Finding Data Protection Choices To Meet Your Disaster Recovery Priorities

David Hill & Dan Tanner

Mesabi Group LLC (Hill)

ProgresSmart (Tanner)

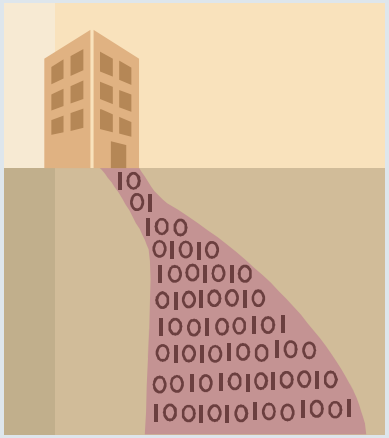
Agenda

- Why is data protection is important to you?
- Where does data protection fit in?
- How can you protect your data?

Why is data protection important to you?

"The world is not run anymore by weapons anymore, or energy, or money; it's run by little 1s and 0s, little bits of data."

— spoken by the character Cosmo (played by Ben Kingsley) in the movie *Sneakers*

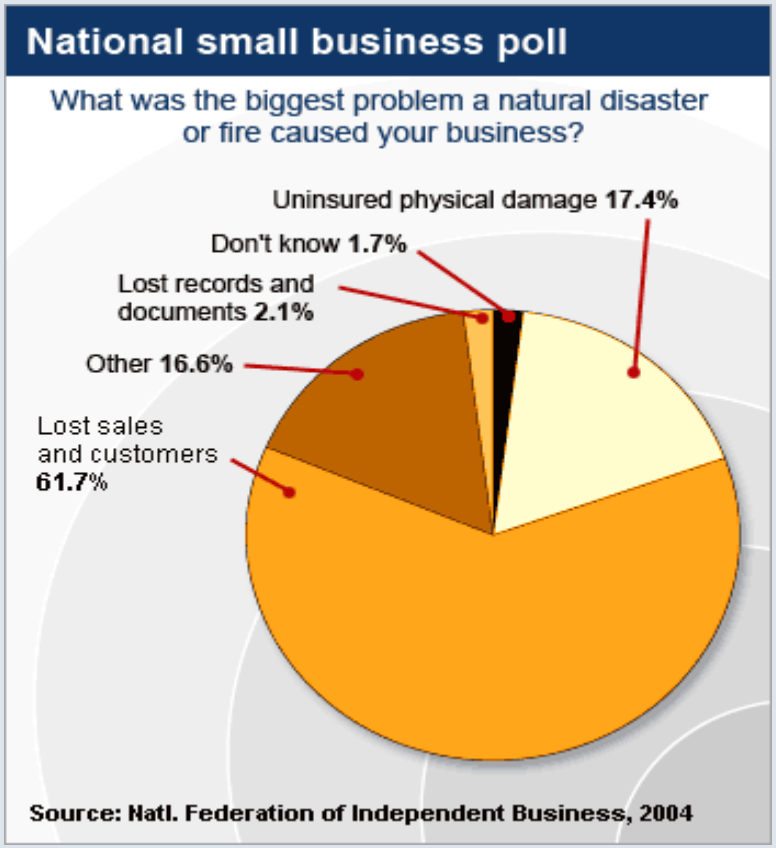


Why Should You Care? -1

- What happens if you lose the information?
  - **You lose your company's memory**
    - Who owes money to us
    - Who we owe money to
    - Who are my customers, prospects?
    - How is my company doing?
    - What regulatory problems do I need to deal with?

Why Should You Care? - 2

- What happens if you lose ability to process the information?
  - You lose competitive advantage
    - Everyone else can do business, you can't
  - **You lose customers**
  - You can't to respond to what's going on
  - You can't connect your information to the environment
    - You don't know any information about your customers, partners, suppliers when they call



