



Evaluating Disk Backup with Deduplication: *Case Studies and Industry Information from the Frontlines*

Michael Krieger, Moderator
Robert Stevenson, TheInfoPro
Marc Crespi, ExaGrid Systems

July 30, 2009
2:00 p.m. Eastern / 11:00 a.m. Pacific
45 minutes



Hosted by



Today's Agenda

- Key IT insights from TheInfoPro's latest storage study
- Impact of deduplication now and in the future
- Important considerations when evaluating a solution for disk backup with data deduplication
- What role deduplication plays in achieving cost-effective backups that are fast and reliable
- Real-world customer examples of using disk backup with deduplication to greatly improve backup operations.
- Questions

Robert Stevenson
Managing Director of Storage Research
TheInfoPro



THEINFOPRO
The Voice of the Customer

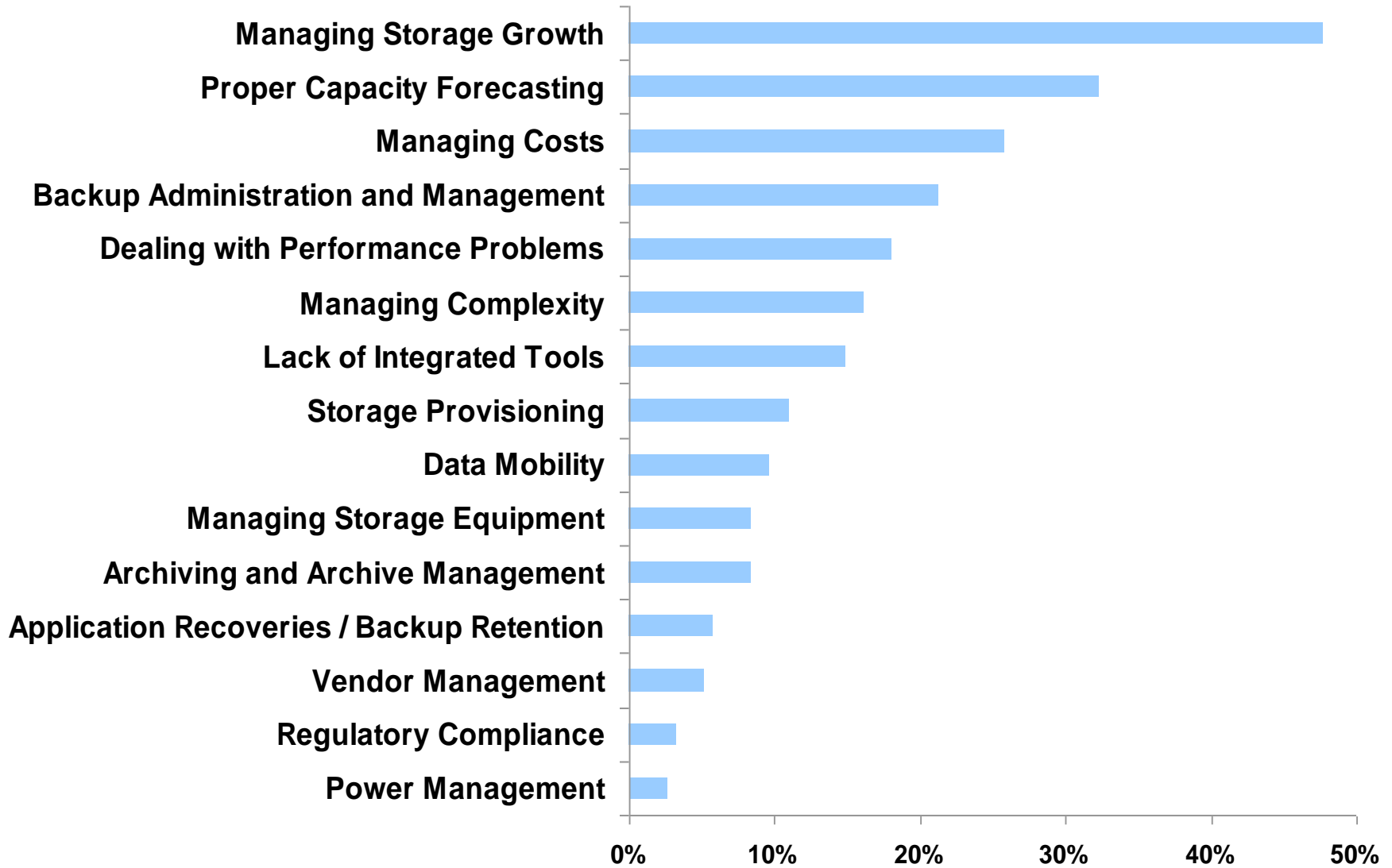
Deduplication – Adoption Moving Forward at a Rapid Pace

TheInfoPro™ Storage Study- Wave 12

TIP Network - IT Professionals (sampling)



Top Storage Professionals' Pain Points

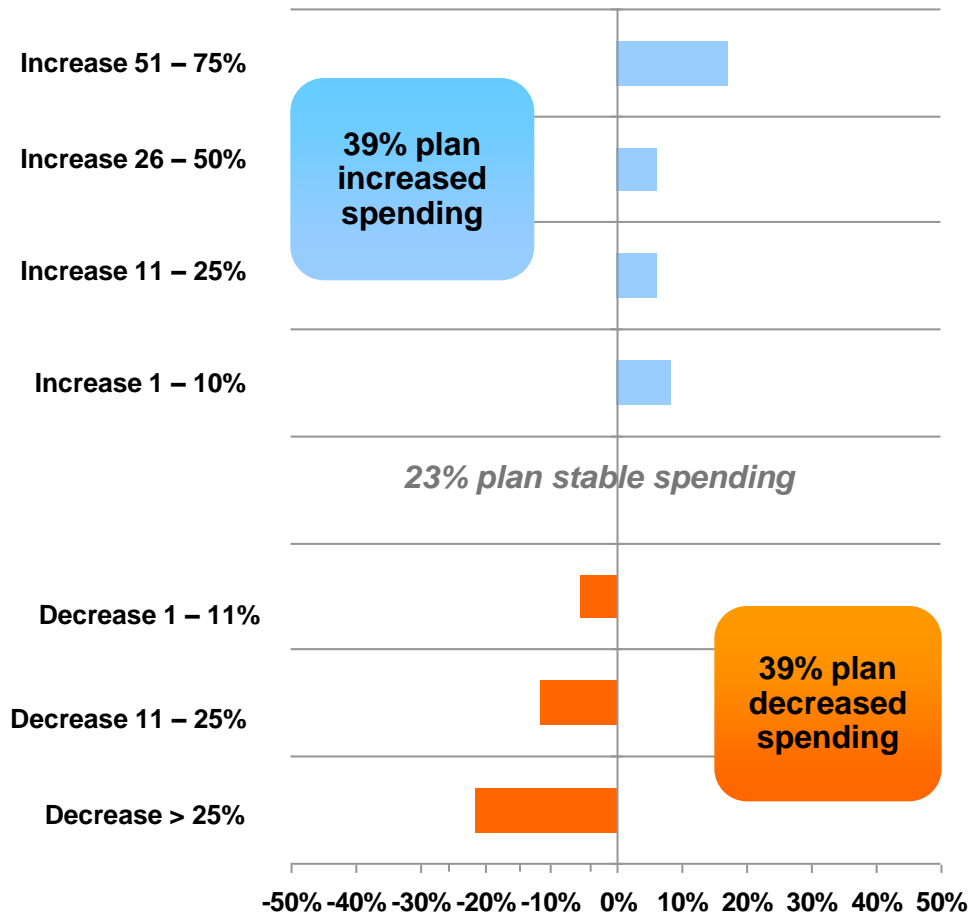


(6/10/09): F1000 Sample. n=155. * Note that due to multiple responses per interview, total may exceed 100%. Others n=3 with no answers

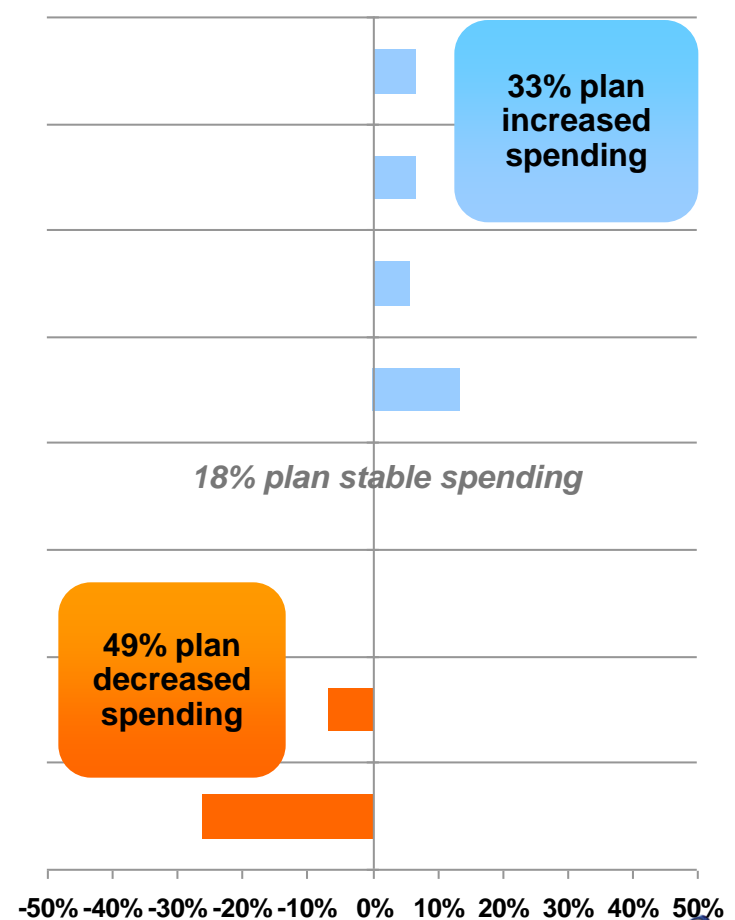
Change in Storage Budgets

How did your 2008 Storage budget spending change compared to 2007? Compared to 2008, how do you expect your Storage budget spending to change in 2009?

