



“ For over a decade Coolspirit have been supplying the UK’s top organisations with storage products and solutions so be assured we will meet your requirements head on.

It’s all about getting things right first time, quickly and simply! ”

**Damon Robertson**  
Coolspirit Ltd

#### Our address

24 The Bridge Business Centre  
Beresford Way  
Chesterfield  
S41 9FG

#### Get in touch

Call us on: 01246 454222  
Email us: [web@coolspirit.co.uk](mailto:web@coolspirit.co.uk)  
Find us: [View location map](#)  
Web: [www.coolspirit.co.uk](http://www.coolspirit.co.uk)

#### Office hours

mon - thurs 8:30am - 5:30pm  
fri 8:30am - 5pm  
sat - sun Closed

“ Boost your storage buying power...  
use ours! ”

Buy with confidence from  
Coolspirit your authorised  
FalconStor Partner

**FalconStor®**  
Software

# Dedupe Competitive Positioning

July 2010 Update

**FalconStor**<sup>®</sup>  
Software

# Agenda

- Positioning
- Key Differentiators
- Competitor Update
  - Data Domain
  - IBM Protectier
  - Sepaton
- Source Deduplication

# Positioning

# FalconStor Deduplication Solutions

## FalconStor® Virtual Tape Library (VTL)



- Large existing tape environments
- Disk seen as tape interface
- Extremely fast backup
- Scales to petabytes
- SAN infrastructure
- Physical tape export
- Data replication

## FalconStor® File-interface Deduplication System (FDS)



- Disk seen as file share (CIFS/NFS)
- Backup, Archive, Migration, Data Movement Applications
- Fast backup
- Scales to terabytes
- LAN infrastructure
- Data replication

# Five Reasons Enterprises Continue to use VTL

- **Keep backup traffic on FC SANs.**
  - Enterprises back up tens to hundreds of terabytes (TBs) of data nightly and weekly.
  - FC SANs facilitate these high levels of throughput - FC protocol is specifically designed to handle the block-based, sequential nature of backup traffic - higher throughput and faster backups.
- **Overhead of NFS and CIFS.**
  - CIFS and NFS processing overhead over an Ethernet LAN, faster than tape, slower than FC
- **Server overhead.**
  - Sending data to a NAS target requires that a server read data, encapsulate in TCP/IP packet and send over Ethernet
  - Creates unacceptable processing overhead for busy application servers that run 24x7, with limited backup windows
- **Fast, successful backups, not deduplication ratios, are the top priority.**
  - While enterprise shops are thrilled if they can achieve high deduplication ratios; fast, successful backups are still the first priority.
- **Better backup software and tape support.**
  - Backup software supports both NAS and VTL but many enterprise organizations still plan to use tape as part of their overall data protection strategy in some capacity. Using NAS solutions, data typically has to be re-read (and reprocessed) by the backup software before it can be stored to tape.
  - Using VTLs, the data is already stored in a format that is conducive to copying the data from disk to tape and back again. This task is executed and performed by the FalconStor VTL

# Why FalconStor VTL - Focus on Key Differentiators

- True Global Deduplication
- Scales capacity or compute cycles as needed
- FalconStor® Single Instance Repository (SIR) performance scales up to 256 TB of managed capacity per node
- FalconStor VTL scalability through clustering:
  - Up to 8 node cluster with a single point of management
  - Manages up to 32 petabytes of data in one single cluster
  - Up to 12 GB/s data ingest rate
- FalconStor SIR scalability through clustering :
  - Up to 4+1 cluster with a single point of management
  - Up to 256TB of capacity per node, equivalent to 5PB per node
  - Up to 2 GB/s of data deduplication speed
- HA – No Single Point of Vulnerability
- Seamless Tape Media Management – Direct Tape Out

