Dell DR Series Appliance Cleaner Best Practices

Dell Engineering
June 2016

A Dell Best Practices Document

Revisions

Date	Description
July 2016	Initial release

THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.
Copyright © 2016 Dell Inc. All rights reserved. Dell and the Dell logo are trademarks of Dell Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

Table of contents

Revisions 2

Intended audience 4

1 Deduplication and cleaning 5

2 How to determine if the cleaner is keeping up 7

2.1.1 Using the stats --cleaner command 7

2.1.2 Understanding cleaner statistics 8

3 Cleaner adjustments & best practices 9

3.1 Adjusting the cleaner runtime and schedule 9

3.1.1 Cleaner considerations 11

3.1.2 Scenario 1: Cleaner analysis and adjustments 11

A Additional cleaner commands 13

A.1 Create a cleaner schedule 13

A.2 Examine the cleaner 13

A.3 Modify the cleaner schedule 13

A.4 Running the cleaner manually 14

B How to run the cleaner 16

C Estimating cleaning work currently in progress 17

D If the cleaner runs behind 19

E Cleaner monitoring and adjustment flow chart 24

Intended audience

The DR appliance cleaner is process that is configured to run efficiently and effectively out of the box with no tuning or adjusting required. Only in extreme cases will the DR cleaner possibly may need calibration. This document is intended only for DR appliances which are exposed to the following extreme cases:

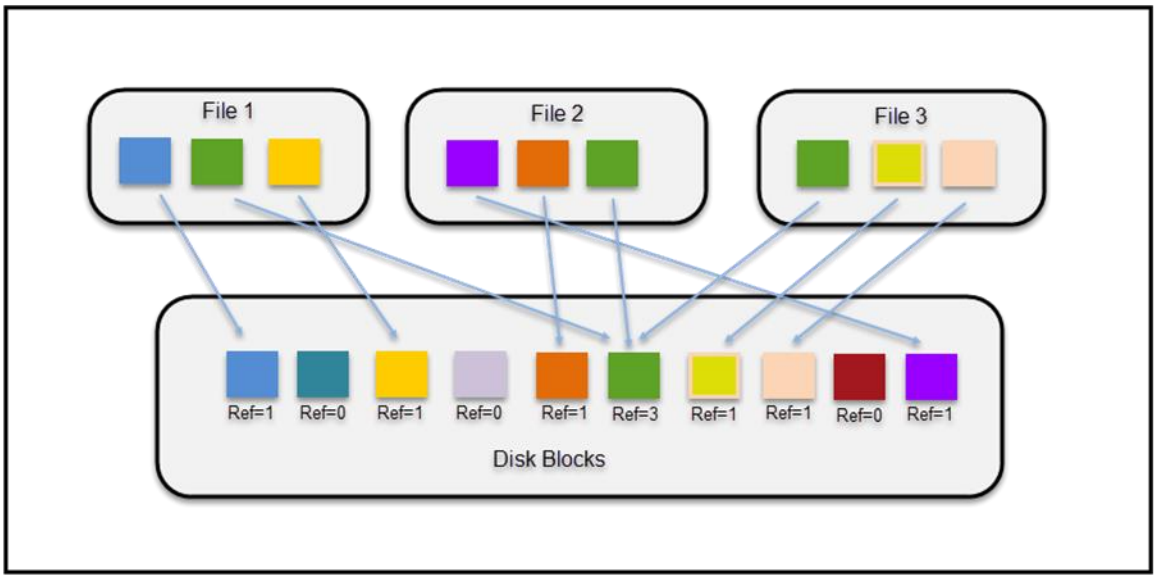
- Ingesting or deleting 100TB's or more per week
- Ingest occurs 24x7
- When poor savings is experienced or reported
- Full cleaner pass not finishing once per week

1 Deduplication and cleaning

Data deduplication refers to a technique for eliminating redundant data across all files in a data set. Large amounts of data can be significantly reduced saving costs and resources. The DR implements a variable block sliding window deduplication engine to produce industry leading results.

Each file stored in a DR appliance is a blockmap consisting of pointers to its chunks of data saved on the filesystem. When a file is stored in the DR appliance, a chunking process is applied such that the file is segmented into variable sized chunks. Each chunk is fingerprinted and examined in the DR's deduplication dictionary to see if it has been discovered before. If the DR system has encountered the chunk already, the files block map is updated to point to the already existing chunk in the system and the chunks reference count is increased. If a chunk is unique to the DR system, the chunk's fingerprint is inserted into the dictionary, the chunk is written to the filesystem, the files blockmap is updated to point to it and the chunks reference count is set to 1. Thus, files stored in the DR might point to unique blocks, non-unique blocks that are shared with other files in the system, or a combination of unique and non-unique blocks.

When a file is deleted, the files chunk reference counts are decremented by 1.



The DR cleaner plays a critical role to a DRs capacity because it updates chunk reference counts and reclaims space when chunk references are equal zero.

DR appliances are shipped with the cleaner setting to run automatically during idle time. In most cases, this setting allows more than enough time to update chunk references and reclaim space. In extreme cases where the DR cleaner is not permitted complete a full pass at least once per week, the suggestions in the documentation should be applied to the cleaner settings.

This document explains how to determine if the cleaner is keeping up, how much is required for the cleaner to finish, how to adjust the cleaner schedule and runtimes, how to determine what the cleaner is doing and best practices on what to set.

