



# **Driving Innovation Through the Information Infrastructure**

**SPRING 2011**



# ***Reduced Complexity with Next- Generation Deduplication Innovation***

Sean R Kinney  
Director, HP StoreOnce  
Hewlett-Packard



# Data Storage Priorities 2010

**Data backup, Capacity growth, Disaster Recovery**

Top three data storage priorities in 2010:

1. Improving data backup (48%)
2. Adding capacity (37%)
3. Enhancing Disaster Recovery (36%)

SearchStorage.com Storage Priorities, Survey of 360 respondents, fielded December 2009



# Remote Office/Back Office (ROBO) Data Protection Challenges

- 1/5 of large companies have over 50% of data in ROBO (IDC, 2009)
- 1/3 of large companies have 20-50% of data in ROBO
- Mission and Business Critical data exists at these sites – without mission critical data protection
- Data Center managers do not have control at these remote sites
- Backups are not being done or are being done by non-IT personnel



# TCO & Economics Driving Deduplication Adoption

- Deduplication adoption is about the elasticity of demand
- Do things that were previously not affordable



# The Key Benefits Of Data Deduplication

## 1. Keep more data on disk - typically 20x more backup data

- Allows longer retention of backup data on disk before archiving off, or purchasing more backup storage
- Improve service levels and business performance – data is available on disk for longer for fast recovery of lost or corrupt files from multiple backup points
- More efficient storage utilization – effectively reduces cost per gigabyte for backup data

## 2. Enables network efficient replication

- Cuts the costs of data transmission through lower bandwidth lines
- Makes remote backup and disaster recovery more affordable for a wider range of businesses
- Reduces the overhead required at remote sites required for backup, media management and off-site vaulting



# Idemitsu Petroleum Norge

## Backup to disk solution

“Idemitsu was in need of a simple, cost-effective backup to disk solution capable of handling complete backup history for more than 30 terabyte of production data and capable of handling a 100 % growth of these data.

“With backup speed of over 3 terabyte/hour and capable of storing up to around 1 petabyte of backup history in an appliance only using 10u in a rack, we saw the opportunity to completely eliminate tape as our primary backup media.

With a stipulated appliance cost (including 5 year 24/7 support) being very low per logical terabyte, we have also found the “deduplication” solution to be very attractive from an isolated cost perspective. This applies even when comparing the cost of the “appliance” to the cost of physical tape.”

Mats Bergersen, IT Systems Administrator, Idemitsu Petroleum Norge AS

Idemitsu Petroleum Norge AS is a subsidiary of the Idemitsu Group, one of Japan's largest energy corporations with 79 affiliated companies and over 8000 employees worldwide. As an E&P company we have been active in operations offshore Norway since 1989. We own stakes in several valuable producing assets and have experienced remarkable success in exploration in the past few years.



# Data protection for a converged infrastructure

Human error



System failure



Natural disaster



Sabotage

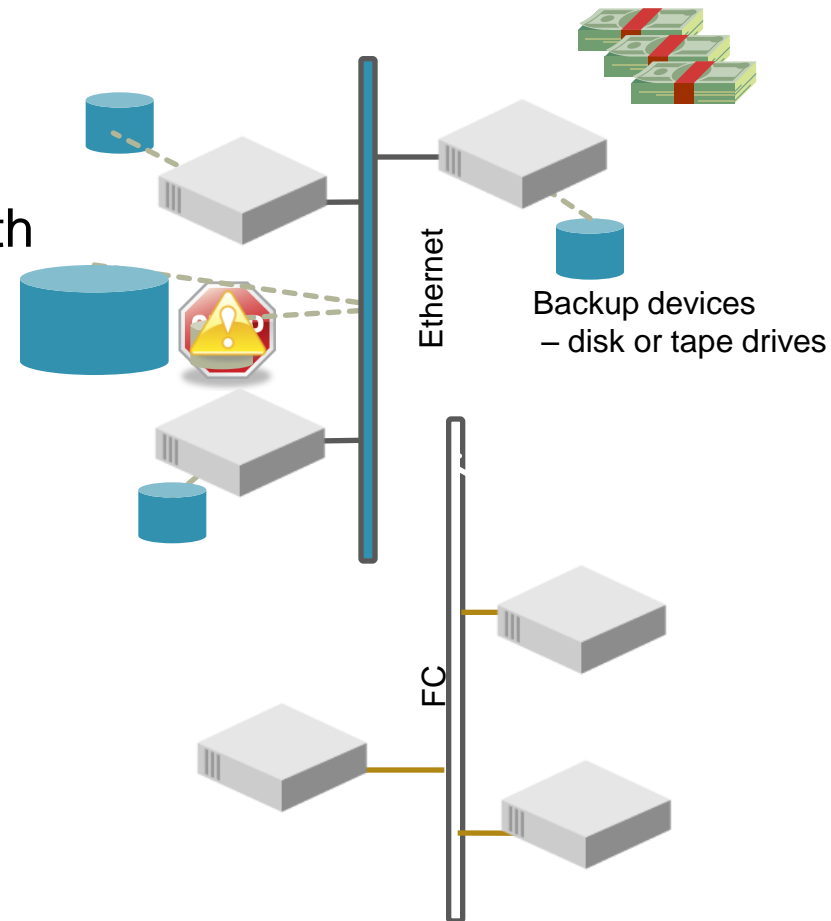


Back up is all about restore



## Problem: existing backup solutions are expensive and complex to manage

- Ongoing investment in data protection devices to keep pace with data growth
- Unpredictable storage growth can lead to complex, distributed configurations
- Overheads in managing multiple tape and/or disk backup processes and devices
- Human error creeps in with complexity
- Potential for SAN performance bottlenecks





