



GTUG

Why using Deduplicated-storage

Fernand Lussier

VP Research and Development

Nonstop File type

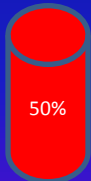
Hypothesis of simulation

- 1. Dynamic file:** 1 or 2 % of dynamic file change every day. And represented 50% of the data. Ex Cardholder master file
- 2. Static file:** represent 20% of the data . Doesn't change during the 90 day simulation. Ex. OS , obeyfile , programs , configuration
- 3. Semi-static file:** represent 30% of the data . 7 days of delay are kept on disk. Ex. Logfile



20%

Static files : OS, program, obey file, configuration file,...



50%

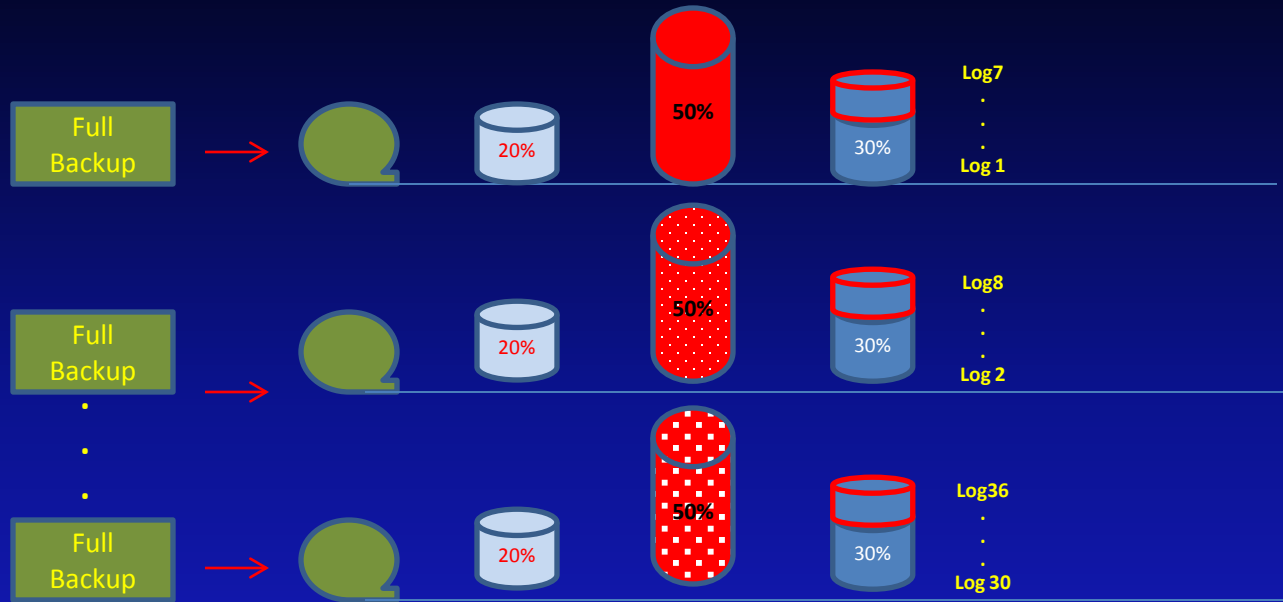
Dynamic files : data base table, master file