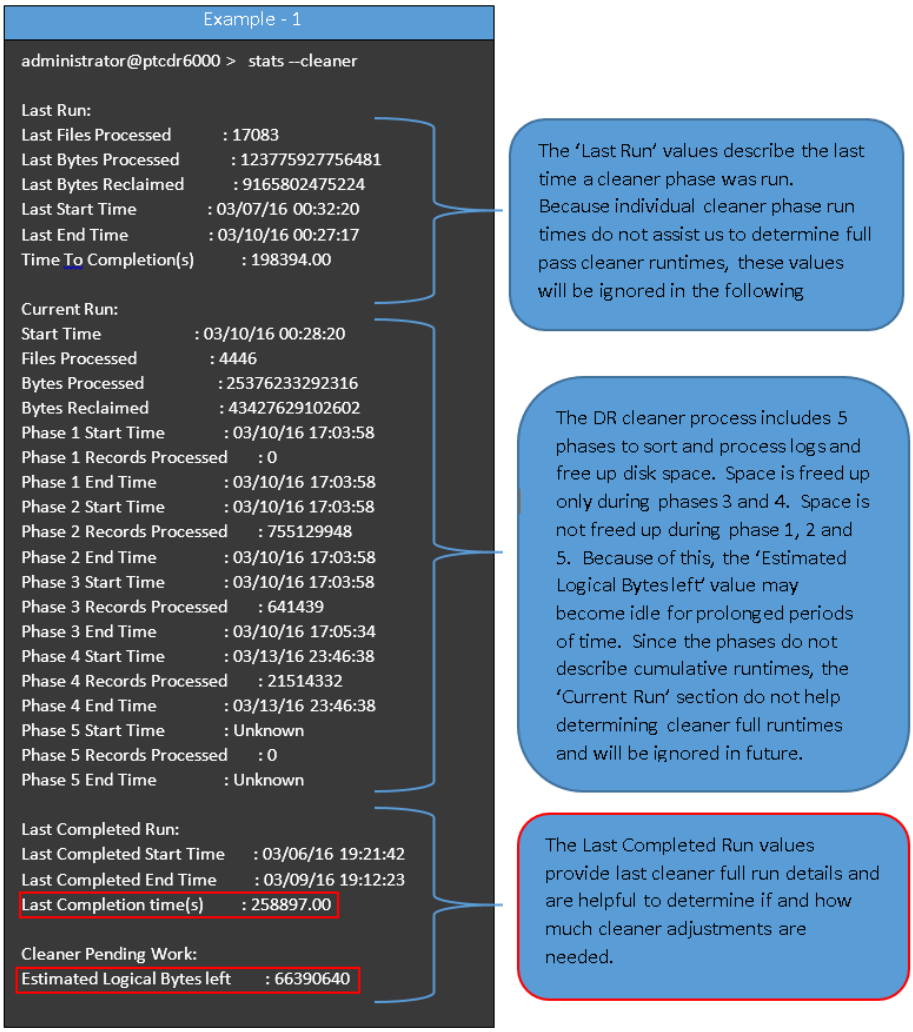
Note: To use this document, ensure the DR appliance is running 3.2.0.2 or 3.2.6.1. Build 3.2.0418.1a and future releases can also be used with this document.

2 How to determine if the cleaner is keeping up

Dell recommends that a full cleaner pass compete once every 7 days or less. To determine if the cleaner is keeping up, a quick inspection of the cleaner statistics 'Last Completion time(s)' variable is described below.

2.1.1 Using the stats --cleaner command

The **`stats --cleaner`** command displays the current running cleaner progress and the amount of time taken to complete its latest full pass. The 'Last Completion time(s)' value indicates that the cleaner is keeping up when it is less than or equal to 7 days.



Note: If the cleaner is running and it is not reporting to reclaim space, that does not mean there is an issue or problem with the cleaner. When cleaner is running, somecleaner phases do not reclaim data space. Only when the cleaner is executing in phases 3 and 4 is space reclaimed and the 'Estimated Logical Bytes left' is updated.

2.1.2 Understanding cleaner statistics

The 'Last Completion time(s)' variable provides the latest full pass cleaner runtime. In Example-1 above, the cleaner completed its latest full pass consuming 258897.00 seconds or 2.99 days. To convert from seconds to days, take the total seconds and divide by 86,400 (60 * 60 * 24 = 86,400). Thus, 258,897 seconds / 86,400 = 2.99 days. Because the cleaner completed its latest job within a week's time frame (7 days), the cleaner is not falling behind.

3 Cleaner adjustments & best practices

Under rare and extreme cases, the DR cleaner may run longer than 7 days. If so, apply the following cleaner best practices in order when adjusting cleaner runtimes.

- 1. If the cleaner is set to run automatically (default setting) schedule the cleaner to run for 40 or more hours per week.
- 2. Schedule the cleaner to run in long durations instead of smaller short durations.
- 3. If the cleaner does not finish in 40 hours, schedule cleaning to run in parallel with replication.
- 4. If the cleaner does not finish in 40 hours, schedule cleaning to run during low periods where there is not a full backup load.

