



# **Functional validation - Silly rename fix**

## **NetApp Solutions**

NetApp  
October 20, 2023

This PDF was generated from <https://docs.netapp.com/us-en/netapp-solutions/data-analytics/kafka-nfs-functional-validation-silly-rename-fix.html> on October 20, 2023. Always check docs.netapp.com for the latest.

# Table of Contents

- Functional validation - Silly rename fix . . . . . 1
  - Validation setup . . . . . 1
  - Architectural flow . . . . . 2
  - Methodology of testing . . . . . 2

# Functional validation - Silly rename fix

[Previous: NetApp solution for silly rename issue in NFS to Kafka workload.](#)

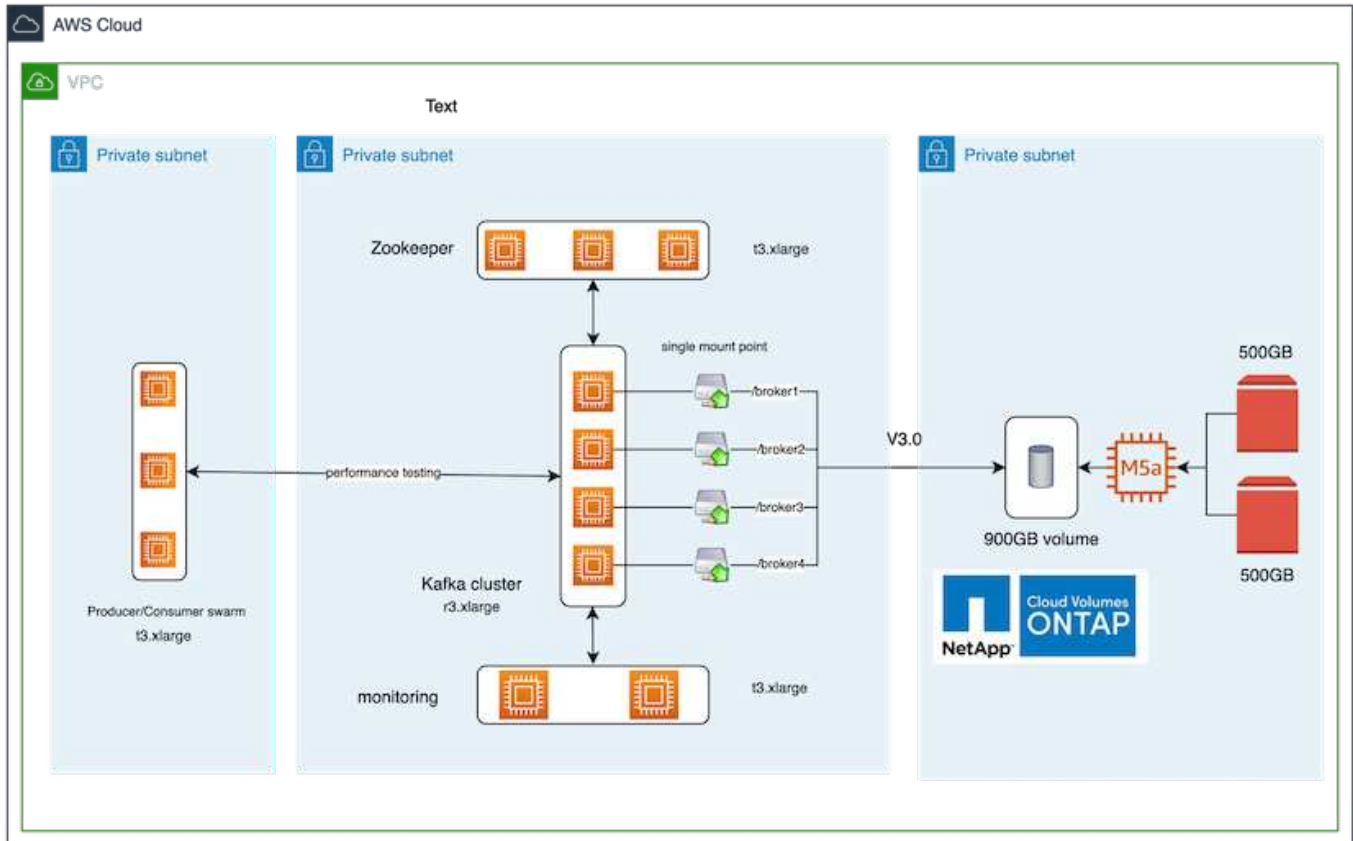
For the functional validation, we showed that a Kafka cluster with an NFSv3 mount for storage fails to perform Kafka operations like partition redistribution, whereas another cluster mounted on NFSv4 with the fix can perform the same operations without any disruptions.

