

IBM Spectrum Protect Ver 8.1

At CS Department

UCY

July 2017

By Maria Tsiolakki

IBM Spectrum Protect at Computer Science Department

We currently running TSM ver 6.3 on Computer Science Department. We intent to upgrade TSM now known as IBM Spectrum Protect (referred as IBM SP) to the latest version, which is ver 8.1. Our intension is to implement Multisite Disk Solution, since the traditional server-tape solution is not efficient any more.

In Multisite Disk Solution, data that is backed up from the primary site, Site A, is replicated to a server at the secondary, disaster recovery site, Site B. If a disaster occurs at Site A, such as failure of the server, you can use the server at Site B to recover systems. Alternatively, you can use the server at Site A to restore primary storage pool data at Site B, such as after a disk storage failure at Site B.

The new software recommends to have the software installed on server and storage pools to be located on disks, especially if you make use of the new storage pool type, named as Container Storage Pools. Client nodes we contact the server through schedules for the daily backup.

Solution 1. = server + tape library,

Go traditionally. Meaning that we install the IBM Spectrum v8.1 on cs10329 (this is a new server bought by department on November 2016), attached a new autoloader library, that will be purchased next year. Tapes will be used for offsite storage only, since the backup pools will be on disks. This means that :

1. we will not implement Storage Container, and we will not have space savings (See note at the end of this paper for various notes regarding new terminology)
2. We will have additional cost for the tapes
3. We may need :

1-2 extra 4TB disks for storage pools, = $2 \times 4 \text{ TB} \times 600 \text{ euro} = 1200$

3 x 1 TB for the database = $3 \times 1 \text{ TB} \times 245 = 735 \text{ euro}$

2 x 16 GB extra memory = $2 \times 16 \text{ GB} \times 390 \text{ euro} = 720 \text{ euro}$

Total costs for 2017 = $1200+735+720 = 2655 \text{ euro}$ (all prices are list prices)

4. **Extra cost : Cost of a new autoloader, : ~6000 euro + VAT, for the next year, 2018**

- Cost of HP LTO-6 Ultrium 6.25TB RW Data Cartridge (C7976A) = 40 euro each
- HP LTO-7 Ultrium 15TB RW Data Cartridge (C7977A) = 120 euro each

Option 1 – LTO6	Option 2 – LTO7
573-L2U TS3100 Tape Library Express	573-L2U TS3100 Tape Library Express
5502 2.0M Mini-SAS/Mini-SAS 1x Cable	5502 2.0M Mini-SAS/Mini-SAS 1x Cable
7002 Rack Mount Kit	7002 Rack Mount Kit
8347 Ultrium 6 Half-High SAS Drive	8447 Ultrium 7 Half-High SAS Drive
9825 2.8m Power Cord 250V UK	9825 2.8m Power Cord 250V UK
9848 Rack to PDU Line Cord	9848 Rack to PDU Line Cord
Price : 5.700 Euro + VAT	Price : 5.950 Euro + VAT

