



ProgresSmart



Continuous Data Protection

CDP: A State Analysis

A Tutorial created and presented by:

David Hill & Dan Tanner

Mesabi Group LLC (Hill)

ProgresSmart (Tanner)



ProgresSmart



Agenda

- The need for continuous data protection.
- The way CDP works today.
- Data “forensics”: putting CDP to the test.
- CDP in action.
- Issues with CDP.
- Summary: CDP benefits.
- Q & A



ProgresSmart



What is the need for CDP?

- Trouble, trouble right here in IT city.
 - Your cell phone rings; you hear that a time-sensitive crucial application is down and the downtime meter is running.
- Your crack IT response team springs into action.
 - The problem is **not physical**; your servers, networks, and storage are all responding properly.
 - The problem is **logical**, but knowing that does not help; your application is still down.



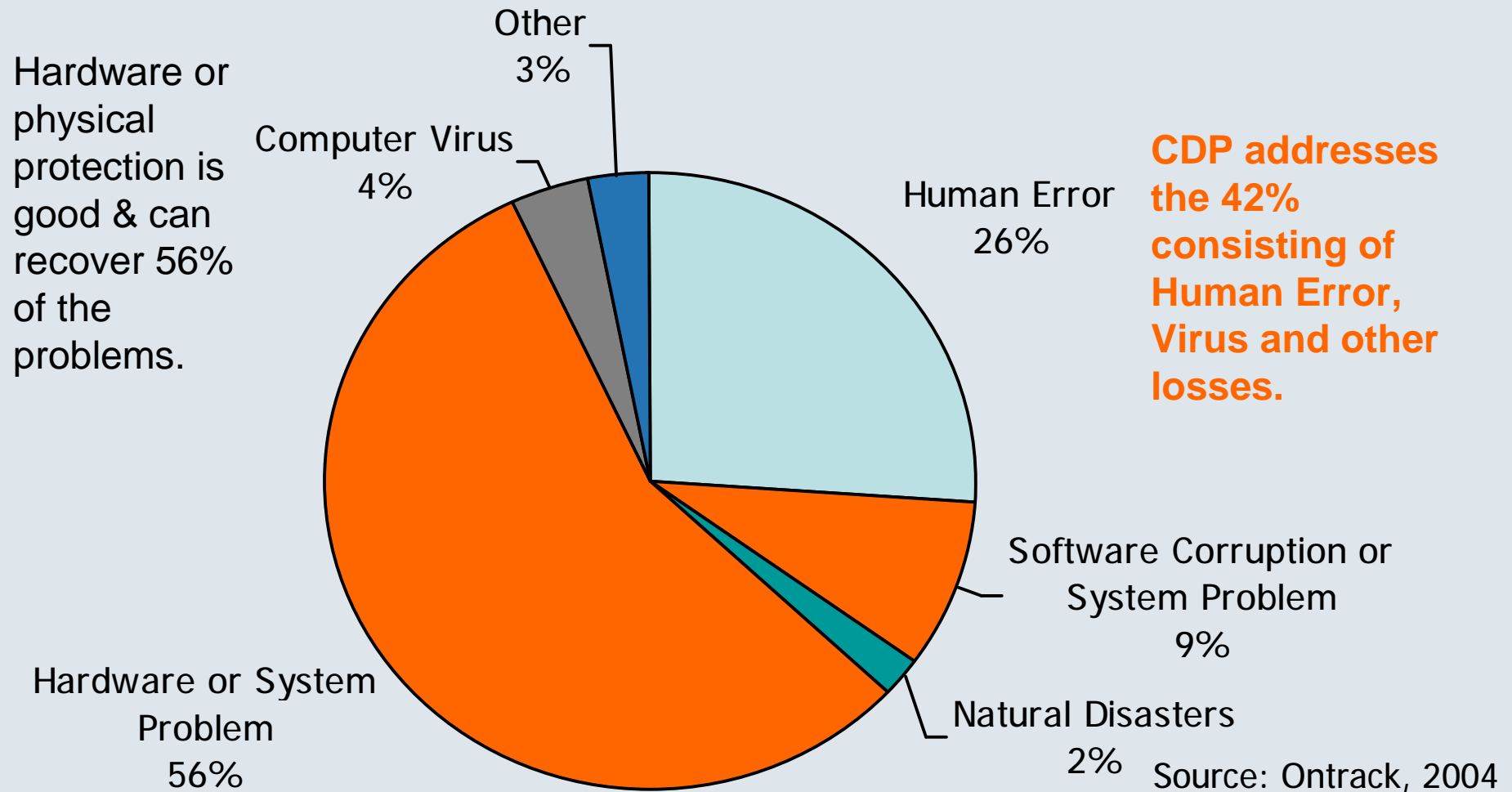
ProgresSmart



Physical vs. logical protection

- **Physical data protection** protects the integrity and functionality of the storage itself, for example, RAID.
- **Logical data protection** protects the data itself from change through unauthorized or erroneous I/O requests.
 - *Examples of logical problems:* database corruption, virus, and human error, such as accidental file deletion, configuration mistakes, and insufficiently-tested application upgrades.

Causes of Data Loss or Downtime





ProgresSmart



So what do you do now?

- Turning to your disk mirror is not the answer.
 - The mirrors are also corrupted, as the problem is logical not physical.
- Snapshots might help, but they are not a complete answer.
 - Finding the right snapshot and minimizing the loss of data may not be easy or possible.
- Restoring from tape (and now disk) is the typical response.
 - But the price may be hours or days of downtime that may be unacceptable as well as a non-zero risk that the restoration process may not succeed at all.



ProgresSmart



Enter continuous data protection

- Continuous data protection (CDP) is an answer to logical data protection problems.
 - If CDP (as software) is implemented as an appliance with a RAID disk array, then physical data protection is a side benefit.
- In contrast to alternatives, CDP delivers the following benefits.
 - Higher availability (seconds or minutes vs. hours or days).
 - Reduced risk (in both the amount of data that can be permanently lost as well in the failure of the restoration process).



ProgresSmart



CDP definition

- CDP is a methodology that continuously captures or tracks data modifications and stores changes independent of the primary data, enabling recovery points from any point in the past. *CDP systems may be block-, file- or application-based* and can provide fine granularities of restorable objects to highly variable recovery points.