MSE 2009 vs. 2008

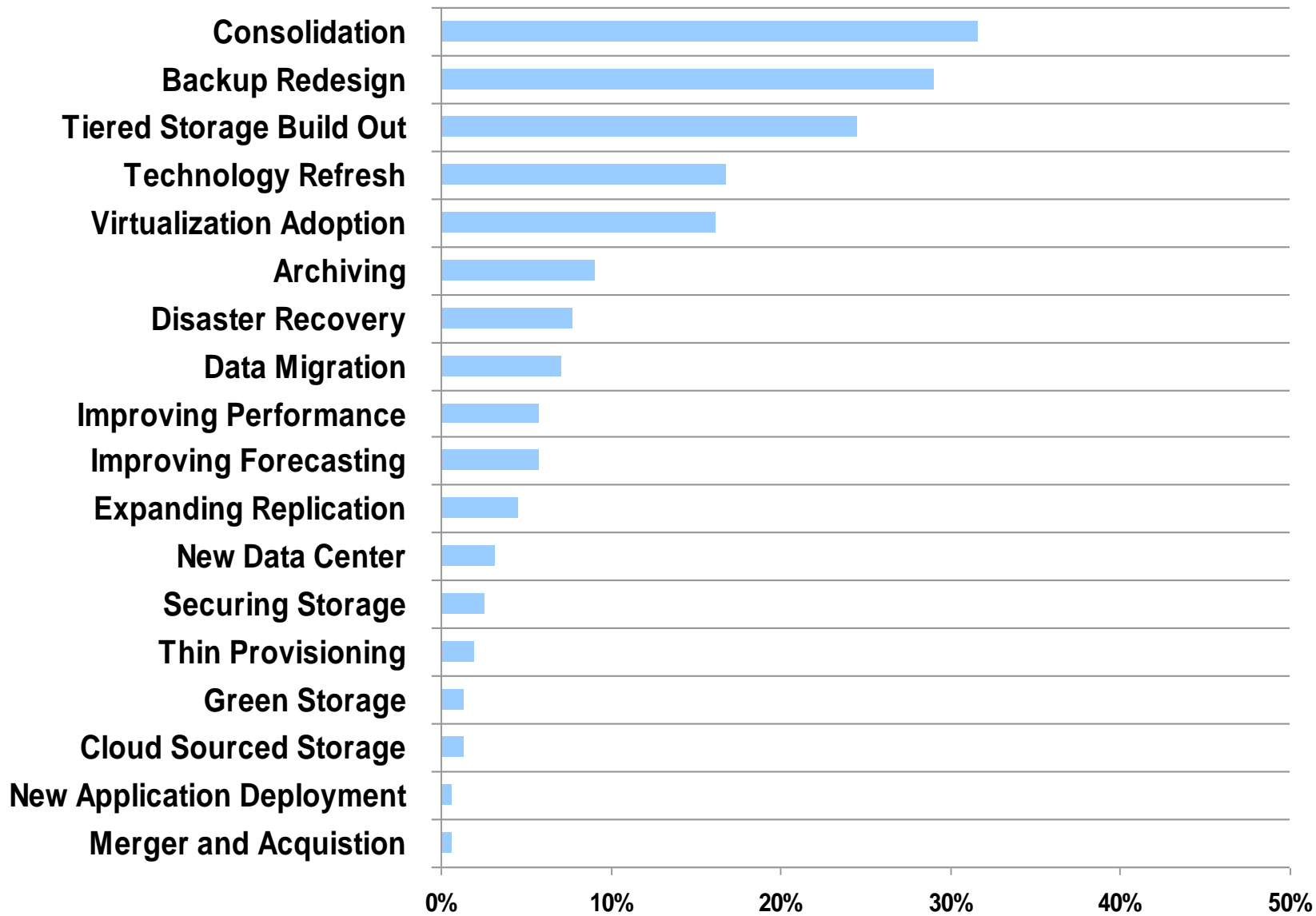


F1000 2009 vs. 2008



(5/22/09): F1000 Sample. Left Chart: n=99. Right Chart: n=153

Top Storage Team Projects for 2009

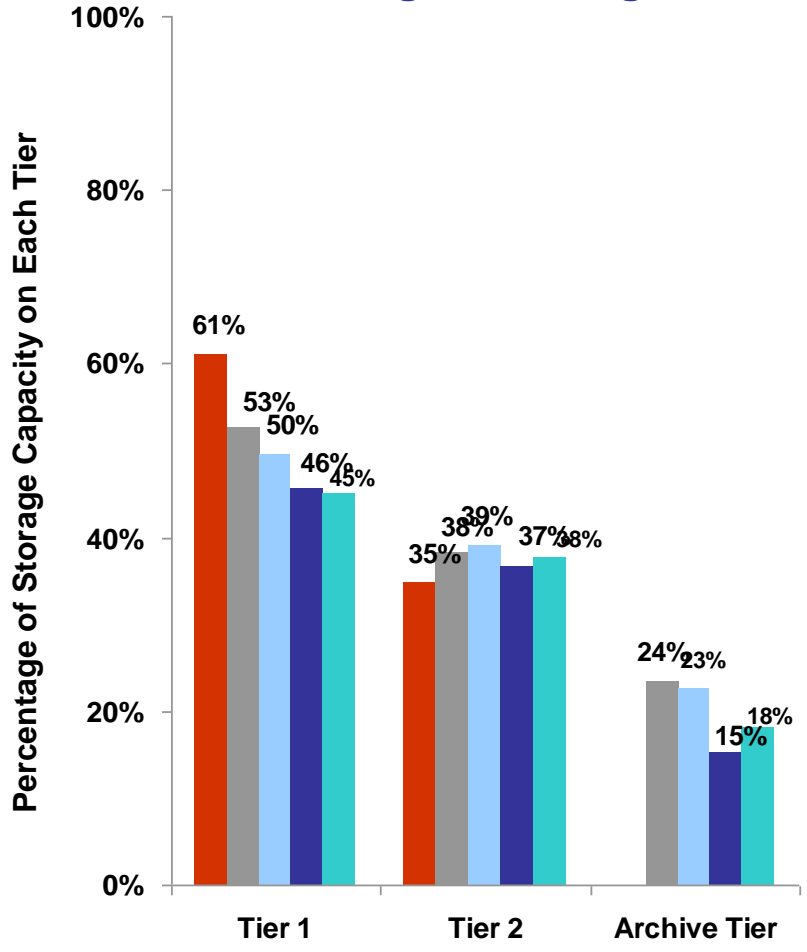


(6/3/09): F1000 Sample. n=155. * Note that due to multiple responses per interview, total *may exceed* 100%

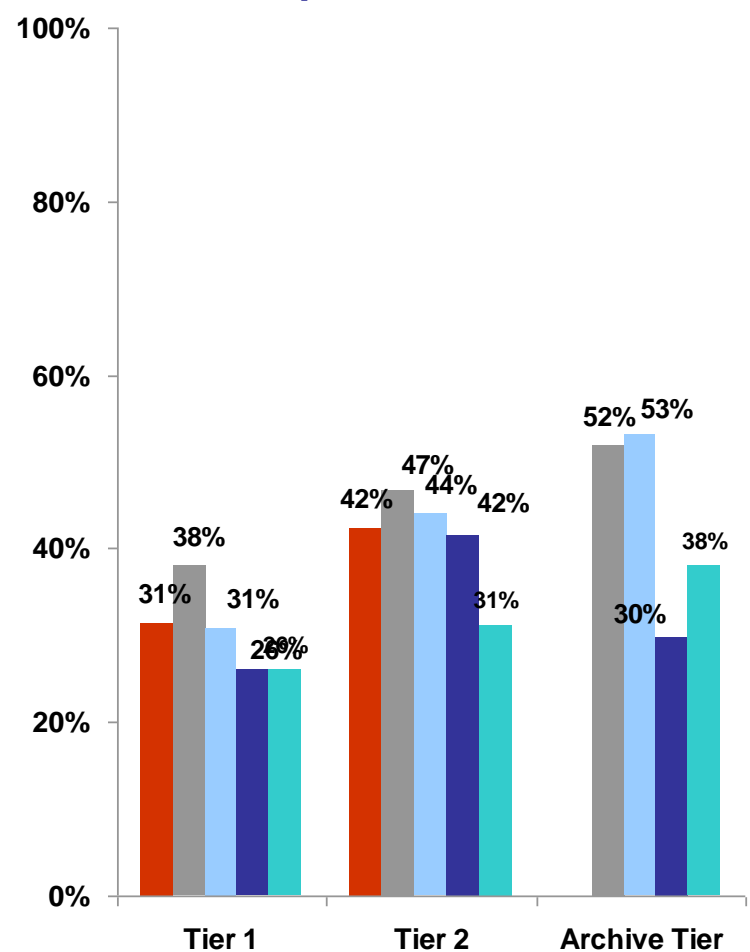
Present Tier 1, Tier 2, and Archive Tier Capacity

What percentage of your Storage capacity is on the following tiers?

