

This is how you create a comprehensive dedupe report of your DDBoost backups

In my document [gener736.pdf](#) I demonstrated how to generate such report from the NMC GUI. I personally do not like the NMC reports in general. The major reason is that one does not know how they will be generated.

On the other hand, there is a report for a specific save set which will tell you exact details about the amount of data before and after the compression. Simply use the command `mminfo -q ssid=<ssid> -S` to achieve that. Here is an example:

```
D:\>mminfo -q ssid=4286059840 -S
ssid=4286059840 savetime=4/15/2021 12:28:14 (1618482494) 19-nwtest.eval.local:D:\5GB
level=full sflags=vF size=5000154140 files=4 insert=4/15/2021
create=4/15/2021 complete=4/15/2021 browse=5/15/2021 23:59:59 retent=5/15/2021 23:59:59
clientid=d7d6416b-00000004-60610392-60610391-00010c00-35df0829
*backup start time: 1618482458;
*NSR_VSS_SHADOWCOPY_SET: {958F3D5B-A9B0-41C9-862E-87F7B2FBB9FE};
*policy action jobid: 32002;
*policy action name: "backup: 1618482494";
*policy name: "TEST: 1618482494";
*policy workflow name: "Backup: 1618482494";
*ss data domain backup cloneid: 1618482494;
*ss data domain dedup statistics: "v1:1618482494:5014205776:5014205776:272721499";
group: TEST;
saveset features: CLIENT_SAVETIME;
Clone #1: cloneid=1618482494 time=4/15/2021 12:28:14 retent=5/15/2021 flags=
frag@ 0 volid= 7869421 file/rec= 0/0 rn=0 last=4/15/2021
D:\>
```

For a 'compression report' you really only need these 3 lines, which I started with bold characters:

```
ssid=4286059840 ...
level=full ...
.....
*ss data domain dedup statistics: ...
```

This is what the coloured values will display:

size=5000154140	The size of the save set in bytes (mminfo: <i>sumsize</i>)
1618482494	The exact time stamp (mminfo: <i>cloneid</i>)
272721499	The size after compression (aka <i>post comp size</i>). This is the amount of disk space the save set needs on the Data Domain

The numbers in green **5014205776:5014205776** represent internal sizes which not of real interest. As you can see, their difference with respect to the save set size is absolutely irrelevant.

If you divide the *post comp size* by *sumsize*, you will get an information about the compression ratio.

So wouldn't it be nice to create a comprehensive report about the compression rate of all DDBoost backups where you exactly know how the values have been created?

For such a job one must run the appropriate report `mminfo -S` **for each single save set**. Then the results must be collected and processed appropriately. This sounds as a perfect job to be automated by a small utility.

Now - I am not a programmer at all. Of course, over my long career I had to write the one or the other program and got some experience with assembler, Fortran, (Turbo-)Pascal and finally with the Windows PowerShell. Unfortunately, 'higher' object-oriented program languages like C (with all its relatives) and Pearl are still too cryptic for me. Therefore I got the idea to modify and extend one of my existing PowerShell scripts that it will be capable to generate such a compression report. The result will be stored in a CSV- and a text file and can now be imported in an Excel table for further processing.

As a foundation I use the framework of a monthly backup report. It has been written in such a way, that it will run an **automatic report for the last completed month**. However, you will be able to override this period with an earlier date (month).

I have organized the script to be structured using the 'top-down' method:

- The subroutines will be defined first.
- At the end, the main program will just start these subroutines in the proper sequence.

At the beginning you will find some routines, which - due to their general nature - can be used by the majority of my other scripts. This includes, among other things ...

- The preparation of necessary functions which are not yet available by PowerShell today, like `touch`
- The automated calculation of the last month,
- The automated creation of uniquely named subdirectories.

Otherwise the subroutines are named with respect to their functionality and do not really need any further description. However, if such is needed, you will find it within the code itself.

Besides the time-consuming collection of the data, the second big task is to extract only the relevant data. This is not an easy job because the `mminfo -S` report in fact contains a bunch of obstacles like ...