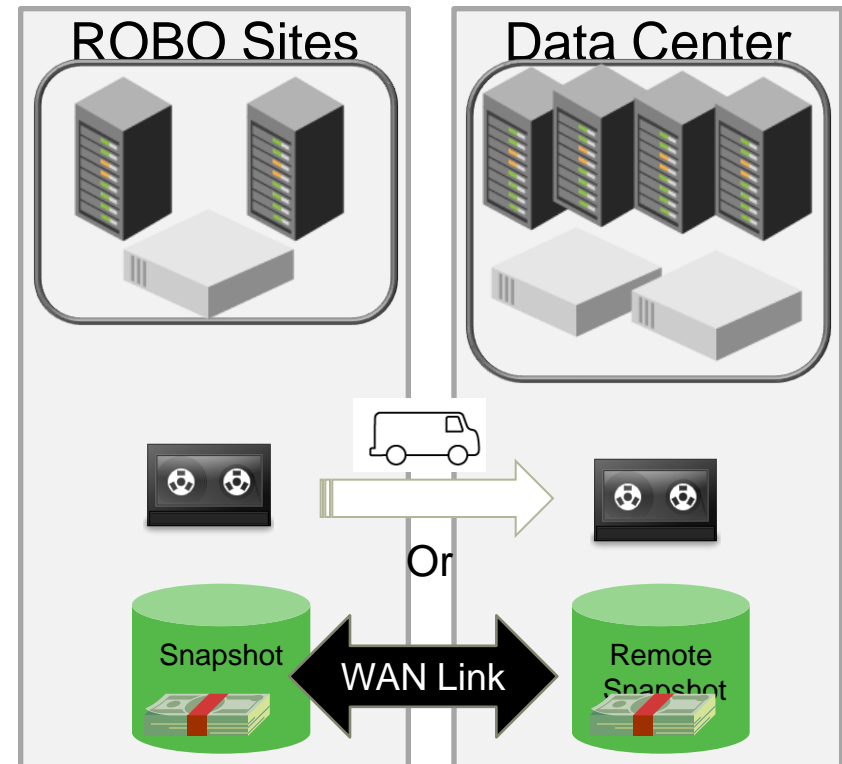
# Impact of disk based backup

Problem	What people do today	How disk-based backup solves the problem
Not meeting backup window	<ul style="list-style-type: none"> <li>• Gamble</li> <li>• Accept lower performance during backup jobs</li> <li>• Report backup failure to management</li> </ul>	<ul style="list-style-type: none"> <li>• Accelerates backup performance</li> <li>• Reduces back up window needs</li> <li>• Improves backup process reliability</li> </ul>
Slow restores	<ul style="list-style-type: none"> <li>• Devote storage administrator resources</li> <li>• Accept lower productivity</li> </ul>	<ul style="list-style-type: none"> <li>• Reduces workload of storage administrators and helps desk operators</li> <li>• Allows quicker time to productivity by offering fast restore time</li> </ul>
Inefficient media usage	<ul style="list-style-type: none"> <li>• Purchase more media</li> <li>• Over-pay for offsite services</li> </ul>	<ul style="list-style-type: none"> <li>• Reduces number of tape copies</li> <li>• Fills media more efficiently</li> </ul>
Perishable solutions (become too small, don't work with new applications, etc.)	<ul style="list-style-type: none"> <li>• Rip and replace solutions</li> <li>• Deploy point solutions for the worst problems</li> </ul>	<ul style="list-style-type: none"> <li>• Scales capacity and throughput independently</li> <li>• Works with almost any backup application</li> <li>• Uses data deduplication, deduplication-enabled replication and automigration</li> </ul>



# Problem: Providing disaster recovery capabilities across the organization, including remote sites

- Limited resources
  - Unreliable backup
  - Difficult to monitor and audit
- Cost of disaster recovery
  - Physically transporting data between sites
  - Expense of traditional replication
- Impact to business productivity and continuity
  - Ability to quickly access and restore lost or corrupt files





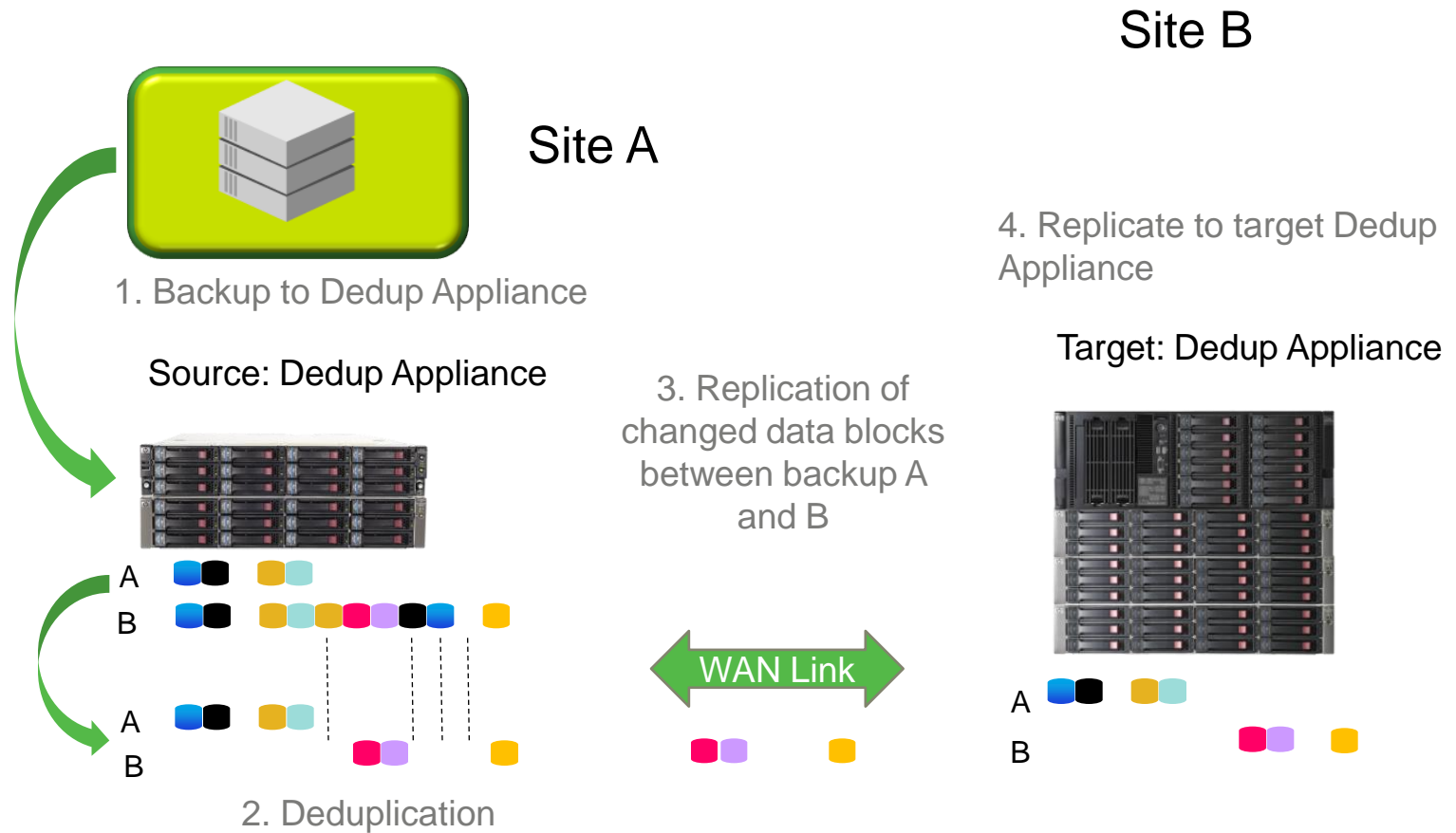
# Why Use Disk Rather Than Tape for a Backup Target ?