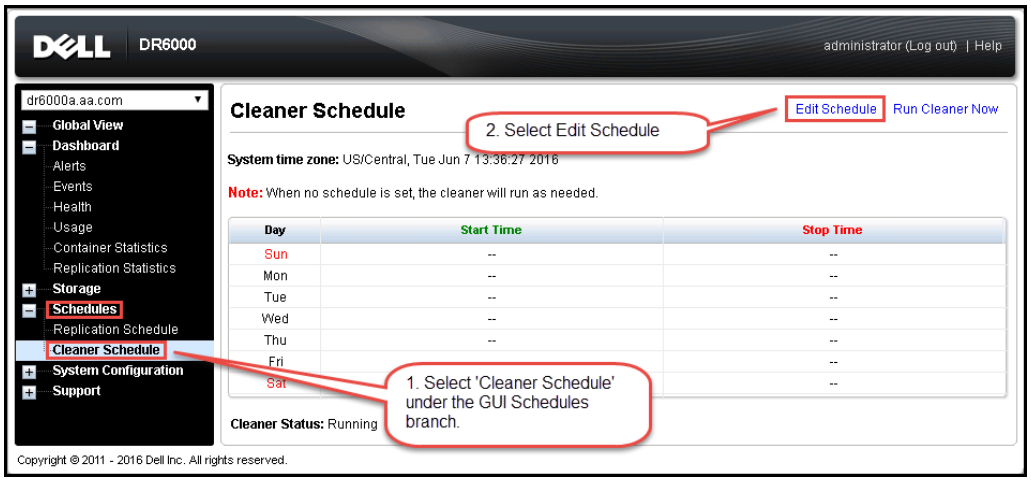
See Appendix-C & D for additional details and examples.

3.1 Adjusting the cleaner runtime and schedule

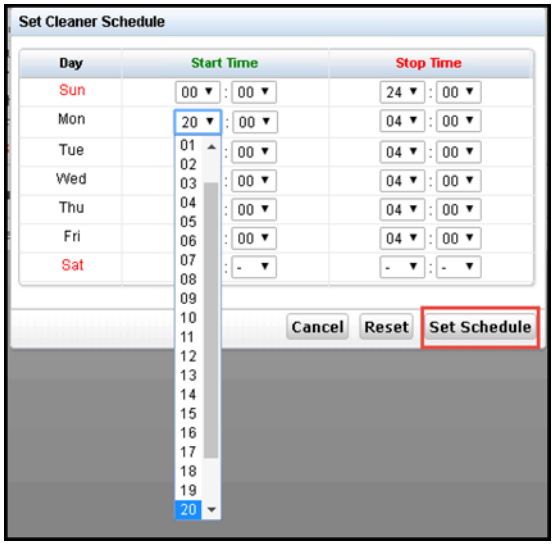
If the value of 'Last Completion time(s)' exceeds 7 days, the cleaner needs adjusting with additional runtime. Begin by scheduling the cleaner to run 40 hours / week during non-peak times. Scheduling can be applied in both the GUI and CLI.

The following example schedules the cleaner to run 40 hours per week:

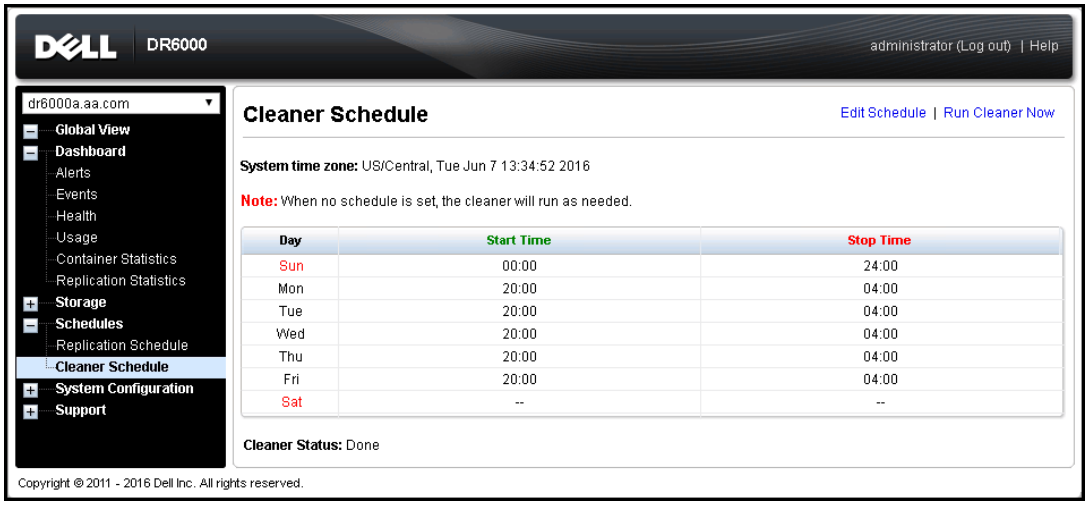
1. On the **Cleaner Schedule** page, click **Edit Schedule**:



2. For each day, apply a cleaner runtime schedule, then click **Set Schedule**.



3. Inspect the newly configured cleaner schedule.



4. After a week (7 days), run the stats --cleaner command to examine cleaner's 'Last Completion Time(s)' value. If the cleaner completed a full pass within 7 or less days, no more cleaner adjustments are necessary. If the cleaner has not yet completed a full pass or has taken beyond 7 days to complete, increase the cleaner schedule runtimes and repeat this step.

