Resiliency for Unstructured Data in an Era of Rising Cyber Threats

2023 Spectrum Scale User Group NYC

September 27, 2023

Shailesh Shenoy
Assistant Dean for Einstein Information Technology
Senior Staff Scientist, Department of Cell Biology
shailesh.shenoy@einsteinmed.edu

Albert Einstein College of Medicine

- Research Intensive Medical School
  - > Our mission is to prepare a diverse body of students to become knowledgeable, compassionate physicians and innovative scientific investigators, and to create new knowledge

- About 1,900 Faculty & 1,000 Students (MD & PhD)
- Funding Primarily from the National Institutes of Health
- Part of Montefiore Medicine Academic Health System

EINSTEIN
Albert Einstein College of Medicine
Montefiore

Business Requirements

- Single Namespace & Multi-Protocol Support
- Mature, Robust, & Innovative High-Performance Filesystem
- Tiering for Cost Management
- Hardware/Software Ecosystem with Simple Scaling & Management
- Hybrid Cloud Capabilities
- Skilled Solution Partners
- Best of Class Support Organization
- High Speed Backup, Archiving and Disaster Recovery

EINSTEIN
Albert Einstein College of Medicine
Montefiore

Data Hygiene Goals

- Tiering: NVMe, HDD, Archive (Single Namespace)
- Encryption at Rest
- High Availability at Each Tier: No Single Point of Failure & Site Resiliency
- Air Gapped Backups Daily
- Disaster Recovery: Separate Daily Backup Copy Offsite with Ability to Restore
- Immutable File System Snapshots
- Independent File System for Protected Research Data
  - PII (Personally Identifiable Information) and PHI (Protected Health Information)

EINSTEIN
Albert Einstein College of Medicine

Montefiore

Storage Scale at Einstein

Storage Scale is the central data repository for **everything**:

- Instrumentation: microscopes, genome sequencers, and others
- Data storage for High-performance computing, including scratch storage
- Home directories and other user information

Storage Scale is **the** platform for data management and sharing:

- POSIX protocol for HPC
- NFS for other data center workloads
- SMB for desktop client access
- Aspera for outside file sharing
- Object for published data sets
- Tiering between Flash, HDD, and tape to balance performance and economics
- Persistent storage for containers

EINSTEIN
Albert Einstein College of Medicine

Montefiore

Data Resiliency

Data Resiliency is the ability to keep all the data available that an we need to continue functioning, even in the face of disruptions such as:

- Natural disasters
- Human-caused disasters
- “Oops!”
- Cyber attacks
- Compromised insider account