Percentage of Storage on a Tier



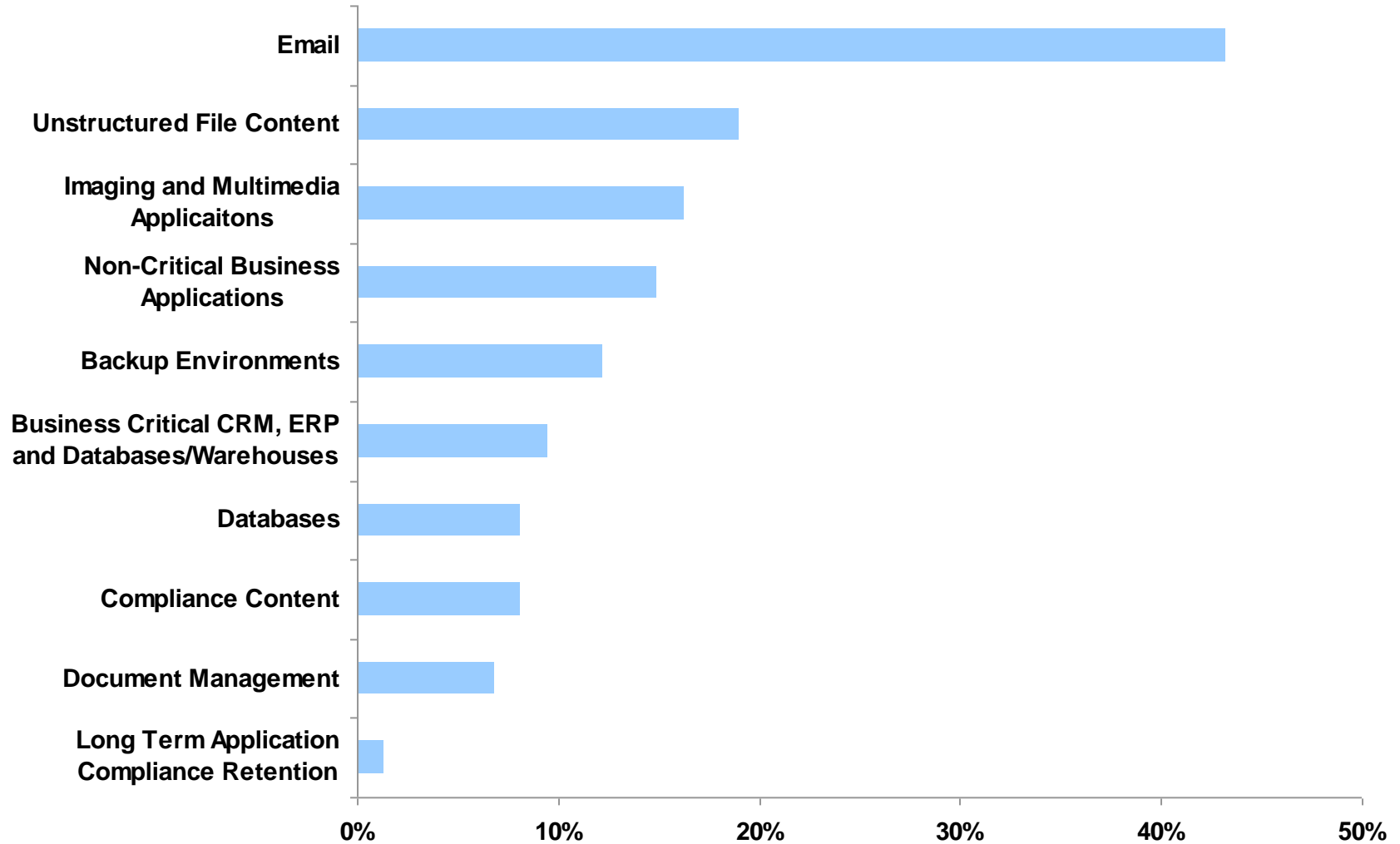
Anticipated Growth of Tiers



(5/29/09): F1000 Sample. W5 n=153, W6 n=155, W7 n=155, W8 n=153, W9 n=152, W10 n=152, W11 n=140, W12 n=156

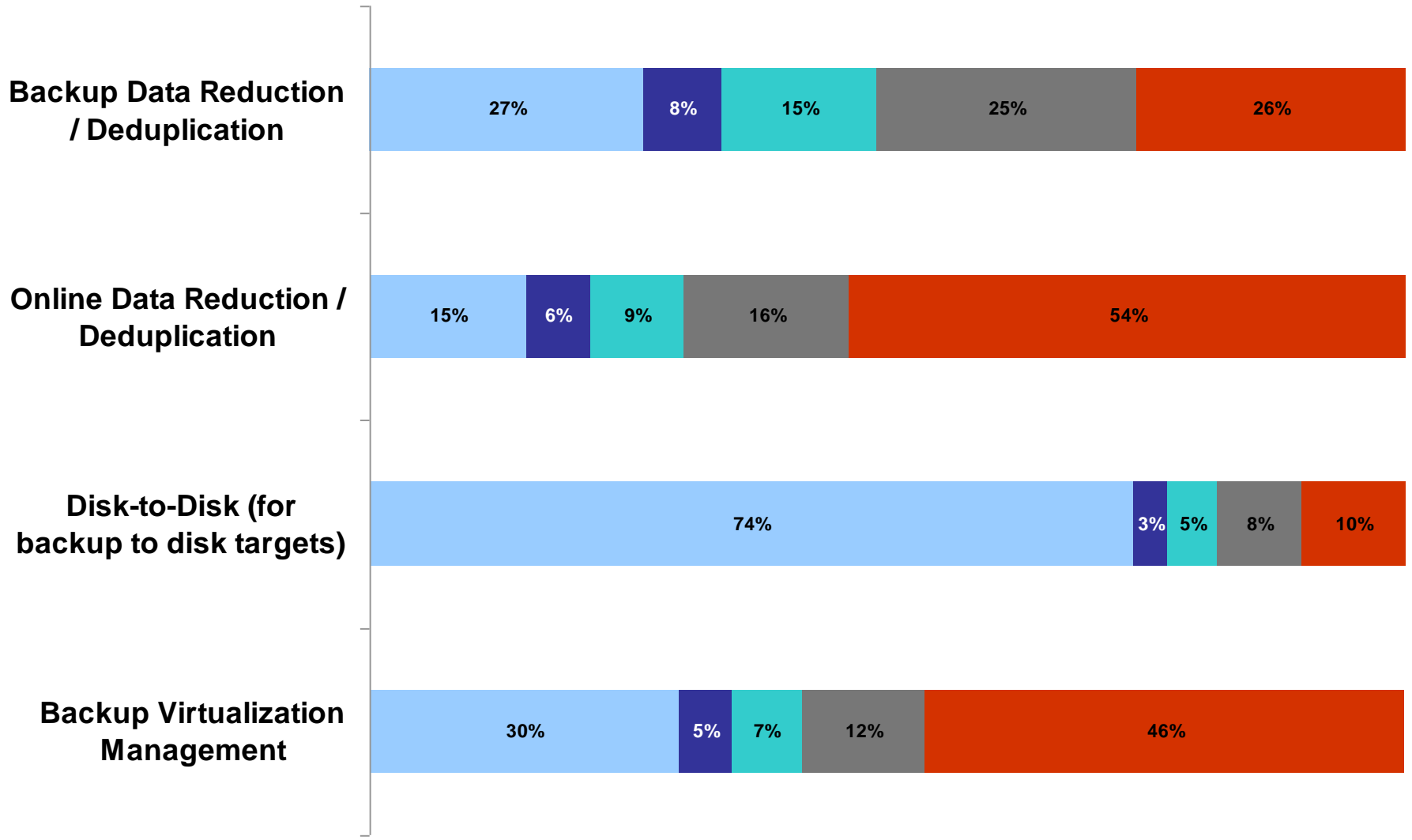
Applications for Archive Tier

What applications are targeted for your archive tier?



(6/5/09): Full Sample. n=74. * Note that due to multiple responses per interview, total may exceed 100%

