17609: Continuous Data Protection Transforms the Game

Wednesday, August 12, 2015: 8:30 AM-9:30 AM
Southern Hemisphere 5 (Walt Disney World Dolphin)

Tony Negro - EMC

Rebecca Levesque – 21st Century Software

About the Speakers

Rebecca M. Levesque is Chief Executive Officer of 21st Century Software and responsible for shaping the company's product strategy and vision. She has over 20 years experience helping hundreds of companies establish resiliency and recoverability strategies from any disruptive event.

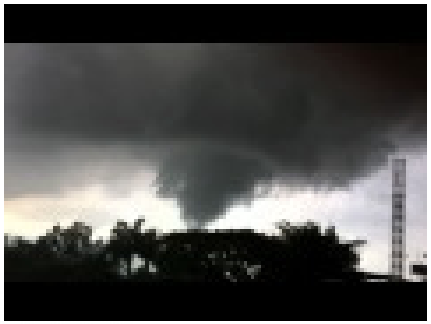
Tony Negro is currently a Corporate Systems Engineer in the Core Technologies Division of EMC Corporation. As a technical resource for VMAX and Mainframe Business Unit, Tony provides in-depth product consulting to customers as well as Field Sales teams worldwide.

About this Session

Near Continuous Data Protection (CDP) is possible but application failover/high availability approaches cannot guarantee batch data consistency and integrity protection when disruptive events strike. If an unplanned outage occurs, does it matter whether it's an operational event or disaster? When minutes count and you have multiple copies and intertwined dependencies, how do you know which copy to use? This session provides insight into what is needed, possible, and attainable.

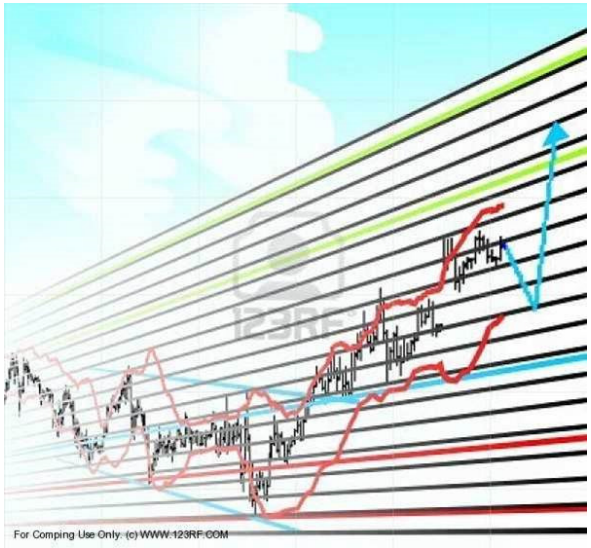
As We Begin Today, Consider This

- After a trio of high profile outages, some industry analysts raised the question of whether such outages are inevitable in today's IT environment?
- Phone companies used to offer 99.999% reliability, which allowed for about five minutes of downtime a year. Today's networks are far less expensive, infinitely more capable and nowhere near as reliable. The NYSE outage lasted about four hours, or ***nearly 50 years*** of allowable downtime using the "five nines" standard.
- Today's reliability problems are a reflection of the complexity and interdependency of computer systems, the pace of change, and insufficient organizational and cultural practices.