- Reliability
- Simplicity
- Aggregated Bandwidth
- New technologies
  - Deduplication
  - Off site data replication using existing network
- Remote Office Backup
- Faster Restore times





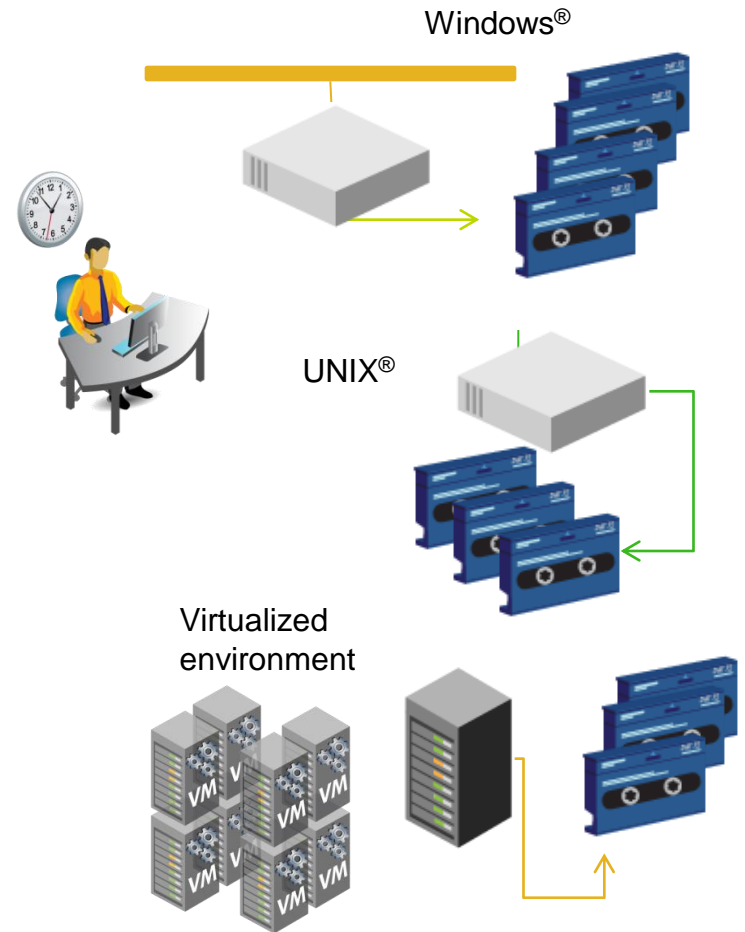
# Data replication that is network efficient





# Before: challenge meeting backup service levels (RPO/RTO)

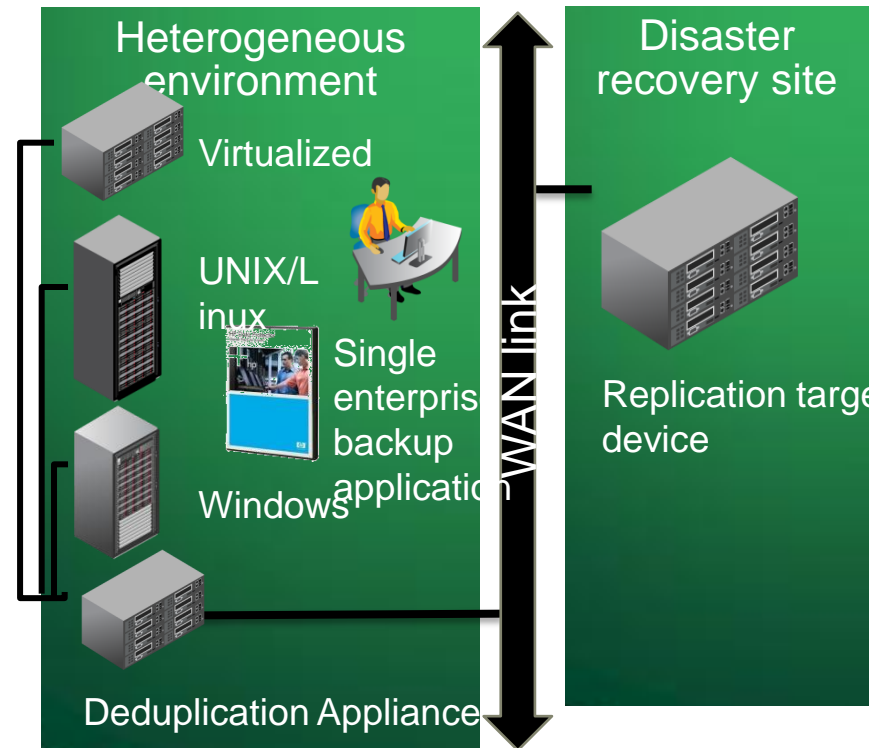
- Time to find and restore a file from a tape can impact productivity.
- Overheads are involved in tracking and managing multiple tapes—including off-site copies.
- Managing backup and restore across different network environments can be complex with virtualized environments in addition to traditional ones.





