IBM Data Engine for Hadoop and Spark - Power Systems Edition: An integrated and optimized solution for high-performance Hadoop-based and Spark-based analytics

Table of contents

At a glance

IBM(R) Data Engine for Hadoop and Spark is a new offering that combines the recently added line of OpenPOWER Linux™ servers designed for big data and analytics with an open source Apache Hadoop and Spark distribution along with optional advanced analytics capabilities.

This solution is ideal for addressing a wide range of new business opportunities such as:

- Retailers can make marketing and merchandising decisions based on historical customer buying patterns and preferences combined with near real-time social media feedback.
- Manufacturers and large institutions can analyze event data from their increasingly instrumented environment to optimize their operations with speed and accuracy.
- Telecommunications and finance companies can combine customer activity with new social media insights to present customers with the best offer matching their personal needs.
- IT and security teams can quickly analyze massive amounts of log data for security breaches or pending issues.

The server, networking, storage, and software components are preintegrated and tested prior to delivery. Services are available to quickly bring the cluster into initial operation.

Overview

IBM Power Systems™ offers a range of solutions for big data, from standard configurations for dedicated Hadoop or Spark infrastructures with the IBM Data Engine for Hadoop and Spark - Power Systems Edition to petabyte-scale, mixed analytics solutions with the IBM Data Engine for Analytics - Power Systems Edition.

Data Engine for Hadoop and Spark offers a range of configurations based on the new storage-dense, analytics-optimized S812LC line of IBM POWER8(R) servers. These servers offer up to 14 large form factor disk drives or solid-state drives and up to 1 TB of memory. These features, combined with the leading multi-threading I/O bandwidth and cache of POWER8 servers, makes them ideal for cost-effective, optimized analytics workloads.

Five node starter configurations are available, handling up to 216 terabytes of raw data and providing over 50 terabytes of usable data in a standard triple replica Hadoop or Spark configuration. Multi-rack configurations are available, providing up to 1.3 petabytes of raw data per rack. Advanced cluster management software is a standard component of the solution.

These cluster configurations can be optionally preloaded with the following leading IBM Analytics software:

- IBM Open Platform with Apache Hadoop is an IBM supported distribution that includes the open source components that are aligned with the Open Data Platform (ODPi) consortium. This includes core Apache Hadoop components along with Apache Ambari for ease of deployment and management. In addition, this distribution includes Apache Spark, which provides extremely fast in-memory analytics.
- IBM BigInsights(R) V4 offers a selection of value-added services that provide significant enhancements beyond the capabilities of the open source stack:
  - BigInsights Analyst supports data center modernization with capabilities to run massively parallel SQL queries directly on the Hadoop cluster, including federation with external data sources, and to perform analytics using intuitive browser-based tools.
  - BigInsights Data Scientist provides advanced analytics capabilities such as BigR for statistical analysis and modeling.

All these packages are available as optional preloaded software, giving you the flexibility to obtain just the required functions, with the ability to add additional capabilities at a later time. All analytics software must be separately ordered.

The cluster configurations and optional preloaded software give you the flexibility to obtain just the required functions, with the ability to add additional capabilities at a later time.

For more information on the IBM Power(R) System S812LC (8348-21C) server, refer to Hardware Announcement 115-125, dated October 5, 2015.

For more information on the IBM Open Platform with Apache Hadoop and IBM BigInsights, refer to Software Announcement 215-259, dated August 25, 2015.

For more information on IBM Data Engine for Analytics - Power Systems Edition, refer to Hardware Announcement 114-176, dated October 6, 2014.

For information on IBM Data Engine for Analytics Release 2, refer to Hardware Announcement 115-115, dated September 8, 2015.

Feature exchange

Not applicable.

Key prerequisites

All analytics software must be separately ordered.

Planned availability date

March 18, 2016

Description

Data Engine for Hadoop and Spark is a fully integrated infrastructure solution with integrated cluster management and analytics software that is optimized for Hadoop-based and Spark-based workloads. It is designed to deliver superior price/performance for these workloads while improving ease of deployment and cluster operational simplicity for clients deploying big data and analytics applications to support their line of business.

The solution is based on a set of standard building blocks that can be tailored to fit the data size, throughput, and scale required for the target analytics scenarios. Spark workloads benefit from large, fast memory and lots of processor threads. Hadoop workloads require large storage capacity, high-speed networks, and a resilient cluster file system.