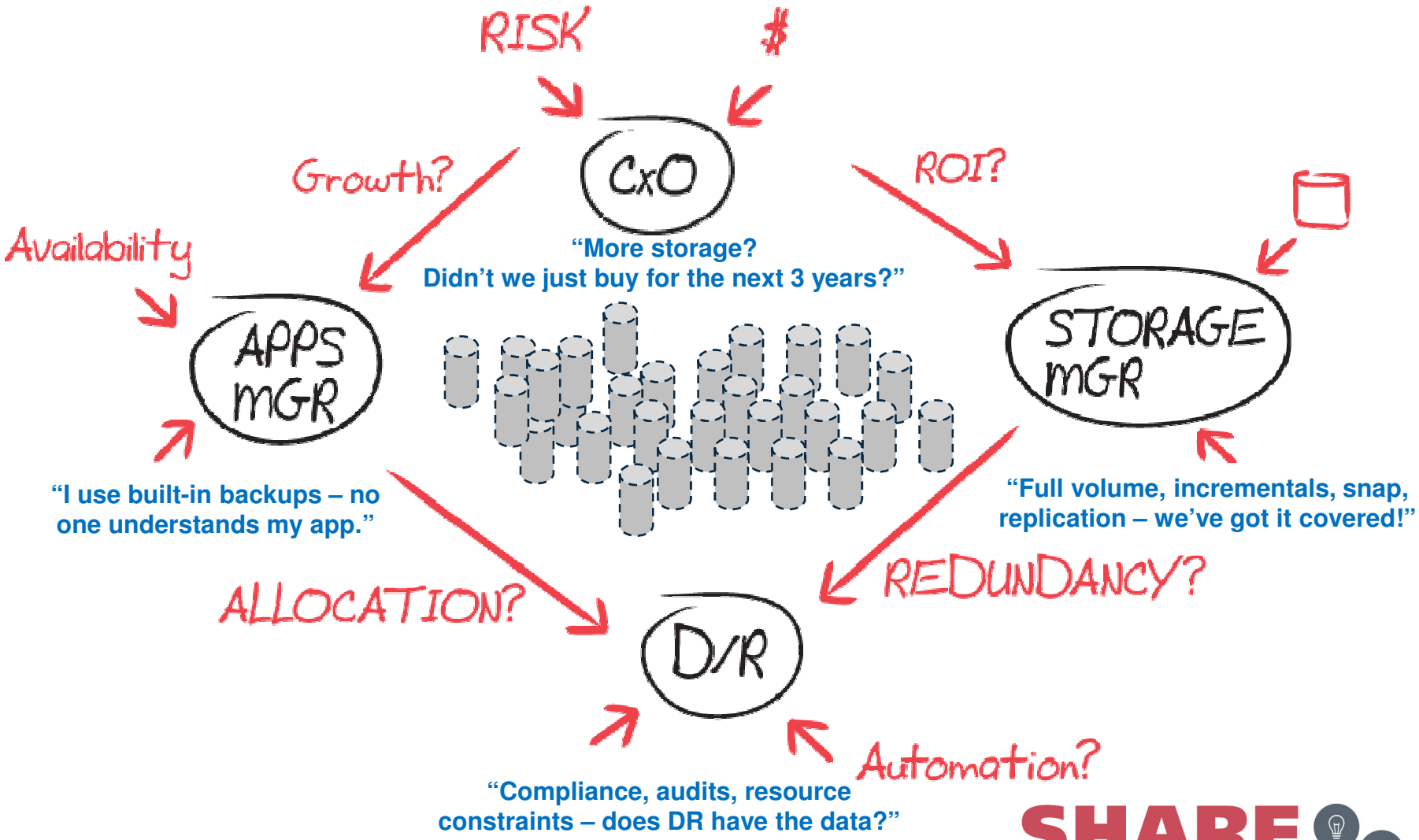


Interesting Facts

- 71% of all Fortune 500 companies have their core business on the mainframe
- 23 of the world's top 25 retailers use a mainframe
- 92% of the top 100 banks use a mainframe
- 10 out of 10 of the top insurers use a mainframe
- More than 225 state and local governments worldwide rely on a mainframe
