<http://www.theselights.com/2010_03_01_archive.html>

**NetApp Reallocate**

On one of our filers we had a weird status on one of our aggregates and all of it's containing volumes. An appreciated opportunity to dive into reallocaton on NetApps filers. Because that's what the status was about:

Aggregate status: redirect

Meaning: Aggregate reallocation or file reallocation with the ”-p” option has been started on the aggregate, read performance will be degraded

Volumes statuses: redirect,active\_redirect

Volume redirect: The volume's containing aggregate is undergoing aggregate reallocation or file reallocation with the -p option. Read performance to volumes in the aggregate might be degraded.  
Volume active\_redirect: The volume's containing aggregate is undergoing reallocation (with the -p option specified). Read performance may be reduced while the volume is in this state.

The explaination is from the [NetApp Cheat Sheet](http://www.datadisk.co.uk/html_docs/netapp/netapp_cs.htm).

We have a few performance issues on this filer and are working on it from different angles, so I decided to not ignore this and look into it.

**What is Reallocation**

Reallocate optimizes the layout of data on disk for “Sequential Read Access”. There are three types of reallocation, volume/lun reallocation, aggregate reallocation and read reallocation. Having said that, the first one comes in two flavors, traditional and physical. When most engineers talk about reallocation they talk about volume/lun reallocation so in this article we'll do the same thing.

**Reallocation**

Reallocation should be run in the following way. First you measure the current layout which will present you with an optimization score between 1(optimized) and 10(very unoptimized). You set a threshold (4 by default), and if the optimization score is above the threshold you run the reallocation process.  
  
Reallocation has two flavors, traditional and physical. The [TR-3929](https://fieldportal.netapp.com/viewcontent.asp?qv=1&docid=33904) guide from NetApp provides an excellent description from these two flavors:

**Traditional Reallocation**

The reallocation process progresses through the file system and moves data blocks by rewriting them when Data ONTAP determines that the layout can be improved. If no improvement is predicted, no data is moved. NetApp Snapshot® data is not moved even when active file system data has been moved to new, optimized locations. Because data is rewritten to disk, if Snapshot copies are used, additional space is required to maintain the copies.

**Physical Reallocation**

The reallocate tool also provides a physical reallocation option. Physical reallocation follows the same process as traditional reallocation; however, instead of completely rewriting data to the disks, the data blocks are moved by changing the physical block location while maintaining the logical block location within the FlexVol® volume. The benefit of using physical reallocation is that no additional space for is required for Snapshot copies, compared to using traditional reallocation.

Note that the best practice is to always use the physical reallocation if possible.

**Aggregate Reallocation**

The [TR-3929](https://fieldportal.netapp.com/viewcontent.asp?qv=1&docid=33904) guide from NetApp also provides an excellent description for aggregate reallocation:

An additional option, -A, is available at the aggregate abstraction. This reallocation method reallocates blocks within an aggregate to improve contiguous free space. The –A option does not reallocate all of the data in the aggregate following the normal reallocation method. It should not be used to improve sequential read performance. Because aggregate reallocation uses the physical reallocation method to move blocks to create contiguous free space, the impacts of using physical reallocate still apply.

Note that the best practice for using the -A option is only when directed to do so by NetApp.

**Read Reallocate**

The [TR-3929](https://fieldportal.netapp.com/viewcontent.asp?qv=1&docid=33904) guide from NetApp also provides an excellent description for read reallocation:  
Read reallocate is a volume option that performs opportunistic reallocation on data to improve performance. Read reallocation uses the normal workload reads along with the read-ahead engine to determine the current layout optimization. If the read was less than optimal, the data will be reallocated to improve the next read of this data. Read reallocate offers both the traditional and physical reallocation methods associated with the reallocate command. Also, because read reallocate uses the existing read workload, it does not require additional scanning or scheduling.

**Reallocation Versus Aggregate Reallocation**

If you're not sure whether you understand the difference between (volume/lun) reallocation and aggregate reallocation I suggest you read this excellent article from Erick Moore. It uses pictures to understand it ;):

<http://www.theselights.com/2010/03/understanding-netapp-volume-and.html>

**The Filer in Place**

The aggregate

storage01\*> reallocate status -v aggr0

Reallocation scans are on

No reallocation status was found for 'aggr0'.

