



Duplicate files

NetApp Solutions

NetApp
September 23, 2021

Table of Contents

Duplicate files 1

Duplicate files

[Previous: Using the XCP Data Mover to migrate large files.](#)

NetApp received a request to find duplicate files from a single volume or multiple volumes. NetApp provided the following solution.

For single volume, run the following commands:

```
[root@mastr-51 linux]# ./xcp -md5 -match 'type==f and nlinks==1 and size
!= 0' 10.63.150.213:/common_volume/nfsconnector_hw_cert/ | sort | uniq -cd
--check-chars=32
XCP 1.5; (c) 2020 NetApp, Inc.; Licensed to Calin Salagean [NetApp Inc]
until Mon Dec 31 00:00:00 2029

176,380 scanned, 138,116 matched, 138,115 summed, 10 giants, 61.1 GiB in
(763 MiB/s), 172 MiB out (2.57 MiB/s), 1m5s

Filtered: 38264 did not match
176,380 scanned, 138,116 matched, 138,116 summed, 10 giants, 62.1 GiB in
(918 MiB/s), 174 MiB out (2.51 MiB/s), 1m9s.
    3 00004964ca155eca1a71d0949c82e37e
nfsconnector_hw_cert/grid_01082017_174316/0/hadoopqe/accumulo/shell/pom.xml
1
    2 000103fbed06d8071410c59047738389
nfsconnector_hw_cert/usr_hdp/2.5.3.0-37/hive2/doc/examples/files/dim-
data.txt
    2 000131053a46d67557d27bb678d5d4a1
nfsconnector_hw_cert/grid_01082017_174316/0/log/cluster/mahout_1/artifacts
/classifier/20news_reduceddata/20news-bydate-test/alt.atheism/53265
```

For multiple volumes, run the following commands:

```

[root@mastr-51 linux]# cat multiplevolume_duplicate.sh
#!/usr/bin/bash

#user input
JUNCTION_PATHS='/nc_volume1 /nc_volume2 /nc_volume3 /oplogarchivevolume'
NFS_DATA_LIF='10.63.150.213'

#xcp operation
for i in $JUNCTION_PATHS
do
echo "start - $i" >> /tmp/duplicate_results
/usr/src/xcp/linux/xcp -md5 -match 'type==f and nlinks==1 and size != 0'
${NFS_DATA_LIF}:$i | sort | uniq -cd --check-chars=32 | tee -a
/tmp/duplicate_results
echo "end - $i" >> /tmp/duplicate_results
done

[root@mastr-51 linux]# nohup bash +x multiplevolume_duplicate.sh &
[root@mastr-51 linux]# cat /tmp/duplicate_results

```

Next: [Specific date-based scan and copy of data.](#)

Copyright Information

Copyright © 2021 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.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

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.