- 9 of the top 10 global life and health insurance providers process their high-volume transactions on mainframe



Copies… how many? Which one is needed?

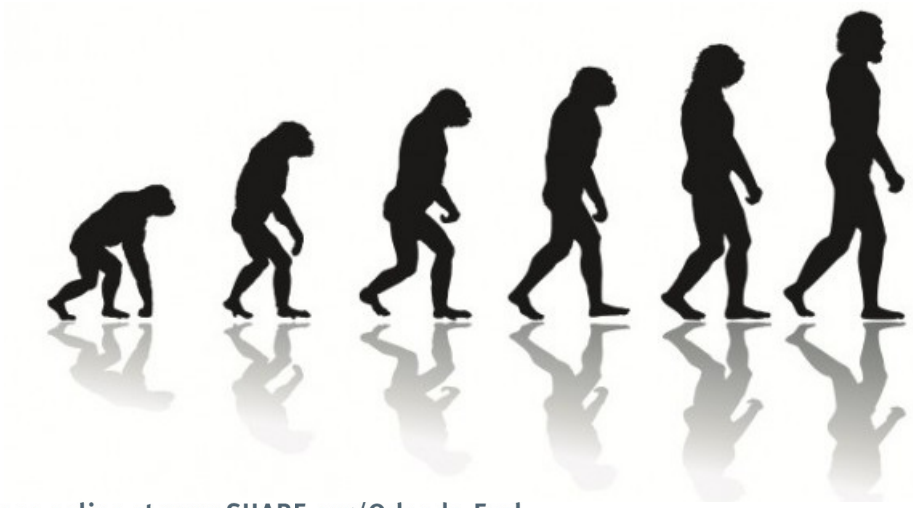


Faster RPO Made Possible through Evolution of Backup

Surveys show organizations want a faster recovery point objective (RPO) and recovery time objective (RTO) from their IT teams. It wasn't until backup products evolved to meet those windows that it became possible.