# Why Global Deduplication

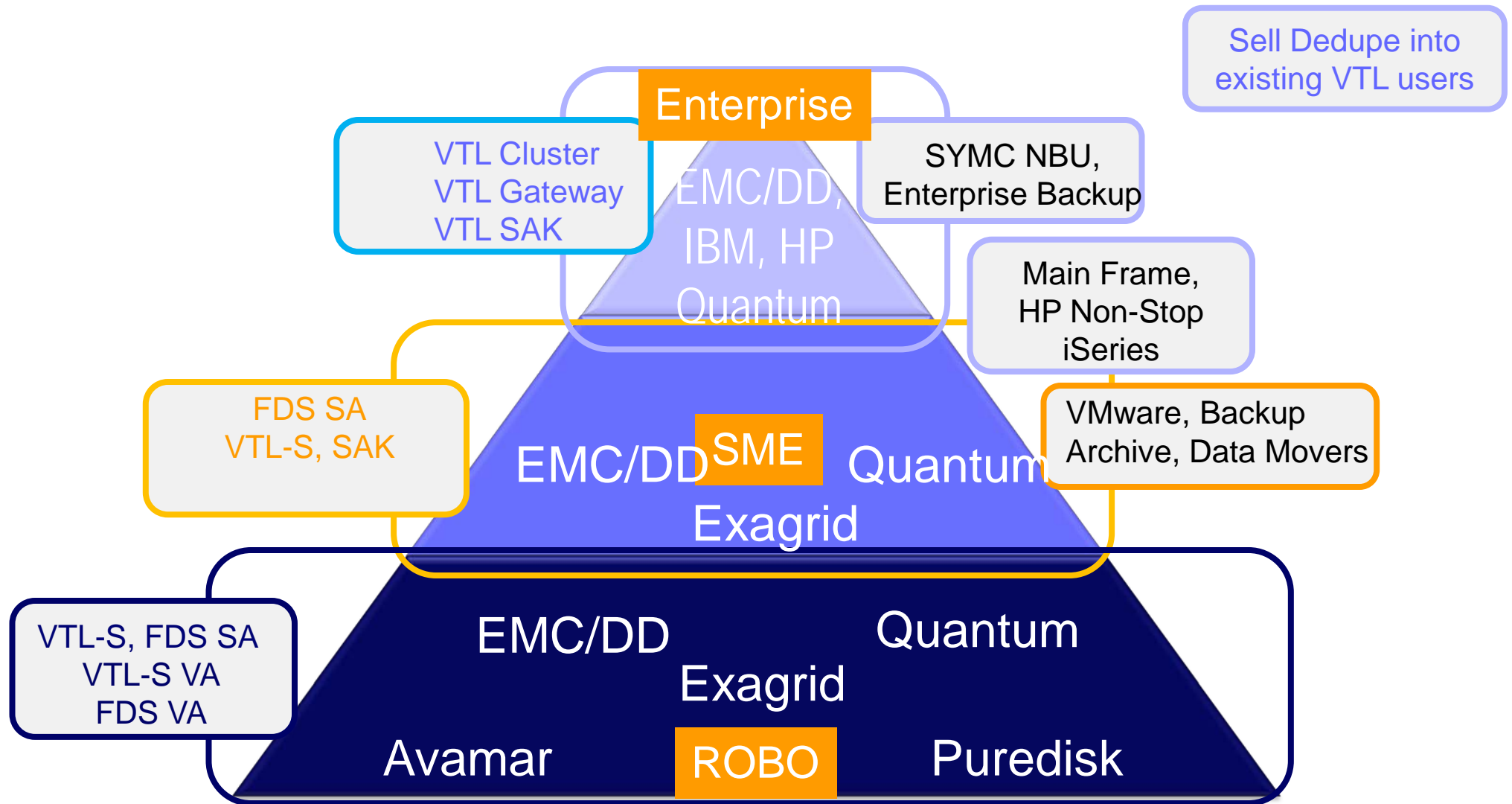
- **Reduced Cost**
  - Increase overall dedupe ratio
  - Increase efficiency, lower physical storage cost
- **Ease of Implementation**
  - Silo's increase management cost and impact administration overhead
  - Eliminate need to load-balance backup across individual appliances
- **Improves backup/recovery**
  - Even large single nodes do not perform as well as multiple global dedupe nodes
  - Throughput, not capacity causes customers to purchase more silo appliances
  - Inline dedupe requires scheduled defragmentation, impacts restore performance
- **Better integration with existing backup process**
  - Doesn't matter which node does backup or dedupe
  - Without global dedupe, need to introduce a process to divvy up large backup jobs
- **Scalability**
  - Better scalability as customer data and performance requirements grow
  - No concern about where and what you need to do with a backup as you scale
- **Availability**
  - FalconStor VTL includes HA, no concern about single point of vulnerability

