

Modernizing Data Protection with IBM Spectrum Protect

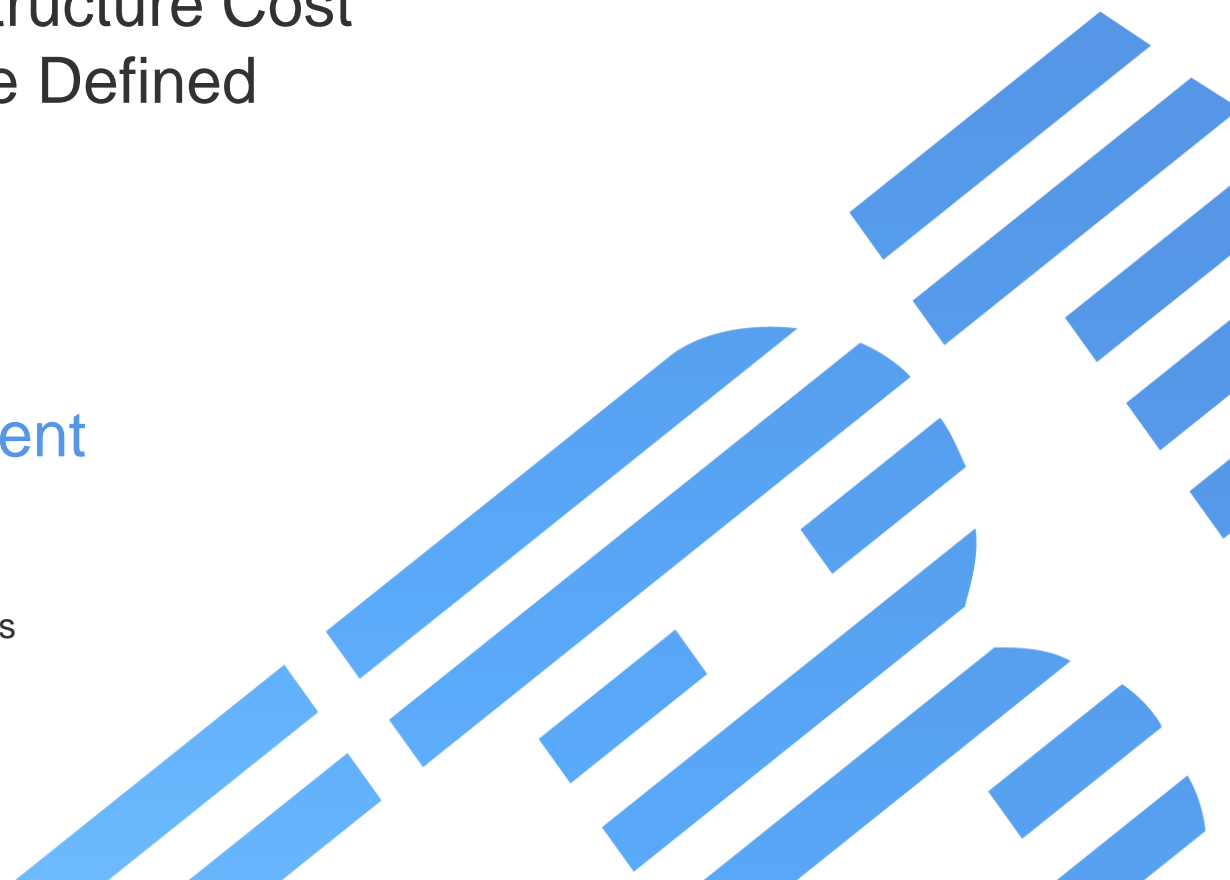


¡Converse con un consultor!

Reduce Storage Infrastructure Cost by Leveraging Software Defined Deduplication

IBM Spectrum Protect
Restore, Reuse, Reinvent

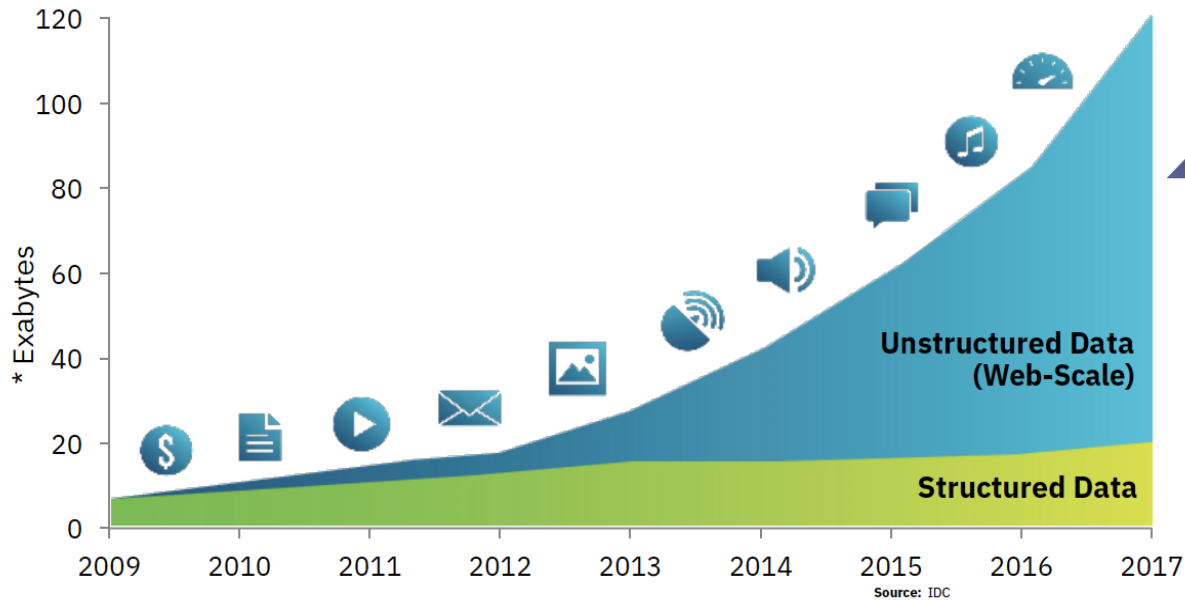
Ing. Gastón Llahi
Software Defined Storage Technical Sales
IBM Systems Storage Software
Expert Certified IT Specialist
llahi@ar.ibm.com



Clients face explosive growth in Unstructured Data



- *Unstructured data: Mobile applications, Internet of Things, Social Media, Analytics*
- *Problem: Legacy Storage Designed for Systems of Record lack the scale and cost we need*



Unstructured data growth of 60–80% per year creates Web-Scale storage needs

At 80% growth rate, we need a 44% cost decrease in Storage to maintain flat budgets

- *Structured Data grows too, the need there is for dramatic improvements in performance (latency) and cost*

Client compression

- Files compressed by client before transmission
- Conserves network bandwidth and server storage



Client

Incremental forever

- After initial backup, file is not backed up again unless it changes
- Conserves network bandwidth and server storage

Client Side Deduplication

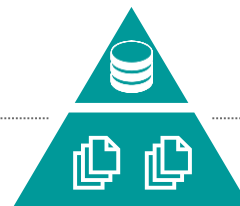
- Reduces network traffic by deduplication data before transfer
- Sends to server's dedupe pool



Server

Device compression

- Compression performed by storage hardware
- Conserves server storage



Storage Hierarchy

Appliance deduplication

- Deduplication performed by storage appliance (VTL or NAS)
- Conserves server storage

Server Side Deduplication

- Duplicate data only stored once
- **Out of band** dedupe to disk pool
- **In-line** dedupe to directory or cloud pools (RECOMMENDED method)



Server side compression

- Files compressed by server after in-line dedupe for containers
- Conserves network bandwidth when writing to cloud

Deduplication is an integrated part of Spectrum Protect

Comparisons of Data Reduction Methods



	Client Compression	Server-side Compression	Incremental Forever	Server-side out-of-band deduplication	Server-side in-line deduplication	Client-side deduplication
How data reduction is achieved	Client compresses files	Server compress data extents	Client only sends changed files	Server eliminates redundant data chunks post-ingest	Server eliminates redundant data extents on ingest	Client and server eliminate redundant data chunks
Conserves network bandwidth?	Yes	No	Yes	No	No	Yes
Data Supported	Backup, Archive, HSM, API	Backup, Archive, HSM, API	Backup	Backup, Archive, HSM, API	Backup, Archive, HSM, API	Backup, Archive, API
Scope of data reduction	Redundant data within same file on client node	Redundant data within same extent	Files that do not change between backups	Redundant data from any files in storage pool	Redundant data from any extent in container	Redundant data from any files storage pool
Avoids storing identical files renamed, copied, or relocated on client node?	No	No	No	Yes	Yes	Yes
Removes redundant data for files from different client nodes?	No	No	No	Yes	Yes	Yes
Storagepool supported	All	Containers	All	File device class	Containers	Containers and File device class
Version supported in	Available prior to v5	7.1.3+	Available prior to v5	6.1+	7.1.3+	6.2+