Solution 2 = Multisite Disk Solution Guide

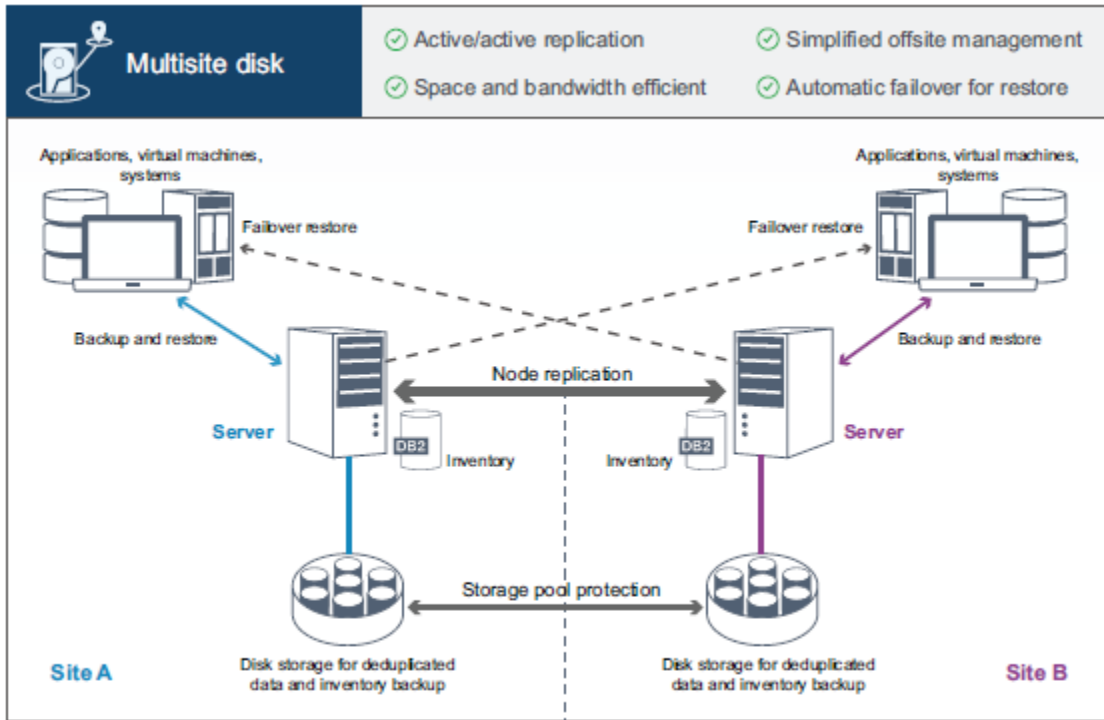


Figure 1. Multisite disk solution

With this solution, we will have an IBM Spectrum Protect server and all clients will backup to disk storage pools on the primary server. A second server, located at 148 room, will act as failover server, where as, all data from the primary server, will be replicated every night to the second server. In case of failure of the primary server, all clients will contact the second server, for their backup/restore tasks. There is also an option to have some clients to backup to primary server, and some clients to backup to the failover server, and then the two server will replicate each other the newly data. We may decide later for this issue. In addition we may add an autoloader library, to the second server, as a further step to have offsite tapes, but this is a further step, not necessary. Also we will need extra budget for this.

<p>Upgrades needed for cs10329 , for Phase-1 Option 1: (budget of 2017)</p> <p>3 x 1 TB@7200 RMS to store the database and logs = $3 \times 1 \text{ TB} \times 245 = 735 \text{ euro}$ 2 x 16 GB extra memory = $2 \times 16 \text{ GB} \times 390 \text{ euro} = 720 \text{ euro}$ (all prices are list prices)</p> <p>Total cost : 1455 euro – We may face performance problems because of slow disk</p>	<p>Upgrades needed for cs10329 : Option 2: : (budget of 2017)</p> <p>3 x 480 GB HPE 480GB SATA 6G MU SFF SC DS SSD = $3 \times 870 = 2610$ 2 x 16 GB Ram = $2 \times 16 \text{ GB} \times 390 \text{ euro} = 720 \text{ euro}$ Total cost : 3330 euro (all prices are list prices)</p> <p>Disks recommended by Antonis- Newcytech</p>
---	---

Phase 1 (by the end of 2017)

IBM Spectrum Protect (formerly TSM) will be installed at Computer Science Department during the Phase-1. The installation will be done on a new machine (cs10329), which is an HP DL180.