## Validation setup

The setup is run on AWS. The following table shows the different platform components and environmental configuration used for the validation.

Platform component	Environment configuration
Confluent Platform version 7.2.1	<ul style="list-style-type: none"><li>• 3 x zookeepers – t3.xlarge</li><li>• 4 x broker servers – r3.xlarge</li><li>• 1 x Grafana – t3.xlarge</li><li>• 1 x control center – t3.xlarge</li><li>• 3 x Producer/consumer</li></ul>
Operating system on all nodes	RHEL8.7or later
NetApp Cloud Volumes ONTAP instance	Single-node instance – M5.2xLarge

The following figure show the architectural configuration for this solution.



## Architectural flow

- **Compute.** We used a four-node Kafka cluster with a three-node zookeeper ensemble running on dedicated servers.
- **Monitoring.** We used two nodes for a Prometheus-Grafana combination.
- **Workload.** For generating workloads, we used a separate three-node cluster that can produce to and consume from this Kafka cluster.
- **Storage.** We used a single-node NetApp Cloud volumes ONTAP instance with two 500GB GP2 AWS-EBS volumes attached to the instance. These volumes were then exposed to the Kafka cluster as single NFSv4.1 volume through a LIF.

The default properties of Kafka were chosen for all servers. The same was done for the zookeeper swarm.

## Methodology of testing

1. Update `-is-preserve-unlink-enabled true` to the kafka volume, as follows:

```
aws-shantanclastreca1-aws::*> volume create -vserver kafka_svm -volume
kafka_fg_vol01 -aggregate kafka_aggr -size 3500GB -state online -policy
kafka_policy -security-style unix -unix-permissions 0777 -junction-path
/kafka_fg_vol01 -type RW -is-preserve-unlink-enabled true
[Job 32] Job succeeded: Successful
```

2. Two similar Kafka clusters were created with the following difference:
  - **Cluster 1.** The backend NFS v4.1 server running production-ready ONTAP version 9.12.1 was hosted by a NetApp CVO instance. RHEL 8.7/RHEL 9.1 were installed on the brokers.
  - **Cluster 2.** The backend NFS server was a manually created generic Linux NFSv3 server.
3. A demo topic was created on both the Kafka clusters.

Cluster 1:

```
[root@ip-172-30-0-160 demo]# kafka-topics --bootstrap-server=172.30.0.160:9092,172.30.0.172:9092,172.30.0.188:9092,172.30.0.123:9092 --describe --topic __a_demo_topic
Topic: __a_demo_topic   TopicId: 2ty29xfhQLq65HKsUQv-pg PartitionCount: 4      ReplicationFactor: 2   Configs:
min.insync.replicas=1,segment.bytes=1073741824
Topic: __a_demo_topic   Partition: 0      Leader: 4      Replicas: 4,1   Isr: 4,1      Offline:
Topic: __a_demo_topic   Partition: 1      Leader: 2      Replicas: 2,4   Isr: 2,4      Offline:
Topic: __a_demo_topic   Partition: 2      Leader: 3      Replicas: 3,2   Isr: 3,2      Offline:
Topic: __a_demo_topic   Partition: 3      Leader: 1      Replicas: 1,3   Isr: 1,3      Offline:
```