All of these data reduction methods conserve storage pool space

Single Spectrum Protect Server - Blueprints



Built and tested in-house, using the specified HW and configuration

Active-Active replication-based data protection

Simplify deployment by providing “building blocks” for success

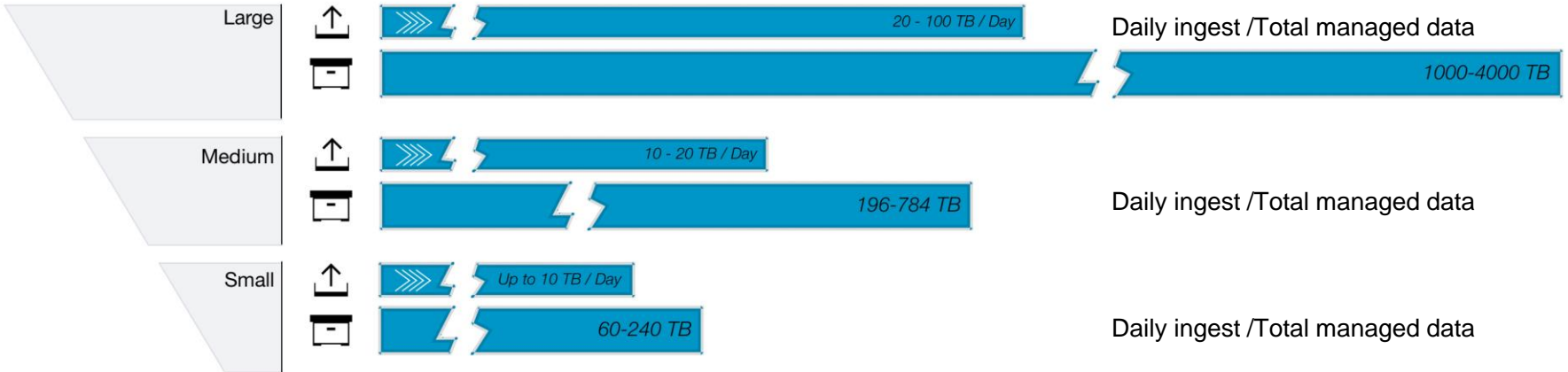
⬆️ Daily backup (ingest) workloads representing real-world clients and data

🗄️ Database and storage pool capacity to support significant amounts of deduplicated and compressed data

Using available Storwize and Elastic Storage Server (ESS) depending upon platform and size

Intel (x86) Based for Linux and Windows

Power Based for AIX and Linux



on developerWorks:

<http://ibm.biz/IBMSpectrumProtectBlueprints>

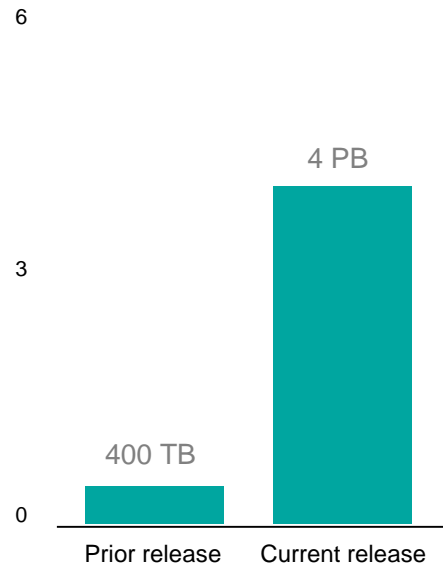


Reduce Costs: IBM Spectrum Protect Deduplication



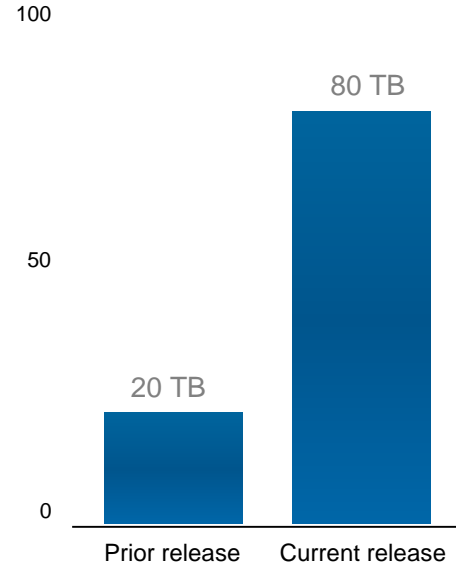
Capacity

Data protected per backup server (PB)



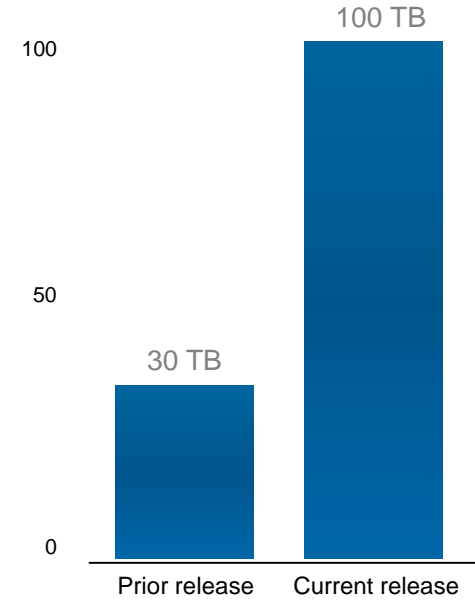
Server Dedupe

Daily workload per backup server (TB)*



Client Dedupe

Daily workload per backup server (TB)*



* IBM internal tests. Using Spectrum Protect version 7.1.3 Includes ingest, deduplication, replication and maintenance

Intelligent data protection in a software defined world

"IBM Spectrum Protect allows us to keep pace with and manage data growth seamlessly and in ways we've never been able to before."

Piedmont Healthcare



Each Spectrum Protect Server Capable of protecting up to 4 PB total managed data...

Using up to 1 PB storage for deduplicated & compressed data

Using up to 6 TB catalog to record and manage the protected data



What

Protect data for databases, mail servers, hypervisors, file servers, and workstations

Protecting what runs your enterprise...

Oracle — SAP Hana — VMware
 Microsoft — SQL — Exchange — Windows — Hyper-V
 DB2 — Domino — Unix — Linux



How

Efficient data reduction utilizing snapshots, progressive incremental, deduplication, and compression

Flexibility for today and your tomorrows...

Policy-Based — HW & SW Snapshots — Hot-Warm-Cold DR
 D2D — D2C — D2D2T — D2T — D2D2C
 HW & SW Replication — WAN Optimization — On/Off Premises Cloud



Where

Multi-site data availability with active-active replication-based architecture and heterogeneous storage flexibility using disk, tape, or cloud.