# Throughput not capacity causes forklift upgrades

- Greater than 10 – 20 TB requires Global Deduplication
- 10 – 20 TB over 12 hours = 230 – 460 MB/sec performance
- Example: 140 TB Data Center
  - 7 servers, 20 TB each
  - 12 hour backup windows
  - Distributed weekly full
  - Incremental 6 servers (average 5 – 10%) = 1- 2 TB/server = 6 – 12 TB
- Nightly backup requirement 26 – 32 TB/night
- Performance requirement 700 – 888 MB/sec (2.5 – 3.2 TB/hr)
- This would require 3 – 4 Data Domain 690 or 880 appliances

# Competitive Overview

# Dedupe Positioning



# Competitive Overview

Vendor/Product	VTL, NAS,	Direct Tape Out	Appliance, Gateway, Software	De-dupe	Inline or Post-process	Global Dedupe	Dedupe Method	OST	Deduped Replication
Falconstor - VTL	VTL	Yes	Software, Appliance, Gateway	Yes	Both	Yes	Hash	Yes	Yes
Falconstor - FDS (NAS)	NAS	No	Software, Appliance, Gateway	Yes	Both	No	Hash	Yes	Yes
Data Domain	VTL & NAS	No	Appliance & Gateway	Yes	Inline	No	Hash	Yes	Yes
Exagrid	NAS	No	Appliance	Yes	Post-process	No	Delta Diff	Yes	Yes
GreenBytes	NAS	Yes	Software	Yes	Inline	Yes	Hash	Yes	Yes
IBM ProtecTier	VTL	No	Appliance	Yes	Inline	Yes (2 nodes)	Delta Diff	No	Yes
Quantum DXi	VTL, NAS	Yes	Appliance	Yes	Both	No	Hash	Yes	Yes
SEPATON	VTL	No	Appliance	Yes	Post-process	Yes (5 nodes max)	Delta Diff	Yes	Yes
PureDisk Storage Unit (Only works with NBU)	OST	N/A	Software	Yes	Inline	No	Hash	Yes	Yes

# Competition - Performance

Backup & dedupe rates for a 12-hour backup window			
Vendor	Ingest Rate (MB/s)	Dedupe Rate (MB/s)	Caveats
Falconstor - VTL	12000	2000	8 node cluster, 4 node dedupe
Falconstor - FDS	400	500	Max Performance; 600 Series CIFS/NFS
Data Domain	330	330	Max performance VTL across 2 x 4 Gbps FC ports
Exagrid	900	450	10 node cluster
Greenbytes	950	950	Max Performance, GX4000 w/dual 10GbE
IBM/Diligent	900	900	2 node cluster, requires FC or XIV disk
Quantum/Dell	500	500	Adaptive mode
SEPATON/HP	3000	1500	5 nodes with global dedupe

# Data Domain

# New Q2 Data Domain Systems and Software

Available Q2 2010

## Systems

- New Data Domain Global Deduplication Array
  - New Dual Cluster Data Domain deduplication storage system
- Data Domain DD880 expansion
  - Double the capacity from 71 TB to 142.5 TB usable

## Software

- New Data Domain Boost software
  - Distributing parts of deduplication process to backup server (Symantec OST only)
- New Data Domain Encryption software
  - Encryption of deduplicated data at rest
- Data Domain Replicator software enhancements
  - One-to-many and cascade replication

# DD880 Update

- 2x capacity
  - Up to 142.5 TB (up to 7.1 PB logical)
  - Support for up to 12 shelves
- Field Upgrade Kit for existing DD880

## FalconStor Position:

- Upgrade for DD 880 Only
- MSRP on ES20-16TB is \$60,000
- Adding 6 Shelves adds \$360,000 to cost
- MSRP for DD880 (12 Shelves) \$1,115,000 includes OST and Replication, no VTL
- DD Logical Capacity calculated at 50:1; normal 20:1 is 2.84 PB
- DD reserves 30% of usable disk for sequential write space, they do not write to reclaimed blocks
- **Sell FalconStor VTL/SIR Cluster as competitive option**



**DD880 targeted at backup data sets between 30 – 125 TBs**

# Data Domain Global Deduplication Array

- Global deduplication and single namespace across two DD880 controllers
  - Speed: Up to **12.8 TB / hour** (3.5 GB/s)
  - Usable capacity: Up to **285 TB**
  - Logical capacity: Up to **14.2 PB**
  - Up to **270** concurrent write streams
  - As replication target, up to **270** originators