3.1.1 Cleaner considerations

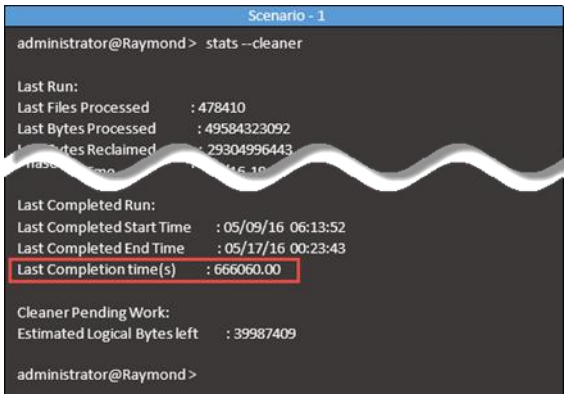
1. Additional cleaner runtime is required as more data is ingested or begins to age out.

2. When the cleaner is triggered to run by manual, scheduled or forced methods, the cleaner is run at the same priority as ingest and replication.

3.1.2 Scenario 1: Cleaner analysis and adjustments

Raymond has been successfully running his DR appliance at his computing center for the past 3 years. Because his DR has recently been reporting poor savings and performance is slow, he feels his cleaner is not completing within the Dell recommendation of a 7 day timeframe.

Inspect Raymond's cleaner results are shown below:

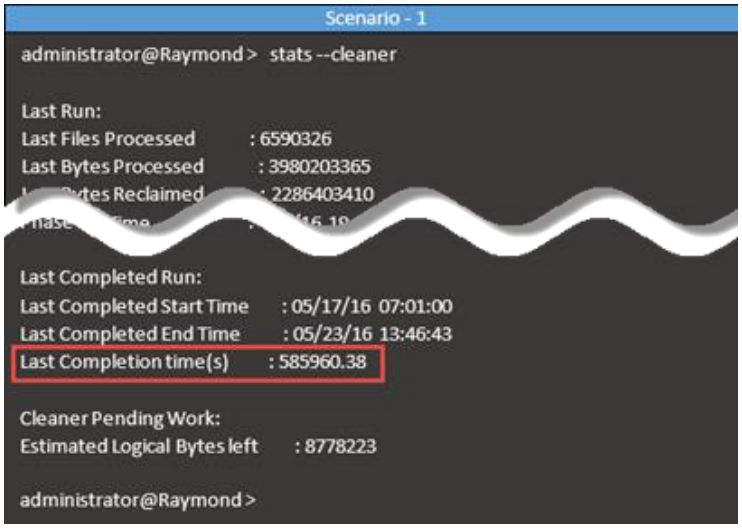


Since Raymond's cleaner is not completing within 7 days, (666,060 / 86,400 = 7.7 days) he will schedule his cleaner to run for 40 hours during the week.

Raymond's Current DR Schedule of Events:

Weekdays	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Full																									
Incremental	X	X	X	X	X	X																	X	X	X
Replication																									
Cleaner													X	X	X	X	X	X	X	X					
Weekends	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Full	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Incremental																									
Replication																									
Cleaner																									

Raymond would like to determine if his adjusted cleaning schedule is adequate for his environment, so he waits one week before he analyses his cleaner runtimes. Raymond then runs the stats --cleaner command.



Raymond adjusted cleaning schedule successfully completes in 6.78 days which meets Dells recommended full cleaner pass of 7 or less days.

A Additional cleaner commands

A.1 Create a cleaner schedule

Using the CLI:

```
schedule --add -day <Day of the Week> --start_time <HH:MM> --stop_time <HH:MM> --cleaner
```

Example

```
administrator@dr6000a > schedule --add --day Sunday --start_time 06:00 --
stop_time 22:00 --cleaner
Successfully updated Cleaner schedule.
administrator@dr6000a >
```

A.2 Examine the cleaner

Using the CLI:

schedule --show --cleaner

Example

```
administrator@dr6000a > schedule  --show --cleaner
Cleaner schedule:
                Start     Stop
Monday          06:00     17:00
Tuesday         06:00     17:00
Wednesday       06:00     17:00
Thursday        06:00     17:00
Friday          06:00     17:00
Saturday        --        ---
Sunday          --        --
administrator@dr6000a >
```

A.3 Modify the cleaner schedule

Using the CLI:

```
schedule --add --day Sunday --start_time 06:00 --stop_time 22:00 --cleaner
```

Example

```
administrator@dr6000a > schedule  --add --day Sunday --start_time 06:00 --stop_time 22:00 --cleaner
Successfully updated Cleaner schedule.
administrator@dr6000a >
```

```
schedule --delete --day <Day of the Week> --cleaner
```

Example

```
administrator@dr6000a > schedule --delete --day Sunday --cleaner
Successfully updated Cleaner schedule.
administrator@dr6000a >
```

Other cleaner schedule help is defined by running the following command:

```
Schedule --help
```

A.4 Running the cleaner manually

The cleaner can be manually run and is a onetime run and will complete a full cleaner pass very quickly as it does not pause for any DR activities until the cleaning process has completed.