The study indicated that **most users expect to recover crucial applications in faster than two hours, while most IT staffers said they cannot deliver recovery sooner than four hours**.

To meet these ***tighter RPO and RTO windows requires that secondary data be positioned in a way that allows the application to access it faster***.



The State of Backup

Backup and recovery is one of the oldest and most frequently performed data center operations

Organizations are seeking ways to easily, quickly and cost-effectively ensure data is appropriately protected

Data protection approaches of the past may no longer suffice in meeting current — much less future — recovery requirements

Adopting new technologies and products show an increased willingness to augment or even completely switch backup/ recovery providers to meet increasing service-level needs

Why the gaps, what goes wrong?

- Checkpoints of consistent states are used to enable rapid recovery to known good states (such as before a patch was applied to an OS, or the last time a database was reorganized) to ensure application consistency of the data and to minimize the number of log transactions that must be applied.
  - What if you don't have logs and journals – batch?
  - Have good data but it is not in sync with other application dependencies (ripple effect)
  - Have too much good data – which do I use?
- Frequent snapshots alone do not equate to near CDP - achieving 100% or near-100% availability requires a top-down approach into engineering availability into the application, infrastructure and operating architectures.

(Gartner, 2014)

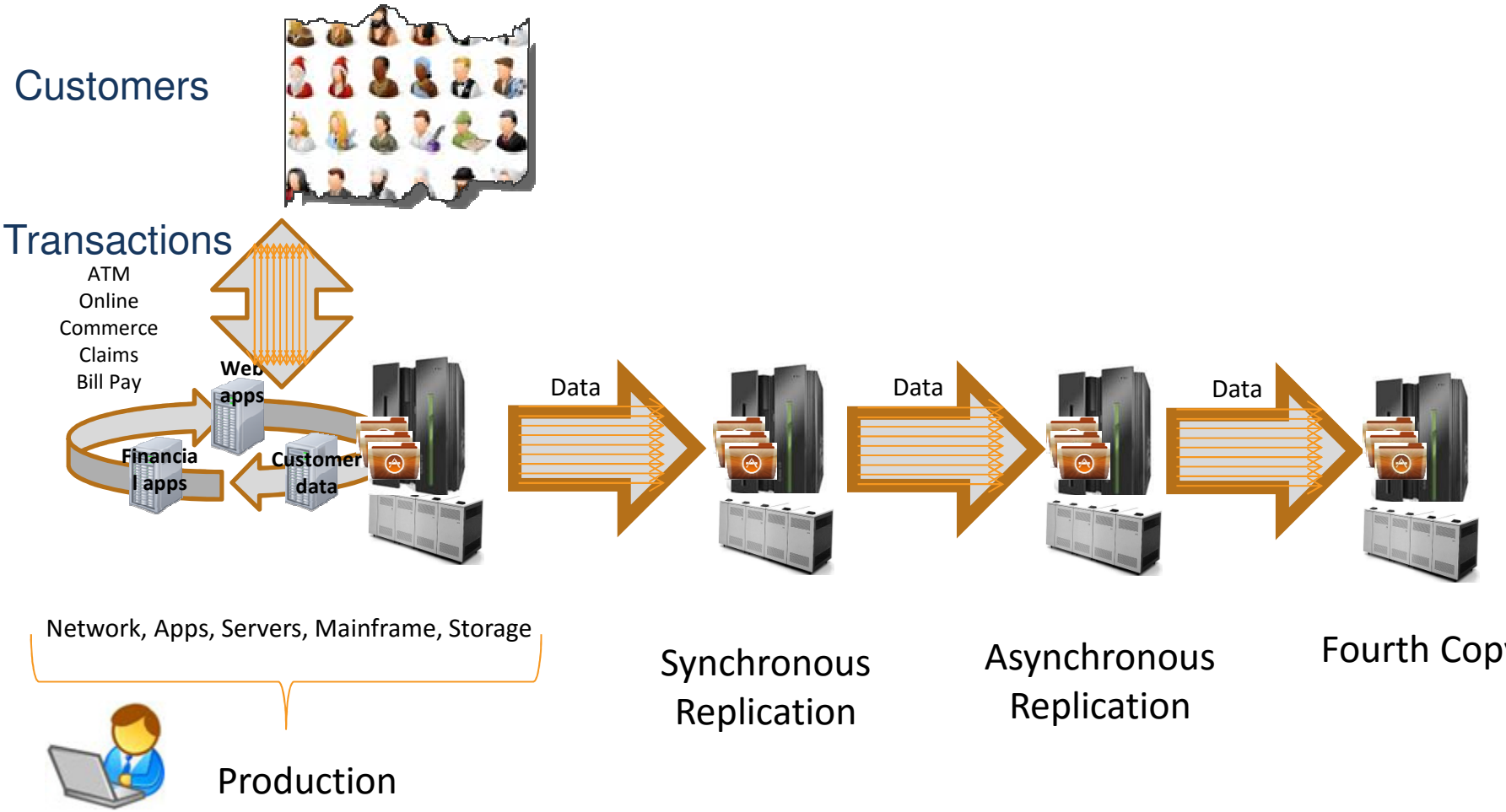
Backup Challenges & CDP

Definition

- Continuous data protection (CDP) is an approach to recovery that continuously, or nearly continuously, captures and transmits changes to files or blocks of data while journaling these changes.
- This capability provides the option to recover to many more-granular points in time to minimize data loss, and enables arbitrary recovery points. Some CDP solutions can be configured to capture data either continuously (true CDP) or at scheduled times (near-CDP).
- The difference between near-CDP and regular backup is that backup is typically performed once a day, whereas near CDP is often done every few minutes or hours, providing many more recovery options and minimizing any potential data loss
- CDP can replace traditional backup, but in most testing scenarios it's only as good as your ability to get backed up data off-site.

The Perfect World

(where nothing goes wrong)