# After: Keep more data on disk.

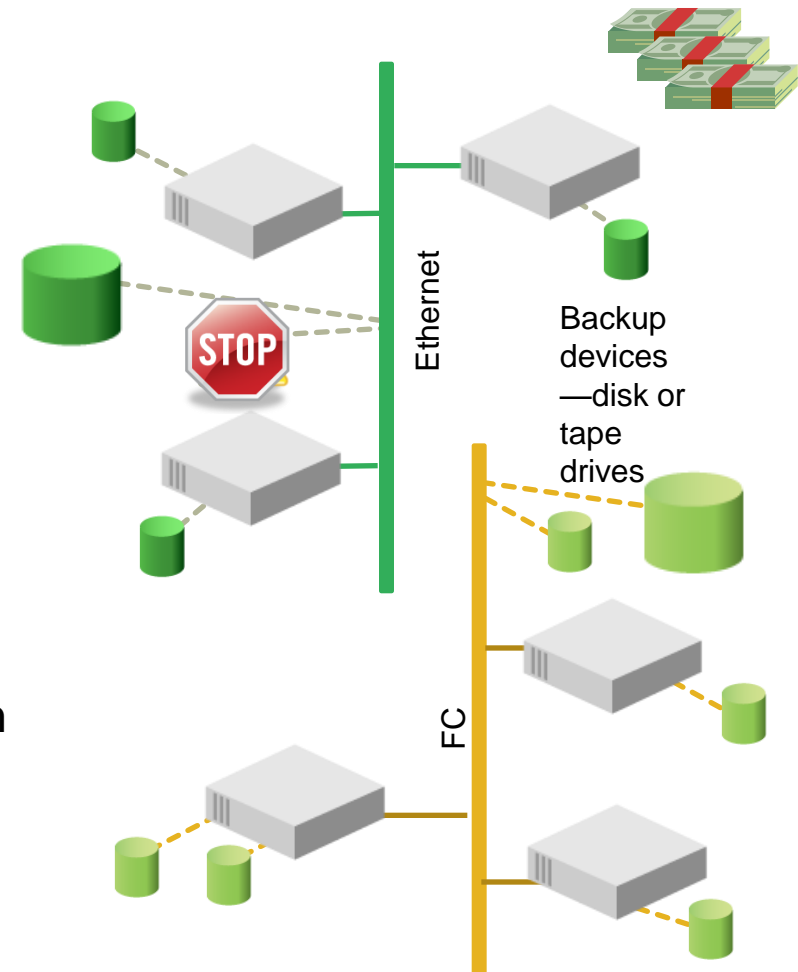
- More data on disk longer.
- Disk-based appliances have faster restores than tape.
- Repurpose tape libraries for longer term archive.
- Deduplication brings the economics of usable disk closer to tape.





# Before: complex backup solutions

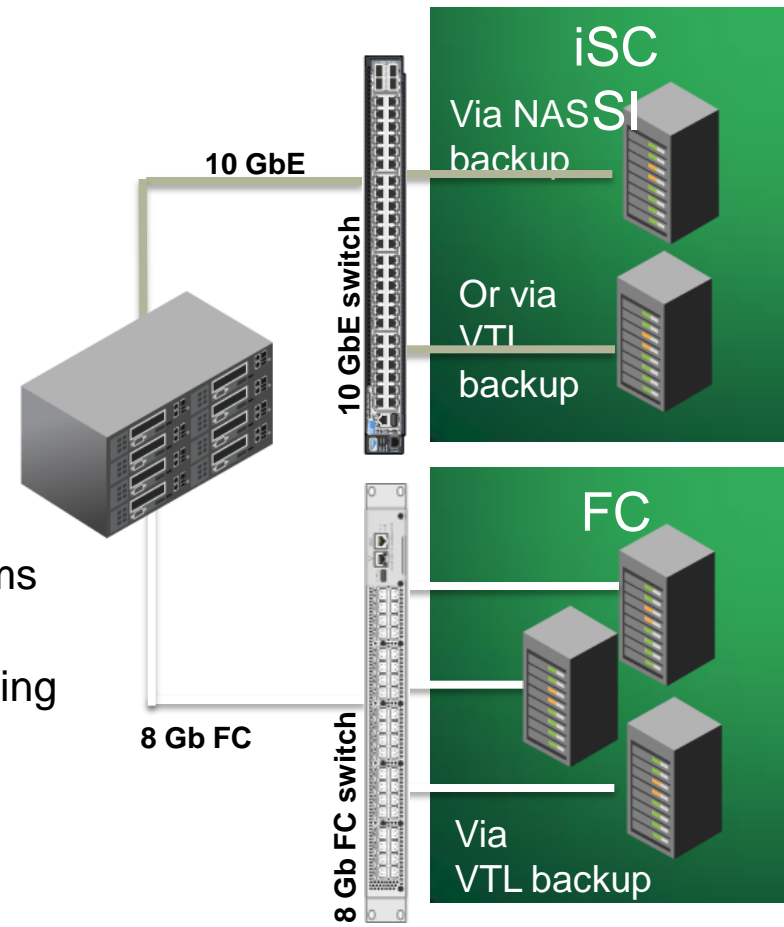
- Ongoing investment in data protection devices to keep pace with data growth
- Unpredictable storage growth that can lead to complex, distributed configurations
- Overheads in managing multiple tape and/or disk backup processes and devices
- Human error that creeps in with complexity
- Potential for SAN performance bottlenecks