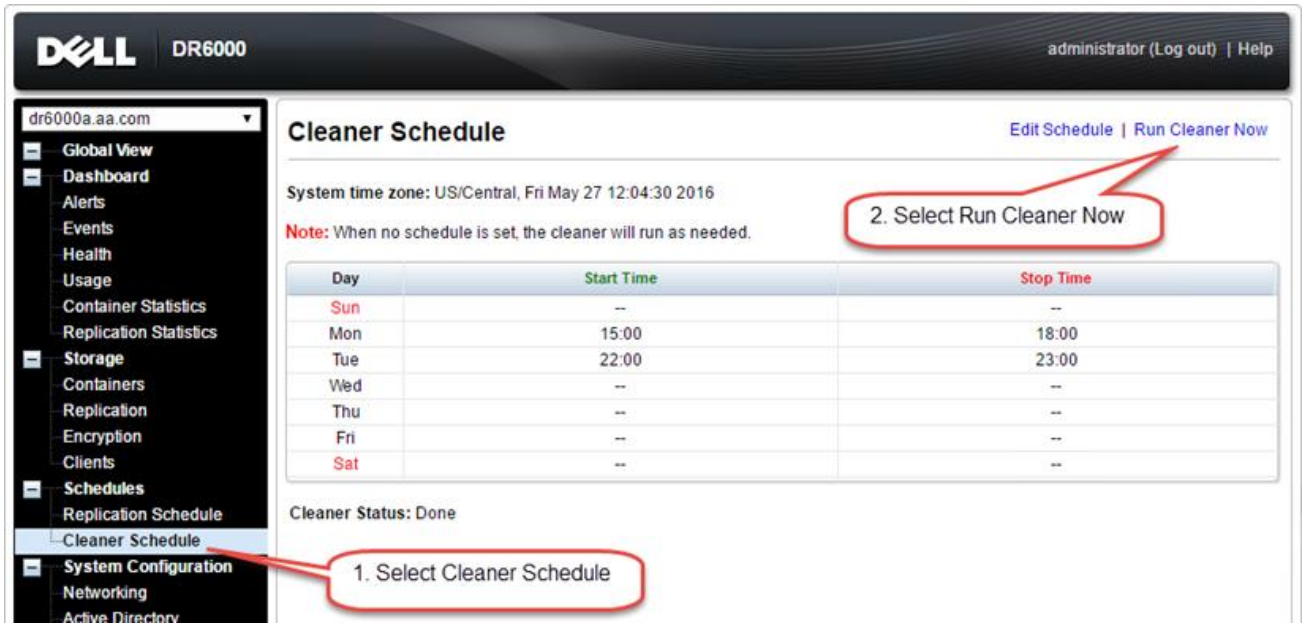
Using the CLI:

```
maintenance --filesystem --reclaim_space
```

Example

```
administrator@dr6300a > maintenance --filesystem --reclaim_space
Successfully started  cleaner.
administrator@dr6300a >
```

Using the GUI:



B How to run the cleaner

There are multiple ways in which the cleaner process is run:

- **Scheduled** – The cleaner is triggered to begin processing and continues until the cleaning work has been completed or the scheduled window ends. If cleaning has completed within the scheduled window, it will not start again until the next scheduled window. For large cleaning jobs, the cleaner will span multiple schedule windows picking up where it left off until the cleaning job has completed. Note: The cleaner job does not pause except in-between cleaning schedules.
- **Manual** – When the cleaner process is manually started, it will continuously run until the cleaning process has finished. Note: the cleaner will not pause until the cleaning job has completed.
- **Automatically (default)** – The cleaner process automatically starts if it detects that there are:
  - No running data ingest
  - No running replication processes
  - Two minutes of system idle time since the latest data ingest

  Once replication or data ingest begins, the cleaner pauses until the three conditions for auto start become true again. It does not matter if a cleaner schedule has been defined, the DR cleaner will run during idle times when there is work to be done. In other words, 50 hours of idle time equals 50 hours of cleaner time.

- **Forced Run** – When the DR is running lower than 1TB of free space or when internal cleaner logs (PL) are queued to 1 day of cleaner work the cleaner is triggered by the DR to begin. Here the cleaner will not pause until a full cleaner pass has been completed.

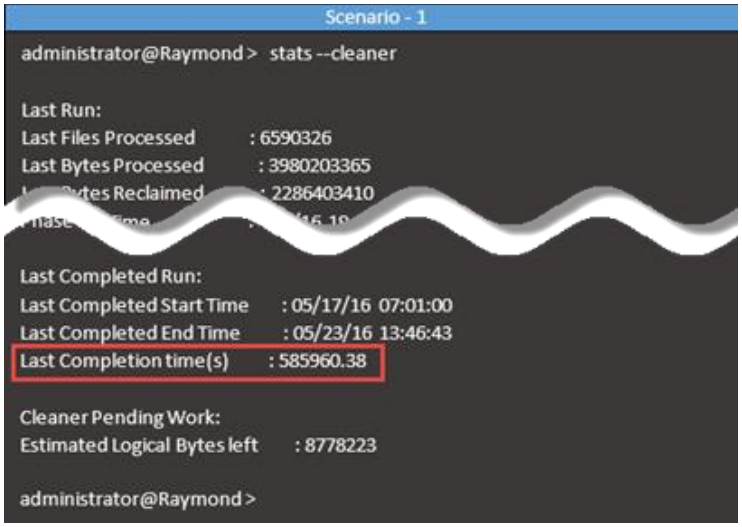
At a minimum, the DR appliance will need to run through a full cleaner pass at least once per week. Dell recommends the cleaner process to be run 6-8 hours per day or 40-50 hours per week. Running the cleaner for long contiguous periods of time is more efficient than many short periods of time. For example, the cleaner will free up disk space more quickly if it is run for 16 hours each day for three days, instead of 7 hours each day for the week.

When ingestion and/or replication complete, the cleaner will automatically start. When replication and/or ingestion stops, the cleaner will automatically resume.

If deleting data to free up disk space as quickly as possible is desired, manually running the cleaner is recommended.