## FalconStor Position:

- Only works with 10GE + Symantec OST (NetBackup and Backup Exec); Networker will be added H2
- Requires DD Boost - must be loaded on every media server
- No HA – no customer wants to expose 285TB to single point of vulnerability
- No VTL support
- No CIFS/NFS Support
- Works with Backup Only – Symantec NetBackup and Backup Exec; Networker (H2 2010)
- Only works with DD880 – Requires disk shelf expansion by 2
- Logical Repository is based on 50:1 deduplication ratio



# Data Domain - Summary

- The in-line deduplication model can slow backup windows, forcing customers to purchase larger appliances
- No Policy level Deduplication - Data Domain deduplicates everything, regardless of whether it is necessary to deduplicate certain data.
- Published performance figures based on VTL/4GB FC or OST/10G; 10G is only supported on DD660, DD690 and DD880 appliances
- Entry Data Domain appliances support only Ethernet (up to 4 x 1 Gbps ports), no FC or 10GE – I/O throughput is slow. The customer needs to move to DD660, DD690, or DD880 for higher performance, FC, or 10GE connectivity.
- Limited Tape Library and Tape Drive Emulation
  - Up to 64 virtual tape libraries and 128 virtual tape devices
  - Emulates one tape library and three tape formats.

# Data Domain – Cont'd

- Because DD does in-line deduplication, it must also do continuous recovery verification during the initial backup. Continuous recovery verification can have a significant impact on the backup performance, further reducing the backup window and causing customers to purchase larger appliances than required, simply to match the backup window.
- Data Domain Literature recommends multiple appliances “To avoid performance bottlenecks and to obtain optimal Data Domain system performance, multiple Data Domain systems can be deployed to support parallel workload / performance demands.” – *from DD data sheet*
- Data Domain appliance performance tops out at 130 MB/sec (DD140), 190 MB/sec (DD610), 300MB/s (DD630, 660) and 330 MB/s (DD690, 880) maximum throughput – figures based on 4GB FC/VTL.
  - To avoid performance bottle necks customer must purchase multiple or a larger appliance.
  - FalconStor offers multiple options, up to 8 \* 1Gbps or 2 \* 10GE for FalconStor FDS or 2/4/8 Gbps FC, multiple 1\*Gbps or 10 GE for FalconStor VTL-S.
  - FalconStor VTL/SIR Cluster scales to 8 VTL nodes and 4 (+1 Standby) deduplication nodes, providing up to 43 TB/hr backup performance and up to 2000 Mb/sec (7.2 TB/hr) deduplication

# Data Domain Boost



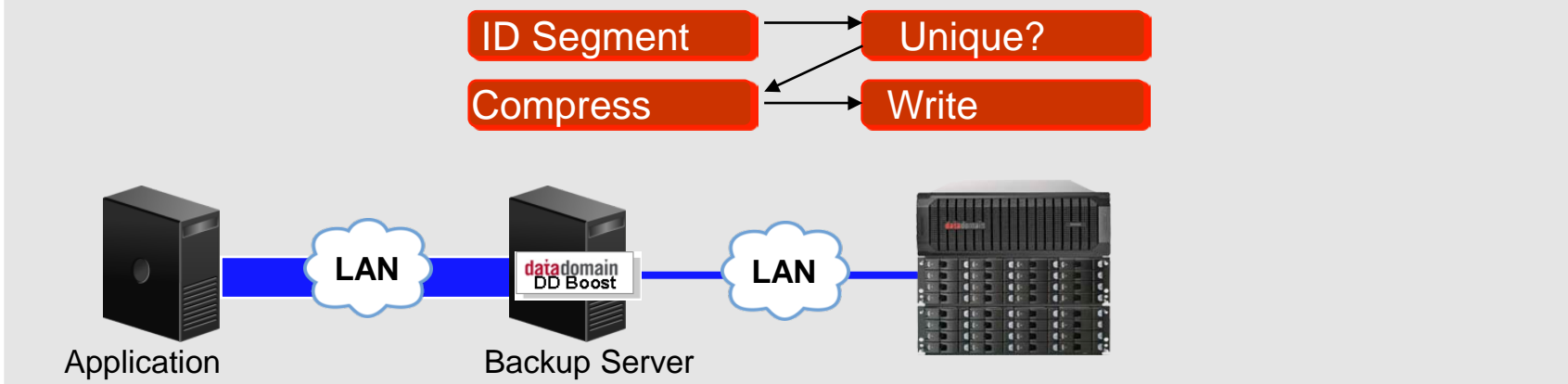
- Distributes parts of deduplication process to backup server
  - Licensable software works across Data Domain portfolio
- Only support specific Backup Applications
  - Symantec NetBackup and Backup Exec now
  - EMC NetWorker support in 2H 2010