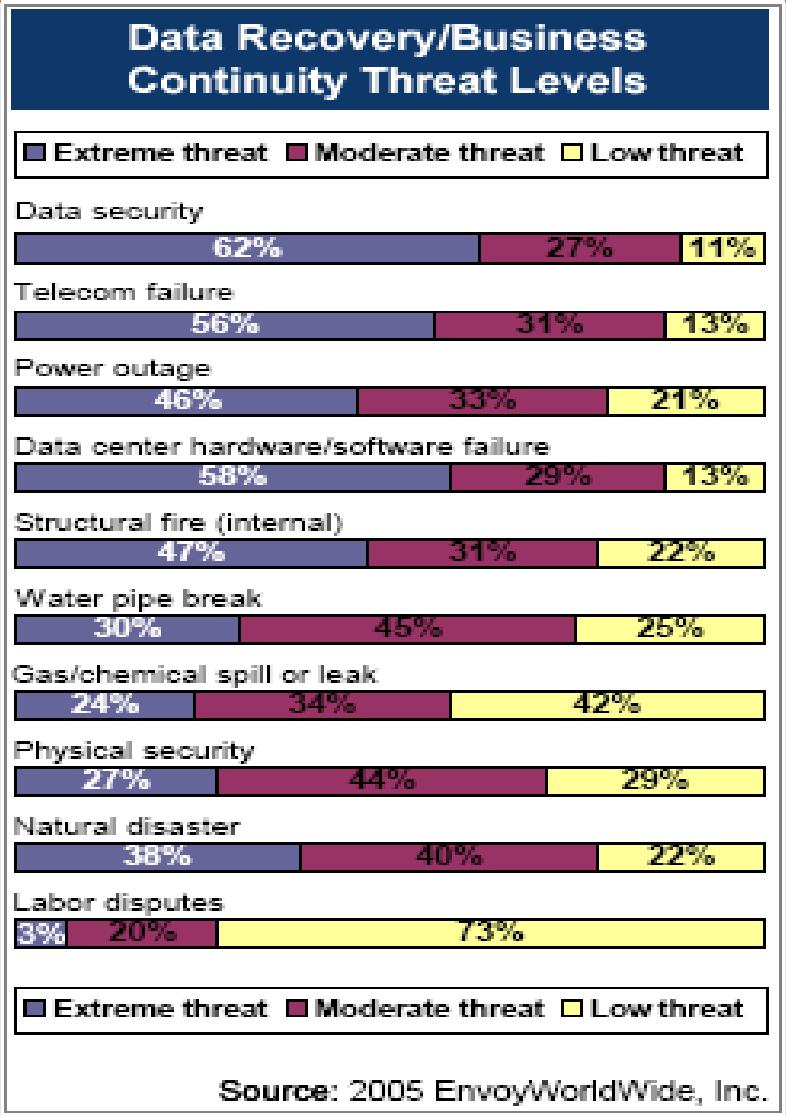
Why Should You Care? - 3

- Why isn't data protection business as usual?
  - The sheer volume of data and its continued growth as well as other demands (such as availability and compliance) puts tremendous strains on data protection technologies
    - The increased value of your data means that you are at increased risk of information asset exposure if the technologies cannot do their job.
  - You have to think of data protection not as a series of point solutions, but rather as a system
    - Think of data protection as an investment portfolio

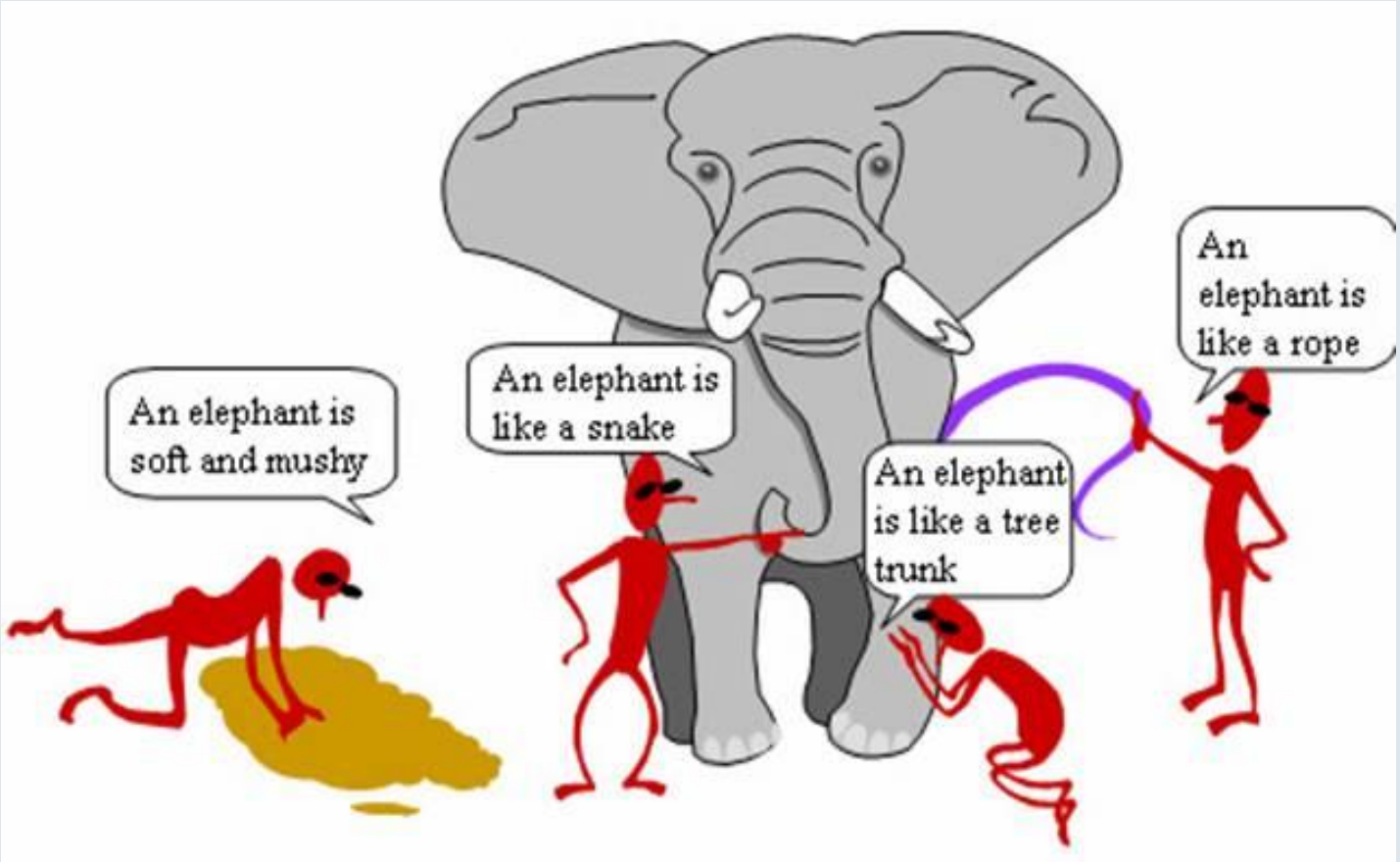
Why Should You Care? - 4

- You are responsible for data protection
  - Yes, we know that everyone is responsible for data protection
    - The backup administrator, IT operations people, CIO, business unit manager, etc., etc.
  - But who else is responsible for looking at the big picture?
    - That's why you're here and why your title may have business continuity, disaster recovery, and/or planning in it