30%

Semi-Static files : daily log keep for several days,...CV

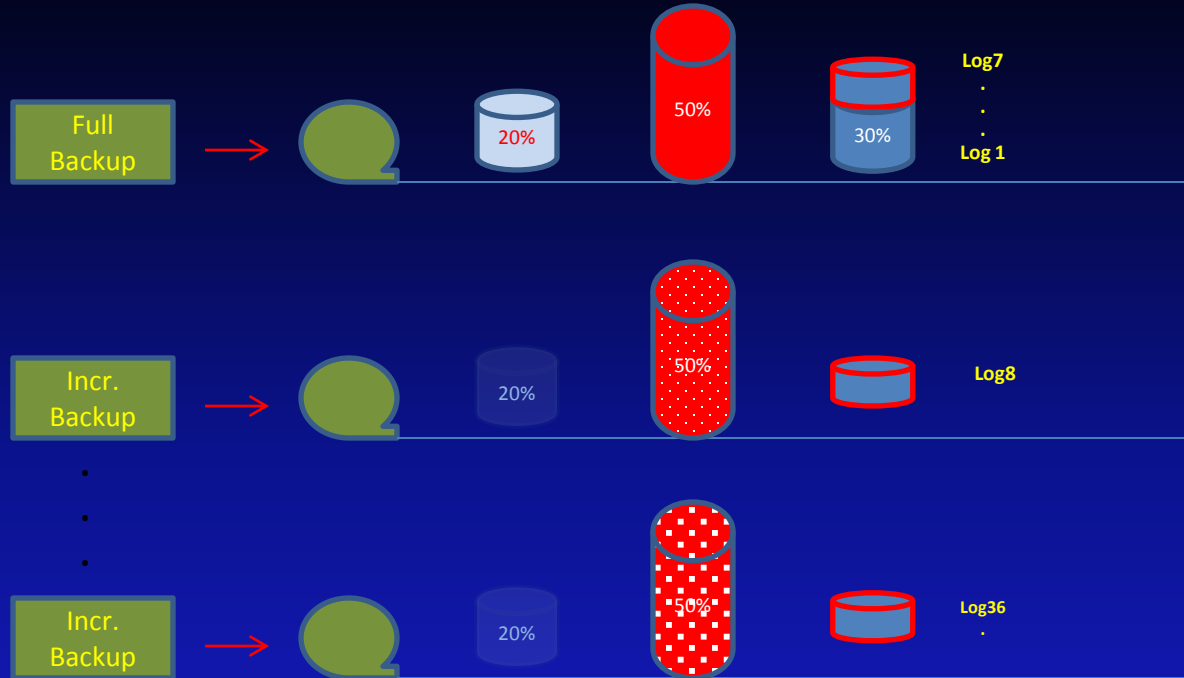
Full backup every day



Day	Original Data(GB)	Compressed data(GB)
1.00	1,000	333.33
2.00	1,000	333.33
3.00	1,000	333.33
30.00	1,000	333.33
Total	30,000	10,000.00

Full volume/subvolume restore of any specific day need a single restore

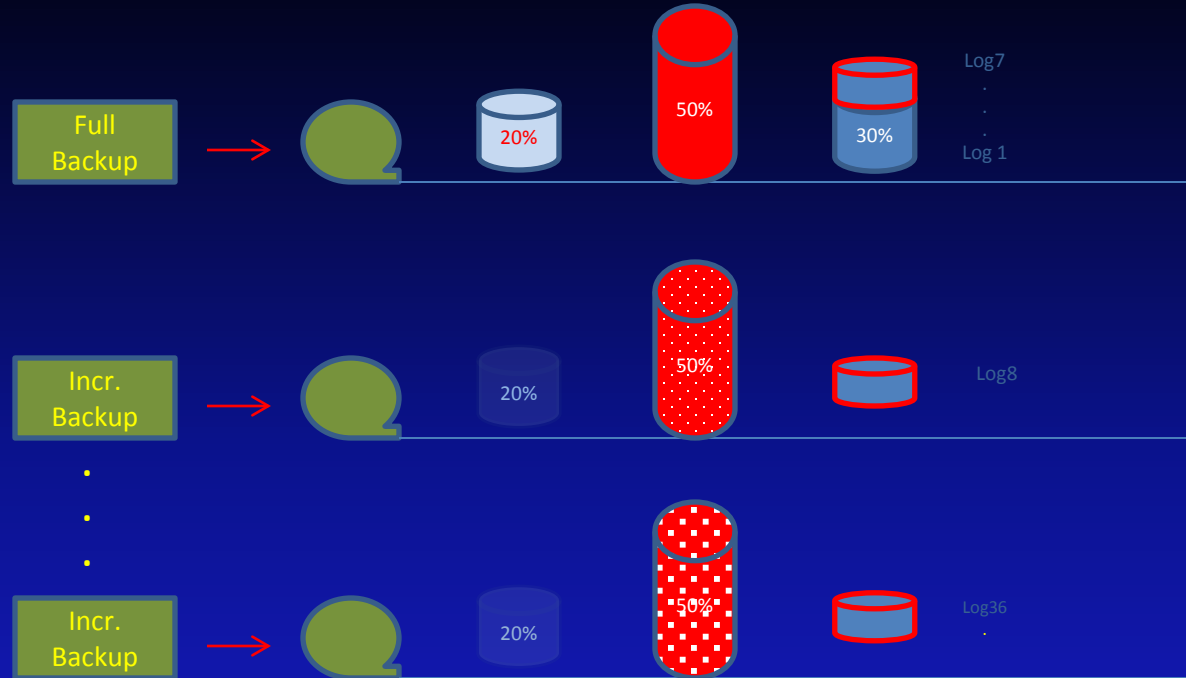
Full+ 29 incremental backup



Day	Original Data(GB)	Compressed data(GB)
1.00	1,000	333.33
2.00	543	181
3.00	543	181
30.00	543	181
Total	16,747	5,582