- different number of lines due to the nature of the save set and multiple clone instances
- leading empty spaces
- gaps of different lengths
- brackets like '(' and ',' used as separators (nsavetime)
- (multiple) colon (:) characters within a single line
- quotation marks (") as separators
- backslashes (\) used to indicate a line continuation

As a result these are the major tasks to be fulfilled:

- Detection of the DDVE save sets
- Creation of the `mminfo -S` reports
- Processing the reports to prepare the extraction of the necessary information
- Data extraction, calculation of the necessary columns and preparation of the output files

And because some of these steps can be very time consuming I am helping the user with appropriate progress feedbacks. Additionally, begin and end of each routines will generate a 0-byte file as time stamps which will allow to verify their duration.

This is how the script will perform for a very small NetWorker server:

```

dd_backup_statistics.ps1

The creation of the DD backup statistic for the last month begins ..

>>>> Depending on the number of save sets this can even take some days! <<<<<

That's why longer lasting steps will display a start and a stop message on the monitor
as well as a time-stamp file in the output directory.
Do not forget to delete them later.

Step 8 - Extract all DD save sets - Start

8.1. Create the 'mminfo' reports - for ALL disk backups
8.2. Create the 'mminfo' reports - for ALL DD disk backups
8.3. Create the 'mminfo -S' reports - for ALL DD disk backups

Number of DD Save Sets : 23

Current SSID : 4027066941
Current SSID : 4010290493
Current SSID : 3792241928
Current SSID : 4164066518