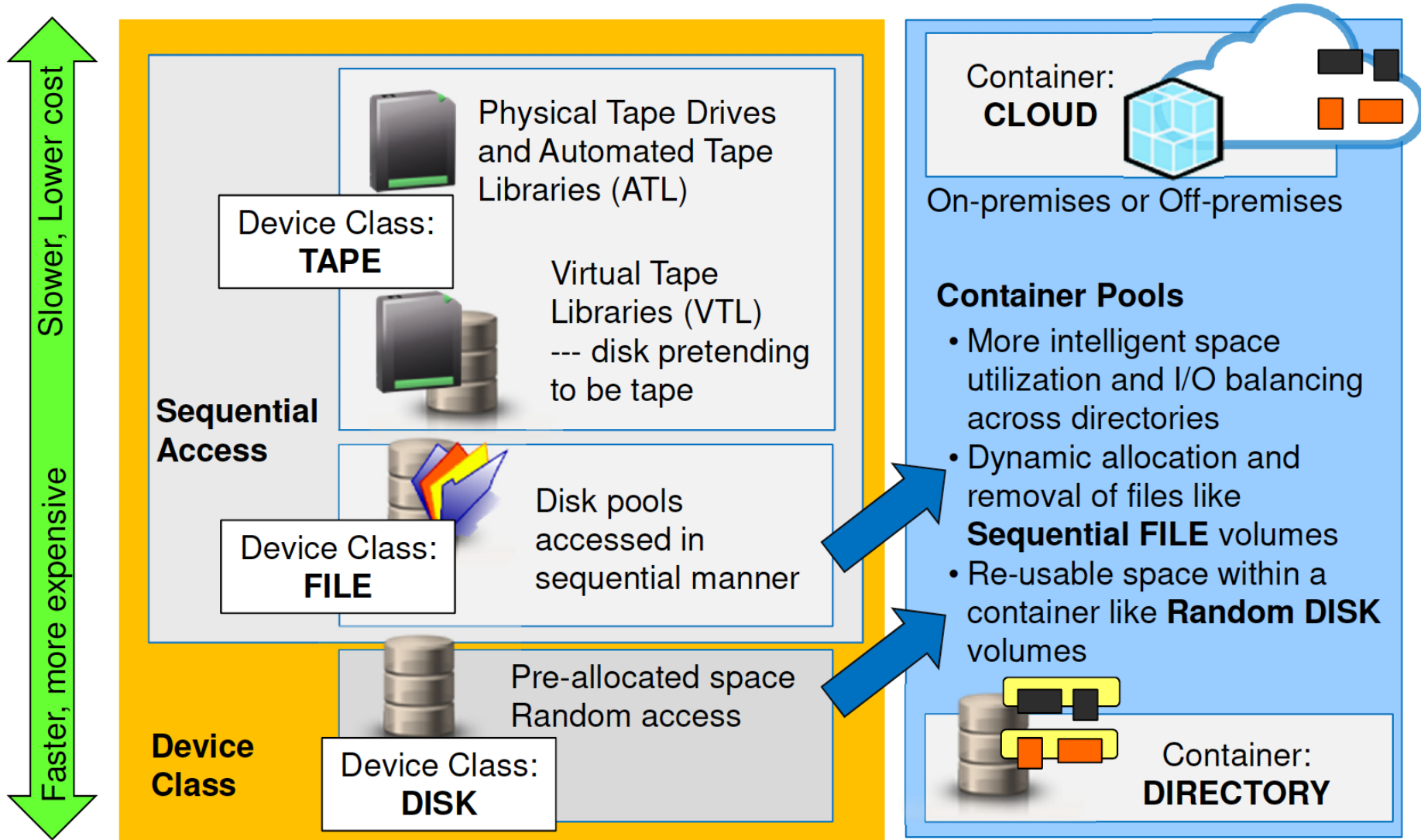
Store Anywhere: Ours, Yours, Theirs...

Storage — IBM — EMC — Hitachi — HP — Oracle — Others...
 Clouds — Softlayer — IBM Cloud Object Storage — Amazon AWS S3
 Deduplicated — Compressed — Encrypted/Secure

D2D and D2C based on container storage pool technology, optimized for deduplication

D2T, D2D2T, and other data protection strategies possible using volume-based storage pools

IBM Spectrum Protect – Device Classes and Containers



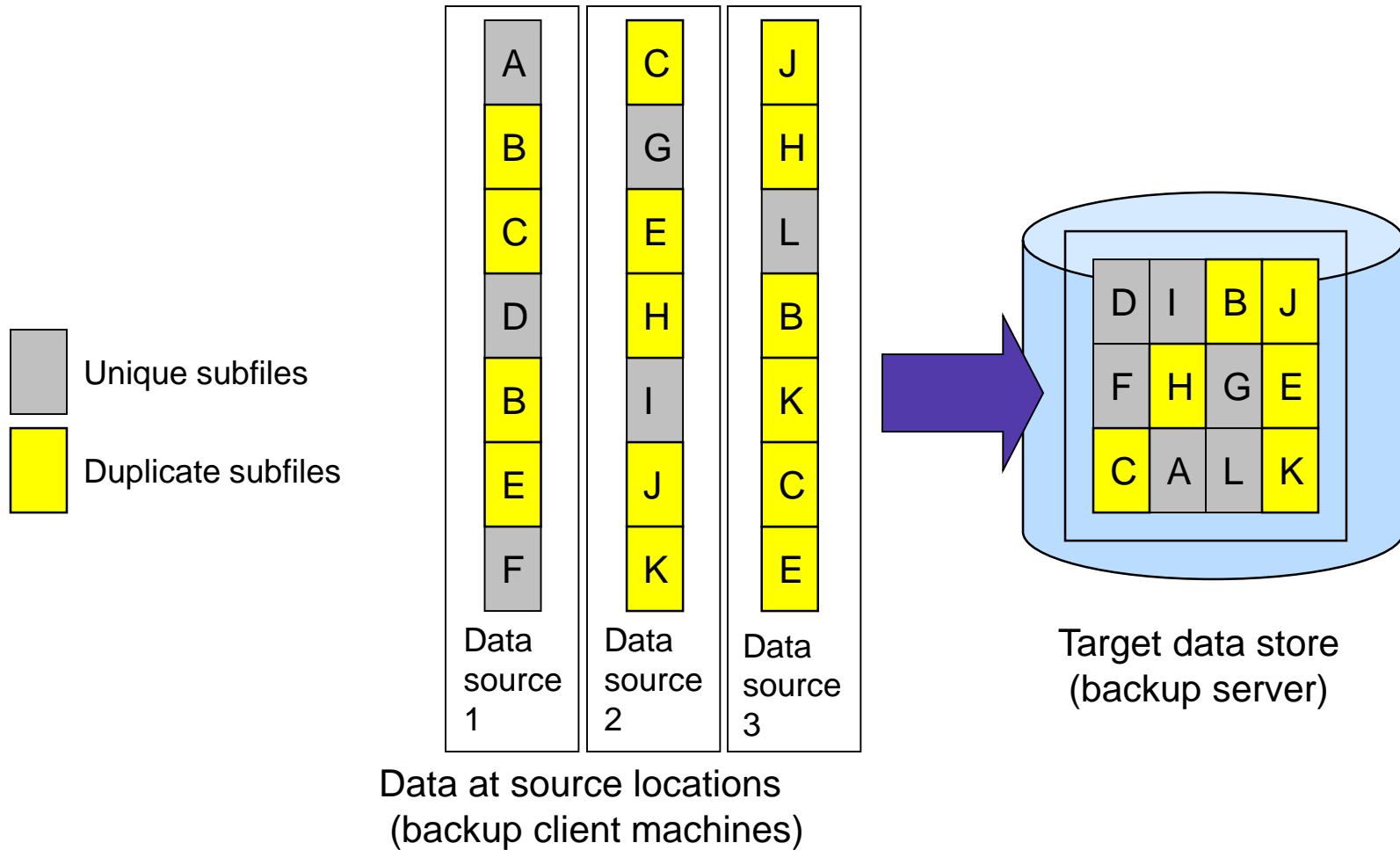
Container Storage Pool Deduplication Compares Favorably to Hardware-based Solutions

- Container storage pool **software-defined data deduplication** compares favorably to hardware-based solutions such as VTL and NAS-based systems
 - Data deduplication is included with Spectrum Protect (**no additional license**)
 - **No dedicated hardware** to buy
 - Allows a single Spectrum Protect server to **scale** to 4 PB of protected data

Use Spectrum Protect Container Storage Pool Deduplication when...	Use Hardware-based Deduplication when...
Managing up to 4 PB of protected data with a single instance	Managing >4 PB of protected data with a single instance or >100 TB daily ingest
Client-side deduplication is needed to minimize network utilization	Deduplicating NDMP data or LAN-free with SAN environments
Sufficient CPU/memory resources exist on to support deduplication (and compression)	Deduplicating across >1 Spectrum Protect server instances

** Can use both in same environment!*

GENERAL DEDUPLICATION CONCEPTS

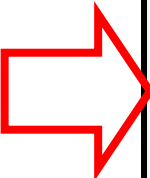
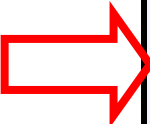


Potential advantages


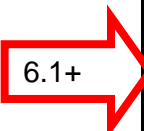
- Reduced storage capacity required for a given amount of data
- Ability to store significantly more data on given amount of disk
- Restore from disk rather than tape may improve ability to meet recovery time objective (RTO)
- Network bandwidth savings (some implementations)
- Lower storage-management and energy costs resulting from reduced storage requirements

Potential tradeoffs/limitations

- Significant CPU and I/O resources required for deduplication processing
- Deduplication might not be compatible with client-side encryption
- Increased sensitivity to media failure as many files could be affected by loss of common chunk
- Deduplication may not be suitable for data on tape because increased fragmentation of data could greatly increase access time

Approach	Advantages	Disadvantages
 Source-side (client-side) Deduplication performed at the data source (e.g., by a backup client), before transfer to target location	<ul style="list-style-type: none">▪ Deduplication before transmission conserves network bandwidth▪ Awareness of data usage and format may allow more effective data reduction▪ Processing at the source may facilitate scale-out	<ul style="list-style-type: none">▪ Deduplication consumes CPU cycles on the file/application server
 Target-side (server-side) Deduplication performed at the target (e.g. by backup software or storage appliance)	<ul style="list-style-type: none">▪ No deployment of client software at endpoints▪ Possible use of direct comparison to confirm duplicates	<ul style="list-style-type: none">▪ Deduplication consumes CPU cycles on the target server or storage device▪ Data may be discarded after being transmitted to the target