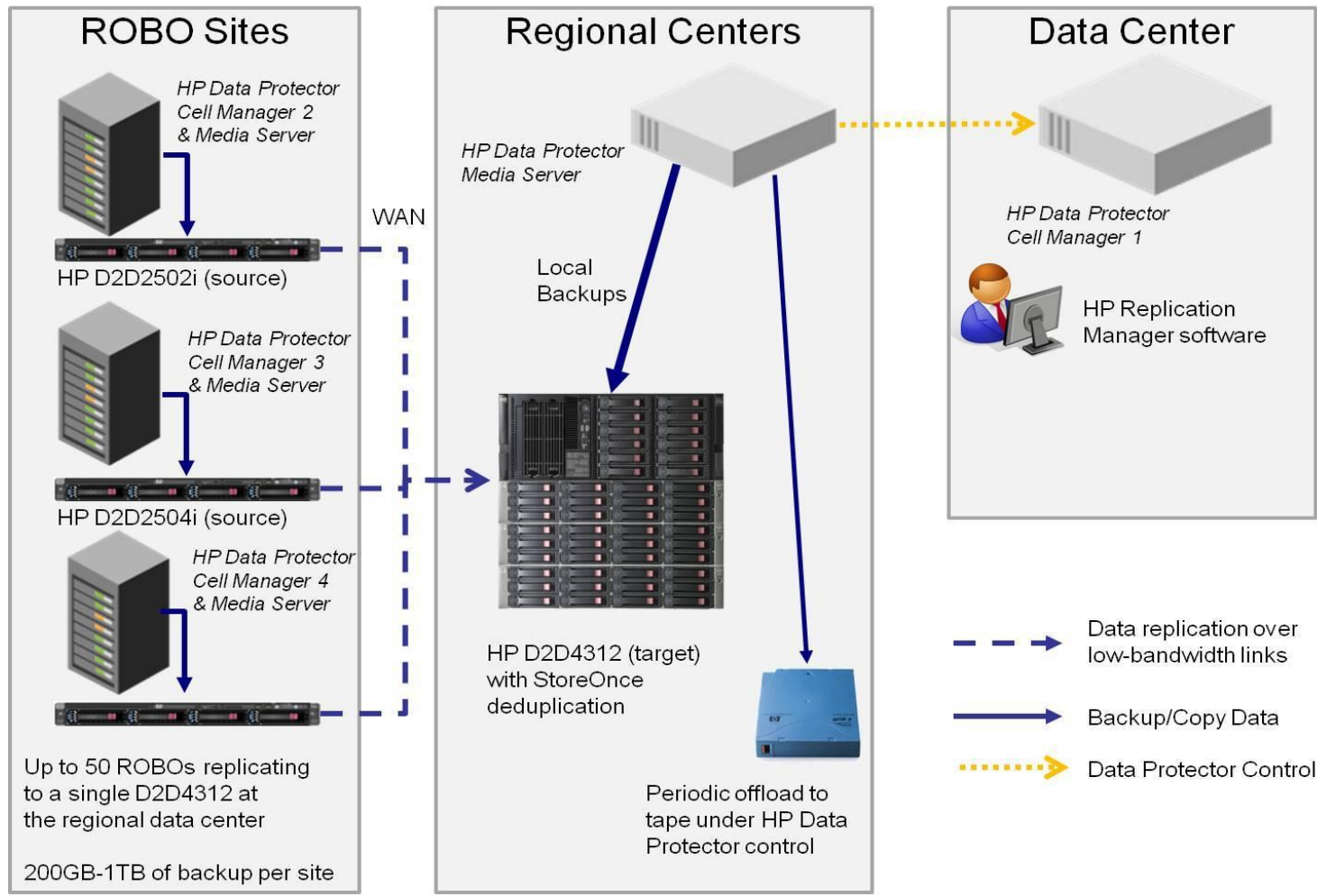
# After: Deduplication Solution automates and consolidates backup

- Save on hardware costs.
  - Get space efficient backup. Deduplication typically retains 20x more data on disk
  - Scale capacity and performance up to PB's.
- Simplify backup.
  - Consolidate multiple backup streams to a single device.
  - Get seamless integration with existing infrastructure.
- Save on resources.
  - Automate backup with little to no maintenance required.





