

Datasheet

Fujitsu Interstage Big Data Parallel Processing Server V1.0

Availability and processing performance enhanced with an Apache Hadoop solution featuring Fujitsu's proprietary distributed file system



Fujitsu Big Data Software

An organizations ability to collect and analyze large volumes of data is fundamental to their capability to predict trends, forecast sales and make crucial business decisions.

In the era of Big Data, the amount of data is increasing exponentially and is becoming more and more diverse. The Collect, Analyze Decide and Act cycle, has become too difficult to process through conventional IT systems and has resulted in the emergence of Big Data technology.

Big Data technology can shorten the, afore mentioned, cycle by performing the analysis in real time, thus making it possible for organizations to engage new opportunities not previously conceived. It is critical that companies can utilize Big Data to better understand their customers and stay ahead of the competition.

Until now, Big Data technology has not been well leveraged. Fujitsu's new Interstage Big Data Software offers reductions in analysis times and an ability to combine Big Data with conventional business data. With Interstage, organizations can benefit from greater efficiencies in enterprise processing and create new business opportunities.

Interstage Big Data Parallel Processing Server

Interstage Big Data Parallel Processing Server is a parallel distributed processing software platform that supports improved data reliability. It combines the Apache Hadoop ^(*) with Fujitsu's Distributed File System.

By combining Fujitsu's Distributed File System with Apache Hadoop, the new solution improves data integrity without requiring data to be transferred to Hadoop processing servers, thereby achieving substantially better processing performance.

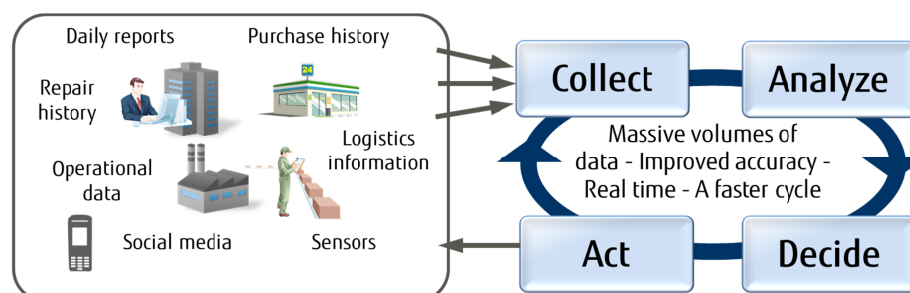
Fujitsu's strong track record in mission-critical enterprise systems supports this technology.

By the Easy setup technology, Organizations enables to reduce time for initial software setup and simplifies operations for scale-out, eliminates setup mistakes.

Usage Scenario

Retail - Previously the data of an individual store was not able to be analyzed and therefore product range decisions were made at the discretion of the store managers.

Big Data technology has enabled the application of batch processing to individual stores. It is now possible to accurately stock the best-selling products leading to higher sales and profits.



*1: Apache Hadoop: Developed and released by the Apache Software Foundation (ASF).

Apache Hadoop is open source software for efficiently performing parallel distributed processing of massive volumes of data.

Features and Benefits

Main features	Benefits
<p>High Performance Distributed File System</p> <ul style="list-style-type: none"> ■ Fujitsu's Distributed File System for Hadoop, which has been proven in mission-critical enterprise systems ■ Data sharing with existing systems eliminates the need to transfer files ■ Easy data access via the standard Linux file interface <p>High Availability</p> <ul style="list-style-type: none"> ■ Built-in high availability management features to realize a secure Hadoop environment <ul style="list-style-type: none"> ➢ Master server automatic fail-over for Hadoop Task Tracker ➢ High availability Distributed File System ➢ Redundant network between the master and slave servers ■ Distributed File System on top of the high-availability storage system. <p>Easy Setup and Scale-out</p> <ul style="list-style-type: none"> ■ Reduces time for initial installation and setup of system. ■ Simplifies operations for scale-out, eliminates setup mistakes 	<ul style="list-style-type: none"> ■ Reduces the time required for Big Data analysis ■ Reduces the time required for enterprise batch processing ■ Simplifies access to Big Data ■ Enhances productivity through flexible and efficient data sharing between applications and Hadoop <ul style="list-style-type: none"> ■ Enables a higher availability Hadoop environment, eliminating the single point of failure ■ Relieves users from the need to deal with the underlying high availability design and configuration ■ Ensures higher availability of data <ul style="list-style-type: none"> ■ Enables easy/automated configuration of high performance and high-availability Hadoop environment ■ Reduces troubles due to setup mistakes ■ Enables easier and quicker scale-out of system to handle more data and/or to reduce data processing time