Note: Source-side and target-side deduplication are not mutually exclusive

Approach	Advantages	Disadvantages
 In-band Deduplication performed during data transfer from source to target <u>RECOMMENDED</u>	<ul style="list-style-type: none">▪ Immediate data reduction, minimizing disk storage requirement▪ No post-processing	<ul style="list-style-type: none">▪ More CPU and memory required
 Out-of-band Deduplication performed after data ingestion at the target	<ul style="list-style-type: none">▪ No impact to data ingestion▪ Potential for deduplication of legacy data▪ Possibility for parallel data deduplication processing	<ul style="list-style-type: none">▪ Data must be processed twice (during ingestion and subsequent deduplication)▪ Storage needed to retain data until deduplication occurs

Note: In-band and out-of-band deduplication are mutually exclusive in Spectrum Protect

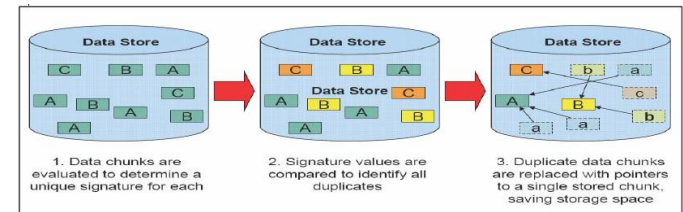
- Used to indicate reduction achieved by deduplication
- If deduplication reduces 500 TB of data to 100 TB, ratio is 5:1
- Ratios reflect design tradeoffs involving performance and reduction
- Actual compression ratios will be highly dependent on other variables
 - Data from each source: redundancy, change rate, retention
 - Number of data sources and redundancy of data among those sources
 - Backup methodology: incremental forever, full+incremental, full+differential
 - Whether data encryption occurs prior to deduplication
- In addition to above variables, some vendors include data reduction achieved by incremental backup and conventional compression
- Deduplication vendors claim ratios in the range 2:1 to 500:1
- Meaningful comparison of ratios is extremely problematic—beware of hype!

SERVER SIDE **OUT-OF-BAND** DEDUPLICATION

OUT-OF-BAND

- 40% increase* in the amount of data available for recovery from disk (compared to TSM 5.5)
 - Adds to the benefits of progressive incremental data capture & data compression
- Effective with data from any source
 - API, backup, HSM, archive, legacy
- Built into Tivoli Storage Manager 6 – *not an add-on solution*
 - No need to update client side SW or API applications
 - No dependency on specific hardware Server-side
 - No required change to clients or legacy data in storage pools
 - V6 client will change to identify meta data vs content
 - Not required, but this will improve the de-dupe process on the server
 - Duplicates across clients can be identified
 - Works with client-side dedupe too
- Out-of-band process**
 - Background process
 - Requires post-backup dedupe and reclamation

Identify DUPLICATES
- Supported with **disk stgpools using devclass=file**
- Uses SHA-1 & MD5 hashing algorithm



THIS IS NOT THE CURRENT RECOMMENDED METHOD OF DEDUPE

* Results will vary depending on user's environment, systems/processors, type of data, expiration policies, data change rates, etc.

SERVER SIDE **IN-LINE** DEDUPLICATION

- Efficient, server-side deduplication method that specifically uses containers (directory or cloud)
- Occurs during client backup or replication operations
- Data is automatically deduplicated as it is ingested into the container
 - No identify duplicates processing required
 - No reclamation or other data movement required to remove duplicate data
- Works in conjunction with Client-Side deduplication
 - Processes data that is not already deduplicated on the client
 - Uses the same data deduplication algorithms as client-side data deduplication and existing server-side data deduplication - hash-based, variable chunk size algorithms.
 - Client-side data deduplication matches data extents that are identified by inline data deduplication

Best practice is to utilize **in-line** dedupe to container stgpools

Extents

7.1.3

- What extents are:
 - Pieces of data stored in a storage pool
 - Can be composed of metadata or file data
 - Can be deduplicate-able or not
 - Can vary in size – Sliding scale window optimized during ingest

- What extents are *not*:
 - Whole objects unto themselves
 - Objects require a list of extents for reconstruction

- What extents do:
 - The means for deduplicating and storing data
 - Allows for easy matching of data – already stored? Discard it!

- What is a non-deduplicated extent?
 - Data TSM can't deduplicate on the server is a non-deduplicated extent

- What is the size of an extent?
 - 50KB is the minimum extent size, 256KB is the target average, 4MB is the upper size
 - For files which are smaller the 50KB, a single extent is used which matches the file size
 - Files smaller than 2KB are not deduplicated
 - Size of writes to stgpools is still 256KB
 - For writes to the DB, the size varies from 4KB to 32KB depending on the table but tuning to 4KB is fine

Enable Server-side **in-line** data de-duplication

- On the Client set:
 - DEDUPLICATION **No**
 - Optional: RESOURCEUTILIZATION >1
- On the Server set:
 - Server-side dedupe is part of container storagepools
 - DEFINE STGPOOL <pool name> **STGTYPE=CONTAINER**
 - UPDATE COPYGROUP <domain_name> <policy_set_name> <class_name> **DESTINATION=<container stgpool>**

Enable Client-side dedupe to work together with server side de-duplication

- Works with **in-line** or **out-of-band** dedupe
- Utilization is optional
 - On the Server set: Register node dedupe=clientorserver or Update node Dedup=clientorserver
 - On the client set: update dsm.opt or dsm.sys with **deduplication yes**

Storage Pool Level

- How efficient deduplication is working on a broad storage pool level
- **QUERY STGPOOL <pool name> FORMAT=DETAILED** (Column: “Deduplicate Data Not Stored”)
- **SELECT STGPOOL_NAME,SPACE_SAVED_MB FROM STGPOOLS**
- Operations Center %saving

Type	Name	Server	Status	Capacity Used	Device Class	Container Type	% Savings
Primary	DEDUPPOOL1	FUSION	Critical	12.7 TB / 616.0 GB	FILEDEV	—	—
Container	DEDUPCONTAINERPOOL	TESLA	Critical	4.6 TB / 12.2 TB	—	Directory	Bytes saved 44.1 TB Ratio 11:1
Primary	DEDUPPOOL1	ION	Critical	5.8 TB / 175.0 GB	FILEDEV	—	—
Container	DEDUPCONTAINERPOOL	DAYTONA	Critical	4.5 TB / 25.5 TB	—	Directory	91
Container	DEDUPPOOL	TAPSRV10	Critical	972.6 TB / 74.6 TB	—	Directory	75
Container	DEDUPPOOL0	TUMBLEWEED	Critical	17.8 TB / 35.0 TB	—	Directory	38

Node Level

- How efficient deduplication is working on data provided by a single node
- **GENERATE DEDUPSTATS**
- **QUERY DEDUPSTATS**
- **DELETE DEDUPSTATS**
 - Records are kept indefinitely
 - Each record is one row in a new table
 - Can use Delete to remove one of the records

```

SERVER1> generate dedupstats dedupcontainer zeus
ANR0984I Process 42 for GENERATE DEDUPSTATS started in the BACKGROUND at 15:32:47.

SERVER1> query dedupstats
Date/Time      Storage Pool Name  Node Name  Filespace Name  FSID  Dedup Pct  Total Data Protected (MB)
-----
2015-09-17, 15:32:47  DEDUPCONTAINER    ZEUS       \\server1\c6    1     49.947     200

SERVER1> query dedupstats f=d
Date/Time: 2015-09-17, 15:32:47
Storage Pool Name: DEDUPCONTAINER
Node Name: ZEUS
Filespace Name: \\server1\c6
FSID: 1
Deduplication Percentage: 49.947
Total Data Protected (MB): 200
Total Space Used (MB): 100
Non-Deduplicated Extent Count: 172
Non-Deduplicated Extent Space Used: 124,838
Unique Extent Count: 0
Unique Extent Space Used: 0
Shared Extent Count: 774
Shared Extent Data Protected: 209,856,794
Shared Extent Space Used: 104,977,756
    
```

Session Level

- How efficient deduplication worked on an individual session
- **SELECT * FROM SUMMARY**
 - BYTES_PROTECTED: <Bytes that have been protected>
 - BYTES_WRITTEN: <Actual bytes sent from the client>
 - DEDUP_SAVINGS: <Savings from deduplication processing>
 - COMP_SAVINGS: <Savings from client-side compression>

7.1.3+ **Query Dedupstats** displays information about data deduplication statistics