We will install IBM Spectrum Protect ver 8.1, which is the latest version. This version deploys some new features and enhancements, which among others are the following :

1. New type of storage pools, named as Storage Container Pools
2. Data Deduplication that reduce the space needed for storage pool
3. Client compression, that reduce the network traffic
4. Node replication with secondary IBM Spectrum Protect Server
5. Tsm scheduler or CAD

Key Points for the implementation :

1. We noticed that some data are valuable, like /home/faculty, /WebData/. We can extend the data retention period, so as when a file is deleted on the home directory, the latest version of the file on IBM SP, to be kept from 90 days, to 365 days.
2. Based on the above we can extend the retention period for /home/student, from 90 days to 120 days
3. The same for /Mail, instead of 90 days, we can increase the retention period to 150 days

OR

Try to keep different schedules for daily,weekly,monthly,yearly schedules, for nodes, like, pallene,kallimachos, =>
Needs investigation, since that means we will have huge data volume to be backed up, at the end of the week,month, and the year

Operation Center

In order to administer the server, apart from command line interface, there is IBM Spectrum Protect Operations Center Version 8.1.x. We will install on Platon since this is a light application, and meets the specs of platon

IBM Spectrum Protect Database Backup

We need to decide whether the backup of the database will be kept. To keep them locally on the server is not recommended , so either we provide space through ISCSI, or through nfs mount point

Phase 2(by the end of 2018)

During the second phase, we will get a new server, install IBM Spectrum Protect, and replicate data from the Primary server to this server.

Cost for second server, (budget of 2018) – Phase 2

Based on prices of Server, bought on 2016

Server (with E5-26xx, 16 GB Ram, 4 x 4 TB Hard disk) : 3200 euro

Additional: 2 x 16 GB x 130 euro : 260

Hard Drive 480 GB SSD Enterprise x 3 x 480 :1440

Hard Drive 4 TB x 2 x 325 :650

5550 + VAT

We may need extra disks for the Primary server, but this case will be reviewed in 2018

CSTSM and TS3200 Autoloader Library

The machine currently running CSTSM server (cs5294), has been acquired by Department on 2007 (through tender ΥΠΣ 081/06). The server has only 6 places for hard disks. The server can not be used as secondary server for the Multisite Disk Solution, because of its limited hard disk capacity. Make use of iscsi storage and cstsm acting as iscsi initiator, would be a solution, but this needs further investigation.

The backup device was purchased on 2009 and it is in good condition. We may format the server, install TSM ver 7, and redirect backup from pcs nodes. This means an extra cost for us, because we may have to administer one more server, BUT....

Machine : cs10329

Model: HP DL180

Size of IBM Spectrum Database : Small. (Daily amount of new data to backup < 6 TB per day, as per IBM classification of IBM Spectrum Database)

Estimated size of the database : < 1 TB,(the size of the database can not be estimated exactly)

Estimated size of archive logs : < 100 GB

Estimated size of active logs : < 100 GB

OS Supported : Centos 7.x is supported

Hardware Requirements

Minimum Recommend by IBM	Owned by CS Department	Needed To buy - Cost
CPU Requirements : 12 processor cores, 1.9 GHz or faster	Intel Xeon E5-2623v4 (2.6GHz/4-core/10MB/85W)	Although our server is below recommendation our daily backup is < 80 GB daily, so the server will not be overloaded
Hard disk for Database: 1. IBM recommends SSD hard disk. 2. Separate the server database, the active log, the archive log, and disk storage pools from each other.	6 x 4 TB SATA HDD.	We suggest to buy 3 HDD of 1 TB or 3 SSD of 480 GB. We will apply Raid-5 With the above suggestion we will have 2 TB, of useful space, which is more than enough for the database and logs
Storage for Storage Pool : Total size of our backup is 20 TB for the primary pools and 20 TB for the offsite pool. We will apply data deduplication, which will allow us to save space, even if we have increase of the amount of the data that is backup	6 x 4 TB SATA HDD.	We will apply Raid-5 , so we end-up with 5 x 4 TB= 20 TB total space for storage pool, which is enough for our needs
Memory : 1. IBM recommends 64 GB minimum, but 128 GB is recommended	2 x 16 GB	2 x 16 GB Ram x 390 euro = 720 euro

Key features of IBM Spectrum Protect

1. Container Storage Pool
2. Data Deduplication
3. Data compression

Data Deduplication

More info here :

<https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Tivoli+Storage+Manager/page/Deduplication+FAQ>

Key points :

When should I consider using TSM deduplication?