Disk space is reduce to 55.8% but full volume/subvolume restore of any specific day need a up to 30 restore job (average case will need 15 iterations)

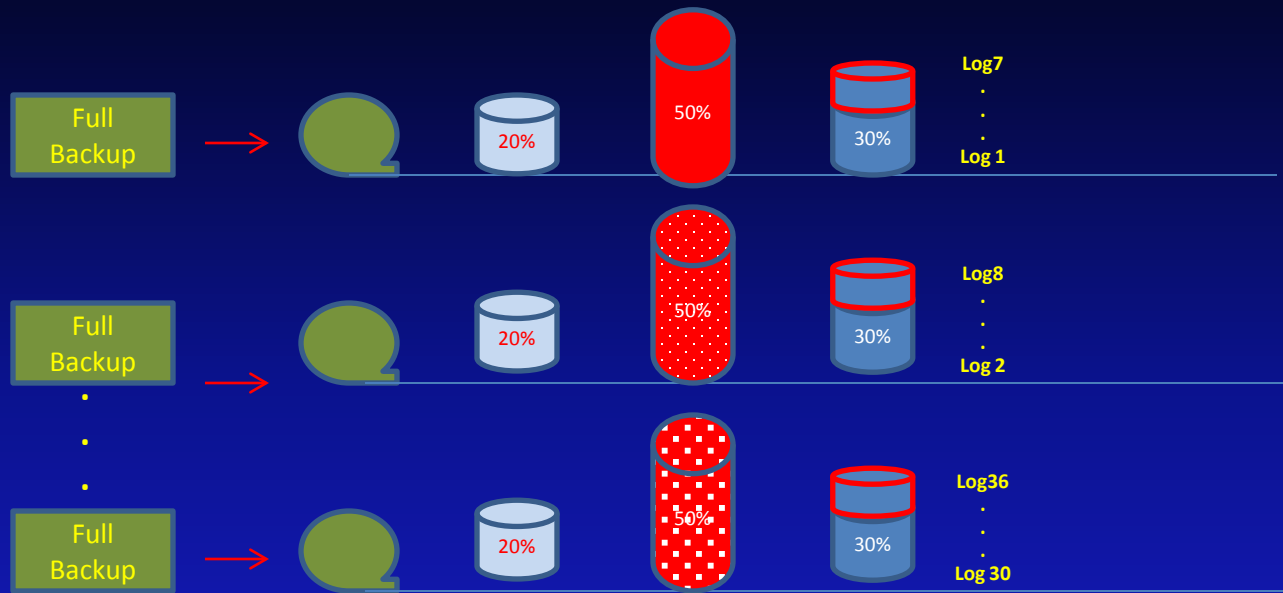
Deduplication - Full+ 29 incremental backup



Day	Original Data(GB)	Compressed data(GB)	With deduplication(GB)
1.00	1,000	333.33	300
2.00	543	181	16
3.00	543	181	16
...
30.00	543	181	16
Total	16,747	5,582	764

With Deduplication storage the disk space is reduce to 7.64% , full volume/subvolume restore of any specific day need a up to 30 restore job (average case will need 15 iterations)