Threats to Data: How They Rank



You Must Understand the Whole Data Protection Elephant and Make It Work



Data Protection — Back to Basics

- **Business continuity** is the mitigation of risk caused by interruption to normal enterprise activities and processes.
  - Business continuity is part of *risk management*, a key functional responsibility of any organization.
- **Data protection** is the mitigation of the risk of loss or damage to an enterprise's data on either a temporary or permanent basis.
  - Data protection is an essential and mandatory part of business continuity.

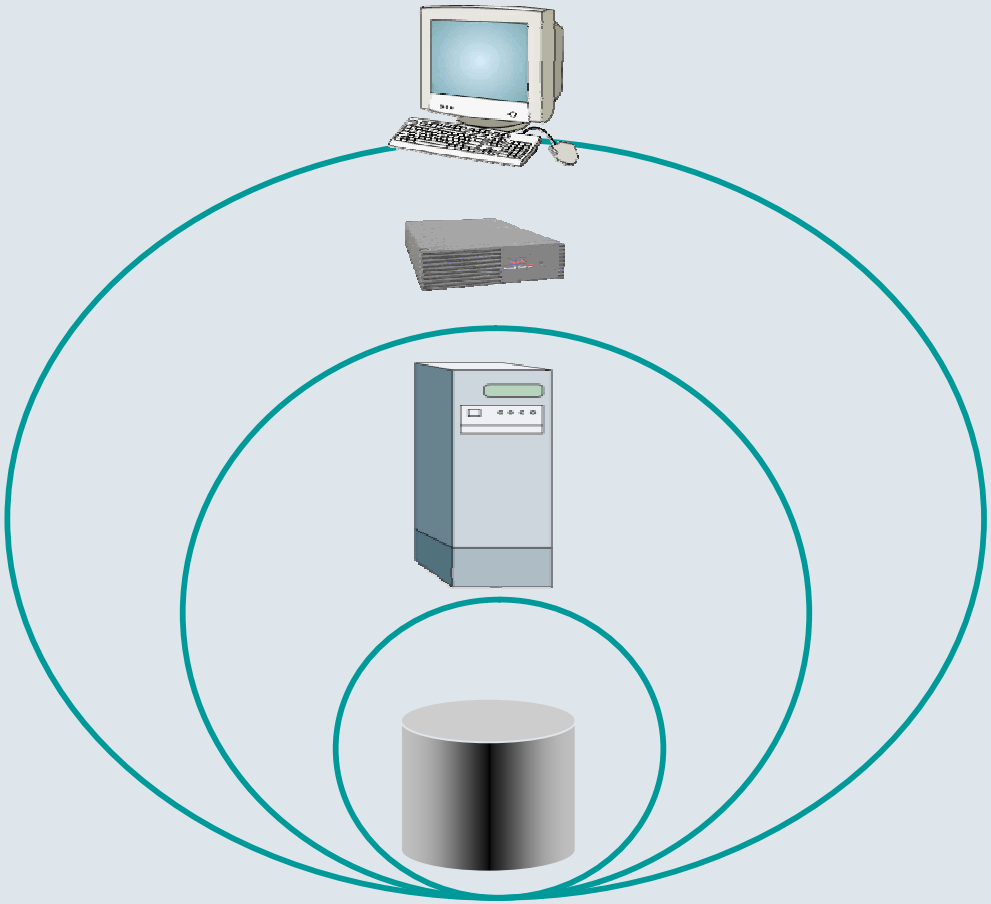
Everything must work or Business Continuity won't

People
ex: knowledge workers,
such as data mining specialists
and IT professionals

