



Data fabric powered by NetApp for big data architecture

NetApp Solutions

NetApp
October 20, 2023

This PDF was generated from <https://docs.netapp.com/us-en/netapp-solutions/data-analytics/hdcs-sh-data-fabric-powered-by-netapp-for-big-data-architecture.html> on October 20, 2023. Always check docs.netapp.com for the latest.

Table of Contents

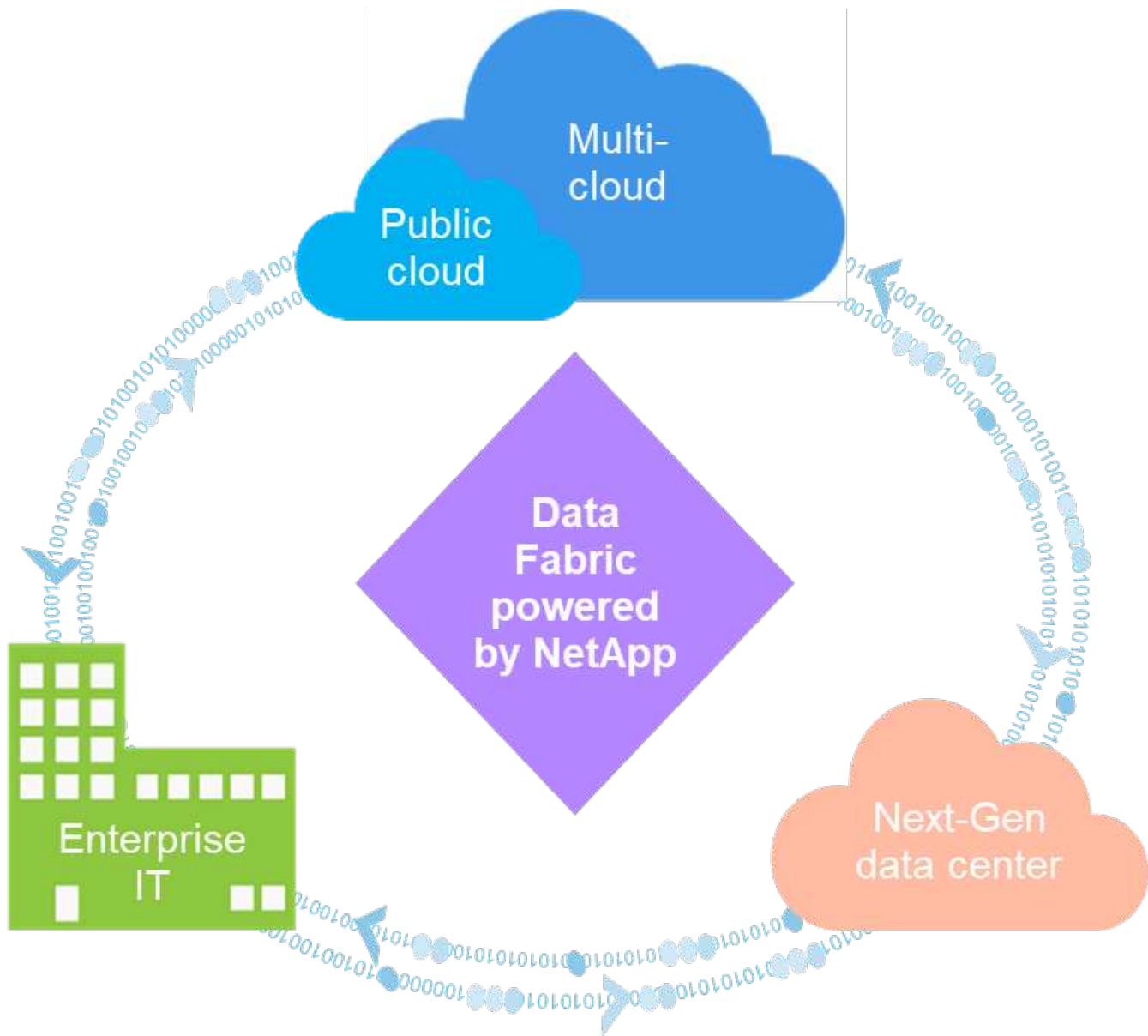
Data fabric powered by NetApp for big data architecture	1
Proven data fabric customer use cases	1

Data fabric powered by NetApp for big data architecture

[Previous: Solution overview.](#)

The data fabric powered by NetApp simplifies and integrates data management across cloud and on-premises environments to accelerate digital transformation.