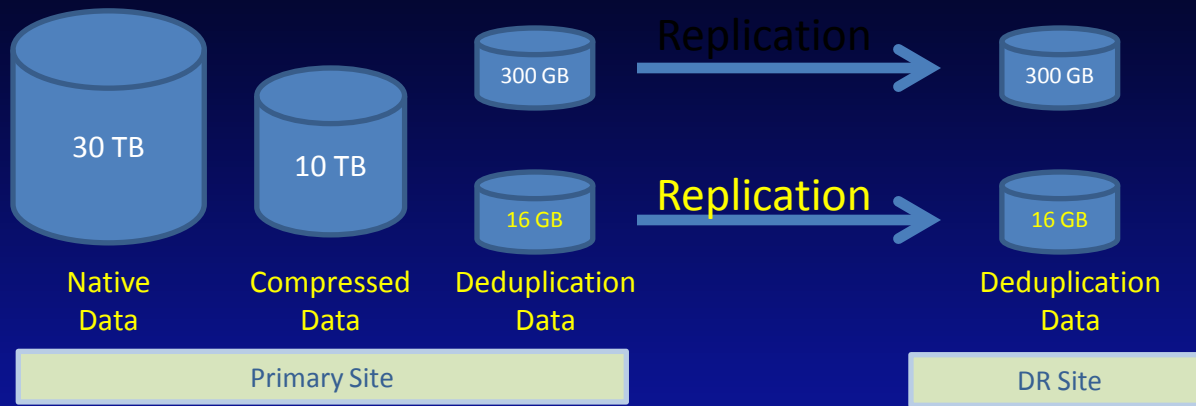
Deduplication - Full backup every day



Day	Original Data(GB)	Compressed data(GB)	With deduplication(GB)
1.00	1,000	333.33	300
2.00	1,000	333.33	16
3.00	1,000	333.33	16
30.00	1,000	333.33	16
Total	30,000	10,000.00	764

With Deduplication storage the disk space is also reduce to 7.64% and any volume/subvolume restore will need a single restore iteration

Deduplication and offsite replication



	Data(bytes)	compression rate	Compressed data(bytes)	Transmission time (in hrs)			
				T1	T3	OC1	OC3
Initial	1,000,000,000,000	3	333,333,333,333	617.3	20.7	17.9	6.0
Subsequent	1,000,000,000,000	3	333,333,333,333	617.3	20.7	17.9	6.0

	Data(bytes)	Dedup rate	Compressed data(bytes)	Transmission time (in hrs)			
				T1	T3	OC1	OC3
Initial	1,000,000,000,000	-	300,000,000,000	555.6	18.6	16.1	5.4
Subsequent	1,000,000,000,000	62.5	16,000,000,000	29.6	1.0	0.9	0.3

Full+ 29 incremental backup

Other impact

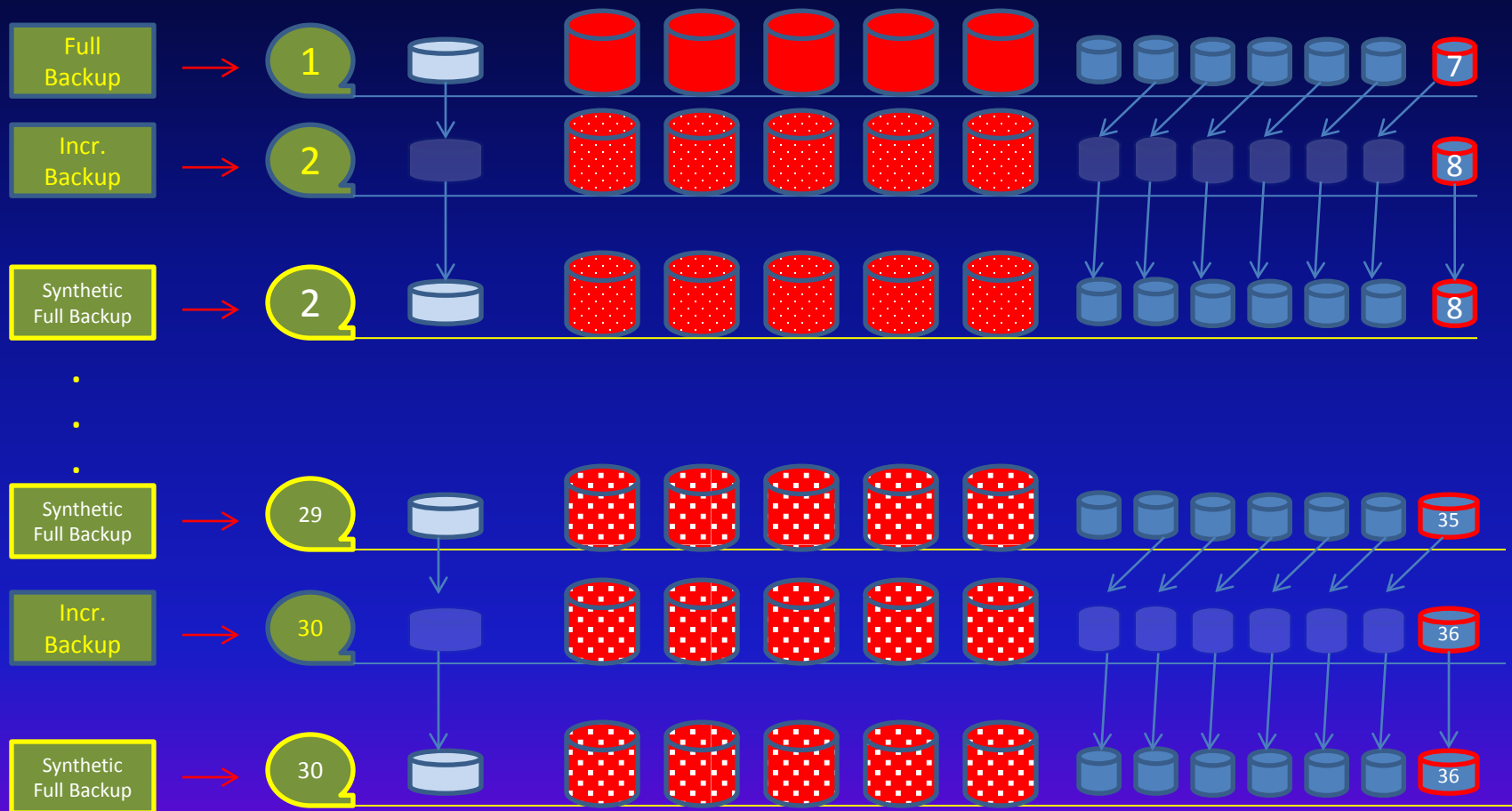
Even if with deduplication, we don't save more disk space using incremental backup than using full Backup. Incremental approach will save more than 43%:

- CPU usage
- Nonstop Disk I/o
- Windows Disk I/O
- Traffic on FC or SCSI
- Network traffic

Incremental approach, will reduce the Nonstop Backup time window

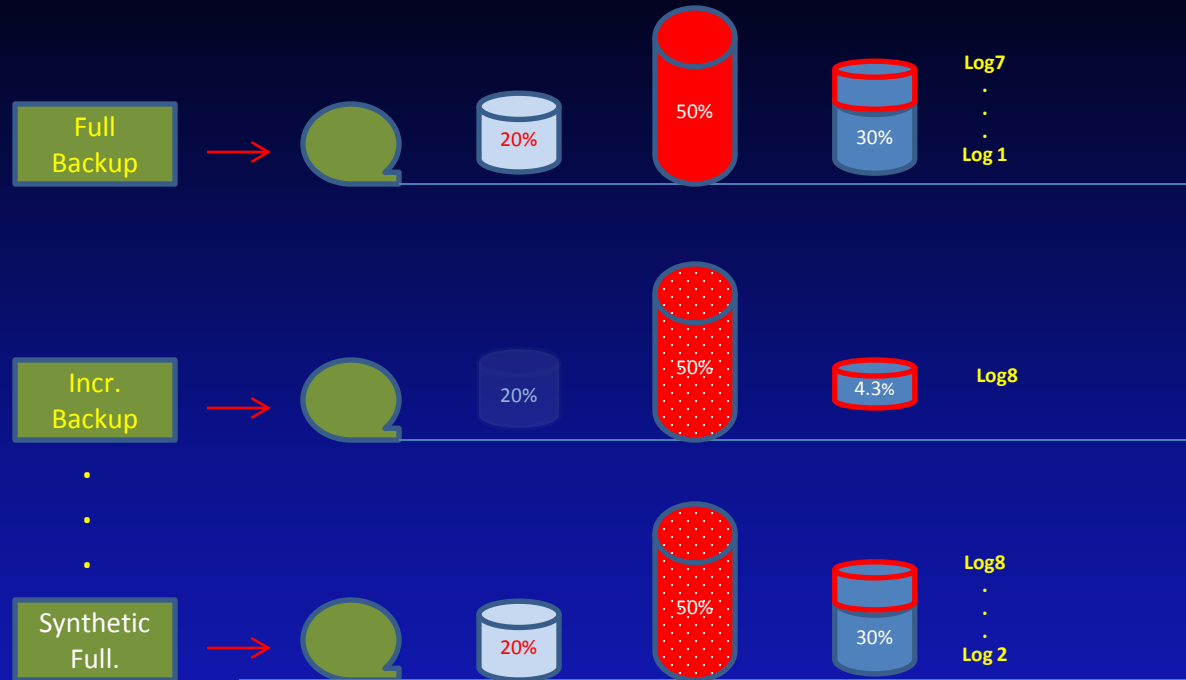
Full+ 29 incremental backup

+ 29 synthetic full backup (lab experimentation)