Topics

High Performance Distributed File System for handling Big Data

In recent years, the amount of data produced in our daily lives continues to grow at an explosive pace. This is due to the spread of technologies such as smartphones, tablet devices, and various kinds of sensors. At the same time, there is also significant demand to use Big Data in corporate business activities—for instance, to discover trends and forecasts from various logs and data sources, and then to leverage such information in new business areas or to improve operations.

Apache Hadoop is considered the industry standard open source software for achieving distributed processing of large volumes of unstructured data. Interstage Big Data Parallel Processing Server combines this in a systematized platform with Fujitsu's enhanced data/operation management and high availability technologies. This enables Big Data and mission-critical systems to be harnessed for use in enterprise information systems.

Interstage Big Data Parallel Processing Server V1.0 enables enhanced availability and processing performance while shortening deployment times. This makes use of Big Data in enterprise systems more feasible.

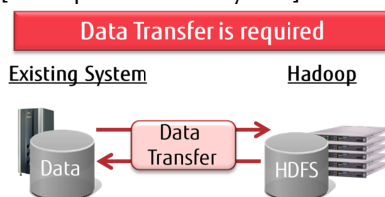
Fujitsu's Distributed File System

In addition to the standard Hadoop Distributed File System (HDFS), this new solution features Fujitsu's proprietary distributed file system (DFS), which boasts a strong track record in mission-critical enterprise systems, enabling the Hadoop system to achieve high performance and high availability.

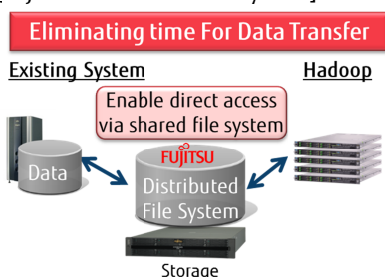
With DFS, data processing can be performed using Hadoop by directly accessing data stored within the storage system. Unlike the standard Apache Hadoop format, which temporarily transfers the required data to HDFS before processing, Fujitsu's software does not require data to be transferred, thereby significantly reducing processing time.

A Fujitsu internal evaluation showed a reduction in data processing time to 1/5 that of HDFS ^{(*)1}. Additionally, efficient use of slave server memory cache enabled 10 times faster I/O performance.

[Hadoop standard file system]



[Fujitsu's distributed file system]



DFS allows applications to access files using a standard Linux interface. Users can employ their existing tools to provide and retrieve data to the Hadoop system, including backup and printing without modification. Thus, the entire system configuration becomes simpler, reducing the time and costs required to tailor applications for Hadoop.

Apache Hadoop's single point of failure can be eliminated through redundant operations. A master server that employs Fujitsu clustering technology enables high availability. Availability is also improved by storing data in the storage system.

Easy Setup and Scale-out

Fujitsu's Easy Setup reduces installation and deployment time. Installation and setup of Apache Hadoop and the underlying Fujitsu software components (e.g., clustering and DFS) are automated. By eliminating numerous error-prone manual operations, a fast, secure setup of a high performance and high availability Hadoop environment is realized.