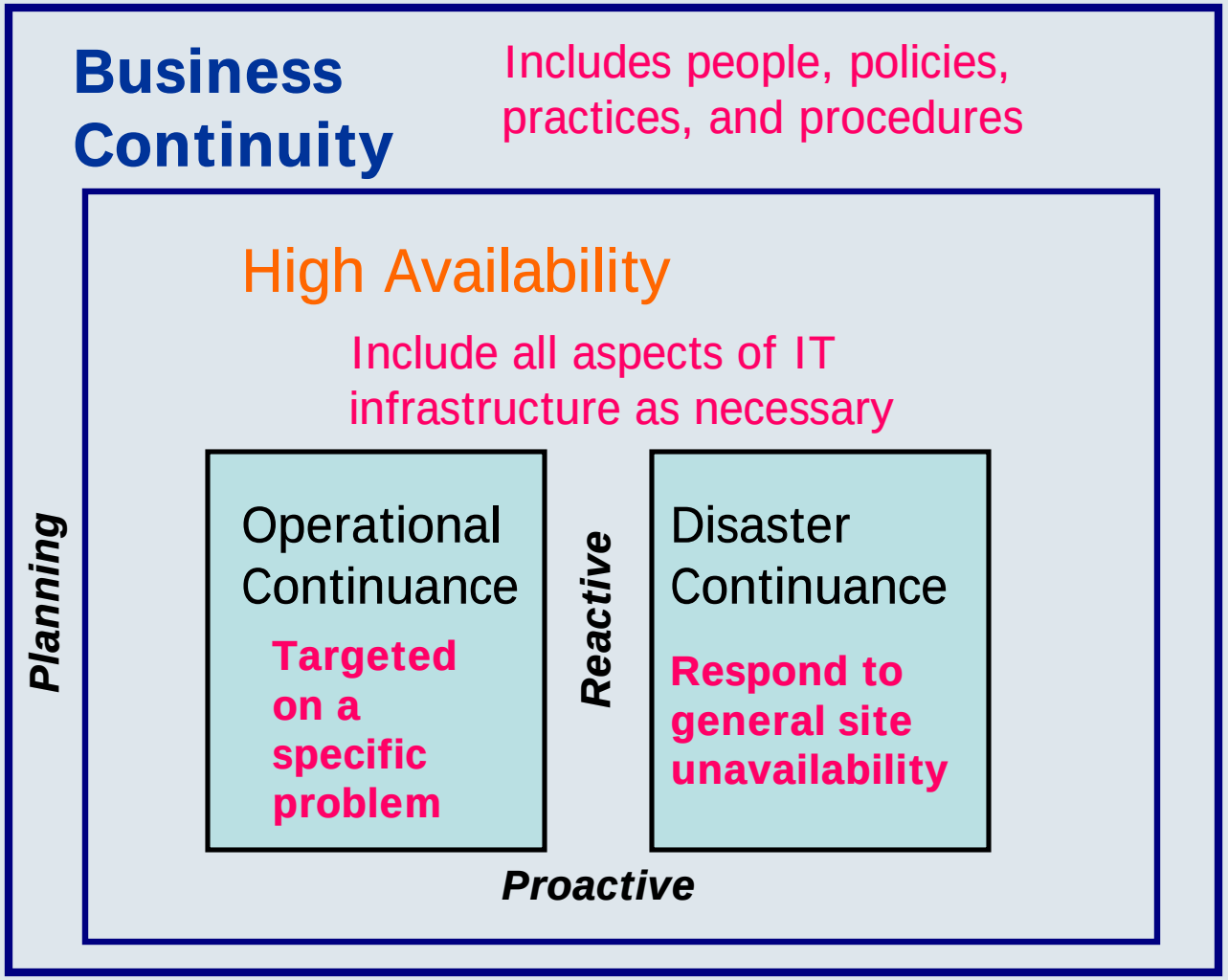
Networks
ex: LAN, MAN, WAN

Application Processes
ex: ERP, CRM

Information
ex: Databases, documents



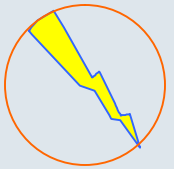
Business Continuity Overview



Business Continuity: More than Disaster Recovery

- **Disaster recovery** occurs only when overall data processing has to be moved from a primary to a secondary site.
- **Operational recovery** is restoration of one or more applications and associated data to a proper operating status at the primary site.

Business Continuity *Keeps Your Business Running*

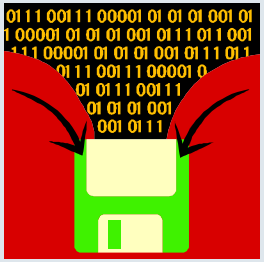


	Operational Recovery	Disaster Recovery
Time Down	• Minutes/year	• Hours to day per event
Infrastructure Threat	• Disk failure • Network congestion • Application performance degradation	• Earthquake • Blizzard • Fire • Flood
Malice	• Computer viruses • Hacking	• Disgruntled employee • Terrorist attack

Physical/Logical Data Protection

- **Physical data protection** — *focuses on storage devices* themselves to recover from dysfunction, failure, or destruction of one or more physical components of a storage system.
  - Protection again disk head crashes, loss of disk sectors, inability to read tape media, etc.
- **Logical data protection** — *focuses on* the protection of the *data itself* (i.e., the bit patterns must retain their designated order and completeness).
  - Protect against viruses, human error, data base corruption, etc.

Simple Data Protection Framework

	Operational Continuity	Disaster Continuity
Physical	Protects against device failure — disk or tape	Protects against the unavailability of the production site
Logical	Protects against viruses, human error, data base corruption, etc.	Protects against propagation of logical problems from local site

All four boxes must be filled in for all applications and data

Availability & Degrees of Protection

- Availability = unplanned downtime per year allowed
  - *High availability* = seconds or minutes per year
  - *Low availability* = hours or days per year
- Data protection comes in degrees (or layers) of protection. Each additional copy = one degree
  - For each degree of protection one failure is tolerable and data can still be recovered
  - Multiple degrees of data protection are necessary; three degrees may be a minimum
  - Degrees of data protection are typically split between higher and lower degrees of availability; when higher availability degrees are exhausted, lower degrees of availability are invoked