The following default server configurations are available with the option to add memory or add additional SSD drives:

- Big Data Standard Data Node, ideal for most typical Hadoop workloads:
  - 1 x POWER8 2.92 GHz 10 Core + 128 GB (32 x 4 GB) DRAM
  - 12 x 6 TB (front) + 2 x 1 TB (rear) SATA HDD drives
- Big Data Memory Rich Data Node, ideal for Spark's memory-intensive workloads:
  - 1 x POWER8 2.92 GHz 10 Core + 256 GB (16 x 16 GB) DRAM
  - 10 x 6 TB SATA HDD + 2 x 960 GB SSD + 2 x 1 TB (rear) SATA HDD
- Big Data Management Node:
  - 1 x POWER8 2.92 GHz 10 Core + 128 GB (32 x 4 GB) DRAM
  - 2 x 1 TB (rear) SATA HDD
- System Management Node:
  - 1 x POWER8 3.32 GHz 8 Core + 32 GB (8 x 4 GB) DRAM
  - 2 x 1 TB (rear) SATA HDD

Five node starter configurations are available, handling up to 216 terabytes of raw data, providing over 50 terabytes of usable data in a standard, triple replica Hadoop or Spark configuration. Multi-rack configurations are available, providing up to 1.3 petabytes of raw data per rack.

High-speed switch configurations are available to deliver fast movement of data at scale required in Hadoop and Spark clusters:

- Management network switch: 1 GB Ethernet (48 x 1 GB and 4 x 10 GB ports)
- Data network switches:
  - 10 GB Ethernet (24 x 10 GB ports)
  - 10 GB Ethernet (48 x 10 GB and 4 x 40 GB ports)

Advanced cluster management software is a standard component of the solution. IBM Platform Cluster Manager provides a cluster administration interface for adding, monitoring, and maintaining cluster nodes.

Optional analytics software products that are available for system preinstall include version 4.1, or later, of IBM Open Platform for Apache Hadoop (and Spark), BigInsights Analyst module, and BigInsights Data Scientist module with Red Hat Enterprise Linux 7.2 for Power.

All analytics software must be separately ordered.

Each solution configuration will, by default, contain the recommended startup services. For more information, refer to the Services section.

Additional services are available to provide additional customization or to deploy initial data and run analytics use cases.

Product number

The following are newly announced features on the specific models of the IBM Power Systems 7014, 7120, and 8348 machine types.

New features available March 18, 2016

Description	Machine Type	Model	Feature number
One CSC Billing Unit	8348	21C	0010
Ten CSC Billing Units	8348	21C	0011
Rack Indicator, Rack #1	8348	21C	4651
Rack Indicator, Rack #2	8348	21C	4652
Rack Indicator, Rack #3	8348	21C	4653
Rack Indicator, Rack #4	8348	21C	4654
Rack Indicator, Rack #5	8348	21C	4655
Rack Indicator, Rack #6	8348	21C	4656
Rack Indicator, Rack #7	8348	21C	4657
Rack Indicator, Rack #8	8348	21C	4658
Rack Indicator, Rack #9	8348	21C	4659
Rack Indicator, Rack #10	8348	21C	4660
Rack Indicator, Rack #11	8348	21C	4661
Rack Indicator, Rack #12	8348	21C	4662
Rack Indicator, Rack #13	8348	21C	4663
Rack Indicator, Rack #14	8348	21C	4664
Rack Indicator, Rack #15	8348	21C	4665
Rack Indicator, Rack #16	8348	21C	4666
Power Cable - Drawer to IBM PDU, 200-240V/10A	8348	21C	6577
Variable Length, Blue Cat5e Cable	8348	21C	ECCG
Variable Length, Green Cat5e Cable	8348	21C	ECCH
Variable Length FIBRE SAN CABLE	8348	21C	ECCK
IBM BigInsights Data Scientist Indicator	8348	21C	EHJR
IBM BigInsights Analyst Indicator	8348	21C	EHJS
IBM Data Engine for Hadoop Solution Specify	7014	T42	EHLA
	7120	24E	
	7120	48E	
	7120	64C	
	8348	21C	
xCAT Server Specify	8348	21C	EHLB
Management Node Specify	8348	21C	EHLC
Hadoop Data Node Specify	8348	21C	EHLD
Spark Worker Node Specify	8348	21C	EHLE
IBM Open Platform with Apache Hadoop Indicator	8348	21C	EHLF
Rack Content Specify: 8348-21C 2EIA	7014	T42	ER1P

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).