Select Storage Backup Technology Trends at F1000



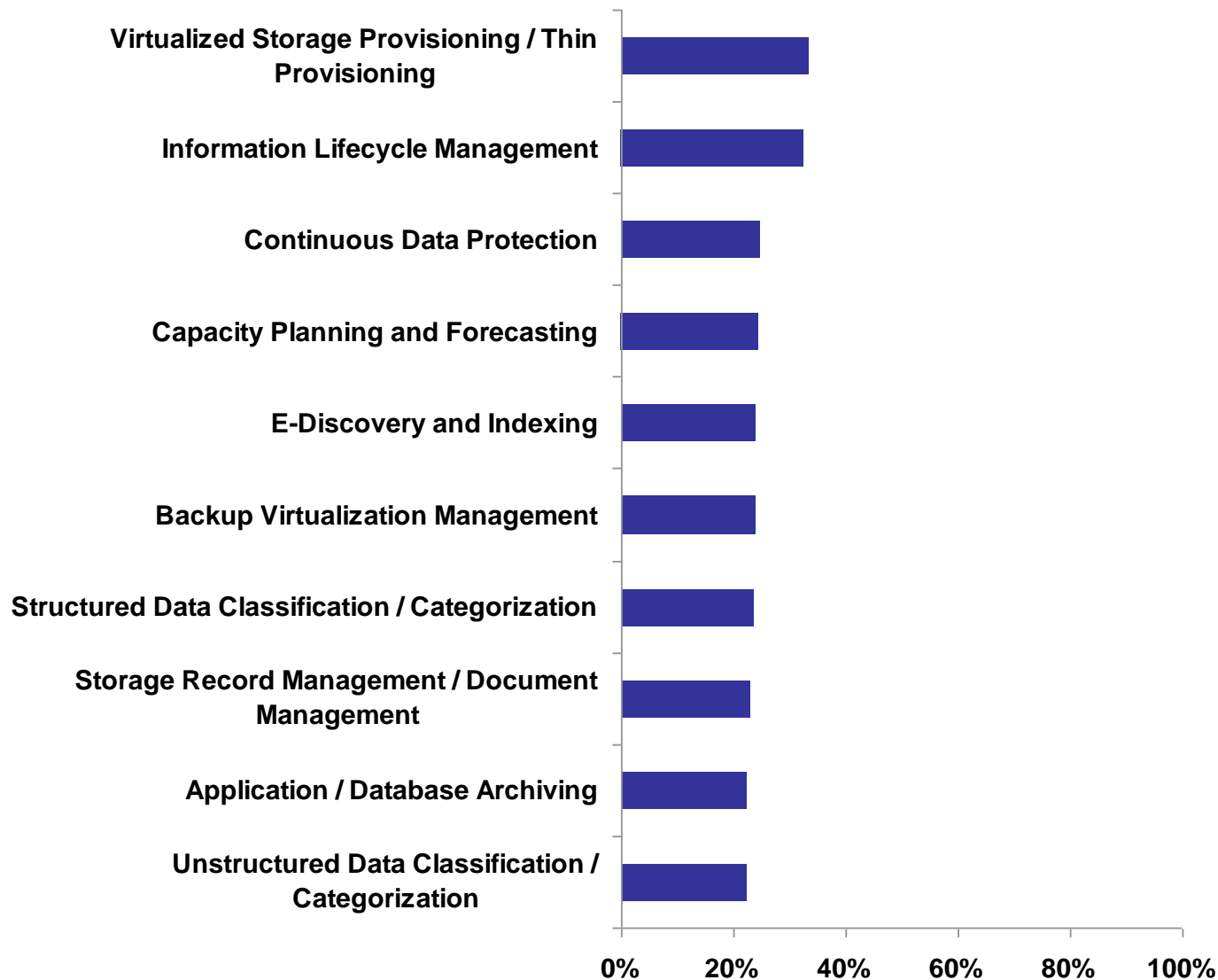
■ In Use Now (NOT including pilots)
 ■ In Pilot / Evaluation
 ■ In Near-term Plan (through Q2 '09)
 ■ In Long-term Plan (Q3 '09 - Q1 '10)
 ■ Not in Plan

Backup and Recovery Technology Heat Index®

(Gauges the Immediacy of User Needs and Planned Spending for Each Technology)

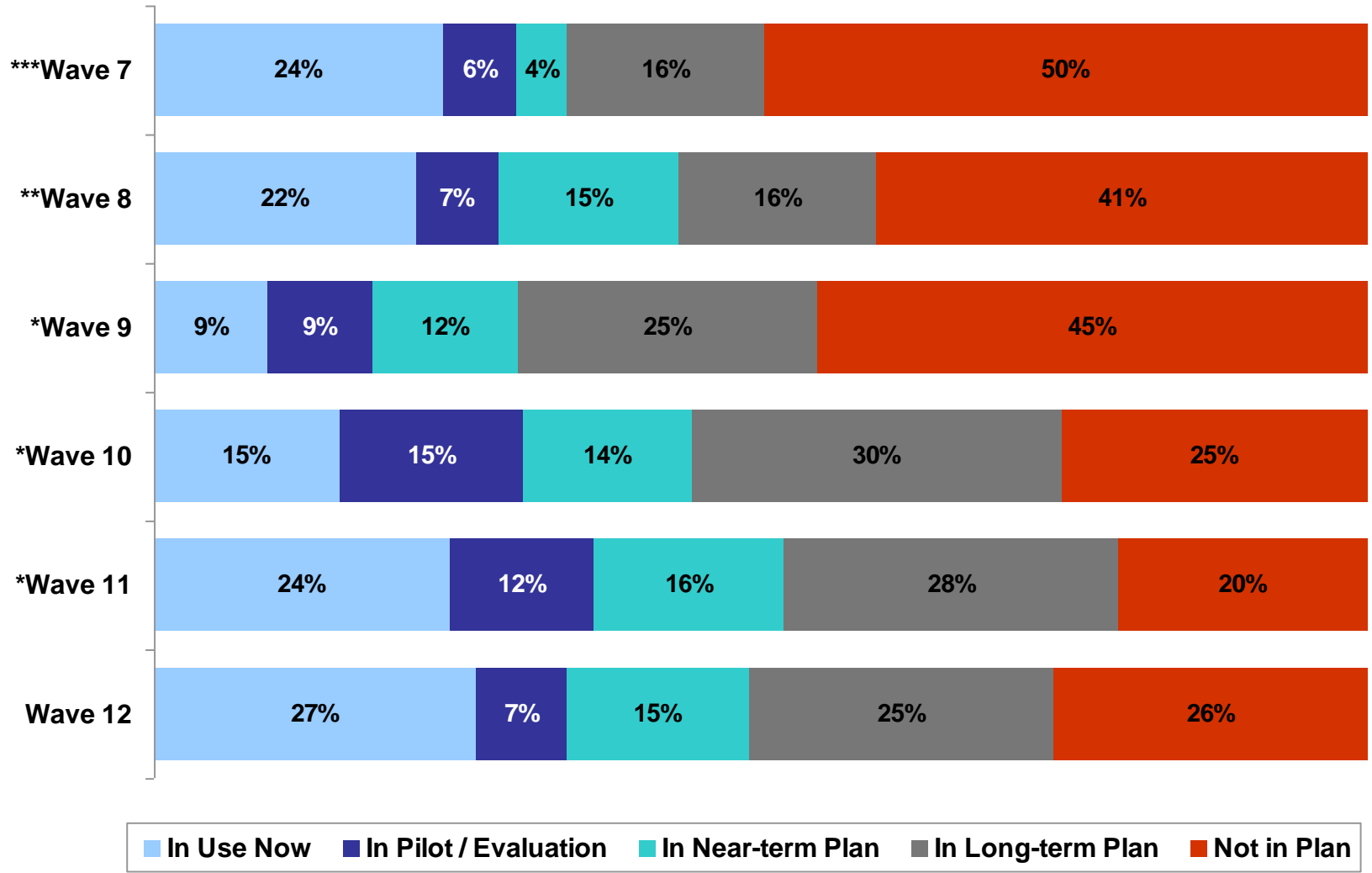
Rank	Technology	Heat Score
1	Backup Data Reduction / Deduplication	100
2	Backup Virtualization Management	60
3	Online Data Reduction / Deduplication	57
3	Email Archiving	57
5	Disk-to-Disk (for backup to disk targets)	56
6	Virtual Tape Library (VTL) for Open Systems	51
7	E-Discovery and Indexing	50
8	Tape Encryption	46
8	Application / Database Archiving	46
10	Continuous Data Protection	38

Top 10 Storage Management Technologies *in Consideration*



Backup Data Reduction / Deduplication – Implementation

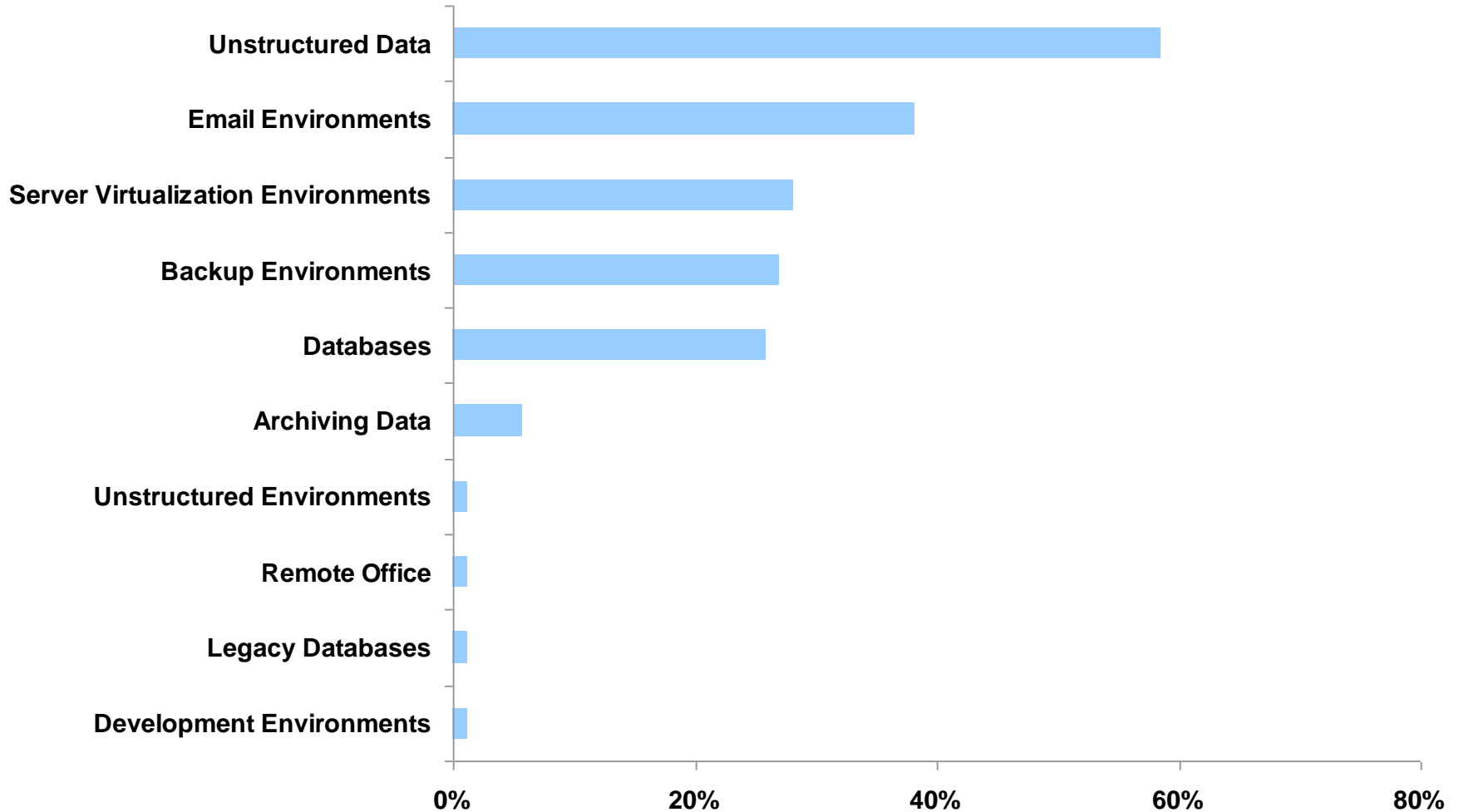
Heat Index Rank: 1
Storage Networking
Wave 12