Current SSID : 4147289302
Current SSID : 4130512086
Current SSID : 4113734870
Current SSID : 4096967345
Current SSID : 4080190155
Current SSID : 4063414013

Save Set 10 of 23

Current SSID : 4046637216
Current SSID : 4029869083
Current SSID : 4013093101
Current SSID : 3996323606
Current SSID : 3979546390
Current SSID : 3962769174
Current SSID : 3945991958
Current SSID : 3929223268
Current SSID : 3912446053
Current SSID : 3895670231

Save Set 20 of 23
    
```

```

powershell.ps1

Current SSID : 3862125072
Current SSID : 3845348922

Step 8 - Extract all DD save sets - Stop

Step 9 - Replace special characters - Start
This may take a few minutes ...
Step 9 - Replace special characters - Stop

Step 10 - Extract DD save set details - Start
This can take some minutes before you see the first output ...
Line 100 of 461
Line 200 of 461
Line 300 of 461
Line 400 of 461
Step 10 - Extract DD save set details - Stop

Step 11 - Separate client & SS names - Start
Save Set 1 of 23
Save Set 11 of 23
Save Set 21 of 23
Step 11 - Separate client & SS names - Stop

Step 12 - Delete variable names - Start
This may last some minutes ...
Step 12 - Delete variable names - Stop

Step 13 - Calculate compression ratio - Start
This may last some minutes ...
Step 13 - Calculate compression ratio - Stop

Done.

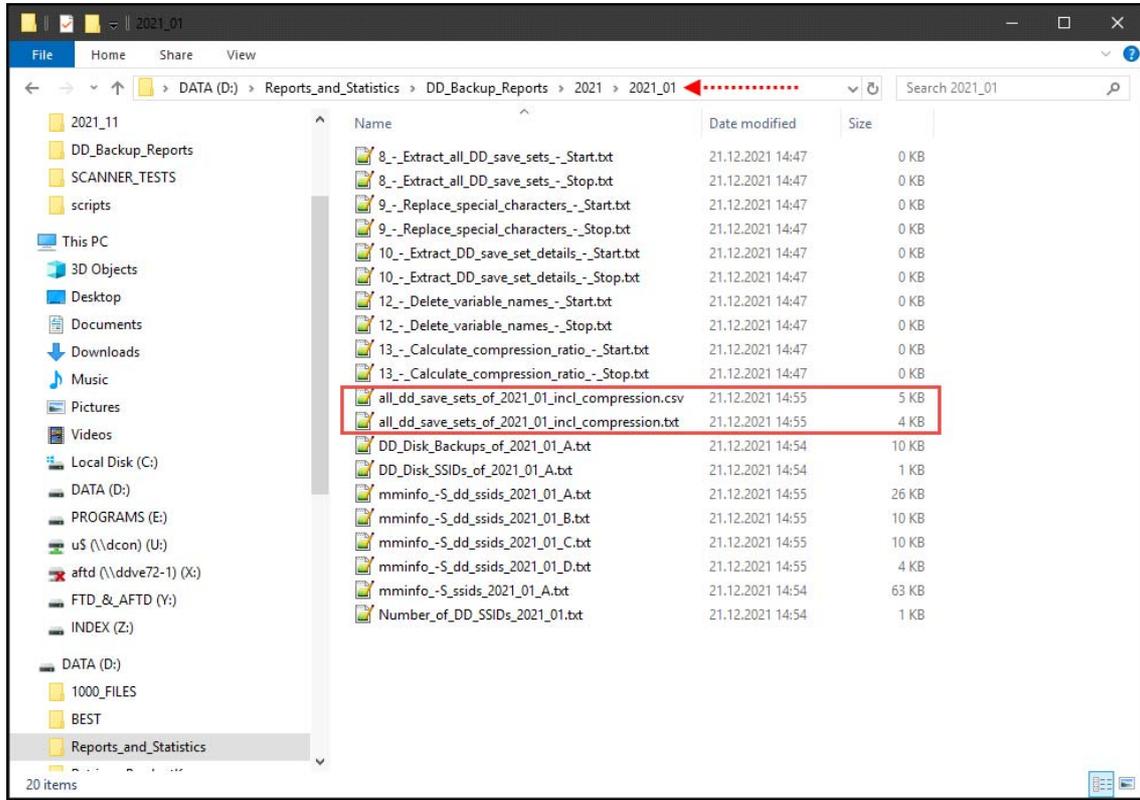
You will find the results in these two files:

D:\Reports_and_Statistics\DD_Backup_Reports\2021\2021_01\all_dd_save_sets_of_2021_01_incl_compression.csv
D:\Reports_and_Statistics\DD_Backup_Reports\2021\2021_01\all_dd_save_sets_of_2021_01_incl_compression.txt

Press any key to continue ...

PS D:\scripts>
    
```