Save CPU cycle with storage with dedup, doesn't use more space ,no complexity for restore
Best of both world !

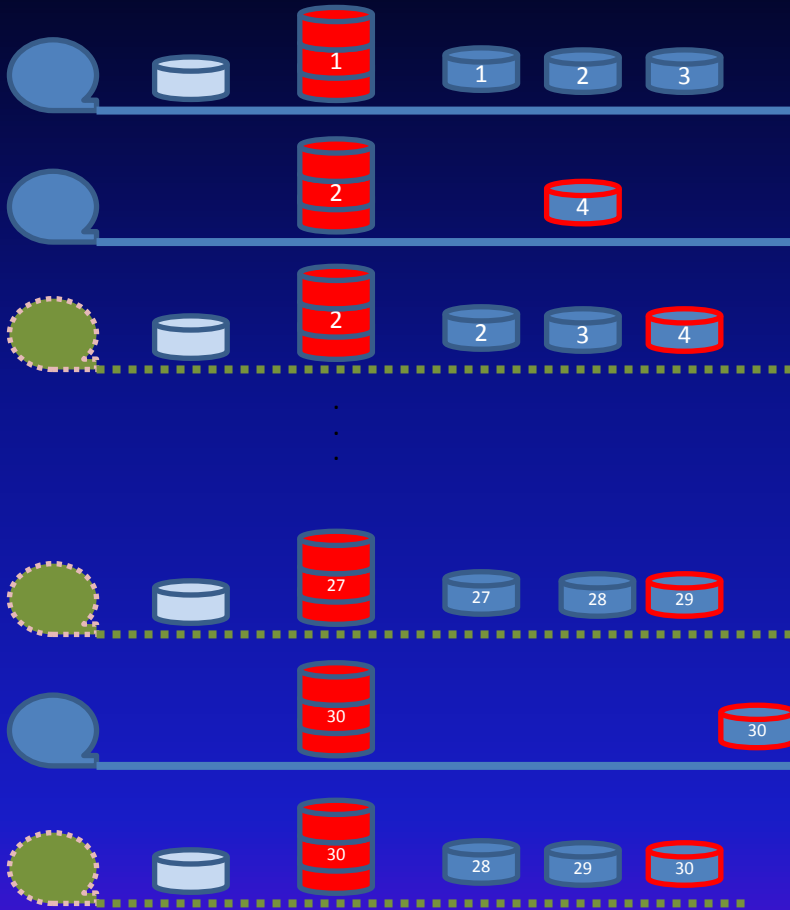
Full+ 29 incremental backup+29 Synthetics



Day	Original Data(GB)	Compressed data(GB)	With deduplication(GB)
1.00	1,000	333.33	300
2.00	1,543	181	16
3.00	1,543	181	16
30.00	1,543	181	16
Total	44,747	14,915	764