Consider using TSM deduplication when the following conditions apply:

- You plan to use a disk-only backup solution (your primary backup storage pool will remain on disk).
- Your priority is to reduce the amount of disk storage required for backup data.
- You have a limited bandwidth connection from clients to the TSM server. In this case client-side deduplication is an appropriate solution.
- You are considering using TSM node replication (available since TSM 6.3).
- Your TSM database is properly sized for deduplication and resides on a high performing disk array (see additional FAQs for examples). Although not required, SSD (Solid State Disk) is recommended for the TSM database.

Q6) How does TSM deduplication affect backup and restore performance?

1. The use of deduplication and compression can result in longer client backup elapsed times compared to backups to a disk storage pool that is not deduplicated. However when the backup network is constrained, backup elapsed times can be faster when using client-side deduplication.
2. The impacts to backup elapsed time can often be mitigated by increasing the number of parallel backup sessions.
3. Restore throughput from a deduplicated storage pool is generally slower when compared to restore from a disk based storage pool that is not deduplicated. However, when compared to restore performance from physical tape, restore from a disk-based deduplicated storage pool can be much faster.

Q8) How do I decide between using TSM's server-side or client-side deduplication?

2. Here are some circumstances when you should consider using client-side deduplication:
 - You wish to distribute the workload across client systems rather than perform deduplication processing in the TSM server.
 - Bandwidth between the client and server is constrained.
3. Here are some circumstances when you should consider using server-side deduplication:
 - CPU resources on the client host system are inadequate to support the additional processing required by client-side deduplication during scheduled backup processing.
- 4.

(Q10) Which TSM features and options are incompatible or not supported with deduplication?

1. Client-side encryption is incompatible with TSM deduplication. However TSM deduplication can be used together with SSL (encryption of data in flight) or encryption by the storage device.
2. LAN-free backup
3. Simultaneous write
4. Subfile backup
5. Client side compression should not be used with *server-side* deduplication (since compressed objects do not deduplicate well). However, client-side compression used in conjunction with *client-side* deduplication can provide an effective means to further reduce storage pool data.

Configuration best practices

(https://www.ibm.com/support/knowledgecenter/SSEQVQ_8.1.0/perf/t_optim_config.html)

Use fast disks for the server database. Enterprise-grade solid-state disks (SSD), with Fibre Channel or SAS interface, offer the best performance.	Use fast, low-latency disks for the database. Using SSD is essential if you are using data deduplication and node replication. Avoid Serial Advanced Technology Attachment (SATA) and Parallel Advanced Technology Attachment (PATA) disks.
Ensure that the server system has enough memory.	Heavier workloads require more than the minimum requirements. Advanced features such as data deduplication and node replication can require more than the minimum memory that is specified in the system requirements document.
Separate the server database, the active log, the archive log, and disk storage pools from each other.	Keep all IBM Spectrum Protect storage resources on separate disks. Keep storage pool disks separate from the disks for the server database and logs. Storage pool operations can interfere with database operations when both are on the same disks. Ideally, the server database and logs are also separated from each other. For details and more tips, see the following topics:
Use at least four directories for the server database. For larger servers or servers that use advanced features, use eight directories.	Place each directory on a LUN that is isolated from other LUNs and from other applications.

	A server is considered to be large if its database is larger than 2 TB or is expected to grow to that size. Use eight directories for such servers.
If you are using data deduplication, node replication, or both, follow the guidelines for database configuration and other items.	Configure the server database according to the guidelines, because the database is extremely important to how well the server runs when these features are being used.