Cluster 2:

```
[root@ip-172-30-0-198 demo]# kafka-topics --bootstrap-server=172.30.0.198:9092,172.30.0.163:9092,172.30.0.221:9092,172.30.0.204:9092 --describe --topic __a_demo_topic
Topic: __a_demo_topic   TopicId: AwQpsZTQShyeMIhaquCG3Q PartitionCount: 4      ReplicationFactor: 2   Configs:
min.insync.replicas=1,segment.bytes=1073741824
Topic: __a_demo_topic   Partition: 0      Leader: 2      Replicas: 2,3   Isr: 2,3      Offline:
Topic: __a_demo_topic   Partition: 1      Leader: 3      Replicas: 3,1   Isr: 3,1      Offline:
Topic: __a_demo_topic   Partition: 2      Leader: 1      Replicas: 1,4   Isr: 1,4      Offline:
Topic: __a_demo_topic   Partition: 3      Leader: 4      Replicas: 4,2   Isr: 4,2      Offline:
```

4. Data was loaded into these newly created topics for both clusters. This was done using the producer-perf-test toolkit that comes in the default Kafka package:

```
./kafka-producer-perf-test.sh --topic __a_demo_topic --throughput -1
--num-records 3000000 --record-size 1024 --producer-props acks=all
bootstrap.servers=172.30.0.160:9092,172.30.0.172:9092,172.30.0.188:9092,
172.30.0.123:9092
```

5. A health check was performed for broker-1 for each of the clusters using telnet:

- telnet 172.30.0.160 9092
- telnet 172.30.0.198 9092

A successful health check for brokers on both clusters is shown in the next screenshot:

```
shantanu@shantanc-mac-0 ~ % telnet 172.30.0.160 9092
Trying 172.30.0.160...
Connected to 172.30.0.160.
Escape character is '^['.
^[

Connection closed by foreign host.
shantanu@shantanc-mac-0 ~ % telnet 172.30.0.198 9092
Trying 172.30.0.198...
Connected to 172.30.0.198.
Escape character is '^['.
^[
```

6. To trigger the failure condition that causes Kafka clusters using NFSv3 storage volumes to crash, we initiated the partition reassignment process on both clusters. Partition reassignment was performed using `kafka-reassign-partitions.sh`. The detailed process is as follows:
  - a. To reassign the partitions for a topic in a Kafka cluster, we generated the proposed reassignment config JSON (this was performed for both the clusters).

```
kafka-reassign-partitions --bootstrap
-server=172.30.0.160:9092,172.30.0.172:9092,172.30.0.188:9092,172.30.
0.123:9092 --broker-list "1,2,3,4" --topics-to-move-json-file
/tmp/topics.json --generate
```

- b. The generated reassignment JSON was then saved in `/tmp/reassignment- file.json`.
  - c. The actual partition reassignment process was triggered by the following command:

```
kafka-reassign-partitions --bootstrap
-server=172.30.0.198:9092,172.30.0.163:9092,172.30.0.221:9092,172.30.
0.204:9092 --reassignment-json-file /tmp/reassignment-file.json
--execute
```

7. After a few minutes when the reassignment was completed, another health check on the brokers showed that cluster using NFSv3 storage volumes had run into a silly rename issue and had crashed, whereas Cluster 1 using NetApp ONTAP NFSv4.1 storage volumes with the fix continued operations without any disruptions.

```
shantanu@shantanc-mac-0 ~ % telnet 172.30.0.160 9092
Trying 172.30.0.160...
Connected to 172.30.0.160.
Escape character is '^['.
^[

Connection closed by foreign host.
shantanu@shantanc-mac-0 ~ % telnet 172.30.0.198 9092
Trying 172.30.0.198...
telnet: connect to address 172.30.0.198: Connection refused
telnet: Unable to connect to remote host
```