The data fabric powered by NetApp delivers consistent and integrated data management services and applications (building blocks) for data visibility and insights, data access and control, and data protection and security, as shown in the figure below.



Proven data fabric customer use cases

The data fabric powered by NetApp provides the following nine proven use cases for customers:

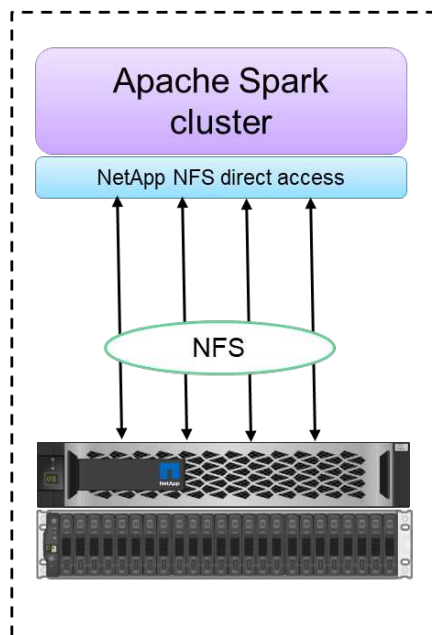
- Accelerate analytics workloads
- Accelerate DevOps transformation
- Build cloud hosting infrastructure
- Integrate cloud data services
- Protect and secure data
- Optimize unstructured data
- Gain data center efficiencies
- Deliver data insights and control
- Simplify and automate

This document covers two of the nine use cases (along with their solutions):

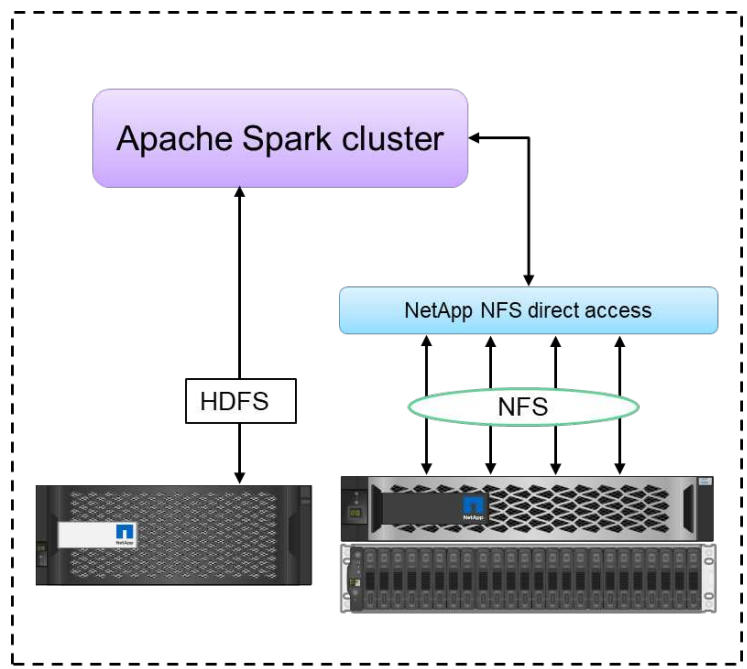
- Accelerate analytics workloads
- Protect and secure data

NetApp NFS direct access

The NetApp NFS direct access (formerly known as NetApp In-Place Analytics Module) (shown in the figure below) allows customers to run big data analytics jobs on their existing or new NFSv3 or NFSv4 data without moving or copying the data. It prevents multiple copies of data and eliminates the need to sync the data with a source. For example, in the financial sector, the movement of data from one place to another place must meet legal obligations, which is not an easy task. In this scenario, the NetApp NFS direct access analyzes the financial data from its original location. Another key benefit is that using the NetApp NFS direct access simplifies protecting Hadoop data by using native Hadoop commands and enables data protection workflows leveraging NetApp's rich data management portfolio.