 - Source: SNIA DMF CDP SIG working definition, April 2005.

Traditional vs. CDP

The Traditional Method of Restoring Data

"I will get you what data I have when I can."



Searching an entire collection of documents ...



"I don't know if this is what you want,
but it is what I have."

Restoring Data Using Continuous Data Protection

"I'll be happy to get it for you.
It will just be a moment."



Extracting a single book
from a single library shelf ...



"Here is what you requested."



ProgresSmart



What CDP does

- CDP systems provide the ability to recover to any point in time (APIT).
 - *The ability to provide access to data as it existed at any point in time is a stark contrast to mirror-splits or snapshot technologies which offer access to single point-in-time (SPIT) copies previously stored.*
- **Recovery Points are virtual**
 - *That means that they are created on demand by reviewing the transaction or journal logs and are thus created in an ad hoc — not a predetermined — manner. (Recovery points are akin to freeze-frame selection from a continuous film.)*



The road to CDP

Event Driven

Continuous



10100010010100111100010010010010101010

- Backup snapshot
- Undo (roll back to previous status)

- Any point-in-time copy

- Searchable file catalog to enable rollback to any point in time

Discrete Snapshots

Virtual Recovery Points



ProgresSmart



CDP is for *active data*

- Non-CDP data protection methodologies may suffice for less than frequently changing data.
- Archiving is the way to handle data that does not (or cannot) change.
- The foregoing implies an Information Lifecycle Management (ILM) activity to “sort” and “channel” slowly-changing &/or archive data.



ProgresSmart



CDP advantages

- The key advantage of CDP is the ability to deliver finely granular and highly variable **recovery points**.
- If data is lost or corrupted, CDP provides the ability to **recover** a clean copy of the affected data, even if the problem is not discovered until hours or days later.
- *A CDP system allows users to rapidly restore a file or database or system to its correct, consistent state — sometimes within seconds, but more typically within minutes.*

What CDP does

A major restoration might be measured in:

Seconds to minutes:



Continuous data protection

Minutes to hours:



Virtual tape library

Standard point-in-time snapshot and replication technologies

Hours to days:



Tape library automation with traditional backup/restore software



ProgresSmart



The Way CDP Works Today

- **Elimination of the backup window**
 - IT managers can quit worrying about backup windows — CDP doesn't need a backup window.
 - Transaction or journal log transactions are continuously captured by the CDP software with little or no impact on operations.