With deduplication 3 Tapevolumes per day doesn't take more space

Another lab experimentation



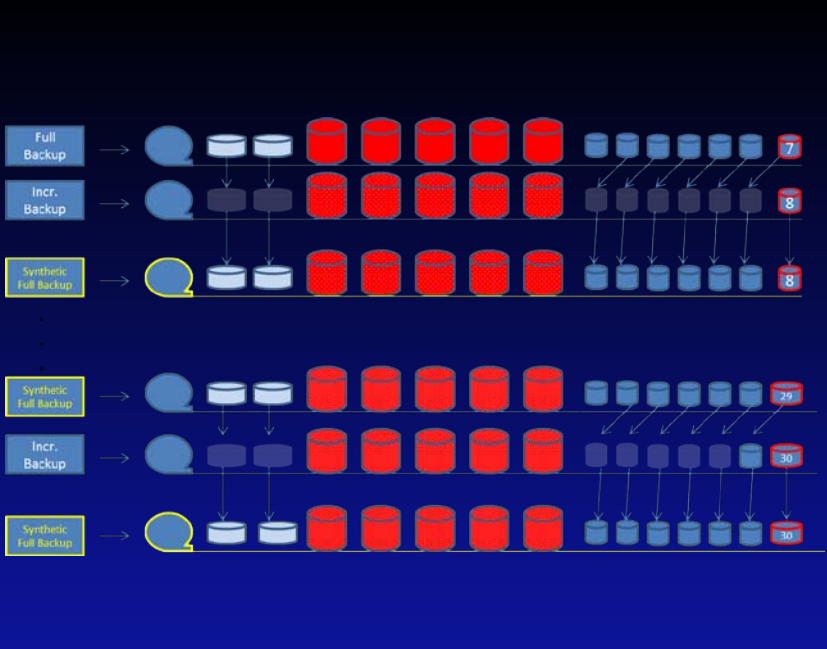
Selection of useful object with file deduplication
Eventually block Deduplication



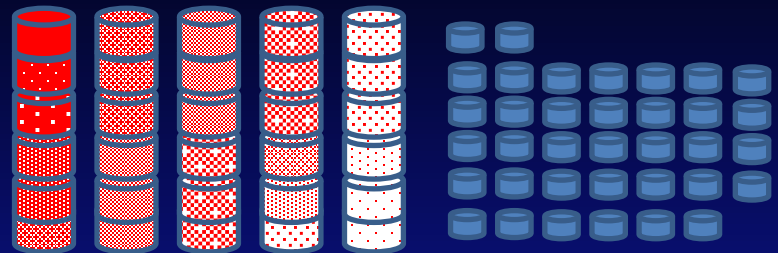
Beginning End

Synthetic Archived





60 GB + 208 GB + 496 GB = .76TB

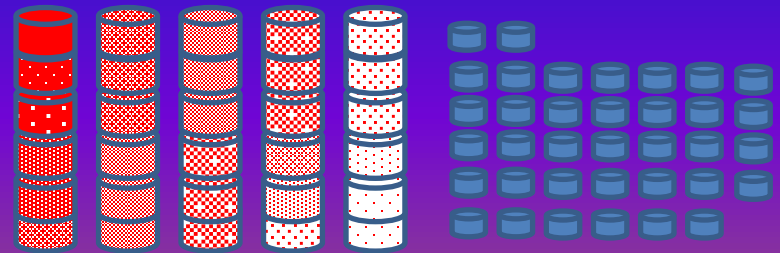


15TB + 1.5 GB = 16.5TB
(Dedup ratio = 22X)

LTO-1	LTO-2	LTO-3	LTO-4	LTO-5	LTO-6	LTO-7	LTO-8
Release Date	2000	2003	2005	2007	2012	TBA ^[6]	TBA
<u>Native</u> Data Capacity	100 GB	200 GB	400 GB	800 GB	1.5 TB ^[7]	2.5 TB ^[8]	6.4 TB ^[6]

Synthetic Archive With no Re-hydration

254 GB + 496 GB = 750 TB



In another ~~word~~ number

	Storage used in GB		
	Uncompress	Compress	Compress+Dedup
Gen0	1000.0	333.3	300.0
Gen 0+x	1000.0	333.3	13.3

Assuming 4% of change at bloc level

So with 1 TB of storage we can keep

- 1 generation if uncompressed

- 3 generations if dedup and compressed

- 54 generations if compressed and dedup

Keeping 7 days, compression&dedup ratio is : 18X

Keeping 30 days, compression&dedup ratio is : 43X

The quiz

Our Nonstop system daily backup are split into two job:

- First job is doing \$SYSTEM.*.* backup, represents **11468** files for a total of **42.8 GB**
- Second job is doing \$DSMSCM.*.* backup, represents **9815** files for a total of **36.6 GB**
- The **daily total** for system backup is **79.4 GB**

Question how many days of backup can fit into that card or that 64 GB USB key?



Hints

- System Backup are compressed 4.9 times
- System Backup are compressed 3.1 times
- Both daily backup fit into 20.5 GB after compression
- Daily incremental is 4.3 GB (5.3% of a full backup)



Win one of those 2,
64GB flash media



- Let your business card with your best guess have the best answer and will win !