- Cluster1-Broker-1 is alive.
  - Cluster2-broker-1 is dead.
8. Upon checking the Kafka log directories, it was clear that Cluster 1 using NetApp ONTAP NFSv4.1 storage volumes with the fix had clean partition assignment, while Cluster 2 using generic NFSv3 storage did not due to silly rename issues, which led to the crash. The following picture shows partition rebalancing of Cluster 2, which resulted in a silly rename issue on NFSv3 storage.

```
/demo/broker_demo_1/___a_demo_topic-1.b31a8dd60fd443b283ffda2ecca9c2b9-delete:
total 40
drwxr-xr-x.  2 nobody nobody  4096 Sep 19 10:37 .
drwxr-xr-x. 246 nobody nobody 32768 Sep 19 10:36 ..
-rw-r--r--.  1 nobody nobody    5 Sep 19 10:22 .nfs0000000025f9008400000045
-rw-r--r--.  1 nobody nobody    0 Sep 19 10:25 .nfs0000000025f91d6800000048

/demo/broker_demo_1/___a_demo_topic-2:
total 832592
drwxr-xr-x.  2 nobody nobody    4096 Sep 19 10:26 .
drwxr-xr-x. 246 nobody nobody   32768 Sep 19 10:36 ..
-rw-r--r--.  1 nobody nobody    5 Sep 19 10:22 .nfs0000000025f91d5500000046
-rw-r--r--.  1 nobody nobody    0 Sep 19 10:25 .nfs0000000025f91fce00000047
-rw-r--r--.  1 nobody nobody 10485760 Sep 19 10:24 00000000000000000000000000000000.index
-rw-r--r--.  1 nobody nobody 848113134 Sep 19 10:24 00000000000000000000000000000000.log
-rw-r--r--.  1 nobody nobody 10485756 Sep 19 10:24 00000000000000000000000000000000.timeindex
-rw-r--r--.  1 nobody nobody    0 Sep 19 10:16 leader-epoch-checkpoint
-rw-r--r--.  1 nobody nobody    43 Sep 19 10:16 partition.metadata
```

The following picture shows a clean partition rebalancing of Cluster 1 using NetApp NFSv4.1 storage.

```
/demo/broker_demo_1/___a_demo_topic-0:
total 710932
drwxr-xr-x.  2 nobody nobody    4096 Sep 19 10:26 .
drwxr-xr-x.  85 nobody nobody    8192 Sep 19 10:37 ..
-rw-r--r--.  1 nobody nobody 10485760 Sep 19 10:25 00000000000000000000000000000000.index
-rw-r--r--.  1 nobody nobody 724167522 Sep 19 10:25 00000000000000000000000000000000.log
-rw-r--r--.  1 nobody nobody 10485756 Sep 19 10:25 00000000000000000000000000000000.timeindex
-rw-r--r--.  1 nobody nobody    0 Sep 19 10:15 leader-epoch-checkpoint
-rw-r--r--.  1 nobody nobody    43 Sep 19 10:15 partition.metadata

/demo/broker_demo_1/___a_demo_topic-2:
total 780016
drwxr-xr-x.  2 nobody nobody    4096 Sep 19 10:35 .
drwxr-xr-x.  85 nobody nobody    8192 Sep 19 10:37 ..
-rw-r--r--.  1 nobody nobody 10485760 Sep 19 10:36 00000000000000000000000000000000.index
-rw-r--r--.  1 nobody nobody 794575786 Sep 19 10:36 00000000000000000000000000000000.log
-rw-r--r--.  1 nobody nobody 10485756 Sep 19 10:36 00000000000000000000000000000000.timeindex
-rw-r--r--.  1 nobody nobody    0 Sep 19 10:35 leader-epoch-checkpoint
-rw-r--r--.  1 nobody nobody    43 Sep 19 10:35 partition.metadata
```

Next: Why NetApp NFS for Kafka workloads?

## 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.