8.1.5+ **Query Dedupstats** produces a summary report of statistics for a specified set of nodes, node groups, and filespaces

- **Node_name** or **node_group_name** and **filesystem_name** or **FSID** display a statistics records for concerned nodes and filespaces only
- **REPortid** can be specified to display a statistics records for concerned **REPortid** only
- ★ **DESCRiption** displays a statistics records for concerned **DESCRiption** only
- ★ **SUMmary** format will obtain a summary report of statistics for a specified set of nodes, node groups, and file spaces which the statistics are generated together. Internally, it will group the statistics records by **REPortid**, and display a summary report for each **REPortid**
- ★ In addition, user can specify a combination of nodes, filespaces and reportId to generate a summary report from a sub set of statistics records with a certain **REPortid**.

```
QUERY DEDUPSTATS dir test *d$ f=d
```

```

Date/Time: 03/01/2018 23:37:33
Storage Pool Name: DIR
Node Name: TEST
Filespace Name: \\win-cu484t0qgu5\d$
FSID: 2
Type: Bkup
Total Data Protected (MB): 72
Total Space Used (MB): 10
Total Space Saved (MB): 62
Total Saving Percentage: 86.76
Deduplication Savings: 0
Deduplication Percentage: 0.00
Non-Deduplicated Extent Count: 36
Non-Deduplicated Extent Space Used: 28,847
Unique Extent Count: 190
Unique Extent Space Used: 75,455,724
Shared Extent Count: 0
Shared Extent Data Protected: 0
Shared Extent Space Used: 0
Compression Savings: 65,488,492
Compression Percentage: 86.76
Compressed Extent Count: 190
Uncompressed Extent count: 36
Encryption Extent Space Used: 0
Encryption Percentage: 0.00
Encrypted Extent Count: 0
Unencrypted Extent Count: 226
Report ID: 3
Description: myrule
    
```

```
QUERY DEDUPSTATS dir f=sum
```

```

Report ID: 3
Description: myrule
Date/Time: 03/01/2018 23:37:33
Storage Pool Name: DIR
Node Name: *
Filespace Name: *
Type: Bkup
Total Data Protected (MB): 640
Total Space Used (MB): 232
Total Space Saved (MB): 407
Total Saving Percentage: 63.66
Deduplication Savings: 0
Deduplication Percentage: 0.00
Non-Deduplicated Extent Count: 63
Non-Deduplicated Extent Space Used: 47,093
Unique Extent Count: 1,999
Unique Extent Space Used: 670,558,523
Shared Extent Count: 0
Shared Extent Data Protected: 0
Shared Extent Space Used: 0
Compression Savings: 426,919,490
Compression Percentage: 63.66
Compressed Extent Count: 1,999
Uncompressed Extent count: 63
Encryption Extent Space Used: 0
Encryption Percentage: 0.00
Encrypted Extent Count: 0
Unencrypted Extent Count: 2,062
    
```

CLIENT SIDE DEDUPLICATION

works with both **in-line** and **out-of-bound** deduplication

Overall Goal:

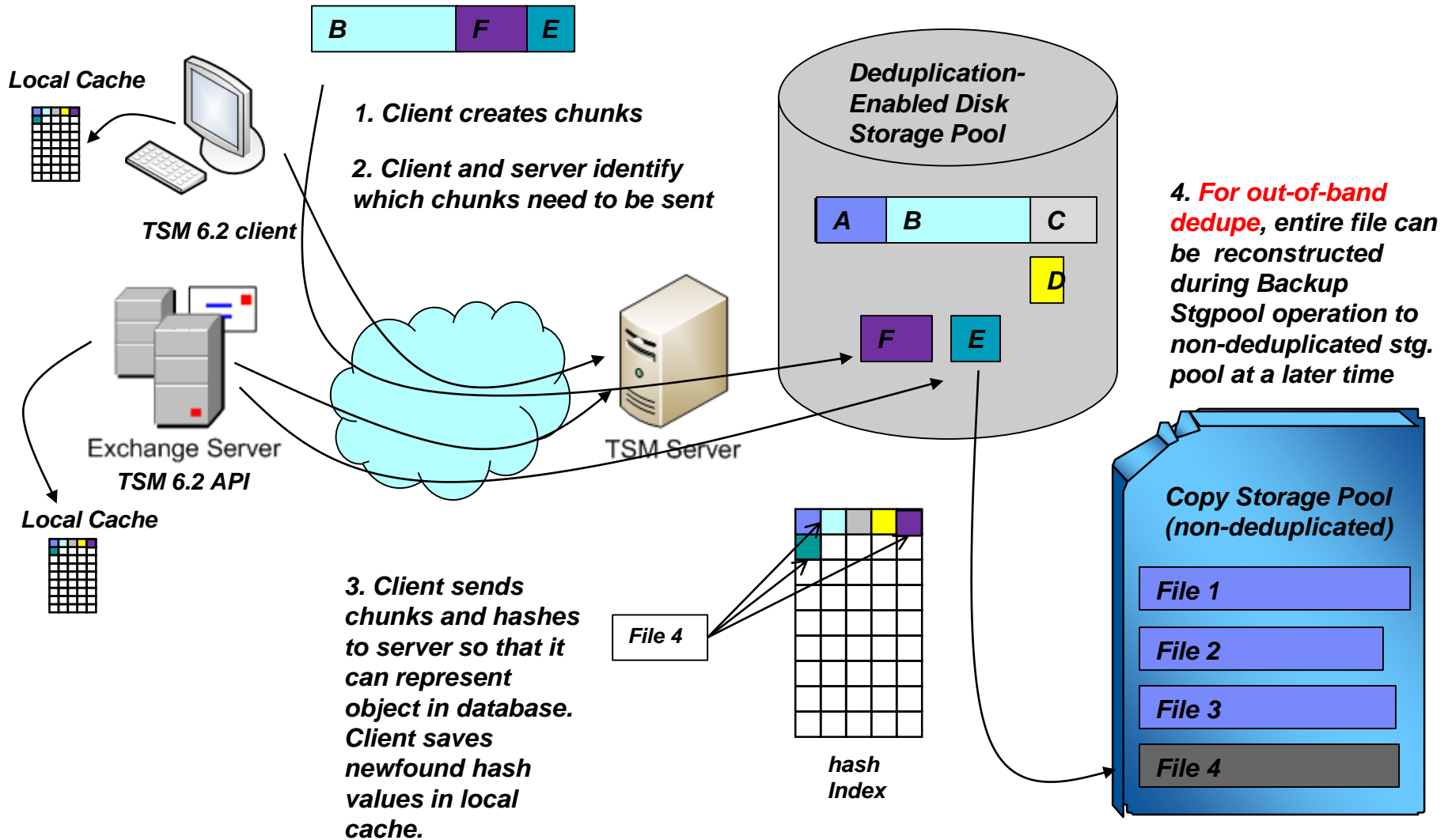
- **Reduce network traffic**
 - Client-side/distributed deduplication
- **While providing:**
 - Optimization of the Deduplication processes
 - Security and data integrity
 - Activity reporting

Implementation:

- **Client-side performs**
 - Data fingerprinting (identification of data extents or chunks)
 - Data digest generation (hashing)
- **Client sends hash and unique chunks to server**
 - Chunks can be compressed before sending

*(API client side dedupe coming in 6.2.x)