Sample Degrees of Data Protection for App. *n*

	Operational Continuity	Disaster Continuity
Physical	**Higher Availability** Degree 1: Local replication Degree 2: Remote replication **Lower Availability** Degrees 3-5: Tape/ disk-based backup	**Higher Availability** Degree 1: Remote dated replication **Lower Availability** Degree 2-4: Vaulted tapes
Logical	**Higher Availability** Degree 1: Continuous data protection Degree 2: Point-in-time copy **Lower Availability** Degree 3-5: Tape/disk-based backup	**Higher Availability:** Degree 1: Remote dated replication **Lower Availability** Degree 2-4: Vaulted tapes

Data Protection Challenges

	Operational Continuity	Disaster Continuity
Physical	**Key Available Technology:** RAID 1, RAID 5, and variants **Key Challenge:** Relatively inexpensive and low performance impact multiple parity RAID	**Key Available Technology:** Synchronous and asynchronous remote mirroring **Key Challenge:** Getting RPO as close to zero as possible over long distances
Logical	**Key Available Technologies:** Point-in-time copy capability and tape **Key Challenge:** Acceptance of continuous data protection technology	**Key Available Technologies:** Vaulting and electronic vaulting **Key Challenge:** Acceptance of dated replication technology as a complement to existing technologies

Goals of Data Protection

- ***Data preservation*** **—** data must be consistent and accurate all the time, and also must be complete within acceptable limits.
- ***Data availability*** **—** the ability of I/O requests to reach a storage device and take the appropriate action.
- ***Data responsiveness*** **—** the ability of I/Os to deliver data to an authorized user according to measures of timeliness that are deemed appropriate for an application.
- ***Data confidentiality*** — data is available only to those authorized.

Two Key Objectives of Data Protection: RTO & RPO

Goal	Objective	Measurement
Data <u>Availability</u>	**RTO (recovery time objective)**	The time required to return an application to a working state after a downtime situation.
Data <u>Preservation</u>	**RPO (recovery point objective)**	The difference between when a failure occurs and the previous time when a set of data was available (e.g. tape from the previous day).

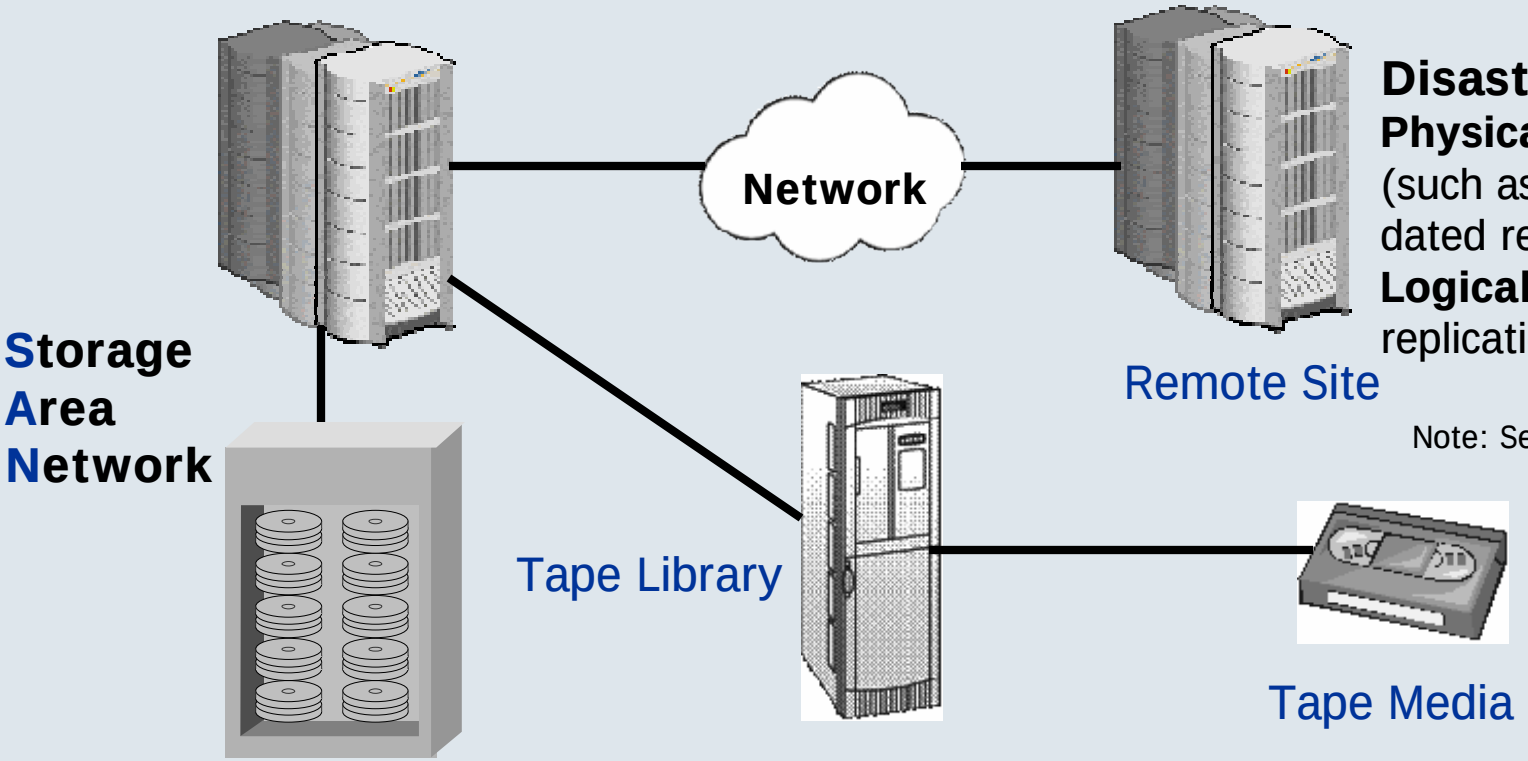
RPO represents **permanent loss of data** (manual reconstruction impossible), **RPO represents** the 1st fallback point — transactions for last 5 minutes, hour, day — it represents the **risk** of a **permanent loss** of only **part of your data**. Operational RPO can't accommodate catastrophe **risking permanent loss** of **all** of your **data**.