Monitoring and maintaining the environment for performance (https://www.ibm.com/support/knowledgecenter/SSEQVQ_8.1.0/perf/t_mon_perf.html)

Linux: Active and archive log space recommendation (https://www.ibm.com/support/knowledgecenter/SSEQVQ_8.1.0/srv.install/c_srv_inst_active_log_space-linux.html)

When you estimate space requirements for active and archive logs, include some extra space for contingencies such as occasional heavy workloads and failovers.

In IBM Spectrum Protect™ servers V7.1 and later, the active log can be a maximum size of 512 GB. The archive log size is limited to the size of the file system that it is installed on.

Use the following general guidelines when you estimate the size of the active log:

- The suggested starting size for the active log is 16 GB.
- Ensure that the active log is at least large enough for the amount of concurrent activity that the server typically handles. As a precaution, try to anticipate the largest amount of work that the server manages at one time. Provision the active log with extra space that can be used if needed. Consider using 20% of extra space.
- Monitor used and available active log space. Adjust the size of the active log as needed, depending upon factors such as client activity and the level of server operations.
- Ensure that the directory that holds the active log is as large as, or larger than, the size of the active log. A directory that is larger than the active log can accommodate failovers, if they occur.
- Ensure that the file system that contains the active log directory has at least 8 GB of free space for temporary log movement requirements.

The suggested starting size for the archive log is 48 GB.

The archive log directory must be large enough to contain the log files that are generated since the previous full backup. For example, if you perform a full backup of the database every day, the archive log directory

must be large enough to hold the log files for all the client activity that occurs during 24 hours. To recover space, the server deletes obsolete archive log files after a full backup of the database. If the archive log directory becomes full and a directory for archive failover logs does not exist, log files remain in the active log directory. This condition can cause the active log directory to fill up and stop the server. When the server restarts, some of the existing active-log space is released.

After the server is installed, you can monitor archive log utilization and the space in the archive log directory. If the space in the archive log directory fills up, it can cause the following problems:

- The server is unable to perform full database backups. Investigate and resolve this problem.
- Other applications write to the archive log directory, exhausting the space that is required by the archive log. Do not share archive log space with other applications including other IBM Spectrum Protect servers. Ensure that each server has a separate storage location that is owned and managed by that specific server.

Directory Container Storage Pool

Q1. What is a container type storage pool? Are there different types of container storage pools?

A container storage pool is a new type of storage pool that is designed specifically for data deduplication. Data is stored in containers and is either deduplicated at the source (client) or inline during the server ingest phase. By using directory-container storage pools, you do not need to post-process the data to deduplicate the data. There are two types of container storage pools:

- Cloud-container: Cloud based storage pools. For the FAQs for the cloud-container storage pool, see [Cloud-container storage pools FAQs](#).
- Directory-container: Directory based storage pools that defined by using the DEFINE STGPOOLDIRECTORY command.

Q2. How are DIRECTORY type containers different from DISK or SEQUENTIAL type volumes?

Directory-containers as the best of both DISK and SEQUENTIAL worlds. Directory-containers are hybrids of the two techniques that allow for advantages of both without some of the disadvantages. Directory-containers are managed like sequential type file volumes but don't have the disadvantage of reclamation. Data extents are stored in a random fashion like DISK volumes but don't have the disadvantage of a static footprint and lack of variable sizing.

Directory-container storage pools

A storage pool that the server uses to store data in containers in storage pool directories. Data that is stored in a directory-container storage pool can use either inline data deduplication, client-side data deduplication, inline compression, or client-side compression. Inline data deduplication or inline compression reduces data at the time it is stored. By using directory-container storage pools, you remove the need for volume reclamation, which improves server performance and reduces the cost of storage hardware. You can protect and repair data in directory-container storage pools at the level of the storage pool.

Restriction

You cannot use any of the following functions with directory-container storage pools:

- Migration
- Reclamation
- Aggregation
- Collocation
- Simultaneous-write
- Storage pool backup
- Virtual volumes