(6/3/09): F1000 Sample. Wave 7 n=147, Wave 8 n=148, Wave 9 n=150, Wave 10 n=151, Wave 11 n=127, Wave 12 n=147.
 *Technology was previously categorized as Deduplication. **Technology was previously categorized as De-Duplication / Capacity Optimized Storage / Single Backup Instance Store. ***Technology was previously categorized as Single Backup Instance Store Software

Applications for Deduplication

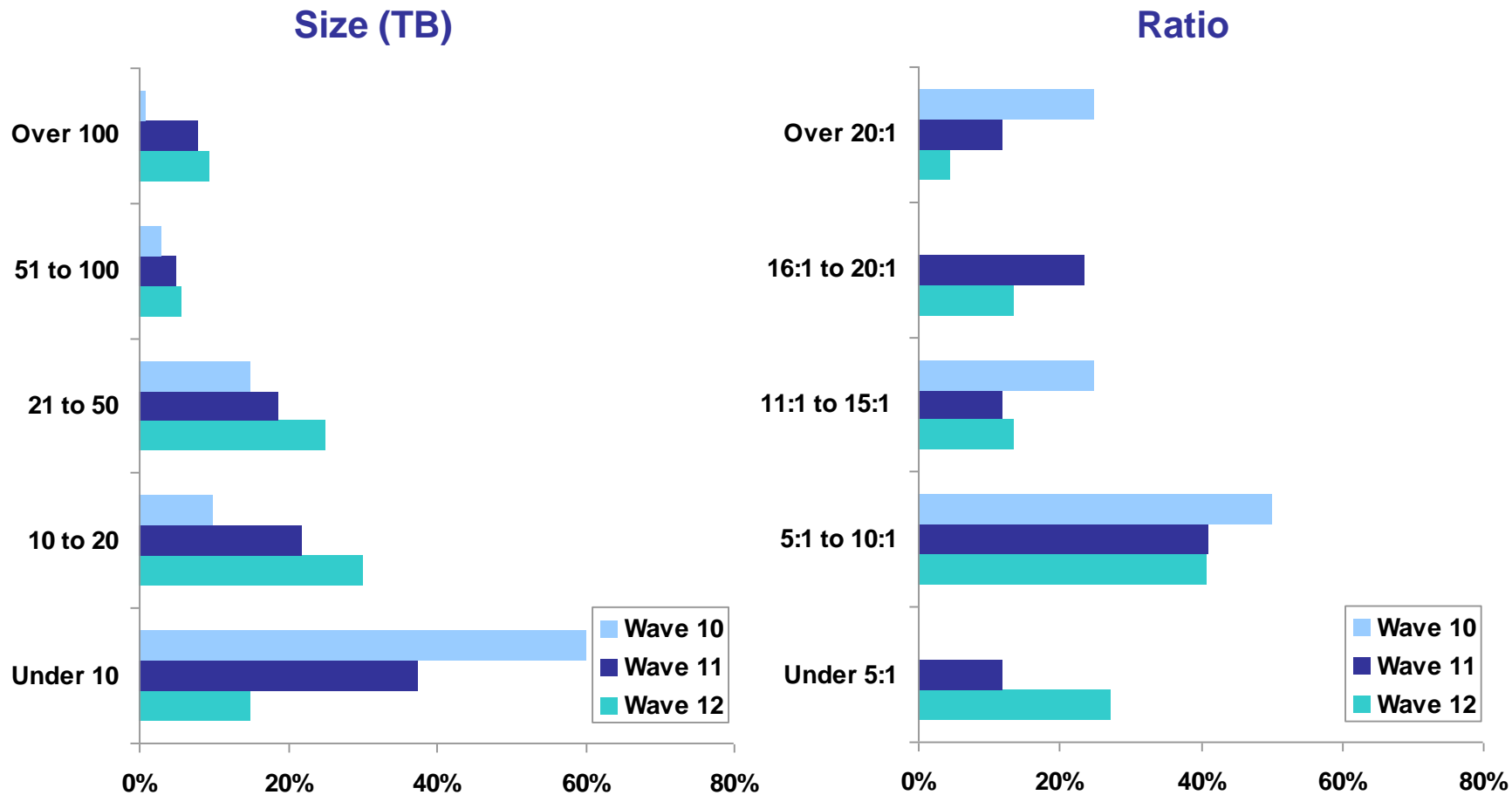
What top three applications are best suited for deduplication?



(6/10/09): Full Sample. n=89. * Note that due to multiple responses per interview, total may exceed 100%

Deduplication Repository Size and Realization Ratio

How large is your current deduplication repository (TB)? What is your deduplication realization (e.g., 7:1, 20:1, etc.)?



(6/10/09): Full Sample. Left Chart: Wave 10 n=10, Wave 11 n=32, Wave 12 n=53. Right Chart: Wave 10 n=4, Wave 11 n=34, Wave 12 n=45