## FalconStor Position:

- DD Boost is an extra license on EVERY Media Server
- This is an admission that their Inline Stream-Informed-Segment-Layout (SISL) architecture is not fast enough or will not scale with CPU performance only, they must use the performance of the OST Media Servers to improve overall deduplication performance. Inline is not fast enough to handle large backup solutions
- Every end user at the Symantec Vision conference stated they do not want to run any additional client software or have to maintain additional client software on each of their media servers.

# What gets distributed?

## With DD Boost



## DD Boost – Distributed deduplication

- GDA automatically directs data to the proper controller
  - Media Server ALWAYS sends data to SAME controller
  - GDA supports load balancing, but must migrate all data across IP interconnect to new controller
- DD Boost client license must be loaded on each media server
  - Offloading the work involved in hashing and compression – loads CPU
  - Requires multiple transmissions across the LAN between Media Server and GDA
  - Only works with Symantec OST
- To improve performance, EMC/DD recommends adding more media servers and backup jobs

# GDA Weakness

- GDA is two separate DD880's
  - A heavy OST client that hashes and software compresses the data and sends it to one of the two boxes based on the first digits of the hash.
  - Forces media servers to perform highly CPU-centric processing.
  - GDA moves the problem to the media server.
  - The problem now rests on the end users shoulders to manage and size every server to ensure it meets appropriate GDA specifications.
  - GDA has assimilated the media servers into its “appliance” realm and pushed processing and management activities onto the customer.
  - Specific to Symantec OST technology on NetBackup and BackupExec.
  - GDA is not an option for any other backup applications
- Points to consider related to the implementation:
  - How do you size the required media server upgrades to meet performance requirements?
  - Will you need to purchase new media servers to gain the processing power required?
  - Who do you call if your performance does not meet expectations?
  - Are you comfortable running such highly CPU intensive agents on your media servers which may already be resource constrained?
  - What happens if you have some applications that backup directly to tape using a LAN free approach such as NDMP or RMAN?
  - What if you are running backup applications like IBM TSM, EMC's own NetWorker or CommVault that do not support OST?
- These are real end user concerns. There are plenty of customers with 20+ media servers who would need to completely overhaul their environment to implement GDA.

# Key Competitive Advantages

- No HA – no customer wants to have single point of vulnerability
- DD Boost - EMC/DD Inline Stream-Informed-Segment-Layout (SISL) architecture doesn't scale and it's running out of steam
- GDA – Only works with DD880
  - No CIFS/NFS
  - No VTL
  - Backup Only
  - Requires DD Boost
- Performance figures – Peak run with 4GB FC or 10GbE OST only

# Data Domain vs FalconStor

	DD140	DD610	DD630	DD660	DD690	DD880	GDA	DDX Array
<b>Speed</b>	450 GB/hr	675 GB/hr	1.1 TB/hr	2.0 TB/hr*	2.7 TB/hr*	5.4 TB/hr*	12.8 TB/hr**	86.4. TB/hr
<b>Logical Capacity</b>	17-43 TB	75-195 TB	165-420 TB	.520-1.31 PB	.710-1.7 PB	2.8-7.1 PB	5.7-14 .2 PB	45.6-114 PB
<b>Raw Capacity</b>	1.5 TB	Up to 6 TB	Up to 12 TB	Up to 36 TB	Up to 48 TB	Up to 192 TB	Up to 384 TB	Up to 3.07 PB
<b>Usable Capacity</b>	0.86 TB	Up to 3.98 TB	Up to 8.4 TB	Up to 26.1 TB	Up to 35.3 TB	Up to 142.5 TB	Up to 285 TB	Up to 2.28 PB

