will be provided when the product is announced.

Description

IBM PureData System for Hadoop can be used with other IBM products for advanced analytics and a more comprehensive information management strategy. It delivers simple-to-use, robust administration and management capabilities, rich developer tools, built-in powerful analytic functions, and the latest open source components from Apache Hadoop and associated projects.

IBM PureData System for Hadoop offers simplicity, flexibility, and consumability in a single integrated system.

Simplicity

IBM PureData System for Hadoop incorporates an integrated management console for the entire system. Through this console, users and administrators can manage almost all aspects of the system; for example, monitoring status of MapReduce jobs, developing and performing analytics, setting user access privileges and permissions, monitoring hardware, data archiving capabilities, and a number of other capabilities.

Flexibility

IBM PureData System for Hadoop integrates with other IBM and third-party products to support advanced analytics and data archiving and transfer, which enables a greater overall information management strategy. It lets you combine traditional data warehouse analysis with dynamic insights from unstructured and semi-structured data by integrating with:

- IBM DB2
- IBM Netezza
- IBM PureData System for Analytics
- IBM PureData System for Operational Analytics
- IBM InfoSphere Streams
- Other popular enterprise data warehouse systems

Consumability

IBM PureData System for Hadoop offers a variety of capabilities to help users address their big data requirements. Key capabilities include:

- Text analytics development - provides an enhanced user experience for developing text analytics applications by delivering easy-to-use navigation of result views, sorting and filtering, and pattern discovery and progress reporting.
- Developer tools - outstanding built-in support for text analytics, support for local mode MapReduce development, easy deployment of applications, and automatic creation of Java™ Database Connectivity (JDBC) connections to Hadoop Hive data sources.
- Machine data analytics development - helps you to analyze various types of machine data, such as system logs, to gain business insights. It:
 - Provides the capability to ingest, parse, and extract machine data.
 - Analyzes the data to generate insights such as patterns and correlations.
 - Provides faceted search capability to more easily find and navigate through large volumes of machine data.
 - Supports visualization of the data and the results of the analysis.
- Social data analytics development - helps you to analyze various types of social media data to gain key insights about consumer sentiment and purchase intent. Then, this information can be leveraged to build applications for lead generation and brand management.
- BigSheets web console - a built-in, spreadsheet-like visualization tool to help users analyze data without MapReduce code. The new chart customization features help make it effortless to access and manipulate data the way you want. New sheets, macros, and readers make it possible to access more data and analyze it in new ways, giving you more control and improved browsing capabilities for Hadoop Distributed File System (HDFS) and Network File System (NFS) files. The addition of new application input parameter types helps make applications easy to run.
- Applications - a new application for Apache HBase helps make it easy for users to access HBase directly from the IBM PureData System for Hadoop web console and view results in BigSheets.

Business Partner information

If you are a Direct Reseller - System Reseller acquiring products from IBM, you may link directly to Business Partner information for this announcement. A PartnerWorld ID and password are required (use IBM ID).

<https://www.ibm.com/partnerworld/mem/sla.jsp?num=213-186>

Statement of general direction

The initial offering will be a full rack, with plans to support both smaller entry points and larger configurations to follow.

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Reference information

For information about IBM InfoSphere BigInsights , refer to Software Announcement [212-442](#), dated November 13, 2012 .

Announcement countries

United States of America (USA) and the USA territory of Puerto Rico.

Trademarks

PureData and BigInsights are trademarks of IBM Corporation in the United States, other countries, or both.

IBM, InfoSphere, System x, DB2 and Guardium are registered trademarks of IBM Corporation in the United States, other countries, or both.

Other company, product, and service names may be trademarks or service marks of others.

Terms of use

IBM products and services which are announced and available in your country can be ordered under the applicable standard agreements, terms, conditions, and prices in effect at the time. IBM reserves the right to modify or withdraw this announcement at any time without notice. This announcement is provided for your information only. Additional terms of use are located at

<http://www.ibm.com/legal/us/en/>

For the most current information regarding IBM products, consult your IBM representative or reseller, or visit the IBM worldwide contacts page

<http://www.ibm.com/planetwide/us/>