One of the volumes:

storage01> reallocate status -v /vol/OS\_Volume

Reallocation scans are on

No reallocation status was found for '/vol/OS\_Volume'.

All of the schedules:

storage01> reallocate status -v

Reallocation scans are on

/vol/vol0:

State: Idle

Flags: whole\_vol,measure\_only,repeat

Threshold: 4

Schedule: n/a

Interval: 1 day

Optimization: 8

Measure Log: n/a

As you can see this schedule only does a measurement of the volume.

When trying to start a measurement:

storage01\*> reallocate measure -o /vol/vol0

A scan already exists for '/vol/vol0', use 'reallocate status' to view current status.

storage01\*> reallocate measure -o /vol/DATAVolume0

Reallocation scan will be started on '/vol/DATAVolume0'.

Monitor the system log for results.

storage01\*> reallocate measure -o /vol/DATAVolume1

Reallocation scan will be started on '/vol/DATAVolume1'.

Monitor the system log for results.

To see what the difference might be I also started the measurement on a volume which is newly created and where all the volumes are moved to. Also, unaligned partitions were aligned:

storage01> reallocate measure -o /vol/ESXACC

Reallocation scan will be started on '/vol/ESXACC'.

Monitor the system log for results.

So in syslog:

Fri Mar 16 14:16:05 CET wafl.reallocate.scan.exists: A reallocation scan for /vol/vol0 already exists.

Fri Mar 16 14:16:32 CET wafl.scan.start: Starting WAFL layout measurement on volume DATAVolume0.

Fri Mar 16 14:38:20 CET wafl.scan.start: Starting WAFL layout measurement on volume DATAVolume1.

Fri Mar 16 15:47:42 CET wafl.scan.start: Starting WAFL layout measurement on volume ESXACC.

Fri Mar 16 15:12:01 CET wafl.scan.start: Starting WAFL layout measurement on volume vol0.

Fri Mar 16 15:13:44 CET wafl.reallocate.check.highAdvise: Allocation check on '/vol/vol0' is 8, hotspot 0 (threshold 4), consider running reallocate.

Fri Mar 16 15:56:13 CET wafl.reallocate.check.value: Allocation measurement check on '/vol/DATAVolume1' is 2.

Fri Mar 16 21:39:46 CET wafl.reallocate.check.value: Allocation measurement check on '/vol/ESXACC' is 2.

Sat Mar 17 04:30:14 CET wafl.reallocate.check.highAdvise: Allocation check on '/vol/DATAVolume0' is 9, hotspot 0 (threshold 4), consider running reallocate.

Note that the vol0 scan started automatically. It comes back every day.

**Stopping the Automatic Measure Only Reallocation Scan**

As you can see with the “reallocate status -v” command there is only one scheduled scan, it's for /vol/vol0 and it's a measure\_only scan:

storage01> reallocate status -v

Reallocation scans are on

/vol/vol0:

State: Idle

Flags: whole\_vol,measure\_only,repeat

Threshold: 4

Schedule: n/a

Interval: 1 day

Optimization: 8

Measure Log: n/a

Note that the scan does not have a schedule, it uses the (default) interval of one day. Note that countdown start after the scan is completed which explains why the scan starts every day a few minutes later. Here you can see that the scan indeed does not have a schedule, you can't delete it:

storage01> reallocate schedule -d /vol/vol0

Reallocation schedule for '/vol/vol0' will removed.

storage01> reallocate status -v

Reallocation scans are on

/vol/vol0:

State: Idle

Flags: whole\_vol,measure\_only,repeat

Threshold: 4

Schedule: n/a

Interval: 1 day

Optimization: 8

Measure Log: n/a

Deleting the schedule does not work since there is no schedule, just an interval

To prevent this scan from running every day we'll have to delete the entire scan. The only way that can be done is with the “reallocate stop” command:

storage01> reallocate stop /vol/vol0

Reallocation will be stopped on '/vol/vol0'.

Monitor the system log for results.

storage01> reallocate status -v

Reallocation scans are on

No reallocation status.

As you can see there are no schedules left.