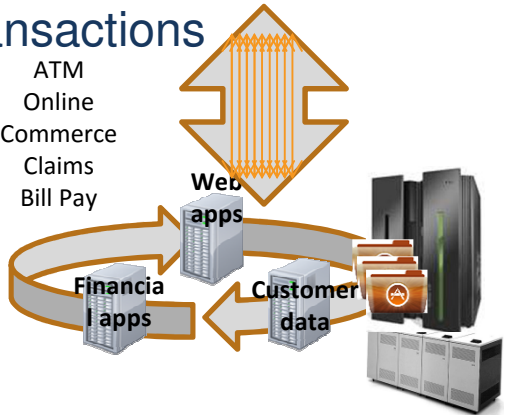


The Real World

Customers

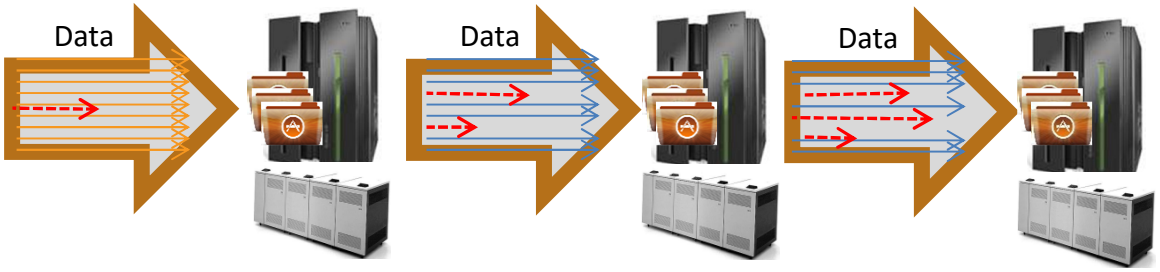
Transactions

ATM
Online
Commerce
Claims
Bill Pay



Batch data doesn't write until job finishes

Cross-dependencies on completed batch jobs become exposure to the business



Network, Apps, Servers, Mainframe, Storage

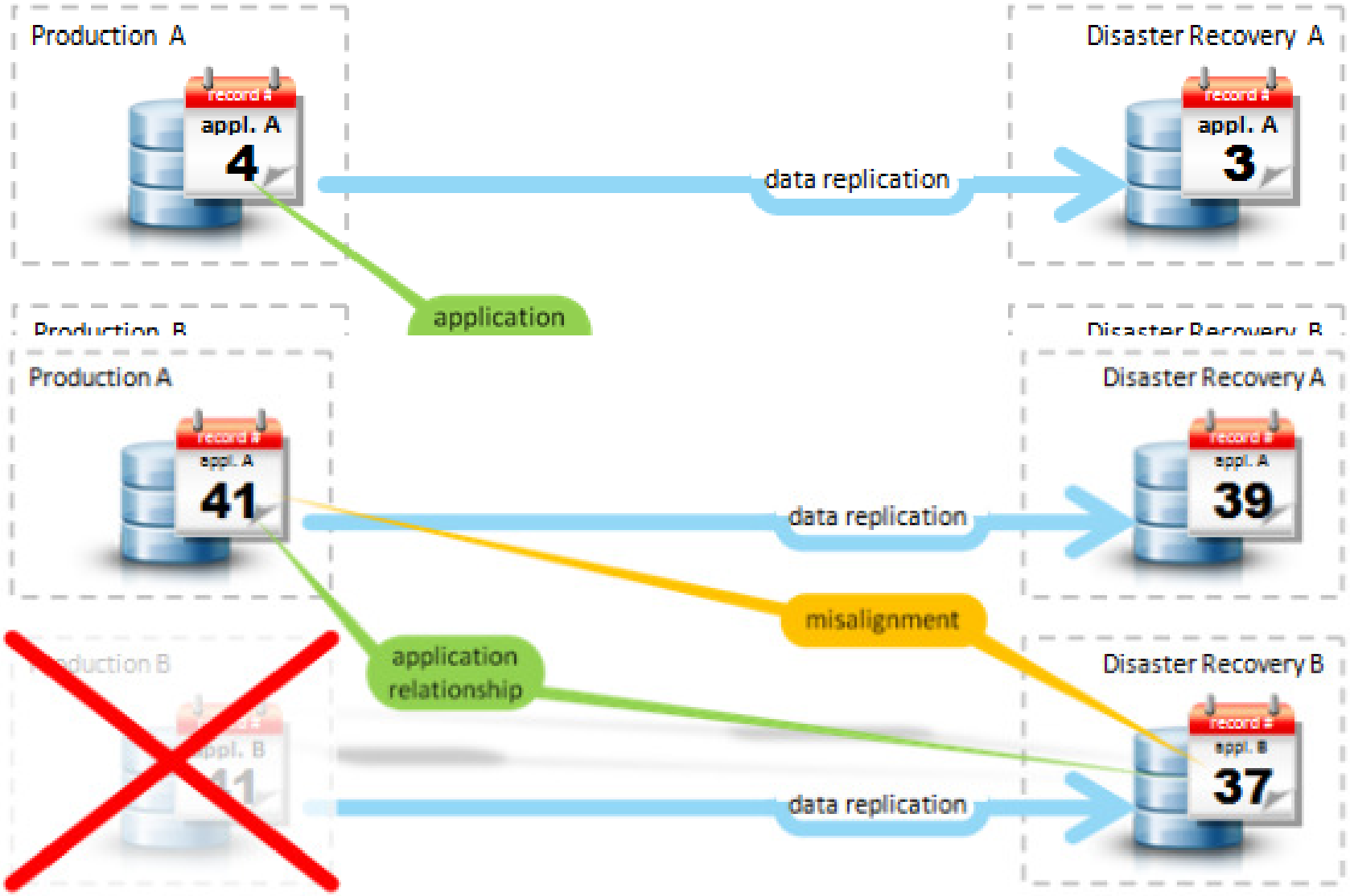
Production

Synchronous Replication

Asynchronous Replication

Fourth Copy

Batch Dependency Problem



