



HDFS and MapR-FS to ONTAP NFS

NetApp Solutions

NetApp
October 20, 2023

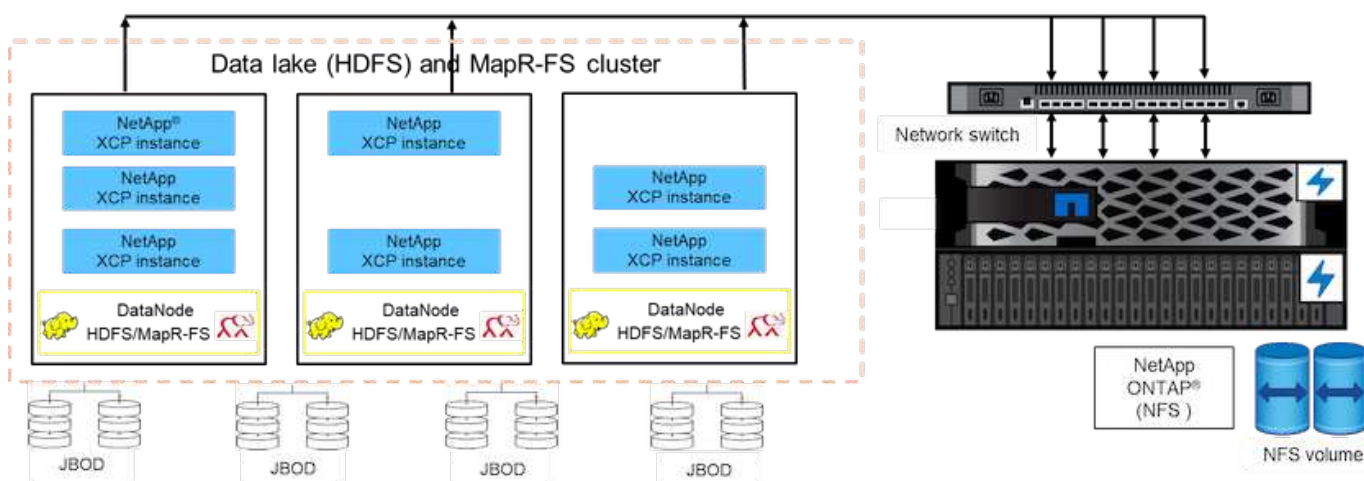
Table of Contents

HDFS and MapR-FS to ONTAP NFS.....	1
Why are customers moving from HDFS and MapR-FS to NFS?.....	1

HDFS and MapR-FS to ONTAP NFS

[Previous: GPFS to NetApp ONTAP NFS.](#)

For this solution, NetApp validated the migration of data from data lake (HDFS) and MapR cluster data to ONTAP NFS. The data resided in MapR-FS and HDFS. NetApp XCP introduced a new feature that directly migrates the data from a distributed file system such as HDFS and MapR-FS to ONTAP NFS. XCP uses async threads and HDFS C API calls to communicate and transfer data from MapR-FS as well as HDFS. The below figure shows the data migration from data lake (HDFS) and MapR-FS to ONTAP NFS. With this new feature, you don't have to export the source as an NFS share.



Why are customers moving from HDFS and MapR-FS to NFS?

Most of the Hadoop distributions such as Cloudera and Hortonworks use HDFS and MapR distributions use their own filesystem called MapR-FS to store data. HDFS and MapR-FS data provides the valuable insights to data scientists that can be leveraged in machine learning (ML) and deep learning (DL). The data in HDFS and MapR-FS is not shared, which means it cannot be used by other applications. Customers are looking for shared data, specifically in the banking sector where customers' sensitive data is used by multiple applications. The latest version of Hadoop (3.x or later) supports NFS data source, which can be accessed without additional third-party software. With the new NetApp XCP feature, data can be moved directly from HDFS and MapR-FS to NetApp NFS in order to provide access to multiple applications.

Testing was done in Amazon Web Services (AWS) to transfer the data from MapR-FS to NFS for the initial performance test with 12 MAPR nodes and 4 NFS servers.

	Quantity	Size	vCPU	Memory	Storage	Network
NFS server	4	i3en.24xlarge	96	488GiB	8x 7500 NVMe SSD	100
MapR nodes	12	I3en.12xlarge	48	384GiB	4x 7500 NVMe SSD	50

Based on initial testing, we obtained 20GBps throughput and were able to transfer 2PB per day of data.

For more information about HDFS data migration without exporting HDFS to NFS, see the “Deployment steps - NAS” section in [TR-4863: TR-4863: Best-Practice Guidelines for NetApp XCP - Data Mover, File Migration, and Analytics](#).

Next: [Business benefits](#).

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.