Configuration 1: NFS as primary storage



Configuration 2: HDFS and NFS in single Spark cluster

The NetApp NFS direct access provides two kinds of deployment options for Hadoop/Spark clusters:

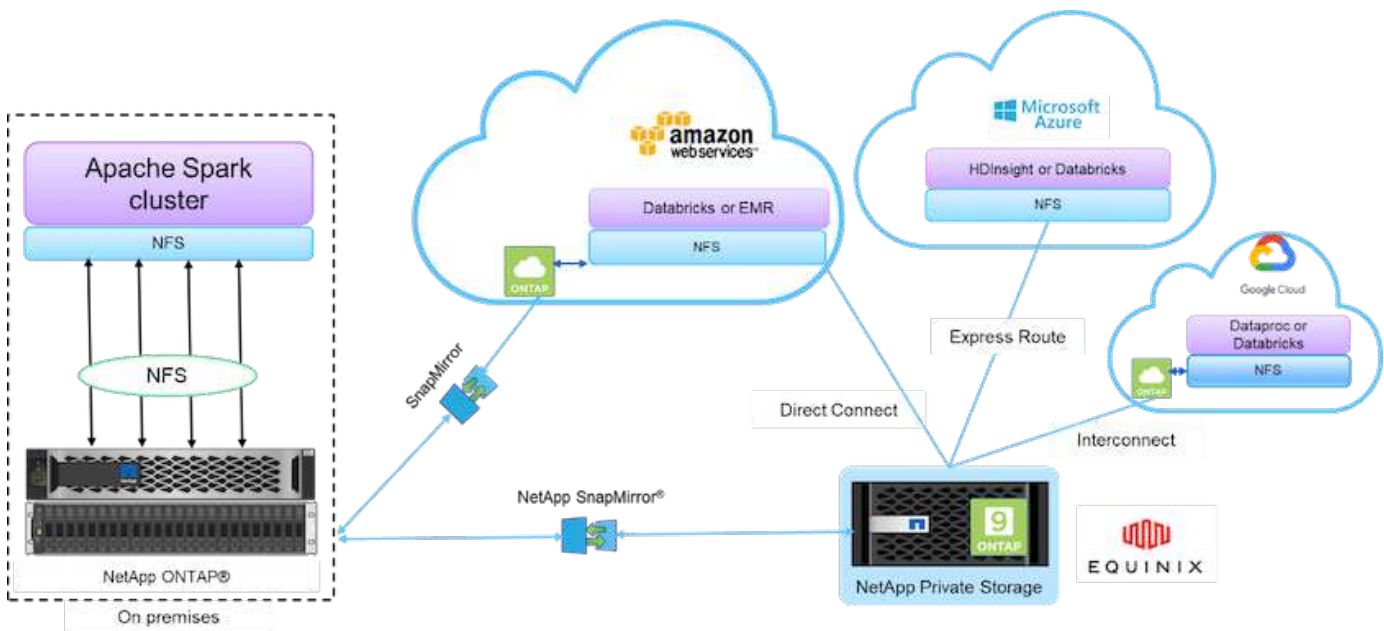
- By default, the Hadoop/Spark clusters use Hadoop Distributed File System (HDFS) for data storage and the default file system. The NetApp NFS direct access can replace the default HDFS with NFS storage as the default file system, enabling direct analytics operations on NFS data.
- In another deployment option, the NetApp NFS direct access supports configuring NFS as additional storage along with HDFS in a single Hadoop/Spark cluster. In this case, the customer can share data through NFS exports and access it from the same cluster along with HDFS data.

The key benefits of using the NetApp NFS direct access include:

- Analyzes the data from its current location, which prevents the time- and performance-consuming task of moving analytics data to a Hadoop infrastructure such as HDFS.
- Reduces the number of replicas from three to one.
- Enables users to decouple the compute and storage to scale them independently.
- Provides enterprise data protection by leveraging the rich data management capabilities of ONTAP.
- Is certified with the Hortonworks data platform.
- Enables hybrid data analytics deployments.
- Reduces the backup time by leveraging dynamic multithread capability.

Building blocks for big data

The data fabric powered by NetApp integrates data management services and applications (building blocks) for data access, control, protection, and security, as shown in the figure below.



The building blocks in the figure above include:

- **NetApp NFS direct access.** Provides the latest Hadoop and Spark clusters with direct access to NetApp NFS volumes without additional software or driver requirements.
- **NetApp Cloud Volumes ONTAP and Cloud Volume Services.** Software-defined connected storage based on ONTAP running in Amazon Web Services (AWS) or Azure NetApp Files (ANF) in Microsoft Azure cloud services.
- **NetApp SnapMirror technology.** Provides data protection capabilities between on-premises and ONTAP

Cloud or NPS instances.

- **Cloud service providers.** These providers include AWS, Microsoft Azure, Google Cloud, and IBM Cloud.
- **PaaS.** Cloud-based analytics services such as Amazon Elastic MapReduce (EMR) and Databricks in AWS as well as Microsoft Azure HDInsight and Azure Databricks.

Next: [Hadoop data protection and NetApp](#).

Copyright information

Copyright © 2023 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.