A different RTO and RPO may be necessary for "normalcy," i.e., operational continuity & disaster continuity that depends upon the severity level of the disaster.

Recovery Point/Time Applicability

- Many organizations, for good reason, have RTO/RPO of availability burned into their service level agreements (SLAs).
- Requirement availability (& low short-term data loss risk) is leading to the greater use of disk-based data protection as a 1st line defense (mirroring & CDP) and 2nd line defense (D2D backup & virtual tape libraries). (Consider: Is availability or data preservation the paramount goal?)
- One measure of data preservation is recovery point objective (RPO) — the amount of data that is exposed to permanent loss. (SNIA suggests 1 minute in a five 9s environment and 10 minutes in a four 9s environment.)
- RPO represents only the 1st fallback design point to which the data protection infrastructure can respond . Additional failures may expose all data to the risk of permanent data loss. (e.g., RPO for a single failure in a RAID 5 array is zero. But a 2nd failure in the array before the 1st failed drive is rebuilt represents potential loss of all the data — which is why there's backup.)

Typical Data Protection Today



Disaster Continuity
Physical: Disk copy (such as remote mirror, dated replication)
Logical: Dated replication

Note: Servers not shown

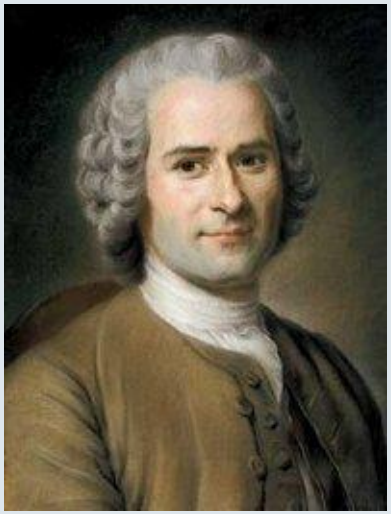
Operational Continuity
Physical: RAID (typically local mirroring or RAID 5)
Logical: point-in-time copy

Operational Continuity
Physical: Individual pieces of tape media
Logical: Individual pieces of tape media not in I/O path

Disaster Continuity
Physical: Tape media kept at remote site
Logical: Tape media kept at remote site

Moving From Present to Future

- Why isn't the way it is today good? **It doesn't scale.**
- We need to start with something and that is data classification. **Because we must treat what's treatable, in the right priority.** That is, we must plan our triage.
- Because the **goals and requirements** for data protection (such as higher availability and greater assurance of data preservation) **are more stringent** as we're now totally data-dependent.



"There are two things to be considered with regard to any scheme.
In the first place, 'Is it good in itself?'
In the second, 'Can it be easily put into practice?'"

Jean-Jacques Rousseau (1712-1778)

Classification is paramount, for separating data into protection categories:

- **Constantly-changing** data (e.g., customer-facing OLTP application data)
- **Frequently-changing** data (e.g., "office productivity data such as spreadsheets, word-processed documents, presentations)
- **Infrequently-changing** data (e.g., your OLTP BOM database, older office productivity data [at some stage in its lifecycle])
- **Data that doesn't change (but which is allowed to)**, such as aged files and reference information – but be aware of active/deep archive requirements
- **Data that can't be allowed to change**, such as compliance-affected data as in e-mail, instant messages, legal/medical records, etc. – but be aware of active/deep archive requirements

Rx: Start to Classify Data

- Separate frequency of access from frequency of updating
- **Active** = frequency of access, i.e. a measure of the level of activity, not a measure of change
- Data that is likely to change and able to be changed is **changeable** data; data that is not likely to change or cannot be changed is fixed content data. Fixed content data may be placed in a repository called an **archive**.
- Data can therefore be separated into two primary categories — **active changeable** and **active archive**
- Data classification is an ongoing process
  - The class of a particular piece of data may change

Understanding Fixed Content

- Over time information becomes frozen, i.e., no longer updated (no more I/O writes), thus becomes **fixed content** (read-only) – an active archive.
- **Fixed content** is likely to be the **majority of** an enterprise's managed digital **data** (Estimates range up to 80%.)
- **Fixed content** can be **managed differently** than data that can change
  - For example, *fixed content does not require traditional backup/restore processes*; it requires a replication process, such as the replication of checks upon receipt to ensure the proper number of copies in a Check 21 compliant environment