\*Performance based on 10G/OST

\*\* Performance based on 2 controllers, 10G, OST and DD Boost

\*\*\* FalconStor performance based on 8 GB FC w/OST

	SA101	SA304	SA311	SA622	SA634	SA668	VTLCLD880	VTL/SIR Cluster
<b>Speed</b>	540 GB/hr	792 GB/Hr	900 GB/Hr	2.7 TB/Hr***	2.7 TB/Hr***	5.4 TB/hr***	10.8 TB/hr***	43TB/hr
<b>Logical Capacity</b>	20 TB	100 TB	220 TB	456 TB	680 TB	1.36 PB	5.1 PB	5.1 PB
<b>Raw Capacity</b>	6 TB	6 TB	15TB	30 TB	45TB	90 TB		
<b>Usable Capacity</b>	1 TB	4TB	11 TB	22.8TB	34 TB	68 TB	Up to 256 TB	256 TB

Sepaton

**FalconStor**<sup>®</sup>  
Software

# Sepaton Overview

- VTL Enterprise Vendor
  - Focused on customers with 50TB or \$1B in revenue
  - Direct sales model, HP OEM
  - Claim 1300 customers (700 HP)
  - Venture funded, last round \$15.5 M in March, 2009
  - Average Sale, \$360,000
- Position performance, ease of use, scalability
- Deduplication – Post Processing, Delta Difference
- Up to 16 Nodes, Grid Architecture
- Content-aware – primary backup application, NetBackup and TSM
- Newest version of software, v5.4 includes:
  - Secure multi-tenancy and cloud optimized data protection
  - Secure Erasure enables complete data destruction
  - DeltaView Manager and content-aware analytics
    - Manage multiple storage pools in a single SEPATON system
    - Tool for optimized performance, maximize system utilization and resource requirements

# Sepaton S2100-ES2 Enterprise Class

- **S2100-ES2**
  - Grid-based system scales from one node to 16 nodes; 5 Node Maximum per Deduplication Domain
  - DeltaScale Architecture, built on SRE (Scalable Replication Engine) nodes
  - Scales from 10TB to 1.6 PB (with 16 nodes)
  - Ingest speed: 600 MB/s per node (2 TB/hr)
  - Dedupe speed: 300 MB/s per node (1 TB/hr) – Max 1500 MB/sec
  - Maximum 96 virtual devices (libraries or drives) per node
    - 1536 virtual devices (Libraries or Drives) per system (16 nodes)
    - Up to 5.3 million virtual tapes per system (16 nodes)
  - RAID 6 protection against double disk faults
  - Global Deduplication – up to 5 Nodes
  - Enterprise-class, high-availability disk array system (HDS) with built-in redundancy for all components and paths
  - Redundant connections to disk via redundant Fibre Channel switches
  - Redundant Ethernet switches for command and control
  - Entry Configuration starts at \$413,500

# Sepaton S2100-MS2 Deduplication Appliance

- S2100-MS-2
  - 30TB dual-node configuration; starting at \$321,000
  - Targeted at Data Domain 880
  - Scales to 160 TB of physical capacity
  - Can be upgraded to S2100-ES2
  - Max ingest performance - 1,200 MB/sec.
  - Same DeltaScale architecture and SRE nodes as S2100-ES2, limited to a 2-node architecture
  - Content-aware backup – limited backup application support
    - EMC NetWorker, HP Data Protector, IBM Tivoli Storage Manager, and Symantec NetBackup
  - Limited to 14 Tape Library and 14 Tape Emulation formats

# Sepaton Weakness

- Positives:
  - Good and easy scalability; Grid master/slave node architecture if the master node fails the VTL is **dead**
  - Performance is good, not as fast as FalconStor VTL
  - Good set of library and tape drive emulations, including some IBM and STK formats, but limited to 14 Tape Library and 14 Tape Emulation formats.
  - *Sepaton can scale VTL across multiple nodes, add nodes non-disruptively*
- Limitations
  - No HA
  - Sepaton requires twice (2.5x) as much “landing area” disk (Delta Difference Post Processing Dedupe)
  - No direct physical tape connect, requires tape creation though a media server (HP OEM integrates with HP tape systems only)
  - No Gateway, does not support 3<sup>rd</sup> party disk (Dot Hill and HDS OEM)
  - Does not support iSeries or Mainframe
  - Does not support STK ACSLS operations
  - Does not support NDMP direct backup to tape
  - Does not support FC-based replication
  - Does not support encryption of data on tape
  - Content-aware – requires specific backup application drivers, no CommVault
  - Is limited to 4GB Fibre Channel interface only, no iSCSI
  - Restoration of older backup data is very slow – In a POC always get customer to restore from older data set