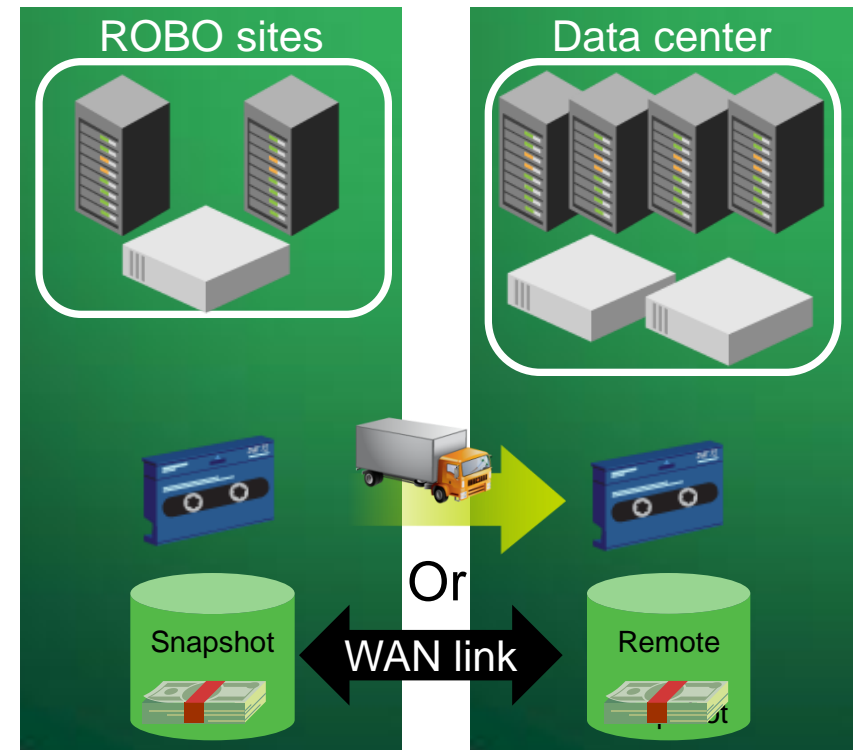
# More Advanced Implementation





# Before: providing disaster recovery

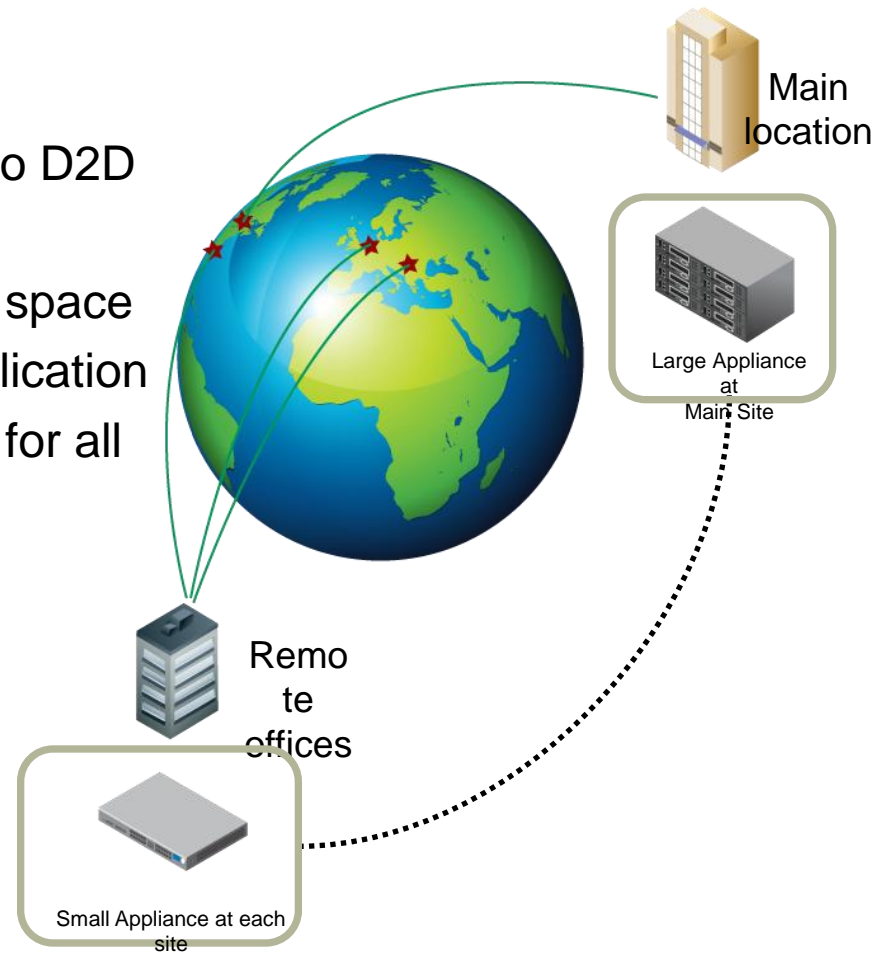
- Limited remote resources
- Inconsistent backups
- Cost of disaster recovery
- Impact to business productivity and continuity





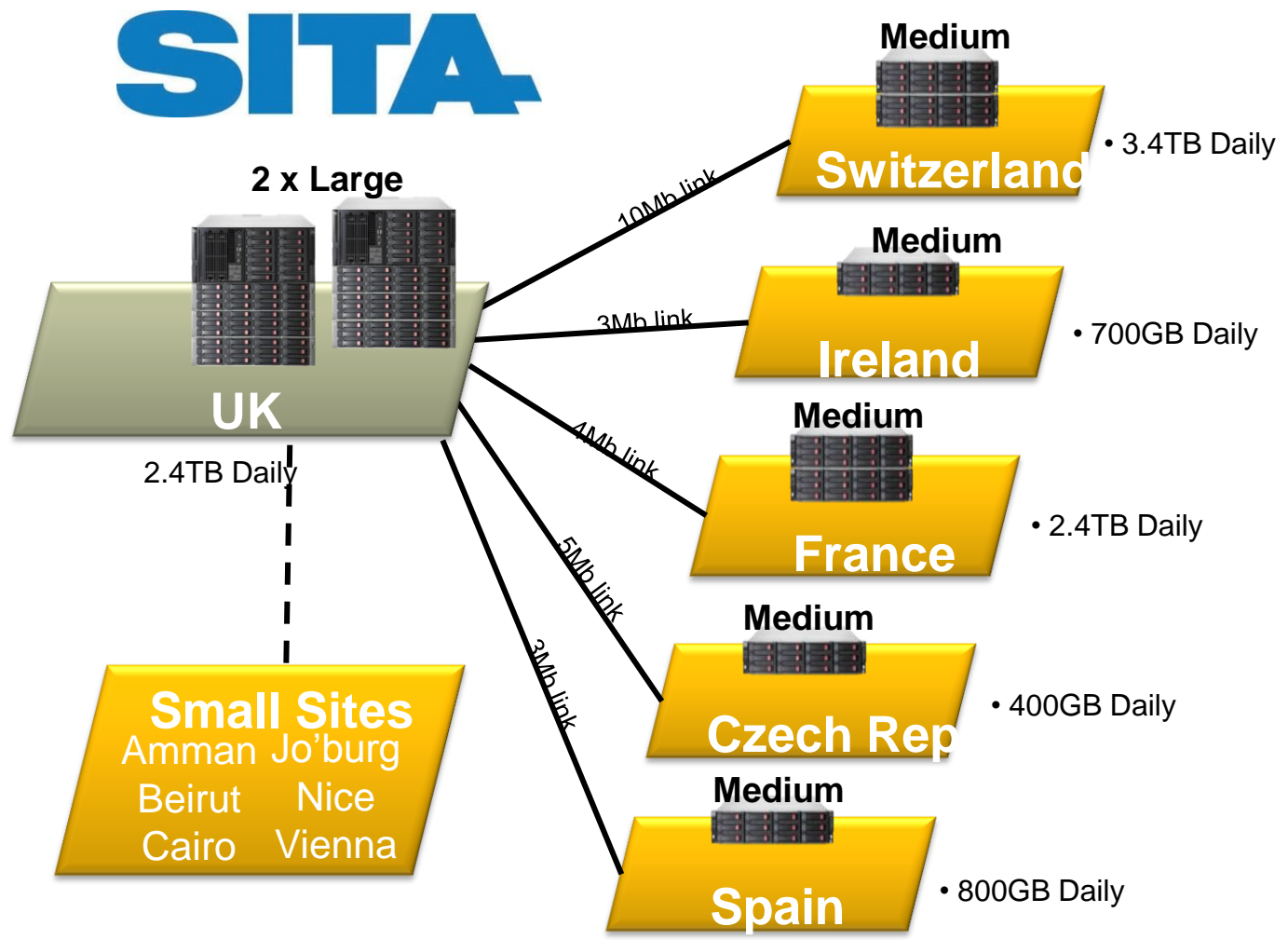
# After: Deduplication & Replication for disaster recovery