Client-Side Data Deduplication - Operation

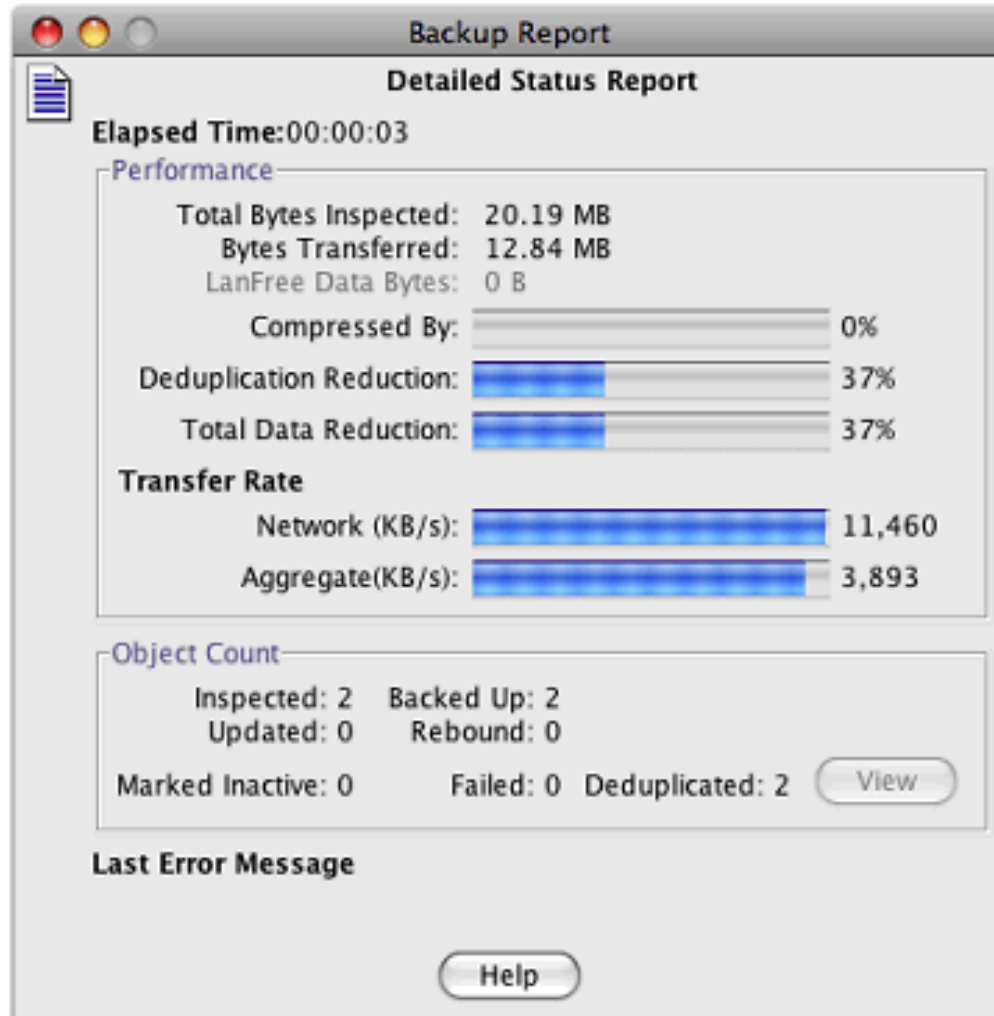


- Client and server must be at version 6.2.0+
- Client must have the client-side deduplication option enabled
- Server must enable the node for client-side deduplication
 - DEDUP=CLIENTORSERVER parameter with REGISTER NODE or UPDATE NODE
- Storage pool destination for the data must have deduplication enabled or be a container storagepool
- File must not be excluded from client side deduplication processing
 - By default all files are included
 - See exclude.dedup Client option for details
- File must be larger than 2 KB

NO Client-side dedupe allowed with:

- LAN-free
- Unix HSM
- Encryption
 - Client-side
 - Files from known encrypted file systems which are read in their raw encrypted format
 - SSL encryption is OK
 - Container stgpool server-side encryption is OK
- API Buffer copy elimination
- Simultaneous Storage Pool Write

➤ *If any of these exist, Client-side dedup does not occur*



COMPRESSION WITH DEDUPLICATION

In-line, server-side LZ4 compression for Container Pools (directory & cloud)

- Increases the amount of available space in a container storage pool
- Does not require additional DB space
- New storage pool level option **DEFINE/UPDATE STGPOOL ... COMPression=Yes|No**
 - Defaults to YES when defining new container pool
- Server containers support LZ4 compression
 - LZ4 is a lossless data compression algorithm that is focused on compression and decompression speed
 - LZ4 is preferred over the LZW compression currently used by the client
 - Choose to use server compression over client compression
 - HW real time compression should not be used
 - When compression turned on, will only compress new data being stored in the container
- Data extents examined as they are ingested, IF not suitable for compression will not be compressed
 - ie metadata extents, data previously encrypted or compressed, or if compression leads to a growing chunk length
- **Dedupe, then compression, then encryption**

Container pool compression is applied after deduplication

CPU impact of compression

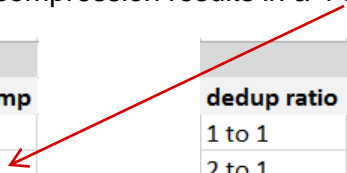
- CPU impact varies depending on how well data responds to deduplication
 - Only data which is not eliminated by deduplication needs to be compressed
 - Largest CPU impact will occur for data which does not deduplicate well
- In the lab, typically seeing an additional 30% CPU cost for using compression
 - Once storage pool I/O becomes a bottleneck, the data savings from compression increase the aggregate backup throughput
- Blueprint systems, have sufficient CPU to perform compression and still reach the published aggregate backup throughput
 - Except Power blueprint where the medium CPU allocation is increasing from 8 to 10 cores
 - Blueprint sizing spreadsheet including CPU calculations & the cost of performing compression, will be available shortly

Overall Space Savings of deduplication + compression

- Combined savings can result in a significant jump in overall storage savings
 - Data with a 2 to 1 dedupe rate and 50% compression results in a 4 to 1 space savings rate

50% compression				
dedup ratio	Dedup %	Comp %	Combined %	dedup+comp
1 to 1	0.0	50.0	50.0	2 to 1
2 to 1	50.0	50.0	75.0	4 to 1
3 to 1	66.6	50.0	83.3	6 to 1
4 to 1	75.0	50.0	87.5	8 to 1
5 to 1	80.0	50.0	90.0	10 to 1
6 to 1	83.3	50.0	91.7	12 to 1
7 to 1	85.7	50.0	92.9	14 to 1
8 to 1	87.5	50.0	93.8	16 to 1
9 to 1	88.8	50.0	94.4	18 to 1
10 to 1	90.0	50.0	95.0	20 to 1

25% compression				
dedup ratio	Dedup %	Comp %	Combined %	dedup+comp
1 to 1	0.0	25.0	25.0	1.4 to 1
2 to 1	50.0	25.0	62.5	2.6 to 1
3 to 1	66.6	25.0	75.0	4 to 1
4 to 1	75.0	25.0	81.3	5.4 to 1
5 to 1	80.0	25.0	85.0	6.8 to 1
6 to 1	83.3	25.0	87.5	8 to 1
7 to 1	85.7	25.0	89.3	9.2 to 1
8 to 1	87.5	25.0	90.6	10 to 1
9 to 1	88.8	25.0	91.6	11.3 to 1
10 to 1	90.0	25.0	92.5	13.4 to 1



- Q stgpool
 - Enhanced to show a compression ratio
- Q DEDUPSTATS
 - Shows compression ratio statistics
 - Compressed data extent count
 - Uncompressed data extent count
 - Compression saving bytes
 - Compression percentage
- SELECT * FROM table STGPOOLS
 - Table field COMP_SPACE_SAVED shows space savings in bytes
- AUDIT CONTAINER
 - No parameter changes, data extents will be de-compressed when they are required for audit

```
Deduplicate Data?: Yes
Processes For Identifying Duplicates:
Compressed: Yes
Deduplication Savings: 2,238 G (90.64%)
Compression Savings: 59,673 M (25.22%)
Total Space Saved: 2,297 G (93.00%)
Auto-copy Mode:
```

```
78MGUI1> q dedupstats
```

Date/Time	Storage Pool Name	Node Name	Filespace Name	FSID	Type	Total Saving Percentage	Total Data Protected (MB)
2016-02-21, 14:11:09	DIRPOOL-1	OPERATIONS.SRVZR.GRANFORP...	\\vmlsvs2\svct	2	Bkup	89.45%	645,913
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-0	/lrgnumfls	4	Bkup	97.634	12,003
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-0	/lrgnumfls	4	Arch	97.62	16,006
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-1	/data	4	Bkup	94.657	77,833
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-1	/data	4	Arch	99.068	116,750
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-1	/lrgnumfls	5	Bkup	97.634	4,001
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-1	/lrgnumfls	5	Arch	97.62	4,002
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-2	/lrgnumfls	4	Bkup	97.634	12,003
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-2	/lrgnumfls	4	Arch	97.62	16,006
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-3	/data	6	Bkup	89.91	39,917
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-3	/lrgnumfls	7	Bkup	97.634	12,003
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-3	/lrgnumfls	7	Arch	97.62	16,006
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-4	/data	6	Bkup	89.805	39,917
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-4	/data	6	Arch	99.103	39,917
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-4	/lrgnumfls	7	Bkup	97.634	12,003
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-4	/lrgnumfls	7	Arch	97.62	12,005
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-5	/data	6	Bkup	89.247	39,917
2016-02-21, 14:11:09	DIRPOOL-1	DIR1-78MGUI1-5	/lrgnumfls	7	Bkup	97.634	12,003

Server-side Compression for in-line dedupe in the Operations Center

7.1.5

Enable, Disable and Visualize Server-side Compression for Container pools

Storage Pools 111 Alerts 999+



Container Critical Primary Critical Copy Normal

Savings	88.1 TB (5:1)
Backed up	109.9 TB
After deduplication	38.4 TB
After compression	21.7 TB