In Summary

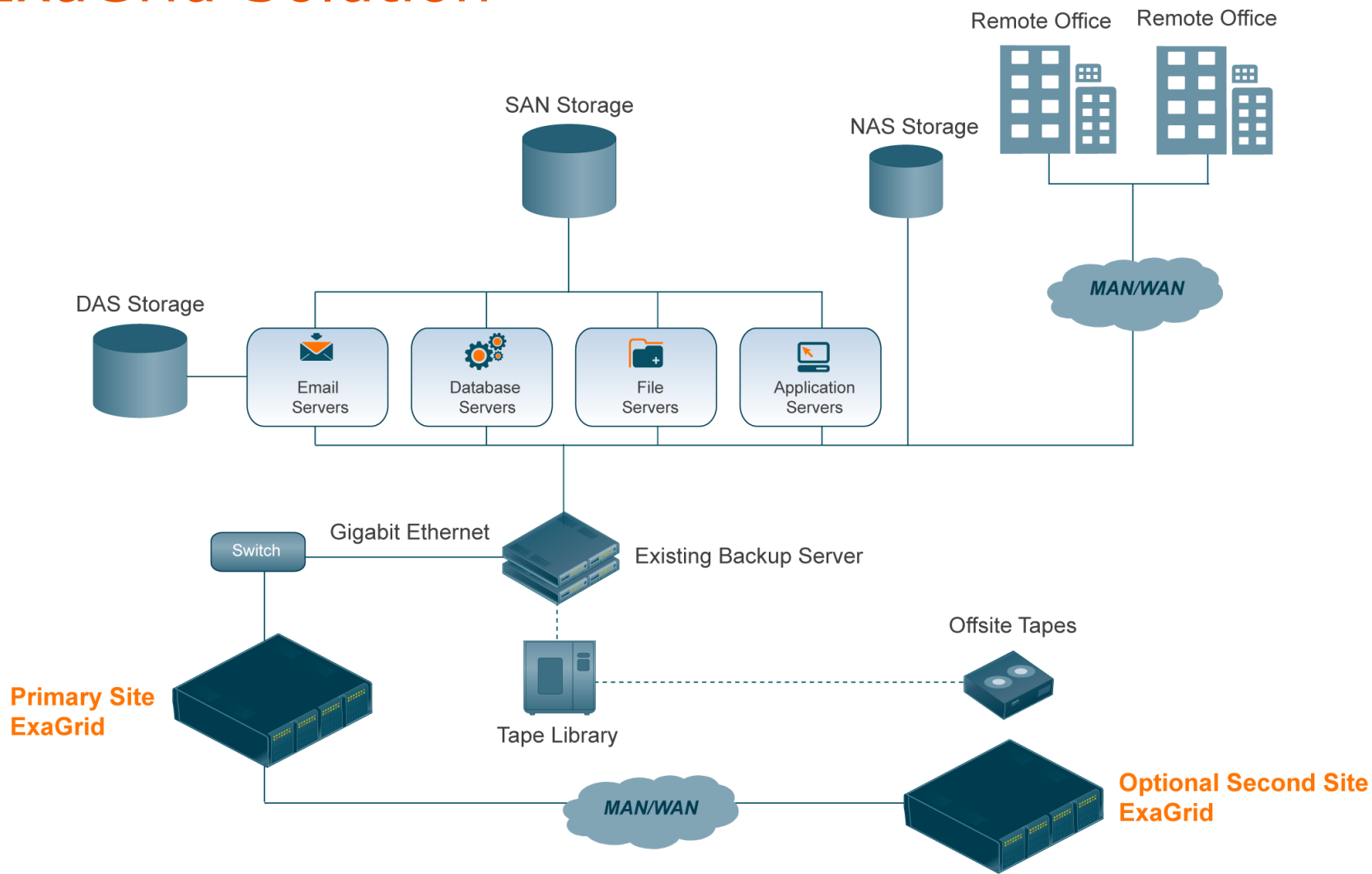
Consolidate Intelligently

**Minimizing Organization
Friction with Data Movement**

**Ensure Backup Scales with
Storage Growth**

Marc Crespi
Vice President of Product Management
ExaGrid Systems

ExaGrid Solution



ExaGrid Appliance

Best-in-Class Components

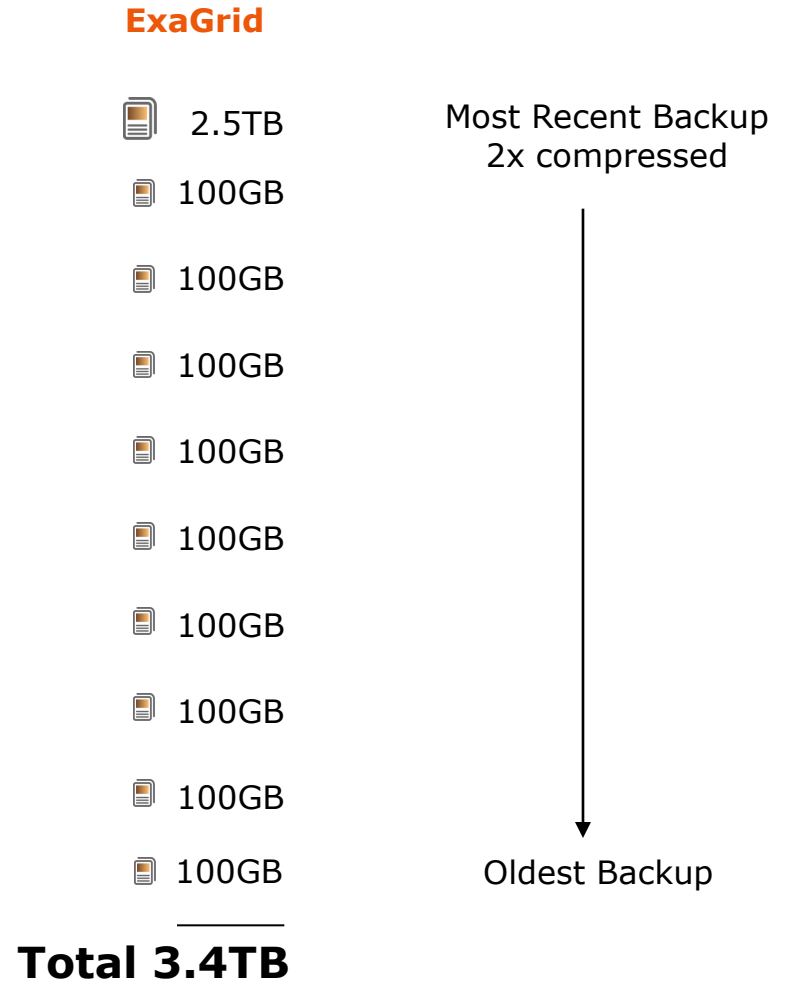
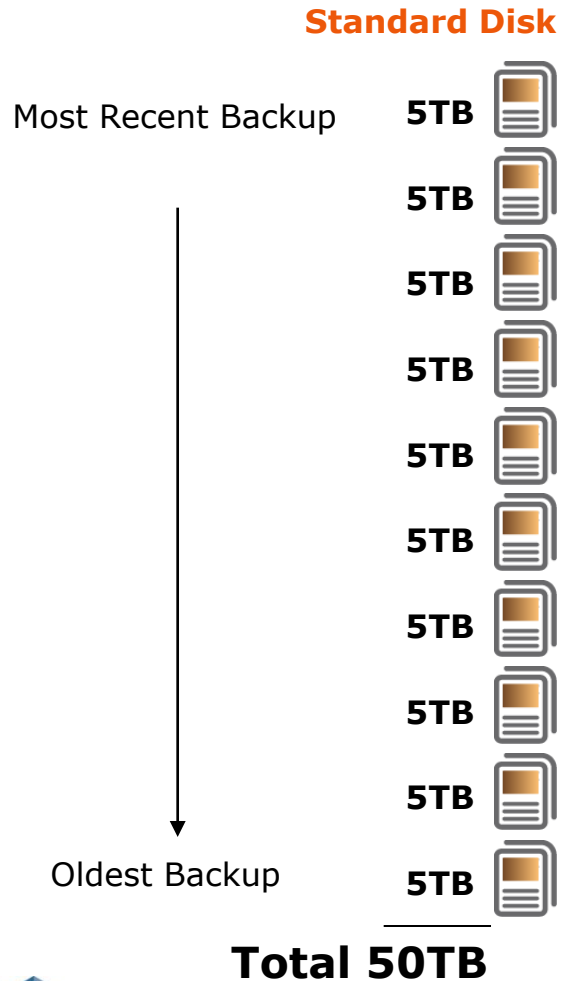
- ❑ Intel processors
- ❑ Seagate or Western Digital SATA drives
- ❑ LSI RAID 6 with hot spare
- ❑ ExaGrid 2nd generation software



EX1000
EX2000
EX3000
EX4000
EX5000

Mix-and-match servers in
a GRID for up to a 30TB
full backup plus retention
per GRID system

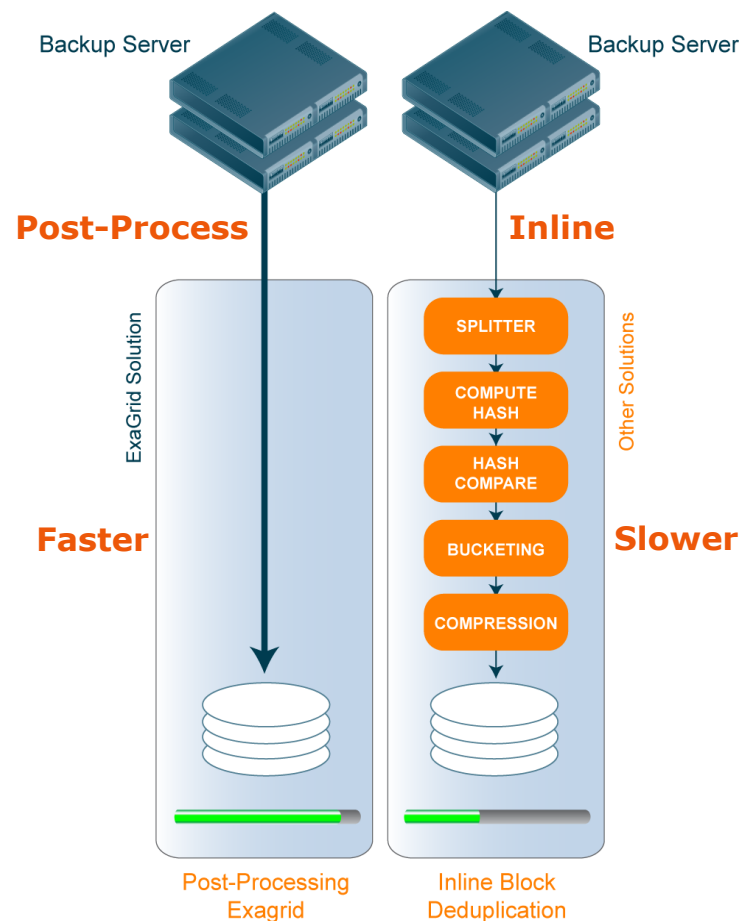
Store Only The Bytes That Change



Post-Process Provides Fastest Backups

Fastest Backups

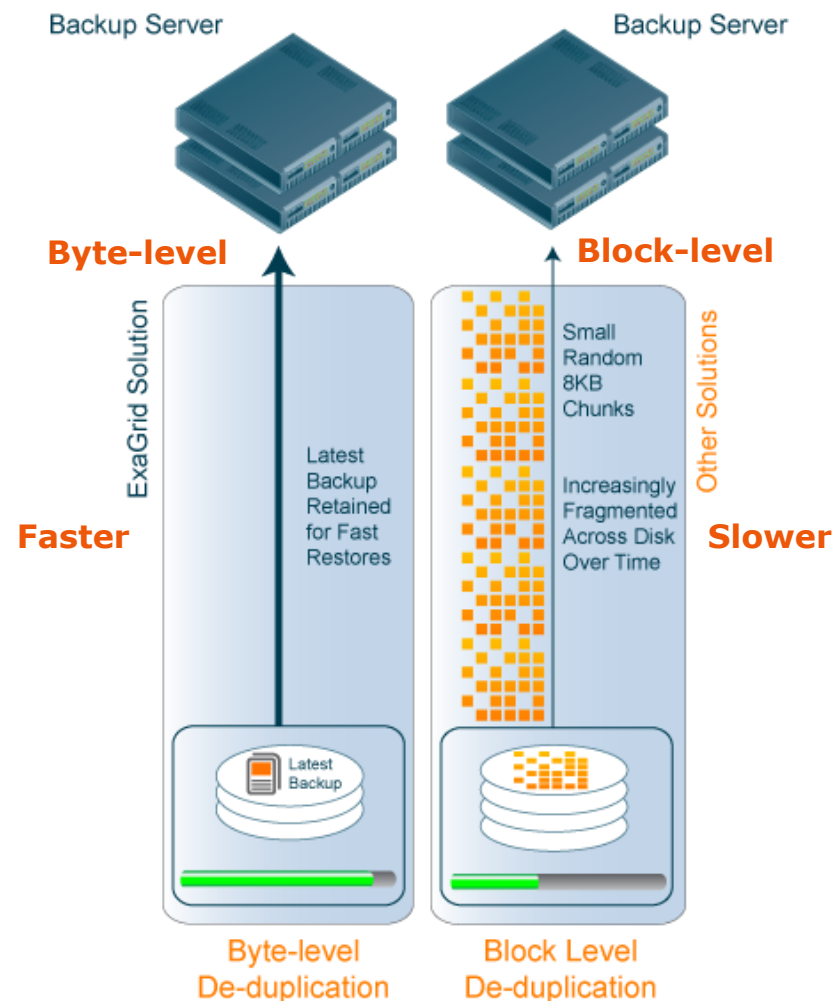
- ❑ Backup directly to disk
- ❑ Utilizes unique landing zone architecture
- ❑ Post process compression and byte level data de-duplication
- ❑ No in-line processes to slow backups down
- ❑ Shortest possible backup window