# IBM Protectier

**FalconStor**<sup>®</sup>  
Software

# IBM ProtecTIER Overview

- Purchased from Diligent on April 18, 2008
- VTL with inline deduplication – appliances and gateways
  - Appliances TS 7650
    - xSeries servers preconfigured with DS4700 storage repositories
    - 7TB, 18TB, & 36TB models – Single node; 36TB dual node
    - Dual 4 Gb/s FC interface
    - HyperFactor deduplication and replication (extra license)
    - Support for 12 virtual libraries, 256 virtual drives, and 128,000 virtual cartridges
    - Up to 500 MB/sec inline deduplication
    - Two node clustered configuration can continue operating due to the redundancy of the nodes but does not offer “failover” or HA
  - Gateways (TS7650G)
    - Support for two server types: 3958 DD3 or 3958 DD1
    - Dual 4 Gb/s FC interface
    - HyperFactor deduplication and replication (extra license)
    - 16 virtual libraries, 512 virtual drives, and 500,000 virtual cartridges (dual node)
    - Up to 1000 MB/sec Inline deduplication (dual node)
    - Dedicated 7680 G for Z-Series Mainframe

# Key Weaknesses of ProtecTIER Systems

- Inline deduplication presents a backup bottleneck, especially with only dual Gigabit IP or dual 4Gb FC ports
- Deduplication is not tape format-aware
- Fibre Channel (FC) only. No support for iSCSI.
- ProtecTIER offers replication – recently introduced (Fall 2009)
  - No global deduplication across sites
  - Licensed by total TB managed
  - No encryption in-flight
  - IP Only; One-to-one only, not bi-directional (many-to-one 1H2010)
- No integration with physical tape
  - No tape export or tape caching
  - No tape duplication
  - No ACSLS support
  - No tape encryption support
- No support for NDMP backups
- No support for NetBackup OpenStorage – 2H2010 (Microsoft Windows, Solaris only)
- *Limited emulation - Up to 16 virtual libraries, 512 virtual drives – total for a two node system*