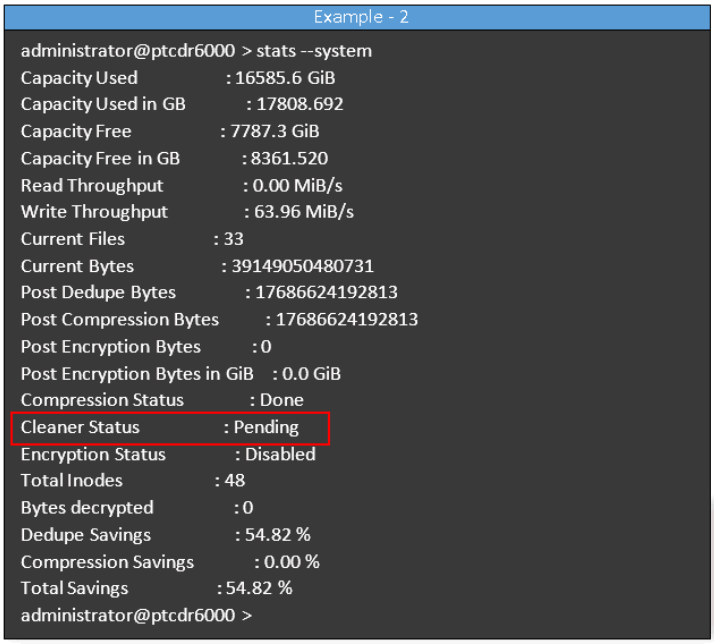
C Estimating cleaning work currently in progress

The 'Estimated Logical Bytes left' value provides current active cleaner progress. It provides an estimate of how much work is left for the current active cleaner to complete a full pass. When this value reaches 0, the cleaner has completed a full pass.



To determine the current state of the cleaner, run the following command:

Stats --system



The DR cleaner operates in three states:

- **Done** – When the cleaner is done and there is nothing more to process.
- **Running** – The cleaner is currently running.
- **Pending** – The cleaner has work to do, but is currently waiting to run.

In Example – 2 above, the cleaner is placed in "Pending" mode because there is cleaning yet to be done, and the DR cleaner is outside of its scheduled window of operation and the DR is not idle, but is busy with ingest or replication.

D If the cleaner runs behind

When a full cleaner pass runs beyond the 7 days recommended by Dell, the cleaner is considered to be in cleaning debt. For example, if a full cleaner pass completes in 9 days, the cleaner has 2 days of cleaning debt and cleaner runtimes need to be immediately increased. Cleaning debt occurs when the DR ingests more data than can be cleaned. Depending on the rate of cleaner debt accumulation, the DR can quickly fill up eventually halting backups. When cleaner runtimes are increased, the cleaning debt can be immediately addressed and eliminated over time.

Follow the cleaner best practices outlined in Section-3 of this document.

Example

Cindy manages backups at a pharmaceutical company which includes multiple remote DR appliances replicating to a production site. Cindy notices that her production site DR appliance is rapidly and unexpectedly running out of free space. Since she has not paid attention to the cleaner process in quite some time, she feels her DR cleaner has fallen behind and is not freeing up disk capacity fast enough.

At her production site, Cindy currently holds a retention policy of 12 weekend full backups and 15 daily incrementals and her backup schedules are described in the following table.

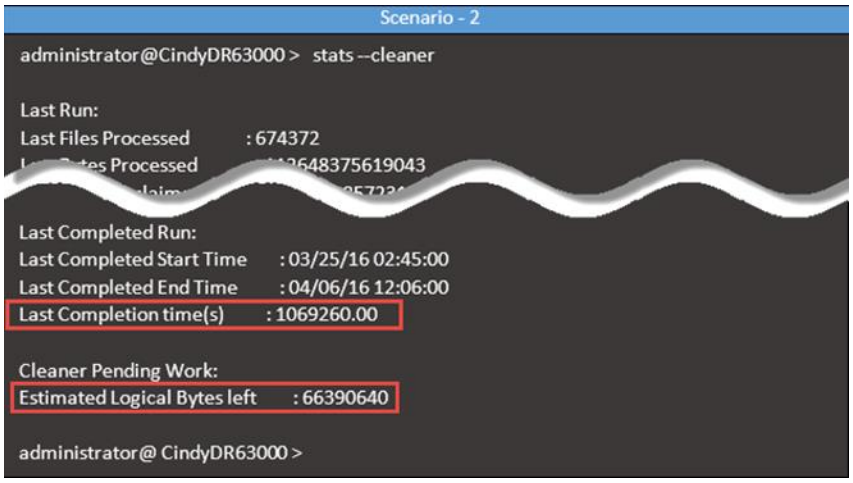
Weekdays (Mon – Fri)	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Full																									
Incremental	X	X	X	X	X	X	X															X	X	X	X
Replication																		X	X	X	X				
Cleaner								X	X	X	X	X	X	X	X	X	X								
Weekend (Sat & Sun)	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Full	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Incremental																									
Replication																									
Cleaner																									

In this scenario, Cindy will calibrate the production site cleaner schedule such that full cleaner passes meet Dell's recommended cleaner completion time of 7 days.

Note: Each of Cindy's remote sites will require a cleaner schedule inspection as well, but in this scenario, only the production site cleaner will be examined.

Cindy examines the current state of her production site cleaner by running the following command:

```
stats  --cleaner
```



Cindy notes that the latest full pass cleaning process required 1,069,260 seconds or 12.38 days (1,069,260 / 86,400) thus, Cindy's cleaner has fallen behind of 5.3 days (12.38 – 7 = 5.3) in cleaner debt. Cindy must increase her cleaning runtimes such that cleaning debt begins to diminish and eventually achieves a full cleaner pass to complete within 7 days.