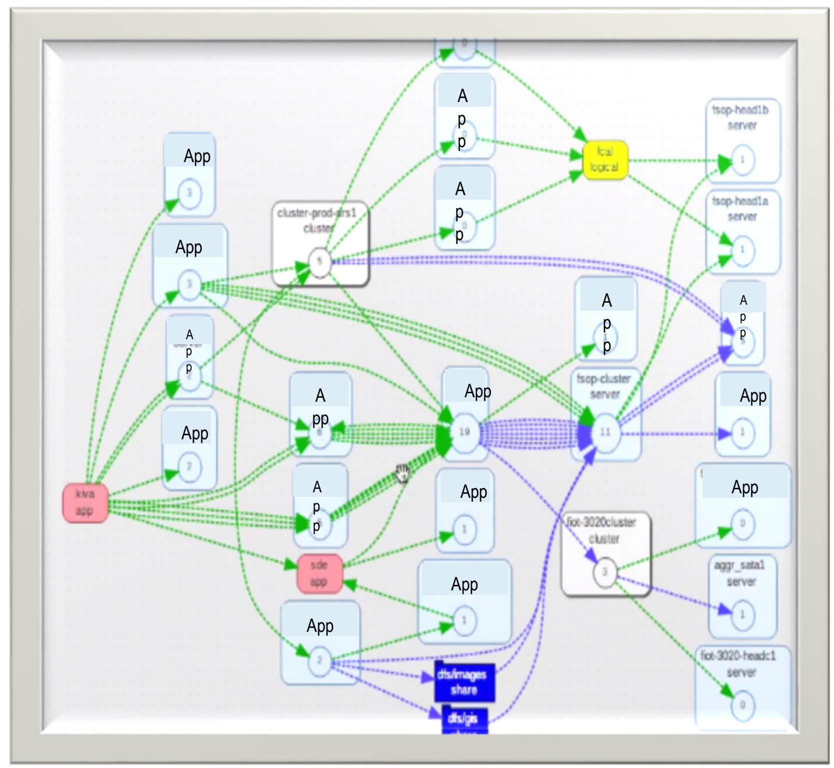
From http://www.smartercomputingblog.com/smarter-computing/business-resiliency-future/

SHARE
Educate · Network · Influence

Understanding Data Consistency:
The Dependent I/O Concept

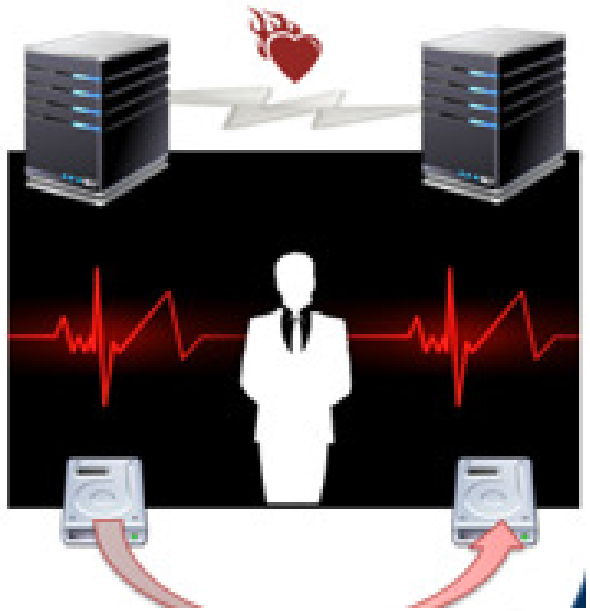
An application concept where the start of one write is dependent on the completion of a previous write

- A logical dependency, not a time dependency
- Examples
  - Data base & log
  - Catalogs, Volume Table of Contents
  - Index & data components