# Key Weaknesses of ProtecTIER Gateway

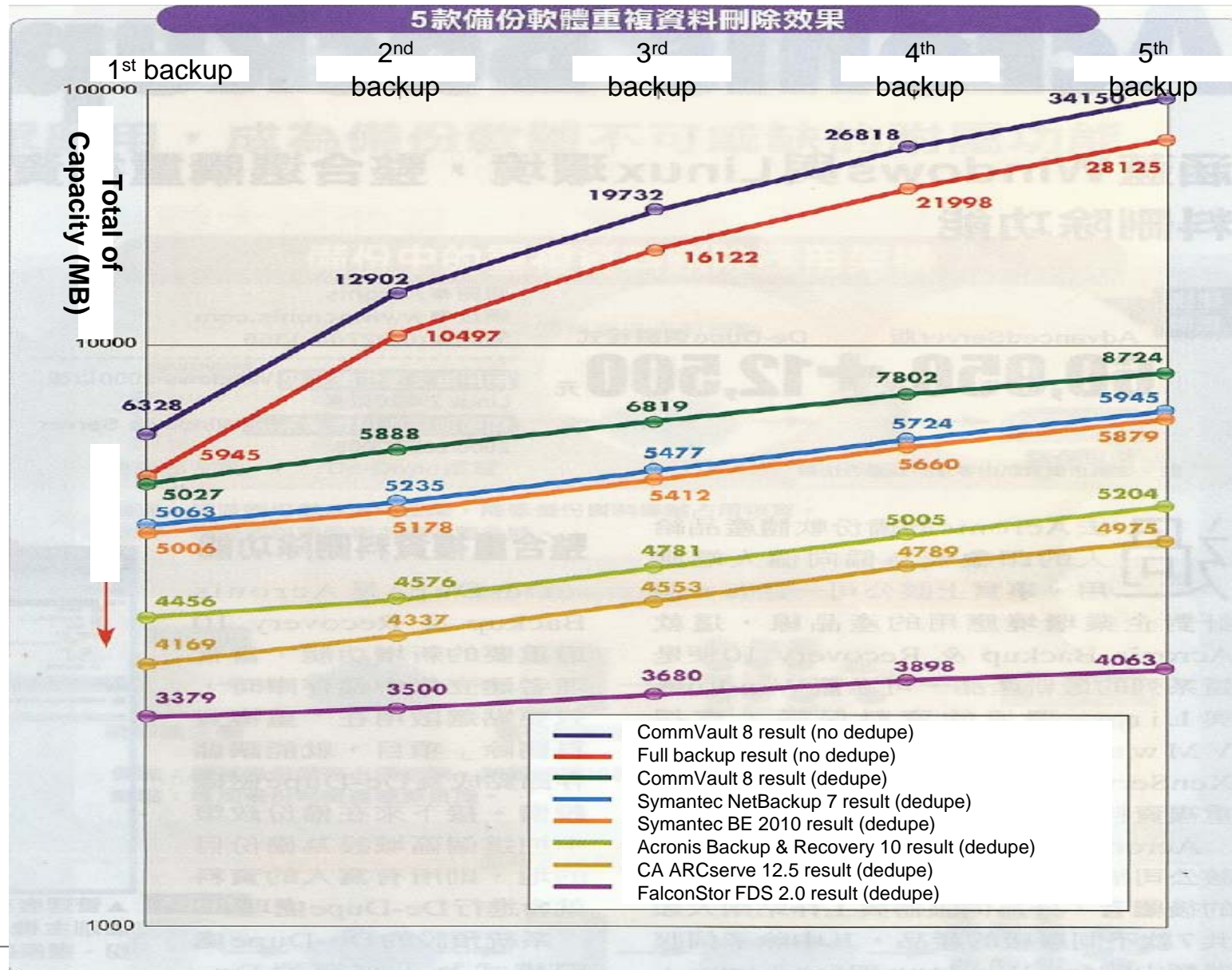
- Limited hardware support. Not a truly open gateway
  - Runs *IBMSystem x®server*
    - 3958 DD3 (IBM System x3850 M2 Type 7233)
    - 3958 DD1 (IBM System x3850 M2 Type 7141)
  - Supports the following storage:
    - IBM System Storage DS8300
    - IBM System Storage DS4800
    - IBM System Storage DS4700
    - XIV
    - HDS AMS1000
    - EMC CX3, CX5, CX7
- Only Supports RAID 5 (user data)
  - Requires RAID 10 for Meta data, (must be SAS or FC drives)
  - No RAID 6 support.
- No support for hardware compression cards
  - Performs CPU Based Lempel-Ziv-Haruyasu (LZH) compression
- Two-node clustered requires (2)3958-DD3 servers
  - A single node configuration can manage up to 32 concurrent jobs
  - Two-node clustered configuration can manage up to 64 concurrent jobs

# Source Deduplication

# Source Deduplication

- Deduplication performed at the source – mainly backup application, Symantec NetBackup 7, Backup Exec 2010, Commvault Simpana 8; CA ArcServe 12.5
- Primary benefit is reduce bandwidth over the LAN, sending only unique data (similar to FalconStor solutions' replication benefit)
- Limited functionality
  - Requires client license on every server or client desktop
  - Typically requires more CPU, RAM, Storage
  - Slow performance
  - Limited to only vendor data, single source – no global deduplication across all data in the enterprise (backup, database dump, archive, VM, etc)
  - Usually still requires a target for longer term retention and consolidation (i.e. Backup Exec 2010 and PureDisk)
- FalconStor FDS offers higher performance, more functionality, better and higher scalability – a more comprehensive solution for all but the very small or remote SMB

# Deduplication test result 5 backup software vendors



# Capacity reduction after data deduplication

Product	1 <sup>st</sup> backup	2 <sup>nd</sup> backup	3 <sup>rd</sup> backup	4 <sup>th</sup> backup	5 <sup>th</sup> backup
Acronis Backup & Recovery 10	13.1%	56.5%	70.4%	77.3%	81.5%
CAARCserve 12.5	18.7%	59.7%	71.8%	78.3%	82.4%
CommVault Simpana 8	20.6%	54.4%	63.5%	71.0%	74.5%
Symantec Backup Exec 2010	2.3%	51.2%	66.5%	74.3%	79.1%
Symantec NetBackup 7	1.2%	50.2%	66.3%	73.9%	78.9%
FalconStor FDS 2.0	35.1%	66.7%	77.2%	82.3%	85.6%