- Hands-free daily backup to D2D
- No maintenance required
- Cost-effective use of disk space
- Network efficient data replication
- Centralized management for all sites





# Real Customer Installation





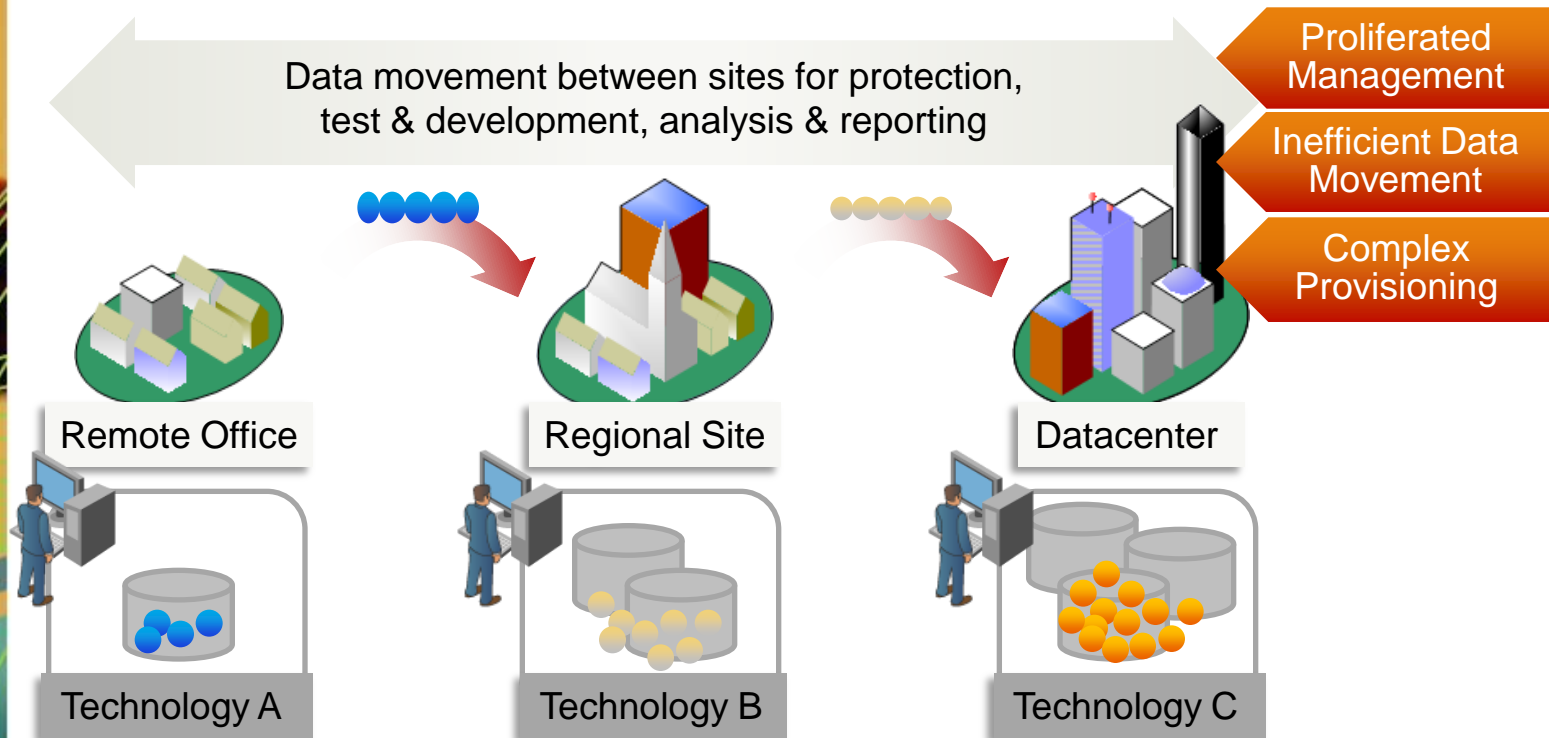


***What's Next?***



# 1<sup>st</sup> Generation Approaches to Deduplication

- All new technologies solve problems and create others
- Point solutions drive complexity