How much should Cindy increase her cleaner schedule? If Cindy's DR appliance free space is critically low, she must aggressively schedule the cleaner to recover disk space as quickly as possible by running the cleaner manually such as the GUI's 'Run Cleaner Now' command. In this example, Cindy is not in a critical free space situation, but will need to increase her cleaner runtime ASAP. Without increasing cleaner runtime schedules, Cindy's production DR can quickly consume the remaining amount eventually entering into a critical state.

Cindy's cleaner debt is substantial because it is ~75% Dell's recommended 7-day full cleaner pass (5.3 / 7 = .75). Thus, she decides to initially increase her cleaning time by ~ 75% in hopes to being to reduce the 5.3 days of cleaning debt and to eventually compete a full cleaner pass within a week. A 75% increase of 50 hours is ~ 88 hours (50 * 1.75 = 87.5), so Cindy decides to increase the cleaning schedule up to 90 hours per week, then observe the impacts to cleaner adjustments. Cindy realizes that the 90 hours of cleaner runtime is just an estimate, so she expects multiple adjustments to calibrate the cleaner schedule correctly.

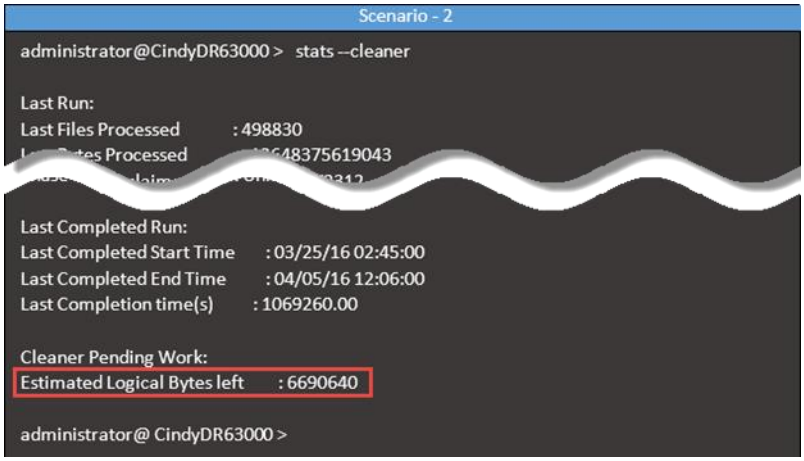
To achieve 90 hours of cleaner run time, Cindy makes her adjustments shown below in the table below:

Weekdays (Mon – Fri)	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Full																									
Incremental	X	X	X	X	X	X	X															X	X	X	X
Replication																		X	X	X	X				
Cleaner								X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Weekend (Sat & Sun)	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Full	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Incremental																									
Replication																									
Cleaner																									

Increased cleaner schedule

Cindy takes note of the value 66,390,640 in the Estimated Logical Bytes left variable and logs it with the date.

After making the cleaner adjustments above, Cindy waits 7 days to determine if additional cleaner adjustments are required. She hopes to reduce the accumulated cleaner debt, and the weeks' worth of aged out data. Cindy now runs the `stats --cleaner` command after a week when the cleaner adjustments were made:

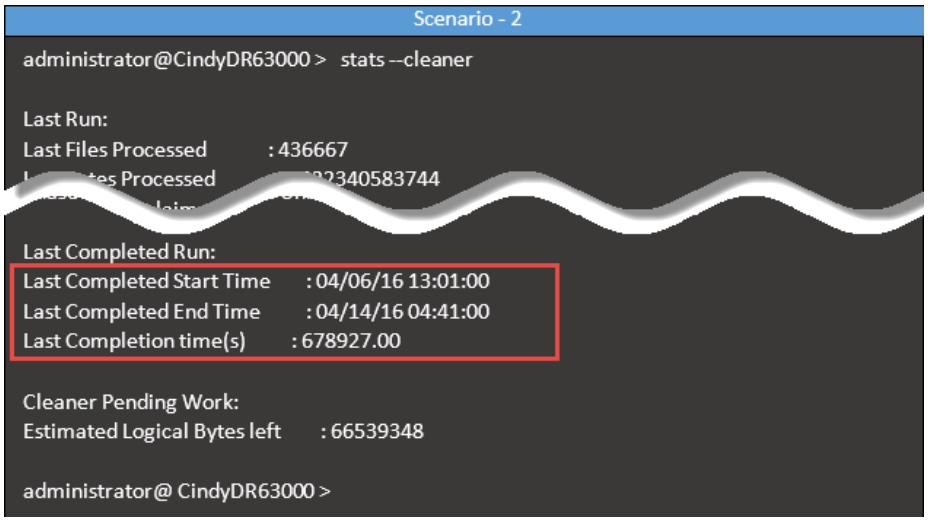


Cindy takes note of the value 6,690,640 in the Estimated Logical Bytes left variable and logs it with the date. Cindy notices that the cleaner has made significant cleaning progress but has not completed a full cleaning pass during the past week.