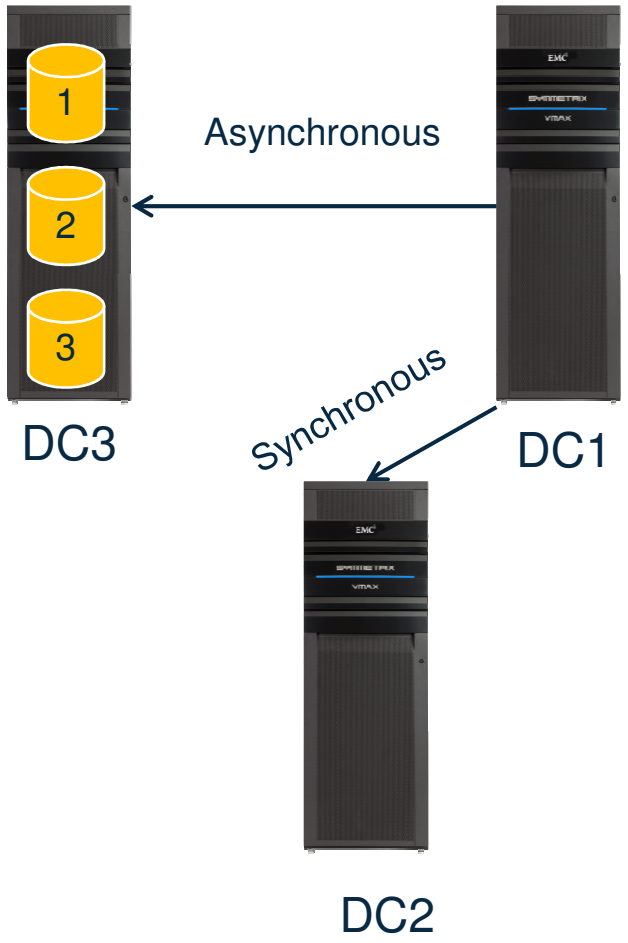
The Dependent I/O Concept

- Works great for DBMS across the Enterprise
- What can be done about logical corruption?
- What can be done for BATCH cycle?
- Point-in-time copies (checkpoints)



Today's Gaps - Physical vs Logical

- Existing solution addresses the physical risk
  - Multiple copies – continuous and point in time
  - Planned and Unplanned outage events
  - EVERYTHING gets replicated
    - Even the logical corruptions!
- Many sites use up to 2 additional full Point in Time (PIT) copies
  - Copy 1: near real time (minutes behind)
  - Copies 2 and 3: new PIT created every 12 hours
  - 24 hours to find corruption
  - Up to 24 hours of data loss



What is needed from a solution?

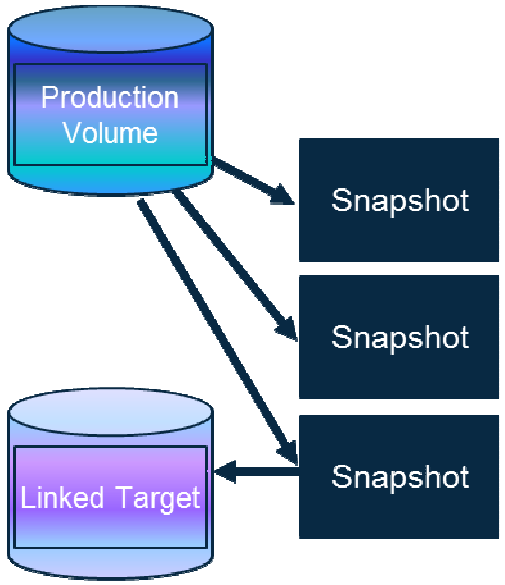
Recovery from loss of data integrity

- Provides a more granular level of application recovery from unintended changes to data stemming from
  - Processing error
  - Malicious intent
  - Human error
- Solutions that focus on the logical
  - Continuous point in time copies
  - Selectable recovery points
  - Automation

- Brings PIT recovery capability to non-database systems
- Provides applications with recovery capability when not 'designed-in'
- Provides cross-application recovery point

Specific Requirements from the Array

- Underlying technologies
  - Easily manage many data copies
  - Reduce impact
    - Target-less SNAPSHOTS
  - Increased agility
    - Multiple (~256) Snapshots per source
  - Ease of use
    - User-defined name/version number
    - Automatic expiration if desired
    - Snap entire Storage Group in single command
  - Consistent scale (>20,000 volumes)

