



## **Use case 3 - Enabling DevTest on existing Hadoop data**

NetApp Solutions

NetApp  
October 20, 2023

This PDF was generated from <https://docs.netapp.com/us-en/netapp-solutions/data-analytics/hdcs-sh-use-case-3--enabling-devtest-on-existing-hadoop-data.html> on October 20, 2023. Always check [docs.netapp.com](https://docs.netapp.com) for the latest.

# Table of Contents

- Use case 3: Enabling DevTest on existing Hadoop data . . . . . 1
  - Scenario . . . . . 1
  - Requirements and challenges . . . . . 1
  - Solution . . . . . 1

# Use case 3: Enabling DevTest on existing Hadoop data

[Previous: Use case 2 - Backup and disaster recovery from the cloud to on-premises.](#)

In this use case, the customer's requirement is to rapidly and efficiently build new Hadoop/Spark clusters based on an existing Hadoop cluster containing a large amount of analytics data for DevTest and reporting purposes in the same data center as well as remote locations.

## Scenario

In this scenario, multiple Spark/Hadoop clusters are built from a large Hadoop data lake implementation on-premises as well as at disaster recovery locations.

## Requirements and challenges

The main requirements and challenges for this use case include:

- Create multiple Hadoop clusters for DevTest, QA, or any other purpose that requires access to the same production data. The challenge here is to clone a very large Hadoop cluster multiple times instantaneously and in a very space-efficient manner.
- Sync the Hadoop data to DevTest and reporting teams for operational efficiency.
- Distribute the Hadoop data by using the same credentials across production and new clusters.
- Use scheduled policies to efficiently create QA clusters without affecting the production cluster.

## Solution

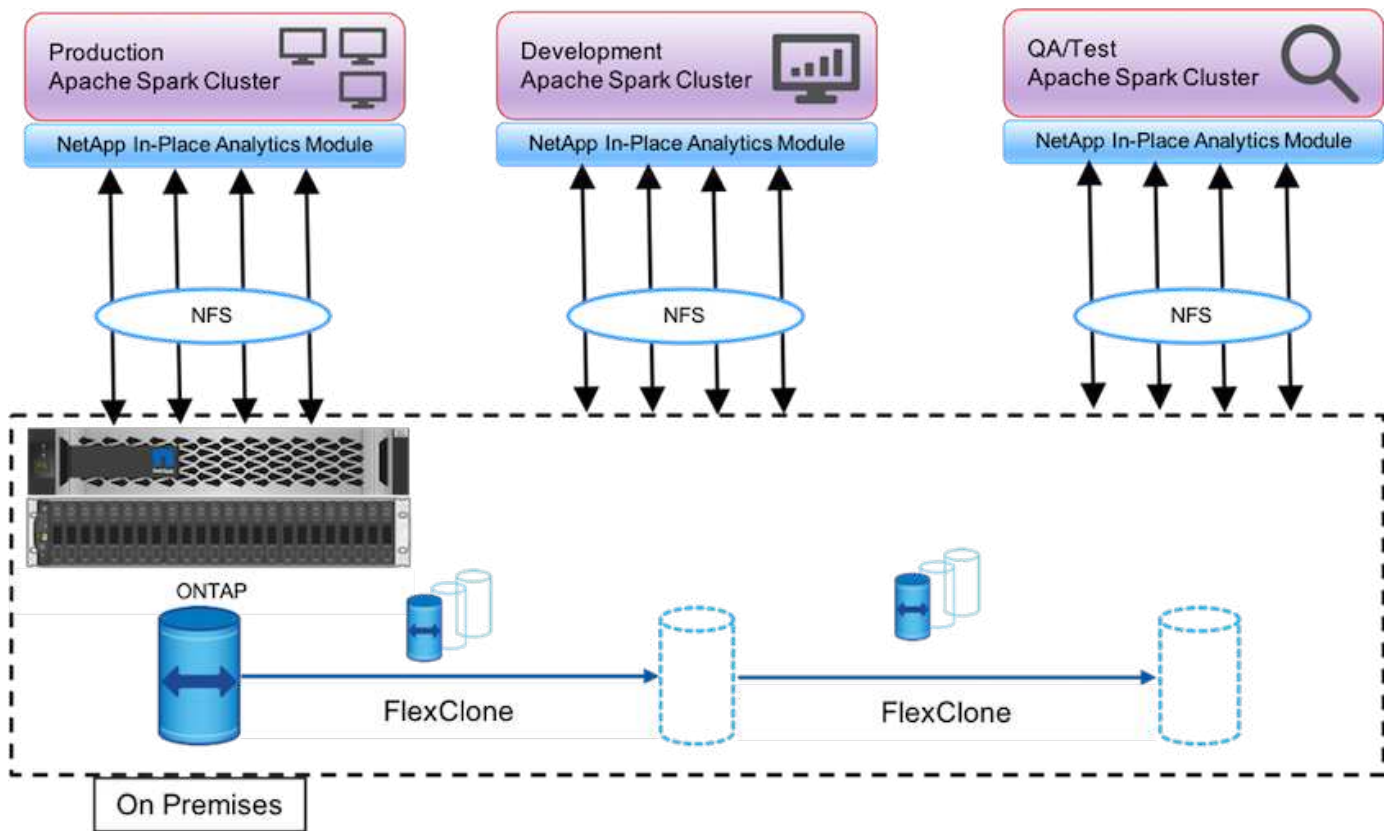
FlexClone technology is used to answer the requirements just described. FlexClone technology is the read/write copy of a Snapshot copy. It reads the data from parent Snapshot copy data and only consumes additional space for new/modified blocks. It is fast and space-efficient.

First, a Snapshot copy of the existing cluster was created by using a NetApp consistency group.

Snapshot copies within NetApp System Manager or the storage admin prompt. The consistency group Snapshot copies are application-consistent group Snapshot copies, and the FlexClone volume is created based on consistency group Snapshot copies. It is worth mentioning that a FlexClone volume inherits the parent volume's NFS export policy. After the Snapshot copy is created, a new Hadoop cluster must be installed for DevTest and reporting purposes, as shown in the figure below. The In-Place Analytics Module accesses the cloned NFS volume from the new Hadoop cluster through In-Place Analytics Module users and group authorization for the NFS data.

To have proper access, the new cluster must have the same UID and GUID for the users configured in the In-Place Analytics Module users and group configurations.

This image shows the Hadoop cluster for DevTest.



Next: Use case 4 - Data protection and multicloud connectivity.

## 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.