Architected for Fastest Restore

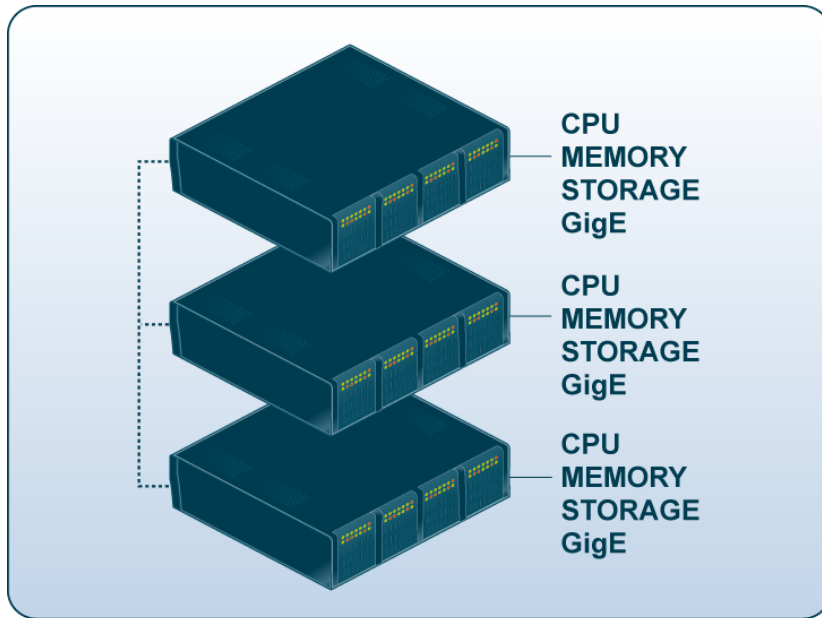
90% of restores are from the latest backup

- ❑ ExaGrid stores the latest backup in its complete form ready for instant restore
- ❑ Restores from landing zone
- ❑ For restores of earlier versions byte-level changes are simply and quickly merged into the latest backup
- ❑ Allow for simultaneous restore jobs
- ❑ Allows fast offsite tape copy
- ❑ Instant DR capability provides fastest possible restores for local and off-site data copies



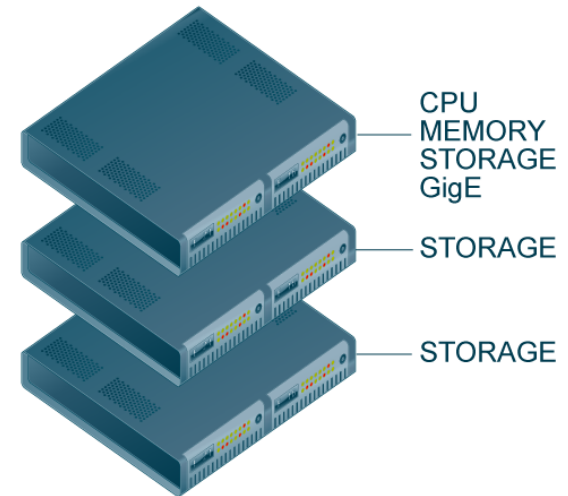
ExaGrid GRID Scalability – Capacity and Performance

ExaGrid



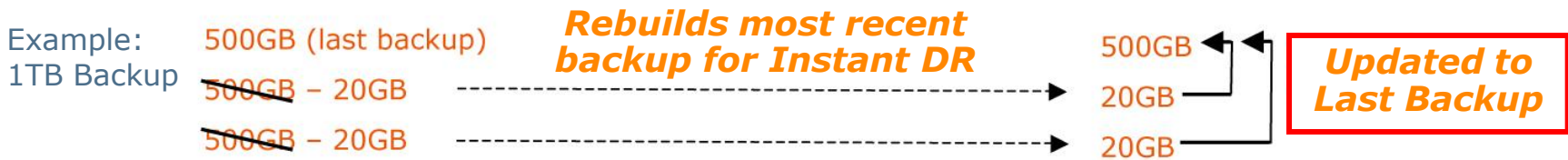
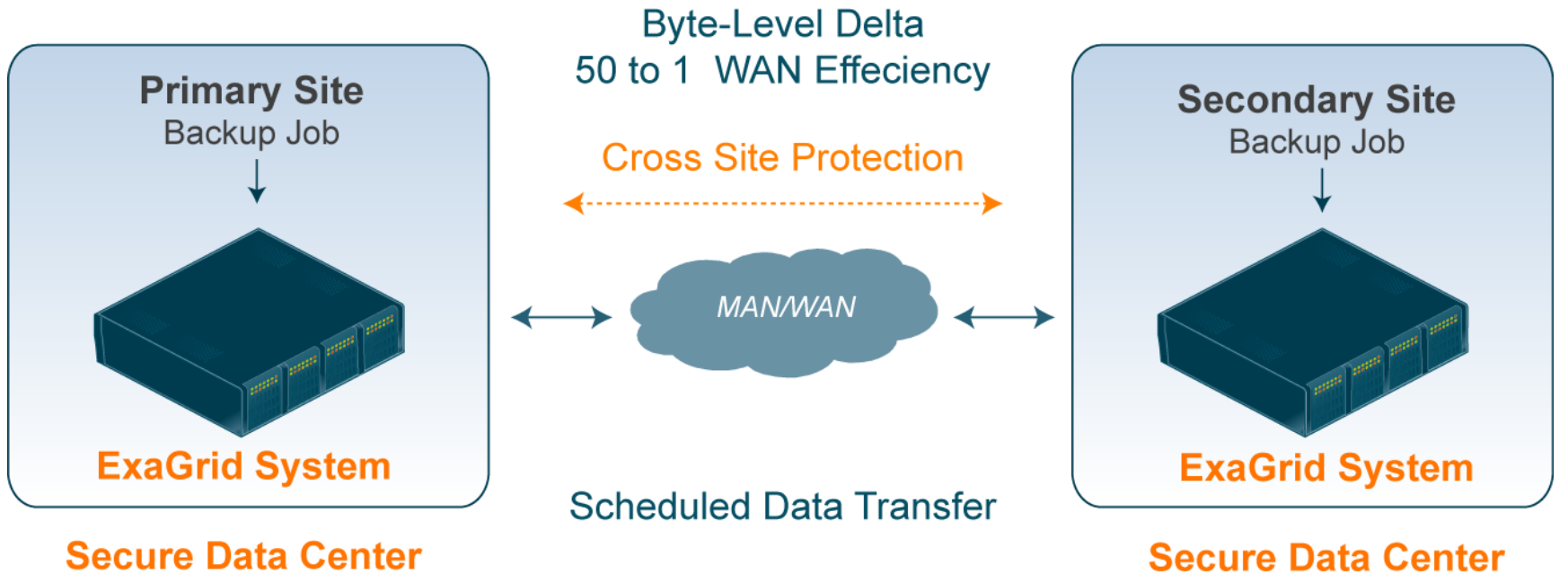
Virtualized GRID Architecture