ProgresSmart



The Way CDP Works Today

- Getting a file version up-to-date only requires mounting the appropriate recovery point in read-write mode & then running the roll-forward sequence (if the recovery point was taken while the database was in backup mode),
- Or, letting the database automatically “play” its journal (if the recovery point was taken while the database was running normally).



ProgresSmart



The Way CDP Works Today

- If the user wants to “restore” the primary disk to the selected PIT, CDP can perform a restore in place operation.
- Data can be accessed as though on primary disk while optimized resynchronization runs in the background.
- No recovery/restore server needed.



ProgresSmart



The Way CDP Works Today

- Typical recovery points are read-only.
- CDP recovery points can be mounted in read-write mode & accessed the same way as primary data. An application needs data in read-write mode to run journal log roll-forward & check data consistency.
- Applications can access CDP recovery points immediately & directly with no time-consuming restore operation.



ProgresSmart



The Way CDP Works Today

- CDP is not free. Additional disk space can range from 0.2x to 2x to larger so storage has to be managed .
- The APIT window is typically configured to 24-48 hours, i.e., maximum time an enterprise believes necessary to provide against typical **logical** data loss problems.
- Recovery points can emulate any retention policy — e.g., hourly recovery points kept a week, dailies kept a quarter, weeklies kept six months, & monthlies for a year.



ProgresSmart



Data “Forensics”: Problem Solving

- Knowing when a data store failure occurred is challenge enough!
- Data corruption (from inadvertently overwriting or deleting parts of data stores, from application error, or by gradual hardware failure) presents a greater challenge.
- The corruption may be discovered only well after the fact.



ProgresSmart



Problem Solving Takes Time

- A logical data protection problem requires time to find out about the problem and to determine the best course of corrective action.
- Continuous data protection therefore does not mean continuous data availability.
 - But CDP is pretty good, as we will see.



ProgresSmart



Making Fixes Using CDP

- You discover that certain unwanted changes happened during the yesterday's transactions from one new employee.
- You point the CDP clock back to before the errors happened.
- You start updating on a minute by minute basis, and between each update, check to see where the data became corrupted.
- You stop, point the clock back to the moment before that corruption, then continue posting changes made **AFTER** the corruption ended.
- This may take a while, but it ensures that all of the errors are corrected.
- Tape restores couldn't do this!



ProgresSmart



SNIA Definitions

- **Recovery Point Objective – RPO**
 - The maximum desired time period prior to a failure or disaster during which changes to data may be lost as a consequence of recovery. Data changes preceding the failure or disaster by at least this time period are preserved by recovery. Zero is a valid value and is equivalent to a "zero data loss" requirement.
- **Recovery Time Objective – RTO**
 - The maximum desired time period required to bring one or more applications and associated data back to a correct operational state.



ProgresSmart



CDP position in data protection

- CDP drives the RTO toward zero.
 - *Realistically from <1 minute to 15 minutes.*
- CDP makes the RPO become “any”.
 - *Realistically from <1 minute (virtually no data loss allowed) to 10 minutes or more.*



ProgresSmart



The Restoration Process

- Typical way to fix a data store corruption: Restore an uncorrupted version of the data store.
- But the *problem* may be finding when the actual corruption occurred (to restore to the PIT just before).



ProgresSmart



The Restoration Process

- Multiple recovery points can be mounted simultaneously (on different servers) to test multiple versions simultaneously.
- An automated test system (such as a binary search) could speed up the process.



ProgresSmart



Restoration Process

- CDP can roll a recovery points forward to a PIT after the recovery point was taken by replaying the write log,
- Enabling a corrupt data store to be restored to the most recent uncorrupted state.



ProgresSmart



CDP in Action

- **Problem:** Current backup/restore process for patient record data (including life critical hospital admission information) as well financial data (not life-threatening, but keeping the CFO happy is important) was impaired.
 - 30 day backup with 60 to 90 day recovery was totally unacceptable.
- **Solution:** Installation of a CDP appliance enables recovery within the minutes that is required.
- **Comment:** Improving the old technology was not an option; only a paradigm shift would do.