Type	Name	Server	Status	Capacity Used	Device Class	Container Type
Container	DEDUPPOOL	TAPSRV08	Normal	21.7 TB / 23.1 TB	—	Directory

DEDUPPOOL

Normal Container

Usage and Configuration

DEDUPPOOL

21.7 TB / 23.1 TB

Savings 80%

Directories 20

DEDUPPOOL

Normal Container

- Summary
- Properties
- Directories

Savings	
Backed up	109.9 TB
Total bytes saved	88.1 TB
Total reduction percentage	80.24%
Total reduction ratio	5:1

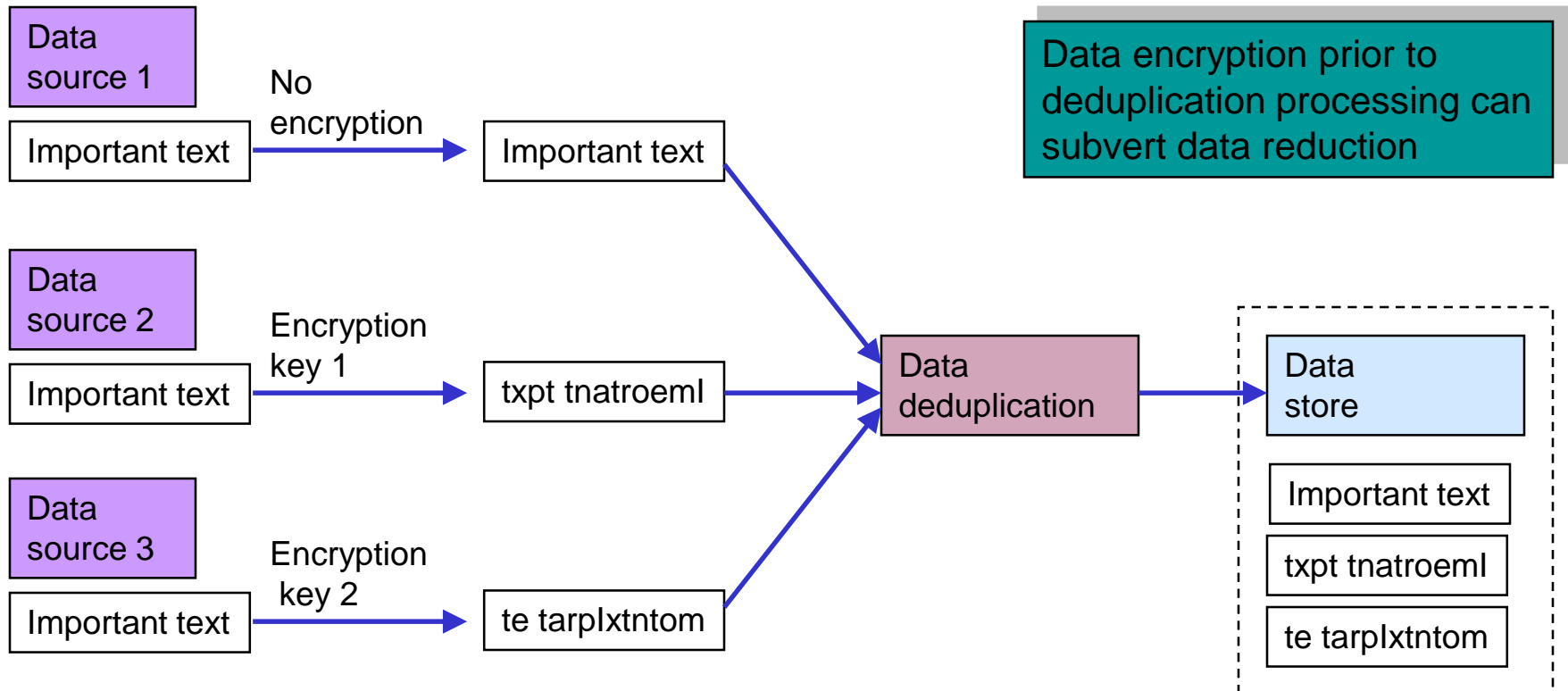
Data Deduplication	
Bytes saved	71.4 TB
Deduplicated	109.9 TB to 38.4 TB
Reduction percentage	65.00%
Reduction ratio	3:1

Compression	
Compression	YES
Bytes saved	16.7 TB
Compressed	38.4 TB to 21.7 TB
Reduction percentage	43.55%
Reduction ratio	2:1

ENCRYPTION WITH DEDUPE

In-line or out-of-band Deduplication and Client-Side Encryption

NOT RECOMMENDED to use CLIENT-side encryption with dedupe!



1. Three data sources have the same text file

2. After encryption, text files do not match

3. Deduplication processing does not detect redundancy

4. Text files are stored without data reduction
NOT DESIRED!

Server Side Encryption for in-line dedupe with Containers

7.1.3

Server encrypts data before it writes it to the directory or cloud container storage pool

- Storage pool encryption was designed to require minimal user interaction – "Set it and forget it"
- Storage Pool Encryption in **cloud containers as of 7.1.3** and **directory containers as of 8.1.2**
 - Administrators want additional protection for data stored outside of their control
 - Encryption is enabled by default for OFF-PREMISES cloud pools, is optional for ON-PREMISES cloud pools
- **Best Practice:** In-line dedupe, then server-side compression, finally server-side encryption
- Encryption is enabled at a storage pool level
 - From cmdline, specify ENCRYPT=YES|NO on DEFINE and UPDATE STGPOOL commands
 - Once encryption is enabled, all new data stored into the pool will be encrypted, existing data is not affected

Node Replication Considerations

Target	8.1.2 (encrypt=yes)	8.1.2 (encrypt=no)	Legacy server
Source			
8.1.2	Encrypted chunk sent as is to target server Non-encrypted chunk is encrypted on target server	Encrypted chunk sent as is to target server Non-encrypted chunk is sent as is to target server.	Server encrypted chunk is decrypted on source server before being sent over Non-encrypted chunk is sent as is to target server
Legacy server	Chunk is encrypted on target server	Chunk is sent as is to target server	

Server Encryption	Client Encryption
Server controls encryption at storage pool level	Client controls encryption at the object level using include.encryption and encryptiontype options
Server manages all of its encryption keys	Client optionally stores keys in tsm.pwd file
Server uses AES encryption with minimum key length of 256-bits	Clients uses AES128, AES256 or DES56 bit
Server signs encrypted data to detect tampering or corruption	Client does not
Data encrypted by the server can be deduplicated	Data encrypted by the client cannot be deduplicated

REPLICATION WITH DEDUPE

Container Stgpool – Protect Stgpool for Replication with in-line dedupe

Protect STGPool type=replserver

- Replicates backup and archive **data** to the target server without the associated metadata
 - Data can be transferred more quickly than with client replication alone
 - Improves replication performance
 - Enables the repair of damaged files on the source server
- For **directory-container** pools, when replication is enabled
- Requires replication schedule to replicate metadata
 - Until node replication runs and the metadata is synchronized, data that was transferred by storage pool protection cannot be restored from the target server
- Should run often enough to keep pace with incoming data
 - Should complete before node replication starts

Limitations on using node replication with a cloud-container storage pool

- Cloud-container storage pool can be a destination pool for target replication server
- Cloud-container storage pool can NOT be a source storage pool
- For redundancy, use replication capabilities available from the cloud storage provider

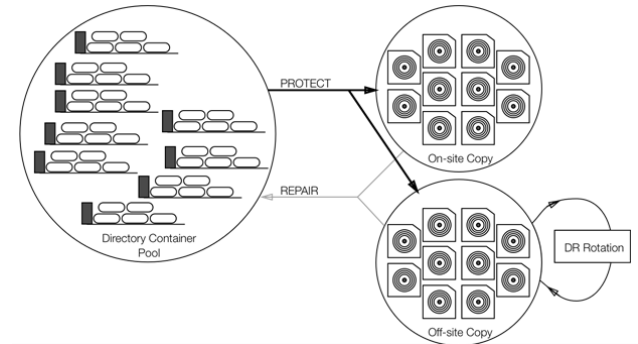
The client protection schedule replicates client data that is stored in DISKCONTAINER on LABSERVER.