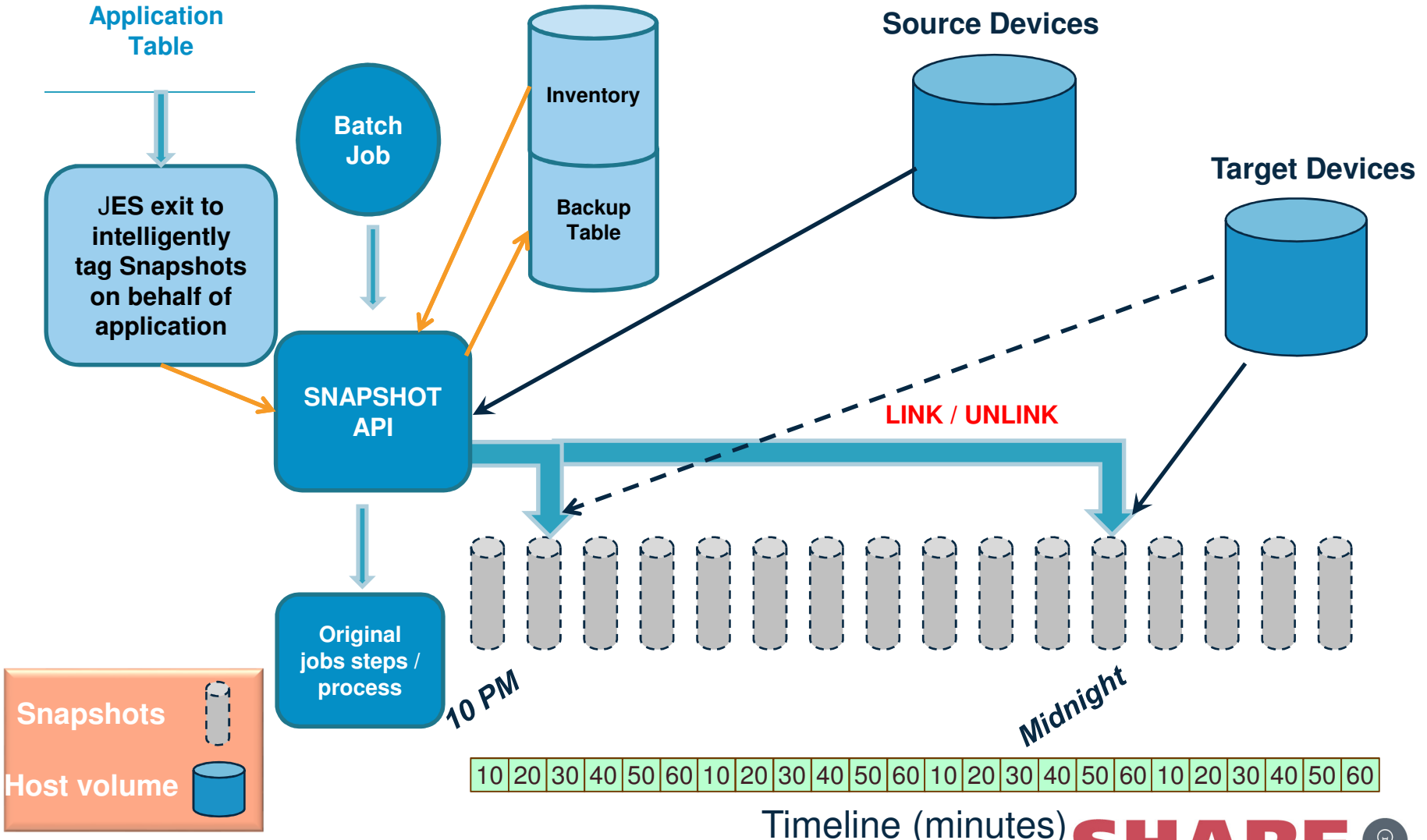


What is needed from Automation side

Recovery from loss of data integrity

- Host automation
- Execute locally or remotely
  - No Asynchronous RPO impact
  - Restore to Secondary from any snapshot
  - Create independent DR test copy
- Frequent creation/deletion of snapshots
  - Fast logical corruption recovery
- Requirements
  - Granularity of minutes
  - Select snapshots to LINK by timestamp
- Space Considerations, monitoring, reporting
- Legacy systems/application, batch – will not be re-engineered from the top down

Versioning Snapshots with Application Awareness



Restoration and Simulation

- Recover to a job or step level from previously snapped data
- Invokes data mover to recover the data necessary to re-initiate compromised job
- Recovery process driven through ISPF interface
- Simulation of the recovery identifies exposures in the recovery of application(s) or entire environment
- Ability to run “what if” scenarios at any time are critical in a world of SLA’s that must be met

Summary

- When downtime and continuous computing is required fault-tolerant rapid recovery application environments are needed.
  - With mainframes this includes both transactional and batch processing.
  - Applications and databases have advanced transaction recovery features, but batch jobs present a different problem.
  - For continuous computing in mainframe environments continuous protection of batch processing systems can also be required.
- Possibilities exist to achieve a more granular overall recovery than we know today
- Lines have blurred between operational, disaster, or synergy between the **next generation controller based snapshots and intelligent host based automation** is needed to bring near-CDP to mainframe applications

Thank you!

Rebecca Levesque

Tony Negro

#SHAREorg

SHARE is an independent volunteer-run information technology association that provides education, professional networking and industry influence.

CELEBRATING 60 YEARS OF SHARE
Influencing IT Since 1955