ProgresSmart



Insurance Company CDP in Action

- **Problem:** After a data corruption problem affecting a crucial application, a tape restore process failed to restore the data.
- **Solution:** The company spent \$50K on temporary labor to recreate the data.
- **Comment:** Wouldn't the \$50K have been better spent on a CDP appliance that could be reused and would have prevented potential customer dissatisfaction?



ProgresSmart



Forbes.com CDP in Action

- **Problem:** Forbes.com wanted the ability to restore data within seconds in the event of loss, corruption, or disaster and a tape restore system could not achieve that goal.
- **Solution:** Snapshots offered only fair to poor recoverability and did not have sufficient forensic analysis capability for failure events. A CDP appliance offered both the necessary recoverability and forensic analysis.
- **Comment:** CDP is like a movie where you can roll back to exactly the right frame; snapshots are like looking through a series of photographs and trying to pick out the closest (and there may be no exact match).



ProgresSmart



Nature's Sunshine: CDP in Action

- **Problem:** Faced with a history of email issues (i.e. inability to restore individual email for Legal Department, email backed up for 2 days), Nature's Sunshine needed a tool to easily backup and quickly recover the huge volume of multi-language email data on their Exchange servers.
- **Solution:** Implemented application-aware CDP software for Exchange that captures each new, modified, or deleted message in real-time, so they always have the latest data for recovery. Nature's Sunshine now performs restores from an entire storage group, mailbox, or public folder, right down to individual email messages, appointments, and contacts.
- **Comment:** Application-aware CDP delivers crash-consistent data and fast RPO & RTO.





ProgresSmart



CDP considerations

- **Cost** — the cost is additive as CDP is likely to be a complement and not a replacement for existing technologies (at least in the short term).
 - CDP software cost.
 - At least 1.5x or more in new online disk capacity (although cost effective disk can be used).
- **Management** — CDP introduces a new process into the mix.
 - Requires a well-thought out strategy to balance storage requirements & application performance with risk.
 - Requires careful planning for how to analyze problems when they occur and how to do testing to ensure that everything is working properly.



ProgresSmart



CDP considerations

- **Perceived Vendor Durability**
 - No major announcements of a CDP product by a large “trusted brand” storage vendor.
 - But several large vendors are active in the Storage Networking Industry Association (SNIA) CDP Special Interest Group (SIG).
 - A number of good startups “make it”.
- **Hesitancy to Employ New Technology**
 - CDP is still an emerging data protection methodology and there are not yet a lot of reference accounts.
 - But the IT industry has a history of small companies delivering very successful products based on technological paradigm shifts.



ProgresSmart



Summary: CDP benefits

- Higher availability minimizes impact on your stock valuation, revenue loss, &/or productivity loss.
- *No more backup windows.* The backup window concept goes away, with it the constant struggle to meet the windows.
- No more long restores. Restore individual files or entire systems in seconds to minutes. No clock racing recovering from major restorations.



ProgresSmart



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 CDP vendors.
- The views we've expressed are not necessarily those of the Underwriters.

Thanks for technical review by: Tony Bautista (Storactive), Charles Curtis (SWC Advisory Board) & Marylies Tauzia (Atempo)



ProgresSmart



Underwriters



Tiered Data Protection for Information Lifecycle Management.



FalconStor Software is a leading developer of network storage software designed to optimize the *storage*, *protection* and *availability* of enterprise data.



Mendocino Software The Recovery Management Company.



ProgresSmart



Underwriters



Unified Data Management for Information Immediacy, Discovery & Continuity.



Revivio, Inc., a leading provider of enterprise continuous data protection (CDP).



Providing Enterprise Data Protection Around the Clock.



STORServer. Everything a business continuity appliance should be.



ProgresSmart



Underwriters



Innovative storage solutions, designed with quality in mind, for reliable backup, archival and disaster recovery needs.



TriAxis, Inc. Your Trusted Storage Advisor.



Just Keep Working



Real-Time Data Protection and Instant Recovery for Mission Critical Servers



ProgresSmart



Thank You! – Questions please.



David Hill



(781) 326-0038

www.mesabigroup.com

davidhill@mesabigroup.com



Dan Tanner

ProgresSmart

(508) 366-7980

www.progresssmart.com

dan@progresssmart.com