Sometime after initial configuration, the system may require more computing power as the size of the data set increases. Fujitsu Easy Setup enables additional Hadoop slave servers to be added easily, allowing pre-constructed system images to be deployed on multiple servers in a single pass for quick system and server deployment with just a few operations. Fujitsu DFS also allows for dynamic disk addition to scale-up the file system.

*1: Measured by Fujitsu

Technical Details

Server Hardware	PRIMERGY RX ^(*1) , ^(*2) , TX ^(*2) Two PRIMERGY RX are required for master server clustering
Operating Systems	Red Hat(R) Enterprise Linux(R) 5.6 (for Intel64) Red Hat(R) Enterprise Linux(R) 5.7 (for Intel64) Red Hat(R) Enterprise Linux(R) 5.8 (for Intel64) Red Hat(R) Enterprise Linux(R) 6 (for Intel64) Red Hat(R) Enterprise Linux(R) 6.1 (for Intel64) Red Hat(R) Enterprise Linux(R) 6.2 (for Intel64)
Storage Hardware	ETERNUS iSCSI connection with 10 Gbit Ethernet is required ^(*3)
Other Software Prerequisites	ServerView Agent for Linux V4.50.12 or later IPMI driver ETERNUS Multi-path driver Interstage Application Server V10
User Interface	English

*1: PRIMERGY RX900 is not supported for this product

*2: One or more (up to four) dual core processors required

*3: FC connection is not supported

More Information

Fujitsu Platform Solutions

In addition to Fujitsu Interstage Big Data Parallel Processing Server V1, Fujitsu provides a wide range of platform solutions. These solutions combine reliable Fujitsu products with the best in services, know-how and worldwide partnerships.

Dynamic Infrastructures

With the Fujitsu Dynamic Infrastructures approach, Fujitsu offers a full portfolio of IT products, solutions and services, ranging from clients to datacenter solutions, Managed Infrastructure to Infrastructure as a Service. How much you benefit from Fujitsu technologies and services depends on the level of cooperation you choose. This takes IT flexibility and efficiency to the next level.

Computing products

www.fujitsu.com/global/services/computing/

- PRIMERGY: Industrial standard server
- SPARC Enterprise: UNIX server
- PRIMEQUEST: Mission-critical IA server
- ETERNUS: Storage system

Software

www.fujitsu.com/software/

- Interstage: Application infrastructure software
- Systemwalker: System management software

More information

To learn more about Fujitsu Interstage Big Data Parallel Processing Server V1, please contact your Fujitsu sales representative, Fujitsu business partner, or visit the Fujitsu website.
www.fujitsu.com/software

Fujitsu Green Policy Innovation

Fujitsu Green Policy Innovation is our worldwide project to reduce the burdens placed on the environment. Using our global know-how, we aim to resolve environmental energy efficiency issues through the use of IT. More information is available at:
www.fujitsu.com/global/about/environment/



**Green
Policy
Innovation**

Copyright

© Copyright 2012 FUJITSU Limited.
Fujitsu, the Fujitsu logo, and Interstage are trademarks or registered trademarks of Fujitsu Limited in Japan and other countries.
Linux is a registered trademark of Linus Torvalds in the United States and other countries.
Apache Hadoop, Hadoop, HDFS, HBase, Hive, and Pig are registered trademarks of the Apache Software Foundation in the United States and other countries.
Other company product and service names may be trademarks or registered trademarks of their respective owners.

Disclaimer

Technical data is subject to modification and delivery is subject to availability. Any liability that the data and illustrations contained herein are complete, actual or correct is hereby disclaimed. Designations used herein may be trademarks and/or copyrights of their respective manufacturers, the use of which by third parties for their own purposes may infringe the rights of such owners.

Contact Information

FUJITSU LIMITED
Address: Shiodome City Center,
1-5-2 Higashi-Shimbashi, Minato-Ku, Tokyo
105-7123, Japan
Website: www.fujitsu.com
2012-10-26 WW EN