04/06/16	66,390,640
04/13/16	6,690,640
Cleaner Progress	59,848,708

Cindy does not adjust the 90 hour per week cleaning schedule and expects a full cleaning pass to complete within the next couple days. After waiting a week (for the next backup to occur), she runs the following command:

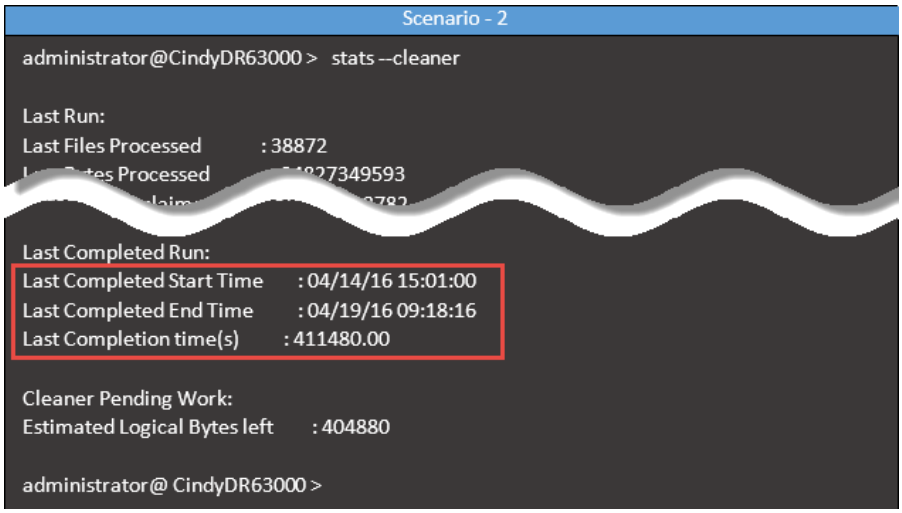
```
stats --cleaner
```



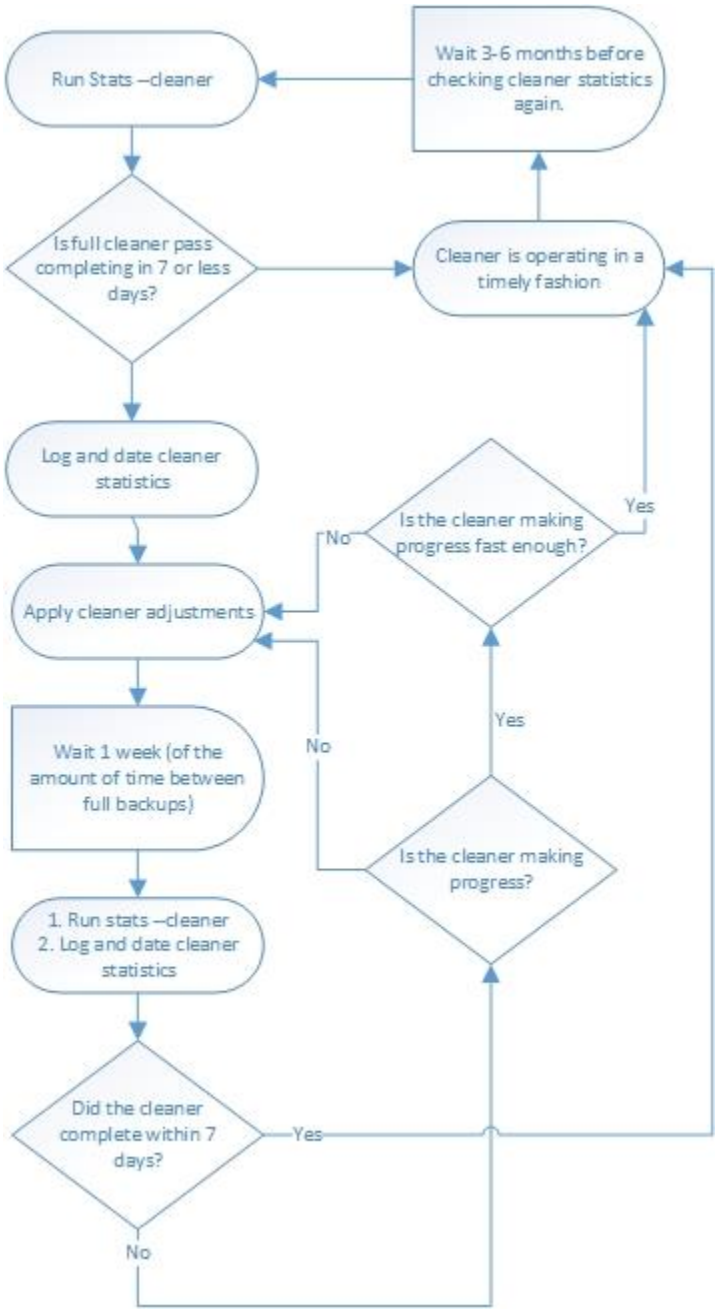
Notice that a full cleaner pass completed in ~ 7.85 days (678,927 / 86,400) which means Cindy's adjusted 90-day cleaner schedule has nearly eliminated the cleaner total run time from 12.38 to 7.85 days. With such big cleaner debt progress and such little cleaner debt remaining, Cindy believes that the current cleaner settings will accomplish Dells recommended full cleaner pass within the next 7 days during next week's backup cycle. Thus, Cindy does not make any cleaner adjustments and waits 7 days to inspect the cleaner results.

Cindy then runs the following command:

```
stats --cleaner
```



Cindy has achieved in setting a successful cleaner schedule as her full cleaner pass has completes in 7 days or less. Her full cleaner pass actually completes in 4.76 days (411,480 / 86,400).

Cindy will wait another 3-6 months and will reexamine if her cleaning runtimes.

E Cleaner monitoring and adjustment flow chart