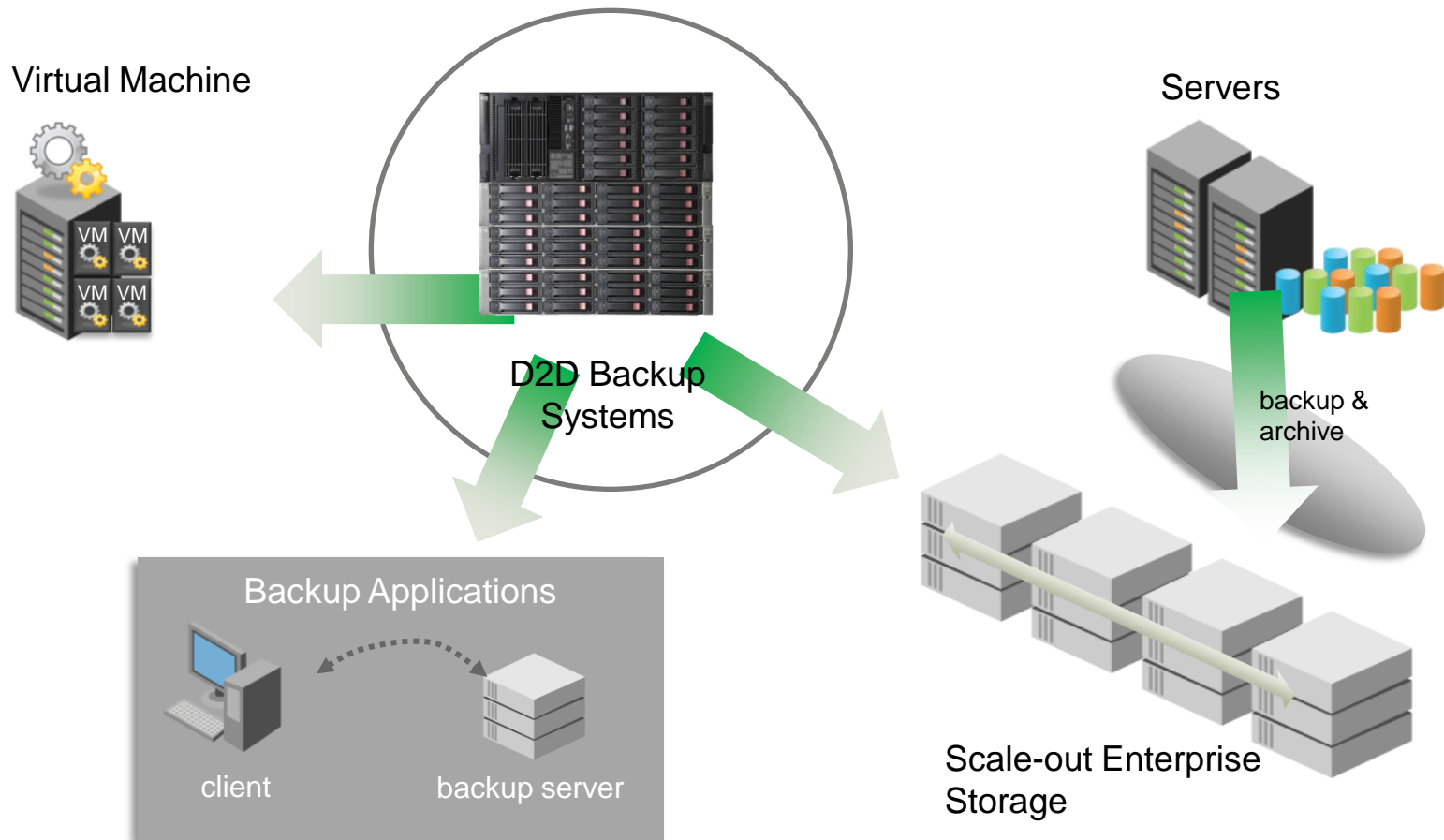




# Tomorrow's View

## Extending the technology

Modular architecture enables flexible deployment and integration

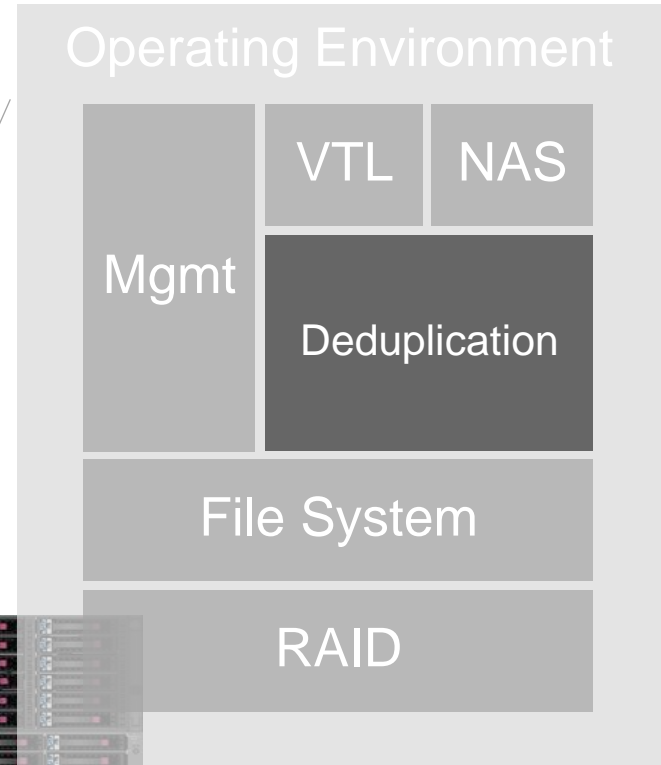




# Portable, modular design

Enables flexible deployment and integration

- Modular 64-bit architecture
- Well-defined interfaces
- Built using standard components
- User space application sitting on a standard file system
- No dependencies on file system, storage, or operating environment





# Flexible Deployment

- An “AND” discussion, not “OR”
- Emulation: VTL and NAS
- Connectivity: FC, iSCSI and Ethernet





# Want to learn more?

HP Booth (Booth #201) for live demos

- Deduplication
- Virtualization
- iSCSI
- Veeam
- ...and more!