**About the Status**

So what now about the status I noticed. It turned out that the relocate statuses stay on the status filed when a reallocate has been run on the aggregate in the past. See [this](https://kb.netapp.com/support/index?page=content&id=2012768&actp=LIST_RECENT&viewlocale=en_US&searchid=1328797690201) netapp knowledgebase article:

If block reallocation has been run on the aggregate, then the aggregate will show the “redirect” status. This status can only be cleared by reverting to a version of Data ONTAP prior to 7.2.3. If block reallocation has not been run on an aggregate, then the “redirect” keyword will not be displayed.  
Flexible volumes within aggregates that have started a block level reallocation (reallocate -A) may show an “active\_redirect” status within 'vol status -v' output. This is only true if there are blocks that have been reallocated, but the redirect scanner (final phase of block level reallocation) has not completed.

So redirect is normal, the active\_redirect not. According to [this thread](https://communities.netapp.com/thread/15696) you could do:

priv set diag

wafl scan redirect volumename

priv set

But I haven't found any confirmation on that.

Right now I'm thinking of opening a support case.

**Status Follow-Up**

I did open a case for this, but it took a long time for IBM (we have a rebranded netapp/n-series) to find the [actual bug report](http://support.netapp.com/NOW/cgi-bin/bol?Type=Detail&Display=395114). So, it is a bug if you want to remove the status run these commands:

priv set diag

wafl scan redirect -a <vol>

**More Information**

**Reallocate Command**

> reallocate

usage:

reallocate on | off

reallocate start [-t threshold] [-p] [-o] [-n] [-i interval] <path> | /vol/<volname>

reallocate start -f [-p] <path> | /vol/<volname>

reallocate start -A [-o] [-i interval] <aggr\_name>

NOTE: -A is for aggregate (freespace) reallocation.

Do NOT use -A after growing an aggregate if you wish to

optimize the layout of existing data; instead use

reallocate start -f /vol/<volname>

for each volume in the aggregate.

reallocate status [-v] [<path> | <aggr\_name>]

reallocate stop <path> | <aggr\_name>

reallocate quiesce <path> | <aggr\_name>

reallocate restart [-i] <path> | <aggr\_name>

reallocate schedule [-d] [-s <schedule>] <path> | <aggr\_name>

reallocate measure [-l logfile] [-t threshold] [-o] [-i interval] <path> | /vol/<volname>

Reallocate parameters:

* -p
  + Executes reallocate by using physical reallocation. Generally recommended.
* -f
  + Executes a forced reallocation. Generally not recommended.
* -o
  + Executes reallocation one time only.
* -n
  + Executes reallocation without measuring the layout first.
* -t
  + Forces reallocate to use a custom threshold.

**Run Reallocate After Adding Storage to Aggregate**

In [this article](http://www.warmetal.nl/netappaggrexpansion) we had to run reallocation after adding disks to the aggregate. Please read the article for performance impact etc. To run reallocate after adding disks issue this command for every volume:

filer01> reallocate start -f /vol/Volume\_1

Reallocation scan will be started on '/vol/Volume\_1'.

Monitor the system log for results.

**Links**

NetApp Cheat Sheet: <http://www.datadisk.co.uk/html_docs/netapp/netapp_cs.htm>   
NetApps TR-3929: Reallocate Best Practices Guide <https://fieldportal.netapp.com/viewcontent.asp?qv=1&docid=33904>   
NetApp Community: <https://communities.netapp.com/thread/6530>   
<https://communities.netapp.com/thread/7456>   
<http://www.wafl.co.uk/reallocate/>   
<http://www.theselights.com/2010/03/understanding-netapp-volume-and.html>   
<https://kb.netapp.com/support/index?page=content&id=2012768&actp=LIST_RECENT&viewlocale=en_US&searchid=1328797690201>   
<http://support.netapp.com/NOW/cgi-bin/bol?Type=Detail&Display=395114>

**Data Patterns**

Reallocate optimizes the layout of data on disk for “Sequential Read Access”. The workload that most benefits from this is “Sequential Reads After Random Writes”.

Examples for these kind of workloads are:

* Online transaction processing databases that have large table scans
* Email systems that use database storage with verification processes
* Host-side backup of LUNs

http://www.warmetal.nl/lib/exe/indexer.php?id=reallocate&1350651795