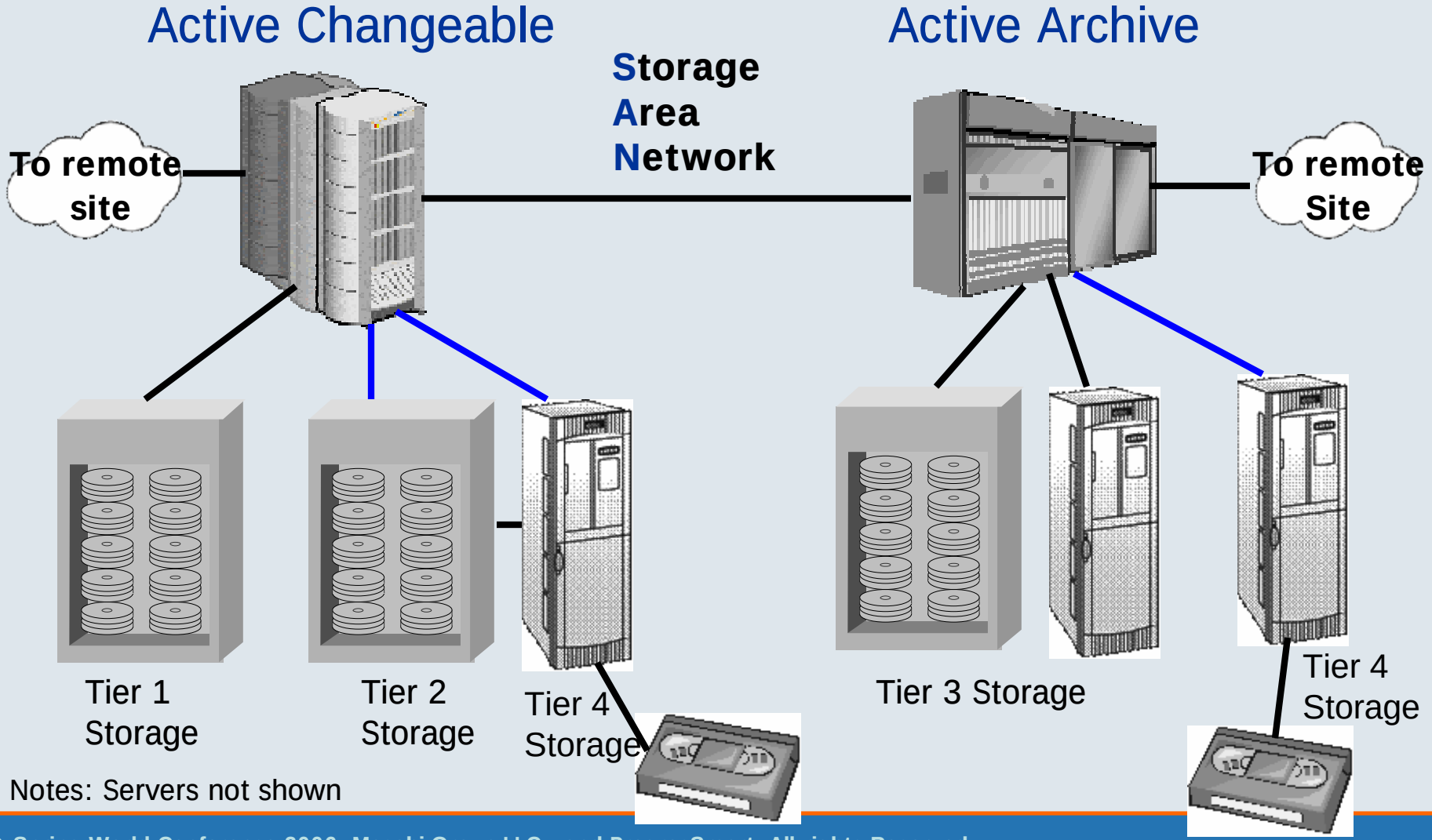
What's the Big Deal? Why Does This Matter?

- Migrating data from active changeable to active archive data reduces data protection requirements for active changeable data
  - Less data to backup means backups and restores don't take as long
  - Less data saves costs in disk systems (such as for remote mirroring)
- Active archive data only needs to be replicated
  - Replication at ingestion into the archive instead of ongoing backup process is easier and less costly
- Active archive data is the only type of data to which data retention policies can easily be applied and compliance — a subset of data retention — is better managed in an active archive
- Different infrastructure, investment, management, and skill sets
- Organization Matters: Backups aren't good archives because
  - Backups are organized by date/time
  - Archives are organized by subject

Introducing ILM and HSM

- Information Lifecycle Management (ILM) is a phrase *du jour* in information technology
  - Not all information is created of equal value and that value may change over the life of a piece of information
  - Classification (which we have just discussed) is an important part of ILM
  - Moving data may involve a new form of hierarchical storage management (HSM)
- Tiering (putting information on different storage platforms – e.g., “tiers”) is an important part of ILM
  - Technologies mature, and their “tiers” change

Modern Data Protection With One Classification



Notes: Servers not shown

“Modern” HSM can be “ILM Lite”

- Use HSM
  - Tackle ILM separately or later, you’re interested in Data Protection
- HSM separates out fixed data
  - Thereby easing backups
- “Modern” HSM
  - Integrates smoothly with back-up & data protection software
  - Not simply algorithmic
  - Not only one-way
  - May preserve file visibility to applications