EINSTEIN
Albert Einstein College of Medicine
Montefiore

17%
of Cyber Attacks are Ransomware

26%
Clients who paid the ransom still could not recover the data

23
days, average recovery after a ransomware attack

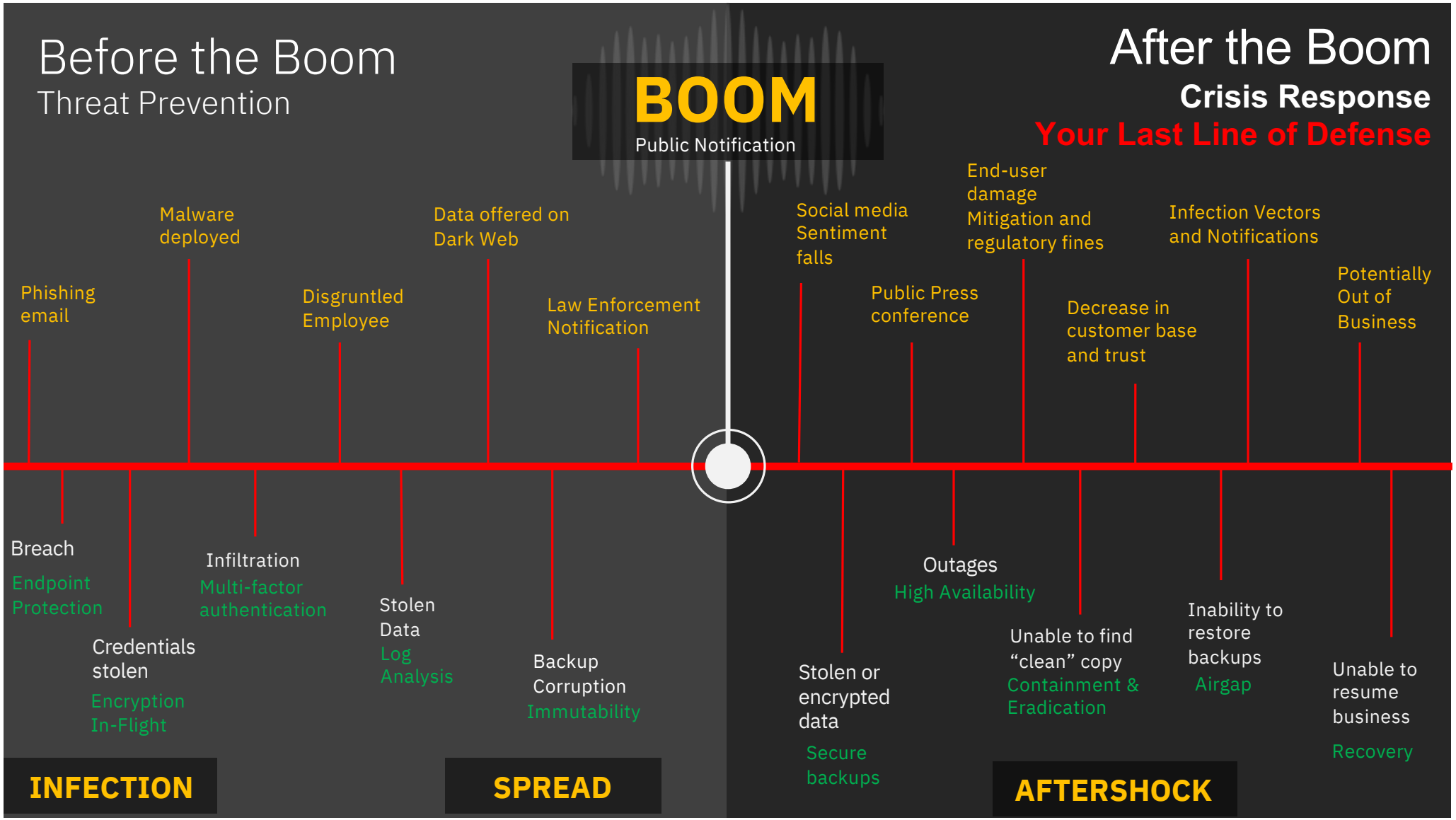
2X
Cyber Attacks YTY

21%
dormant threats, up from 5% YTY

45%
45% of production data affected

Data Resilience

Why should we care?


Before the Boom
Threat Prevention
BOOM
Public Notification
After the Boom
Crisis Response
Your Last Line of Defense
Phishing email
Malware deployed
Disgruntled Employee
Data offered on Dark Web
Law Enforcement Notification
Social media Sentiment falls
Public Press conference
End-user damage Mitigation and regulatory fines
Decrease in customer base and trust
Infection Vectors and Notifications
Potentially Out of Business
Breach
Endpoint Protection
Credentials stolen
Encryption In-Flight
Infiltration
Multi-factor authentication
Stolen Data
Log Analysis
Backup Corruption
Immutability
Stolen or encrypted data
Secure backups
Outages
High Availability
Unable to find "clean" copy
Containment & Eradication
Inability to restore backups
Airgap
Unable to resume business
Recovery
INFECTION
SPREAD
AFTERSHOCK

Business Impacts of Cyber Attacks

BUSINESS IMPACT

$5.2m

Average annual cost of Ransomware

Why Data Resiliency is Important to Life Sciences

Our product is intellectual property.

A data breach would additionally cost:

- Lost opportunities:
  - publications
  - training
  - competitiveness for grant funding
- Institutional Reputation
- Potential Exfiltration of Protected Information (PHI/PII)

EINSTEIN
Albert Einstein College of Medicine
Montefiore

Scale Architecture at Einstein

Data is protected using ACLs:

- Every directory has associated groups controlling access (RW/RO)
- Active Directory controls users' memberships in those groups

Stretched cluster architecture is used for active/active high availability.

Storage Archive is used to tier cold data to tape while maintaining a single name space.

