

With the explosive growth of information volume, data storage, analytics, and mining have become especially vital in every industry. A traditional big data platform that couples storage and compute nodes has bottlenecks, such as resource waste and data silos. Consequently, enterprises urgently need a decoupled architecture for big data platforms. Huawei FusionStorage HDFS uses a fully distributed architecture to integrate hardware, such as HDDs and SSDs, into a large storage pool. Storage nodes are decoupled from compute nodes for on-demand configuration and flexible expansion, eliminating stovepipe deployment and data silos. This vastly reduces investment costs and builds a unified storage platform for data mobility.

## **Elastic, Efficient, and Extensive Scalability Builds a Unified Storage Pool**

FusionStorage HDFS adopts a fully symmetric distributed architecture to enable linear performance growth as more hardware nodes are added without complex resource plans. FusionStorage HDFS can be easily scaled out to thousands of nodes to provide EB-level storage capacity for future storage demands of cloud services. Native primary and standby HDFS NameNodes only support a maximum of 100 million files in a namespace. Conversely, each storage node of FusionStorage HDFS can be used as NameNodes, which allows a single namespace to support ten billions of files and a whole cluster to support trillions.

FusionStorage HDFS connects multiple namespaces with multiple compute clusters. Each compute cluster has an independent authentication system for verification with its corresponding namespace. What's more, there are multiple ways to capitalize on a storage pool's resources: logical data isolation among namespaces, flexible space allocation, and storage capability sharing.

## **Erasure Coding (EC) Compatibility with Native HDFS Semantics Seamlessly Migrates Services**

Unlike S3A object storage that does not support interfaces such as append, rename, flush (), or hflush (), FusionStorage HDFS is fully compatible with native HDFS semantics and allows access using HDFS interfaces without installing any independent plug-ins on compute servers. This allows for effortless service migration and full compatibility with Huawei and third-party big data platforms. FusionStorage HDFS even supports the 22+2 EC scheme with a utilization rate of 91.7%, reducing investment costs and significantly outperforming the native HDFS EC and three-copy mechanism. Additionally, it automatically adjusts the EC scheme for data reliability even in the event of a node fault.

## **Multi-Level Reliability for High System Availability**

FusionStorage HDFS enjoys a fully-redundant deployment for uninterrupted services even if up to four nodes fail simultaneously. It provides system data self-healing. If data is corrupted, multiple nodes automatically and concurrently reconstruct data at up to 2

TB/hour. FusionStorage HDFS supports non-disruptive expansion for storage node capacity and non-disruptive upgrades for storage software. Intelligent I/O flow control implements quick system updates without compromising service continuity. FusionStorage HDFS ensures reliable O&M by providing status reports and preprocessing functions, such as disk checks and scanning.

## Specifications

Item	Description
System Architecture	Fully symmetric distributed architecture
Scalability	3 to 4,096 nodes
Capacity	EB-level
Access Protocol	HDFS
Data Redundancy Protection	Erasur Coding: Supports N + M redundancy protection (M = 2, 3, or 4)
Multi-tenancy	Multi-tenancy and multi-namespace
Authentication Method	Interconnection with Kerberos, IAM, and LDAP
Maximum Files in a Namespace	10 billion
File Name Length in a Directory	8,000 bytes
Quota	Namespace- and directory-level; quotas for capacity and files
QoS	Namespace- and tenant-level; QoS can be set for read bandwidth, write bandwidth, total bandwidth, IOPS, and the number of connections.
Recycle Bin	Data in the recycle bin can be deleted automatically (scheduled) or manually.
Small File Aggregation	The EC utilization rate of small files can exceed 80%.
Live-Network Storage Coexistence	Supports ViewFs HBase enables built-in metadata gateways and HDFS storage on the live network to provide unified namespaces.
Data Self-Healing	Automatic concurrent reconstruction reaching 2 TB/hour
Storage Device	TaiShan 2280 V2/5280 V2 servers in typical configuration
Storage Media	HDDs, SSDs
Network Type	10GE, 25GE

# Huawei FusionStorage HDFS Distributed Big Data Storage



## For More Information

To learn more about Huawei storage, please contact your local Huawei office or visit Huawei Enterprise website: <http://e.huawei.com>.



Huawei Enterprise APP

Huawei IT

Copyright © Huawei Technologies Co., Ltd. 2019. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without the prior written consent of Huawei Technologies Co., Ltd.



## Trademarks and Permissions



, HUAWEI, and are trademarks or registered trademarks of Huawei Technologies Co., Ltd.

Other trademarks, product, service and company names mentioned are the property of their respective holders.

## NO WARRANTY

THE CONTENTS OF THIS MANUAL ARE PROVIDED "AS IS". EXCEPT AS REQUIRED BY APPLICABLE LAWS, NO WARRANTIES OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, ARE MADE IN RELATION TO THE ACCURACY, RELIABILITY OR CONTENTS OF THIS MANUAL.

TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, IN NO CASE SHALL HUAWEI TECHNOLOGIES CO., LTD BE LIABLE FOR ANY SPECIAL, INCIDENTAL, INDIRECT, OR CONSEQUENTIAL DAMAGES, OR LOST PROFITS, BUSINESS, REVENUE, DATA, GOODWILL OR ANTICIPATED SAVINGS ARISING OUT OF, OR IN CONNECTION WITH, THE USE OF THIS MANUAL.

HUAWEI TECHNOLOGIES CO.,LTD.

Bantian Longgang District  
Shenzhen 518129, P.R.China  
Tel: +86-755-28780808

[www.huawei.com](http://www.huawei.com)