Finally, the directory will keep the resulting reports as well as the associated timestamp files :



The next page will show you an the result already imported in an Excel sheet.

Of course you may assign the last number any other name you want.

ssid	savetime	nsavetime	client	ss_name	level	ssflags	size	files	ss_insert	cloneid	raw_size	pre_lc_size	post_lc_size	res_size [%]	compr to [%]
4027066941	02.08.2021 19:24	1627925051	19-client19-1	Z:\100K FILES	full	vF	50542804	100319	02.08.2021	1637672062	50688224	145420	145420	0,3	99,7
4010290493	02.08.2021 19:36	1627925819	19-client19-1	Y:\1_Mio_Files - 1	full	vF	508556792	1003157	02.08.2021	1637672066	509952968	1396176	1396176	0,3	99,7
3792241928	03.08.2021 10:57	1627981063	19-client19-1	X:\10_MIO FILES	full	vF	5045599588	10031548	03.08.2021	1637673150	5060917824	19545158	15566110	0,3	99,7
4164066518	04.09.2021 00:08	1630706899	19-client19-1	Y:\1_Mio_Files - 2	full	vF	508556804	1003157	04.09.2021	1630706901	510129632	509998463	32894205	6,5	93,5
4147289302	04.09.2021 00:08	1630706900	19-client19-1	Y:\1_Mio_Files - 1	full	vF	508556804	1003157	04.09.2021	1630706901	509952980	509952980	24408592	4,8	95,2
4130512086	04.09.2021 00:08	1630706901	19-client19-1	Y:\1_Mio_Files - 4	full	vF	508556804	1003157	04.09.2021	1630706901	509947408	509876686	24571699	4,8	95,2
4113734870	04.09.2021 00:08	1630706902	19-client19-1	Y:\1_Mio_Files - 3	full	vF	508556804	1003157	04.09.2021	1630706901	509963228	509942108	23806091	4,7	95,3
4096967345	04.09.2021 02:49	1630716592	19-client19-1	Y:\1_Mio_Files - 5	full	vF	508556804	1003157	04.09.2021	1630716593	509964824	509938424	24488951	4,8	95,2
4080190155	04.09.2021 02:50	1630716618	19-client19-1	Y:\1_Mio_Files - 6	full	vF	508556804	1003157	04.09.2021	1630716619	509949088	509797118	24423656	4,8	95,2
4063414013	04.09.2021 03:08	1630717692	19-client19-1	Y:\1_Mio_Files - 7	full	vF	508556804	1003157	04.09.2021	1630717692	509960932	509813718	24286235	4,8	95,2
4046637216	04.09.2021 03:15	1630718110	19-client19-1	Y:\1_Mio_Files - 8	full	vF	508556804	1003157	04.09.2021	1630718111	509976108	509824138	25558287	5,0	95,0
4029869083	04.09.2021 05:46	1630727194	19-client19-1	Y:\1_Mio_Files - 9	full	vF	508556804	1003157	04.09.2021	1630727195	509979188	509842209	25463640	5,0	95,0
4013093101	04.09.2021 06:07	1630728427	19-client19-1	Y:\1_Mio_Files - 10	full	vF	508556840	1003157	04.09.2021	1630728429	509976144	509826760	25485106	5,0	95,0
3996323606	04.09.2021 08:15	1630736148	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 1	full	vF	524813304	1003159	04.09.2021	1630736150	526543584	526478738	132711515	25,3	74,7
3979546390	04.09.2021 08:15	1630736149	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 3	full	vF	524813220	1003159	04.09.2021	1630736150	526559432	526559432	131599504	25,1	74,9
3962769174	04.09.2021 08:15	1630736150	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 4	full	vF	524813220	1003159	04.09.2021	1630736150	526543220	526543220	134273322	25,6	74,4
3945991958	04.09.2021 08:15	1630736151	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 2	full	vF	524813220	1003159	04.09.2021	1630736150	526538880	526365000	132666887	25,3	74,7
3929223268	04.09.2021 10:37	1630744676	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 6	full	vF	524813220	1003159	04.09.2021	1630744677	526543864	526183802	134050400	25,5	74,5
3912446053	04.09.2021 10:37	1630744677	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 5	full	vF	524813220	1003159	04.09.2021	1630744677	526544984	526193505	134116220	25,6	74,4
3895670231	04.09.2021 11:01	1630746068	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 7	full	vF	524813220	1003159	04.09.2021	1630746071	526559936	526322223	134182283	25,6	74,4
3878893293	04.09.2021 11:05	1630746349	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 8	full	vF	524813220	1003159	04.09.2021	1630746349	526543584	526267904	133496089	25,4	74,6
3862125072	04.09.2021 13:35	1630755342	19-client19-1	W:\10_MIO_RANDOM\1_Mio_Random - 9	full	vF	524813220	1003159	04.09.2021	1630755344	526541932	526203975	133328118	25,4	74,6

In Excel you can process the data further - filter and sort them as you like for a more specific report.





Please be aware of the following issues.

- Depending on the size of your NetWorker data zone the creation of the save set reports `mminfo -S` can take hours or even days.
mehrere Stunden, ja sogar mehrere Tage dauern.
Please find below the example for 268k DDBoost save sets.

- Such job can of course load your backup server.
Therefore you should consider to use an extra NetWorker server for this purpose. As a well-trained administrator you might use your standby server which you use for your regular ;-) DR tests.

You will find more information in my tip [gener653.pdf](#).

BTW - with the latest NetWorker version 19.5.0.3 you do not need a fake hosts table any longer.

You should rename the program `nsrim.exe`. As a result, the networker server will not be able to delete expired save sets.

- Please provide at least enough RAM for the DR server (minimum 8GB).
- Also provide enough empty space on your NetWorker Volume.
With the DR system mentioned above the media index was almost 500MB - the intermediate output files need - alt least temporarily - more than twice the size on your NetWorker installation volume (see next page).