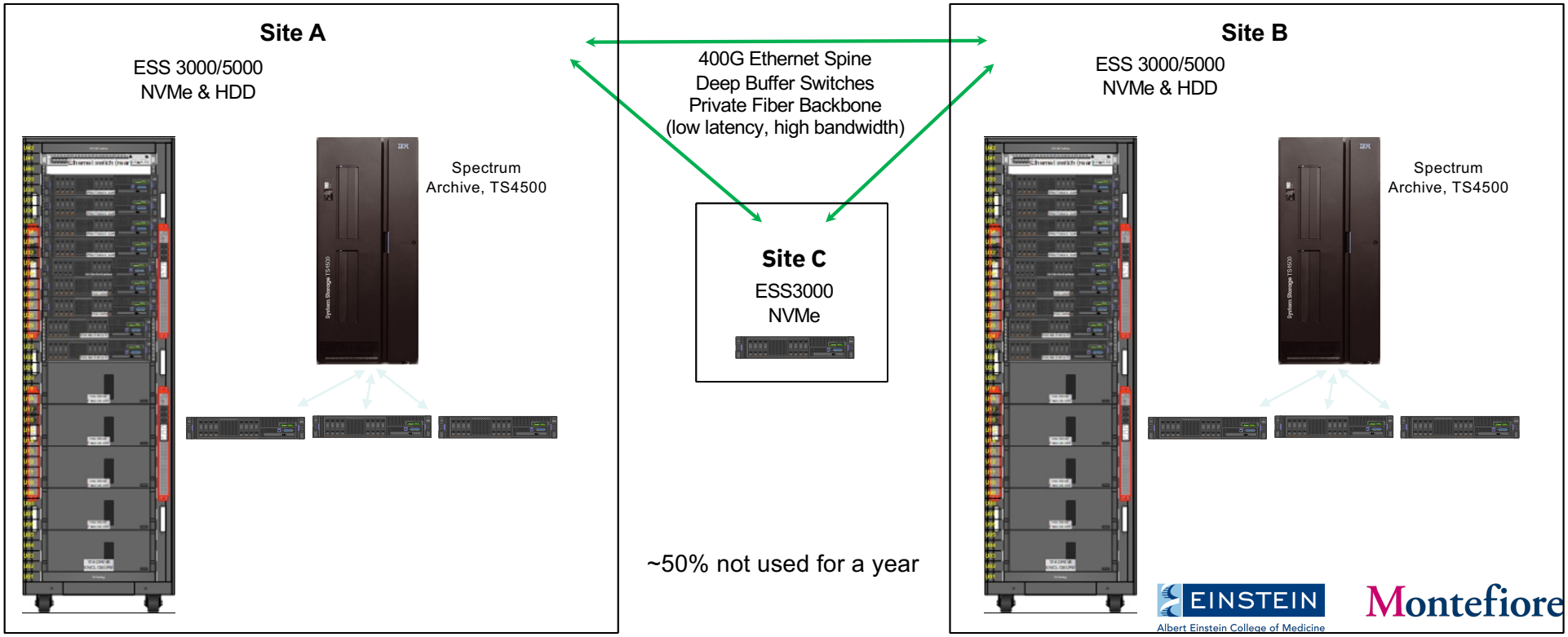
Snapshots are used for recovery points.

Storage Protect is used to make air-gapped back up of data.

EINSTEIN
Albert Einstein College of Medicine

Montefiore

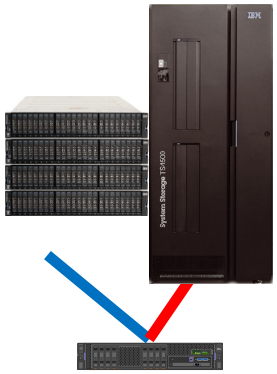
Architecture: Stretch Cluster w/3-Site inode Replication



Architecture: Backup w/Off-Campus DR Copy & Restore

Site C

Backup
Storage Protect, FS7200,
TS4500, ESS3000

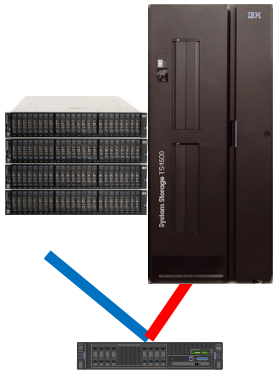


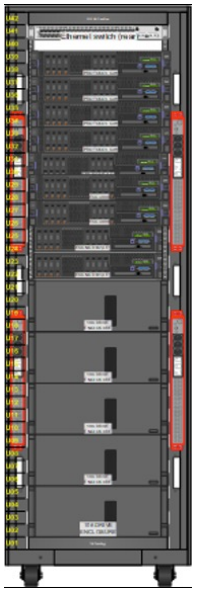
Backup with a Storage Protect Instance so that back-up data is in a different location from the stretch cluster locations

Private, Encrypted Fiber
(low latency, high bandwidth)

Site D

Storage Protect, FS7200,
TS4500, ESS3000/5000
NVMe & HDD





Off Campus:
separate power
grid

EINSTEIN
Albert Einstein College of Medicine

Montefiore

NIST Cyber Resiliency Framework

Identify:
Defining a organizational understanding to build or improve **cyber resiliency plan** – critical assets and strategy

Protect:
Implementing Safeguards to ensure delivery of critical services – protecting against vulnerabilities before they are exploited

Detect:
Detecting occurrence of cyber security events – timely, continuous monitoring, detection processes

Respond:
Taking action regarding a detected event – analysis, **contain**, mitigation, and communication

Recover:
Restore capabilities and services - recovery, improvements, communications

EINSTEIN
Albert Einstein College of Medicine
Montefiore

Integration of Scale with Qradar SIEM and Correlation Engine

- IBM QRadar is ingesting data from File Audit Logs, Network flows, DNS, and Active Directory
- QRadar correlates all these different streams to learn normal access patterns(user behavior analytics) and detect abnormalities such data exfiltration.
- QRadar SOAR will initiate defensive actions such as immutable snapshots.
- Investigating role that Storage Defender will have in this environment.

IBM Storage Protect is used to back up the entire file system

- Additionally, regular restores in a DR location of the critical part of the file system will maintain a point-in-time replica.
- This replica will be ready for use in an emergency.
- Investigating taking snapshots of this replica for additional fast restore points.

EINSTEIN Albert Einstein College of Medicine Montefiore

Questions

- Thank you

EINSTEIN
Albert Einstein College of Medicine
Montefiore