Schedule storage pool protection to improve replication performance and enable automatic repair of damaged files on the source server. [Learn more](#)

Storage pool protection schedules	PROTECT_P (1-2)
Protection pool	CONTAINER
Schedules per day	2
Run time per schedule	1 hour and 30 minutes

- As an alternative to using Replication
 - Can copy data directly from directory-container pools to tape for disaster recovery protection
- In addition to using Replication
 - Tape copy provides additional protection

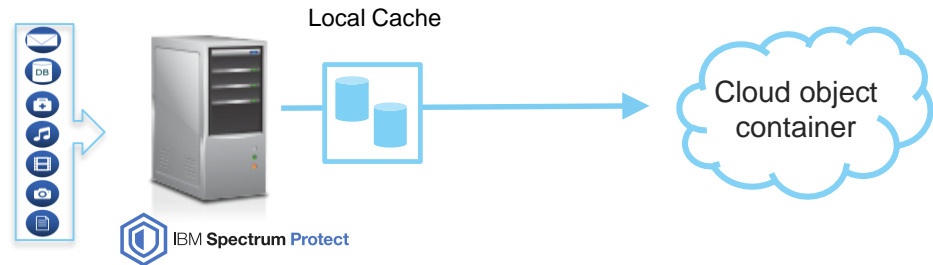


- Replicate directory container stgpools to tape in addition to or instead of another server
 - Uses a new storage pool type: COPYCONTAINER
 - In the documentation COPYCONTAINER is called Container-Copy Pools
 - Copies are created as part of the existing PROTECT STGPOOL command
 - Copies are by extent – just as with the PROTECT STGPOOL replication process
 - Copies are used to repair the directory container
 - Copies can not be accessed directly by the client nodes
 - Copies don't use reclamation processing to maintain tape volume usage
 - Reclamation for copycontainers runs as part of the PROTECT STGPOOL command
- Youtube demo: <https://youtu.be/-sSpRvoh0bk>
- DRM technote: <http://www-01.ibm.com/support/docview.wss?uid=swg27048653>

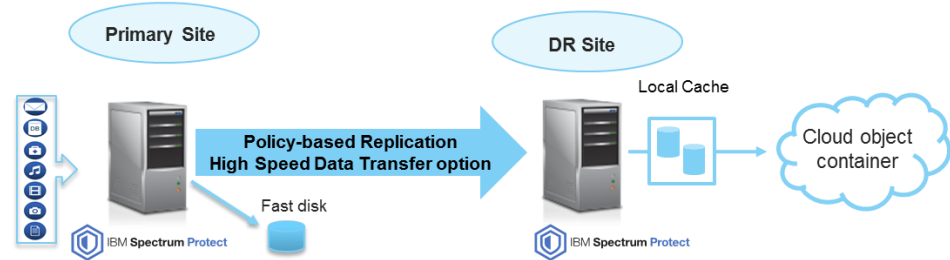
- Copy performed as part of “PROTECT STGPOOL type=local” process
 - Need to run dbbackup after protect stgpool process finishes
- COPYCONTAINER used only for repair of directory container
 - Repair directory container with the “REPAIR STGPOOL SRCLOCation=local” command
 - No direct client restore from COPYCONTAINER pools
 - Can recover the entire directory container pool
 - Can recover only part of the directory container pool
 - Deduplicated extents on tape are copied back to directory container
 - Spectrum Protect DB must be functioning for “repair stgpool” to work
 - Not a self contained backup like backupsets
 - IF DB & directory containers damaged, would first restore TSM DB, then repair directory container from tape copies
 - Client restore can occur only after recovery of the directory container pool, from the container pool
 - Will utilize on-site volumes first & issue ANR3425W to request off-site volumes be brought back
- Space Reclamation performed as part of the “PROTECT STGPOOL type=local” process
 - RECLAIM STGPOOL not valid for the COPYCONTAINER type of pool
 - RECLAIMPROCESS parameter on the storage pool doesn't apply to copycontainers
 - COPYCONTAINER reclamation will not actually read tape volumes
 - First determines all valid chunks on the reclamation target volume
 - Then read the containers files from the container pool that store these chunks
 - Then write them to a new target tape volume

7.1.7

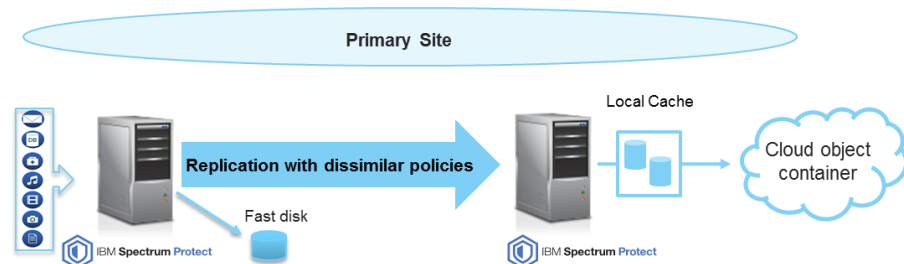
- IBM Spectrum Protect stores primary copy of backup data in object storage



- IBM Spectrum Protect stores secondary copy of backup data in object storage at DR site



- IBM Spectrum Protect stores long term copy of backup data in object storage



SPECTRUM PROTECT PLUS DEDUPE

- The vSnap server can be:
 - on the IBM Spectrum Protect Plus appliance
 - on a separate appliance
 - a physical vSnap installation

- You can encrypt backup data on the vSnap repository at initialization

- After setup, you can enable compression and/or deduplication**
 - <https://www.ibm.com/developerworks/community/wikis/home?lang=en-gb#!/wiki/Tivoli%20Storage%20Manager/page/IBM%20Spectrum%20Protect%20Plus%20Blueprints>

SPP enable vSnap deduplication



	Hostname/IP	Site	Type	Version	Status	Actions
	localhost	Primary	vsnap	10.1.2-9	Ready	Rescan

Manage Backup Storage localhost

- Enable Compression
- Enable Deduplication
- Synchronous Write Mode
- Encryption Enabled

The Encryption setting can only be applied during vSnap initialization. This option is for informational purposes only.

Save

Tools (storage options):

– Manage Backup Storage localhost

- Enable Compression – uses moderate amount of additional CPU resources
- Deduplication – disabled by default because it uses a significant amount of memory relative to the amount of data in the pool
- Synchronous Write Mode – disable only if you have
- Encryption setting must be enabled during vSnap initialization using the Actions menu

IBM Spectrum Protect Drives Results



Reduce Costs, Simplify Management

- Up to 53% reduction in backup infrastructure costs¹ with fewer backup/media servers and deduplication appliances needed.
- No charge software-defined deduplication delivers performance of DD4500 without the cost and complexity².
- Deduplication, compression and incremental ‘forever’ backups reduce space requirements by up to 95%.
- One control point for virtual, physical and cloud backups; application-aware backups, and hardware-assisted snapshots



Manage Data Growth

- One Spectrum Protect server does the work of up to 15 CommVault servers/media agents³
- One Spectrum Protect server does the work of up to 5 NetBackup backup/media servers⁴



Multi-Cloud Enablement

- Cloud backup tier reduces storage costs, improves flexibility
- No cloud gateway appliances needed



Enhanced Business Agility, Improved Service Levels

- Self-service portals for data reuse by dev/test, analytics, application, support, etc teams
- Snapshots integrated with traditional backups, applications and storage systems

1. Based on IBM assessments using Butterfly software.

2. IBM server dedupe is rated at 11.3 TB/hr. EMC rates their DD4500 at 10.2 TB/hr. <http://www.emc.com/data-protection/data-domain/data-domain-deduplication-storage-systems.htm#specifications>

3. Based on CommVault sizing and configuration guidelines, compared to IBM, http://docs.commvault.com/commvault/v10/article?p=system_requirements/commcell_sizing.htm

4. Based on NetBackup 5330 appliance capacity specifications, compared to IBM, <http://www.symantec.com/netbackup-appliance-5300/>

Client compression

- Files compressed by client before transmission
- Conserves network bandwidth and server storage



Client

Incremental forever

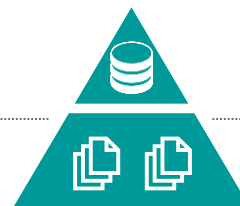
- After initial backup, file is not backed up again unless it changes
- Conserves network bandwidth and server storage



Server

Device compression

- Compression performed by storage hardware
- Conserves server storage



Storage Hierarchy

Server Side Deduplication

- Duplicate data only stored once
- Out of band dedupe to disk pool
- In-line dedupe to directory or cloud pools



Server side compression

- Files compressed by server after in-line dedupe for containers
- Conserves network bandwidth when writing to cloud

Client Side Deduplication

- Reduces network traffic by deduplication data before transfer
- Sends to server's dedupe pool

Appliance deduplication

- Deduplication performed by storage appliance (VTL or NAS)
- Conserves server storage

Deduplication is an integrated part of Spectrum Protect

¡Converse con un consultor!

Vea la grabación del 1º webinar

Vea la grabación del 2º webinar

Vea la grabación del 3º webinar

Thank You