Here is the runtime-screenshot for the server mentioned above with 268k DDBoost backups:

```

dd_backup_statistics.ps1
The creation of the DD backup statistic for the last month begins ..

>>>> Depending on the number of save sets this can even take some days! <<<<<

That's why longer lasting steps will display a start and a stop message on the monitor
as well as a time-stamp file in the output directory.
Do not forget to delete them later.

Step 8 - Extract all DD save sets - Start

8.1. Create the 'mminfo' reports - for ALL disk backups
8.2. Create the 'mminfo' reports - for ALL DD disk backups
8.3. Create the 'mminfo -S' reports - for ALL DD disk backups

Number of DD Save Sets : 268624 ◀.....

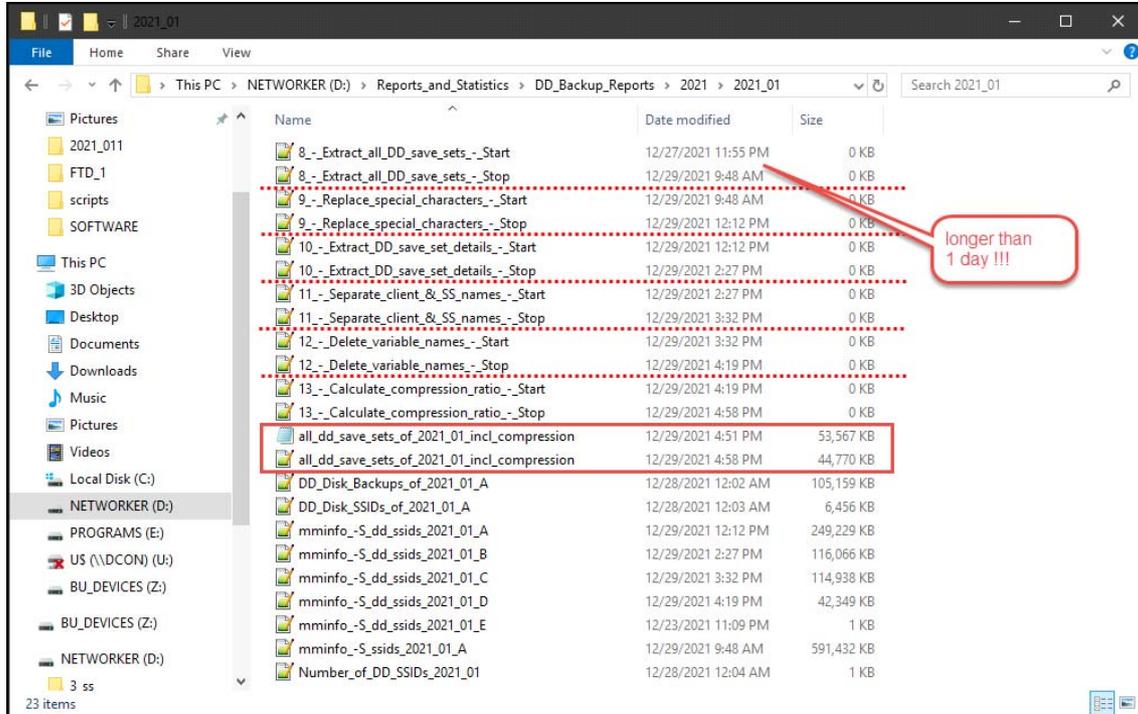
Current SSID : 159326540
Current SSID : 92217696
Current SSID : 4152303983
Current SSID : 4018086258
Current SSID : 3984531827
Current SSID : 4001309043
Current SSID : 4034863474
Current SSID : 3934200179
Current SSID : 3950977395
Current SSID : 3917422963

Save Set 10 of 268624

Current SSID : 3967754611
Current SSID : 3783205240
Current SSID : 3766428025
Current SSID : 3749650810
Current SSID : 3732873595
Current SSID : 3665764733
Current SSID : 3615433087
Current SSID : 3380552079
    
```

And this is what you will find in the output directory:

- Use the timestamp files to verify the duration in your environment.
- For better demonstration I have not deleted the intermediate output files:



Are you interested? - If so, you may receive my script absolutely free. However, I would like to receive a feedback about the number of the interested people. Therefore I decided not to provide it as a downloadable file. But if you sent me an e-mail I will provide it a.s.a.p.

carsten_reinfeld@avus-cr.de