BP Attachment for Announcement Letter 116-011

Publications

Refer to the publications for these products:

- IBM Power S812LC server (8348-21C)
- Networking switches (7120-24E, 7120-48E, and 7120-64C)
- Rack (7014-T42)

To access the IBM Publications Center Portal, go to the IBM Publications Center website.

The Publications Center is a worldwide central repository for IBM product publications and marketing material with a catalog of 70,000 items. Extensive search facilities are provided. A large number of publications are available online in various file formats, which can currently be downloaded.

Services

IBM Systems Lab Services

IBM Systems Lab Services offers a wide array of services available for your enterprise. It brings expertise on the latest technologies from the IBM development community and can help with your most difficult technical challenges.

IBM Systems Lab Services exists to help you successfully implement emerging technologies so as to accelerate your return on investment and improve your satisfaction with your IBM systems and solutions. Services examples include initial implementation, integration, migration, and skills transfer on IBM systems solution capabilities and recommended practices. IBM System Lab Services and Training is one of the for-fee services organizations of IBM's world-renowned IBM Systems Group development labs.

For details on available services, contact your IBM representative or visit the Lab Services website.

IBM Data Engine for Hadoop and Spark basic startup services

IBM Systems Lab Services offers IBM Data Engine for Hadoop and Spark basic startup services that will provide on-site assistance with installation, configuration, and implementation for software and hardware infrastructure on the customer's IBM Data Engine for Hadoop and Spark system.

- As with all IBM Systems Lab Services service requests, IBM Systems Lab Services Sales Specialists (Opportunity Managers) must confirm scope and requirements and create a Statement of Work for these services. Opportunity Managers can be contacted via email (ibmsls@us.ibm.com) or at the Lab Services contact website.
- A signed Statement of Work and IBM Agreement for Services, IBM Customer Agreement, or IBM International Customer Agreement for services is required before services are performed.
- Contact your IBM representative for availability in your country.

Global Technology Services

IBM services include business consulting, outsourcing, hosting services, applications, and other technology management.

These services help you learn about, plan, install, manage, or optimize your IT infrastructure to be an on-demand business. They can help you integrate your high-speed networks, storage systems, application servers, wireless protocols, and an array of platforms, middleware, and communications software for IBM and many non-IBM offerings. IBM is your one-stop shop for IT support needs.

For details on available services, contact your IBM representative or go to the http://www.ibm.com/services/ website.

For details on available IBM Business Continuity and Recovery Services, contact your IBM representative or go to the Resiliency Services website.

Details on education offerings related to specific products can be found on the IBM authorized training website.

Technical information

Specified operating environment

Software requirements

Refer to the feature description section of the Sales Manual for specific software requirements.

Planning information

Cable orders

Cables are provided as features of the system order.

Security, auditability, and control

The customer is responsible for evaluation, selection, and implementation of security features, administrative procedures, and appropriate controls in application systems and communications facilities.

IBM Electronic Services

IBM has transformed its delivery of hardware and software support services to help you achieve higher system availability. Electronic Services is a web-enabled solution that offers an exclusive, no-additional-charge enhancement to the service and support available for IBM servers. These services are designed to provide the opportunity for greater system availability with faster problem resolution and preemptive monitoring.

The Electronic Services news page is a single Internet entry point that replaces the multiple entry points traditionally used to access IBM Internet services and support. The news page enables you to gain easier access to IBM resources for assistance in resolving technical problems.

To learn how Electronic Services can work for you, go to the IBM Electronic Support website.

Terms and conditions

MES discount applicable

No

Warranty period

- Three-year warranty on hardware
- One-year base warranty for software

An IBM part or feature installed during the initial installation of an IBM machine is subject to a full warranty effective on the date of installation of the machine. An IBM part or feature that replaces a previously installed part or feature assumes the remainder of the warranty period for the replaced part or feature. An IBM part or feature added to a machine without replacing a previously installed part or feature is subject to a full warranty effective on its date of installation. Unless specified otherwise, the warranty period, type of warranty service, and service level of a part or feature are the same as those for the machine in which it is installed.