Data Protection Technologies for the New Framework

	Operational Cont.		Disaster Continuity	
	Active Changeable	**Active Archive**	**Active Changeable**	**Active Archive**
Physical	**RAID** **Cloned Point-in-Time Copy** **Tape Automation*** **Virtual Tape Library*** **Continuous Data Protection** **Data Protection Appliance***	**RAID** **Dated Replication** **WORM Tape**	**Synchronous Remote Mirroring** **Asynchronous Remote Mirroring** **Dated Replication** **Vaulting*** **Electronic Vaulting***	**Dated replication Vaulting**
Logical	**Point-in-Time copy** **Tape Automation*** **Virtual Tape library*** **Continuous Data Protection** **Data Protection Appliance***	**WORM Disk** **Guaranteed Uniqueness** **Electronic Locking** **Dated Replication** **WORM Tape** **Compliance Appliance**	**Dated Replication Vaulting*** **Electronic Vaulting***	**Dated Replication Vaulting**

*** Backup/restore software is or might be used in conjunction with this technology**

Now, Let's Consider Disaster Levels

Disaster Levels

Level	Production Data Center	Failover	Length
0) **Operational** "Disaster"	Local data center impacted w/o damage to overall array, but can't meet SLAs: (ex: 2 disks fail in an array, blown transformer, human error deletes system disk, shutdown for coming hurricane.	Failover warm or hot), according to policy.	As short as a minute or much longer.
1) **Repairable** Disaster	Data center (DC) temporarily inoperative due to physical Infrastructure event (ex: fire, flood, hurricane.	Similar. Remote site may need hardening. Set up temp. 3rd site (service?).	Hours, days, or weeks.
2) **Recon-struction** Disaster	DC damaged beyond repair. A new DC must be rebuilt, at original site or elsewhere.	Must build new site of same quality.	Could be months or longer.

Raising the Availability Ante: Your Choices for Lowering Risk

High

Risk

Low

Level	Description	Risk	Local Copy	Remote Copy
None	No additional copies of data	Single failure leads to permanent loss of data	0	0
0) Simple Tape	One copy of data locally	Site destruction leads to permanent loss of data	1	
1) Traditional Tape	One copy locally plus one copy on tape remotely	If local data center is rendered inoperative, time to restore may be extensive	1	1 Virtual
2) Production + Remote Data Center and Tape	Base triad of 3 physically separate points of protection	Much lower risk and much faster time to restore	1	1 Remote Data Center 1 Virtual
3) Production + 2 Remote Data Centers and Tape	Full triad of active data centers + virtual spare on tape	Lowest realistic risk, but time to restore probably not improved much	1	2 Remote Data Center 1 Virtual

Low

Cost & Complexity

High

Notes: 1) Virtual means data on tape can be used, but restoration is needed to move it - tape to disk.
2) Many variations on a theme are possible. This table simply gives some basic examples.

Quality of the Spare

- An alternative data center may not be as capable as the original data center
- It's an emergency spare with less performance and less robustness than production data center
- How do organizations plan quickly and rebuild to a new primary site?

Your "Takeaways" - 1

- You must treat data protection as critical business continuity
- Treat data protection as an investment portfolio
  - Avoid unnecessary risk exposure
  - Allocate available funds to have the maximum impact
- You must protect each piece of data **four ways:**
  - Against **logical** and **physical** attacks
  - Under both **operational** and **disaster** situations

Your "Takeaways" - 2

- You must balance the goals of data preservation and data availability
- You must classify your data
  - Otherwise, everything gets protected the same way
  - The first classification is to separate out fixed content data
- Decide how to address disasters

Our Conclusions

- **You're** responsible for data protection
  - If not you, who? Involvement has to become commitment
- When executive management "gets it"
  - You'll get the funding
- You must make it work – automatically & intrinsically
  - Even when workers can't be educated (Or, would you rather try to change a culture?)
- You must do it now
  - Time is not on your side – disasters are unpredictable, but can be counted on to happen

Presentation Underwriters

- Underwriter contributions funded our research and travel.
- Underwriters did not contribute to this presentation's content (except perhaps as a reviewer of technical accuracy).
- The Underwriters are not necessarily DP vendors.
- The views we've expressed are not necessarily those of the Underwriters.

Gold Underwriter

Revivio, Inc. is a leader in continuous data protection (CDP) solutions. Revivio's Continuous Protection System is a groundbreaking approach to data protection and recovery that allows companies to eliminate backup windows, as well as to restore data instantly, exactly as it existed at any event or point in time, and to recover business applications in just minutes.

Silver Underwriter

Qualstar (Nasdaq: QBAK) manufactures over 30 automated tape libraries spanning capacities from 1 TB to over 340 TB, for backup, archiving and disaster protection in rack mount and free-standing configurations. These rugged, *Simply Reliable* libraries support LTO, AIT, SuperAIT and SDLT tape formats. All models feature Q-Link™ for web browser-based, worldwide remote management.

Bronze Underwriters

Redefining NAS: Next generation features, Accelerated performance, Simplicity

Network Storage Solutions: Optimizing Data STORAGE, PROTECTION & AVAILABILITY

FalconStor is a Spring World 2006 exhibitor.

Thank You.

David Hill: www.mesabigroup.com, davidhill@mesabigroup.com, 782-326-0038

Now available: Mesabi Group Data Protection Report, 2006 Edition

Dan Tanner: www.progressmart.com. dan@progressmart.com, 508-366-7980

David

Dan