Processor Head with Disk Shelves



Additional Disk Shelves

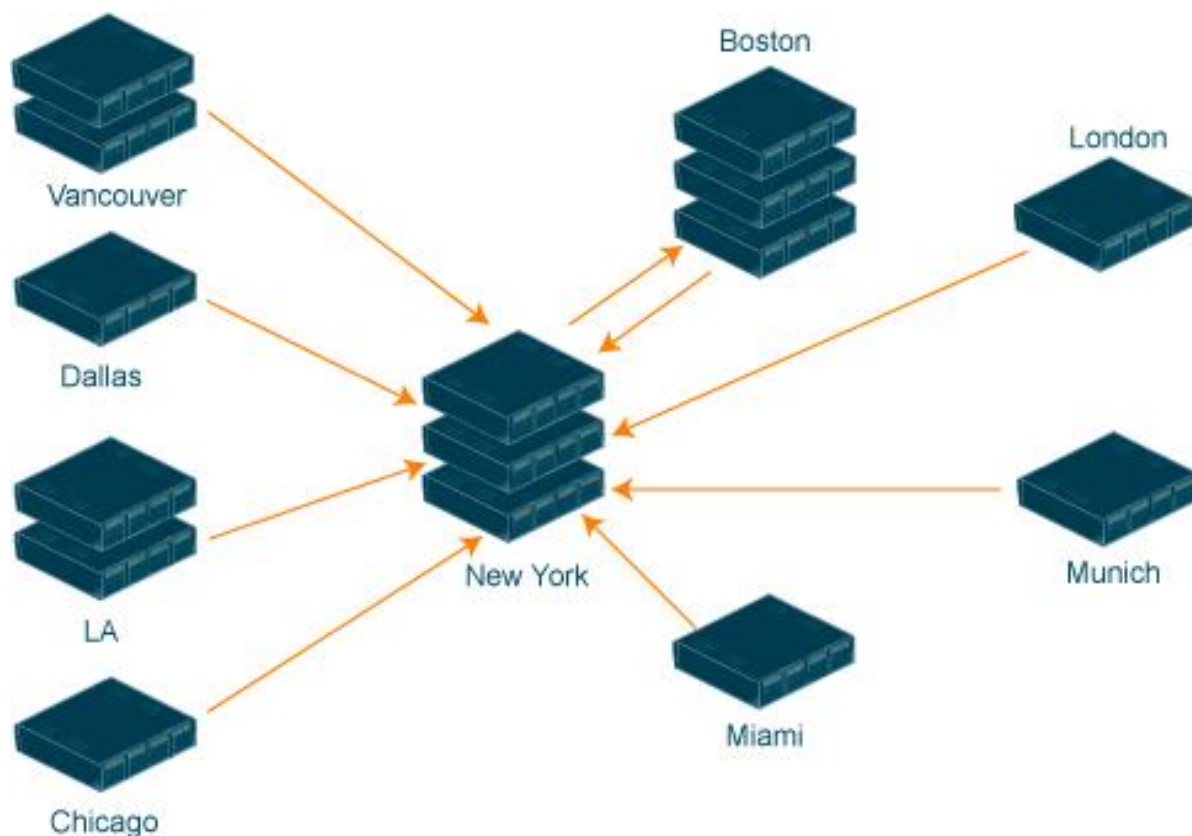
On-site and Off-site Tape Replacement



On-site and Off-site Systems are Identical

For each 1TB full, WAN bandwidth of 3 mbps is required (assuming a 2% byte change rate)

Multi-site Data Center Disk Backup Topology



ExaGrid System Architecture

- 50 to 1 WAN efficiency across all sites
- Manage entire environment from single UI
- Cross-site protection available between major sites
- Reduce costs by consolidating DR copies of backup data

Unified Management Console

The screenshot displays the ExaGrid Unified Management Console interface. On the left is a navigation tree showing a hierarchy of sites: ExaGrid, Site:New York (with sub-items NYC1:EX5000 through NYC6:EX5000), Site:Boston (BOS1:EX5000), Site:Seattle (SEA1:EX1000), Site:Trenton (TRN1:EX2000), Site:Dallas (DAL1:EX5000), Site:Omaha (OMH1:EX1000), Site:Pittsburgh (PIT1:EX3000), Site:Tampa (TAM1:EX1000), and Site:Albany (ALB1:EX4000). A 'Refresh' button is next to the ExaGrid node.

The main content area shows a 'Summary for Site : New York' with an 'Export' button. Below this is the 'Landing Space for Site : New York' section, which includes a table:

Server	Available			Deduplication Status
	Actual	% Available for Next Backup	Maximum	
New York	30,000.00 GB	100%	30,000.00 GB	In Progress

Below the landing space is the 'Retention Space for Site : New York' section, which includes a table:

Server	Retention Space			Total Capacity	Over Capacity	Replication Status
	Available	Consumed	% Available			
New York	23,413.52 GB	4,473.67 GB	89%	27,887.19 GB	0.00 GB	In Progress

Finally, the 'Duplication Summary for Site : New York' section shows a table for 'In Progress' data:

Share	Deduplication Ratio	Total Backup Data	Space Consumed
New York	33.6 : 1	83,774.44 GB	4,473.67 GB
NYC1	✓ 16.1 : 1	16,588.17 GB	1,731.51 GB
SQL1	✓ 7.1 : 1	10,692.79 GB	1,505.61 GB
SQL2	✓ 26.1 : 1	5,895.38 GB	225.90 GB
NYC2	✓ 9.01 : 1	9,155.28 GB	1,016.40 GB
Exchange	✓ 9.01 : 1	9,155.28 GB	1,016.40 GB
NYC3	⚡ 43.78 : 1	9,265.80 GB	211.66 GB
VMWare	⚡ 43.78 : 1	9,265.80 GB	211.66 GB
NYC4	✓ 20.33 : 1	35,766.63 GB	1,759.17 GB
FileSystem1	✓ 30.44 : 1	21,485.09 GB	705.90 GB
FileSystem2	✓ 13.56 : 1	14,281.54 GB	1,053.27 GB

ExaGrid UI

- Single interface across all systems
- Simple dashboard gives quick view

Backup Job Aware Reporting

EXAGRID

Reports Manage Support

Summary Servers Network Status Activity Deduplication Progress Replication

Deduplication Progress Report

Show Backup Jobs : organized by Shares

All Jobs by Share

Name	Size	Last Modified	Status	Deduplication	
				Ratio	Size
Site: London					
Server: tc200s1ob1					
Share: Taneja-1A (Netbackup)					
Backup Job: E:\OfficeSet-shroperf2 on shroperf2					
Instance: (Full) 4-29-2008 10:35am	27846	4-29-2008 10:28am	✓ 4-29-2008 10:43am	119.51 : 1	233
Backup Job: E:\OfficeSet-perfdriver9 on perfdriver9					
Instance: (Full) 4-29-2008 10:43am	27846	4-29-2008 10:28am	✓ 4-29-2008 11:06am	119.51 : 1	233
Backup Job: E:\OfficeSet-shroperf2 on shroperf2					
Instance: (Full) 4-29-2008 12:20pm	27846	4-29-2008 12:11pm	✓ 4-29-2008 12:33pm	2.25 : 1	12375

Reports

- Backup job view ties to backup application
- De-duplication ratio by backup job
- Replication progress and status by backup job

Customer Successes

Medina General Hospital

- 118-bed healthcare facility in Medina, Ohio.
- Employs nearly 1000 to offer 24-hour emergency services and serves more than 31,000 people a year,
- Delivers almost 900 babies annually.
- MGH's Brunswick Campus offers a variety of medical services, such as an Immediate Care Center, an imaging center and physical and occupational therapy



Medina General Hospital's Challenges

- Long standing practice of backing up everything (servers, databases, files, Exchange, etc) to one server then backing that server to tape
- Full backups each weekend and ship tapes offsite for DR
- Backups exceeding window
 - Process over 12 hours to tape
 - Bad tapes would mean restart entire process
- Needed a better solution for disaster recovery

Medina General Hospital's Solution

- Establish off-site DR to disk
- Utilize existing facilities to keep costs low
 - Operates Brunswick, OH campus 15 miles from main facility
 - Dark fiber already in place to exchange radiology files between facilities
- Needed to work with existing backup application (Symantec Backup Exec)
- Chose ExaGrid Systems Disk-based backup

Medina General Hospital's Success and Results

- Weekly full backups reduced from 12 hours to 4
- Differential backups only take 20 minutes
- Achieved HIPAA compliance
- Bonus: ExaGrid's ability to easily expand means MGH can add capacity as regulatory or other needs change
- Combination with Backup Exec is highly cost-effective, helping HGH IT stay within ever-tightening budgets

"Because the ExaGrid compresses our data and deduplicates it at the byte level, we are able to maximize our retention."

*Michael Skrant
MIS Operations Supervisor
Medina General Hospital*

Thomas & King

- 8th largest restaurant franchisee in the U.S.
- Owns and operates restaurants in Arizona, Indiana, Kentucky, Ohio and Pennsylvania
 - 89 Applebee's restaurants
 - 6 Johnny Carino's Italian Grill restaurants
- Over 7,500 employees
- Based in Lexington, Kentucky



Thomas & King's Challenges

- Long backup windows
 - Full backups took 26-30 hours
- Unreliable Backups
 - No confidence in tape for restores
- Tape Management
 - Time consuming and expensive
 - Affected productivity

Thomas & King's Solution

- Disk-based backup with data deduplication
 - Dedicated appliance, independent of SAN
- Seamlessly integrated with existing backup application
- ExaGrid system backed up to tape once a month for archival purposes

Thomas & King's Success and Results

- **Reduced Backup Window**
 - Backup Times Cut in Half
 - *Previously spent 26-30 hours on the weekend backing up data*
- **More Reliable Backups**
 - Full backups completed each and every night
 - Faster restores in just seconds
- **Reduced Reliance on Tape**
 - ExaGrid backed up to tape 1x month
 - Less time spent managing and administering tapes
 - *Previously spent 45 minutes per week or 1 business week per year*
- **Made Backups More Efficient**
 - VMware images backed up directly from Vizioncore's vRanger Pro
 - *3 full VM backups now reduced by over 6:1*

"ExaGrid also fit in well with our existing backup solution, and it took away a lot of the routine, manual tasks that we ordinarily would have to do with tape."

*Jonathan DeMersseman
Manager, Infrastructure
Services
Thomas & King*

Why ExaGrid Disk Backup with Deduplication

- ✓ Disk that is easy to install, use and manage
- ✓ No hardware configuration or complex implementation
- ✓ Greater ability to handle rapid data growth
- ✓ Data de-duplication cuts the cost of disk by up to 95%
- ✓ Dramatically improve performance and reliability of backups and restores
- ✓ Reduce the backup window by 30% to 90%
- ✓ Better offsite strategy for DR (both with tape or tape replacement)
- ✓ Seamlessly scale as data volume or retention grows
- ✓ Better visibility via unified console and backup aware reporting

Questions?

Attendee Services

- Download a copy of today's presentation
- Provide your feedback! Please complete our survey
- View our calendar of upcoming events
- A recorded version of this seminar will be available at www.eSeminarlive.com