Customer setup

Yes

Optional IBM Technical Support Services for hardware installation are available for a fee.

Machine code

No license terms apply

Optional features warranty period

- Three-year warranty on hardware
- One-year base warranty for software

Prices

For additional information and current prices, contact your local IBM representative.

The following are newly announced features on the specific models of the IBM Power Systems 7014, 7120, and 8348 machine types.

Description Machine type 7014	Model number	Feature number	Purchase price	Minimum Initial/ Monthly MES/ Maint. Charge	Both/ Support	CSU	RP MES
Data Engine - Hadoop Solution							
	T42	EHLA			Initial	N/A	No
Rack Specify 8348-21C 2EIA							
	T42	ER1P			Initial	N/A	No

Description Machine type 7120	Model number	Feature number	Purchase price	Minimum Initial/ Monthly MES/ Maint. Charge	Both/ Support	CSU	RP MES
Data Engine - Hadoop Solution							
	24E	EHLA			Initial	N/A	No
	48E				Initial	N/A	No
	64C				Initial	N/A	No

Description Machine type 8348	Model number	Feature number	Purchase price	Minimum Initial/ Monthly MES/ Maint. Charge	Both/ Support	CSU	RP MES
One CSC Billing Unit							
	21C	0010			Both	Yes	No
Ten CSC Billing Units							
	21C	0011			Both	No	No
Rack Indicator, Rack 1							
	21C	4651			Initial	N/A	No
Rack Indicator, Rack 2							
	21C	4652			Initial	N/A	No
Rack Indicator, Rack 3							
	21C	4653			Initial	N/A	No
Rack Indicator, Rack 4							
	21C	4654			Initial	N/A	No
Rack Indicator, Rack 5							
	21C	4655			Initial	N/A	No
Rack Indicator, Rack 6							
	21C	4656			Initial	N/A	No
Rack Indicator, Rack 7							
	21C	4657			Initial	N/A	No
Rack Indicator, Rack 8							
	21C	4658			Initial	N/A	No
Rack Indicator, Rack 9							

```
                        21C     4659                    Initial  N/A No
Rack Indicator, Rack 10
                        21C     4660                    Initial  N/A No
Rack Indicator, Rack 11
                        21C     4661                    Initial  N/A No
Rack Indicator, Rack 12
                        21C     4662                    Initial  N/A No
Rack Indicator, Rack 13
                        21C     4663                    Initial  N/A No
Rack Indicator, Rack 14
                        21C     4664                    Initial  N/A No
Rack Indicator, Rack 15
                        21C     4665                    Initial  N/A No
Rack Indicator, Rack 16
                        21C     4666                    Initial  N/A No
Power Cable Drawer to IBM PD
                        21C     6577                    Both     Yes No
Var lth Blue Cat5e Cable
                        21C     ECCG                    Initial  N/A No
Var lth Green Cat5e Cable
                        21C     ECCH                    Initial  N/A No
Var Length FIBRE SAN CABLE
                        21C     ECCK                    Initial  N/A No
IBM BigInsights Dat Scient Ind
                        21C     EHJR                    Initial  N/A No
IBM BigInsights Analyst Ind
                        21C     EHJS                    Initial  N/A No
Data Engine - Hadoop Solution
                        21C     EHLA                    Initial  N/A No
xCAT Server Specify
                        21C     EHLB                    Initial  N/A No
Management Node Specify
                        21C     EHLC                    Initial  N/A No
Hadoop Data Node Specify
                        21C     EHLD                    Initial  N/A No
Spark Worker Node Specify
                        21C     EHLE                    Initial  N/A No
Open Platform / Apache Hadoop
                        21C     EHLF                    Initial  N/A No
```

Pricing terms

Prices in the following PDF prices link are suggested list prices on day of announcement for the U.S. only. They are provided for your information only. Dealer prices may vary, and prices may also vary by country. IBM list price does not include tax or shipping and is subject to change without notice.

ENUS-116-011-LIST_PRICES_2016_02_09.PDF

Trademarks

Power Systems is a trademark of IBM Corporation in the United States, other countries, or both.
IBM, POWER8, BigInsights, Power, PartnerWorld and Global Technology Services are registered trademarks of IBM Corporation in the United States, other countries, or both.
Linux is a registered trademark of Linus Torvalds 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

Terms of use

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/