Prevention, Detection & Recovery

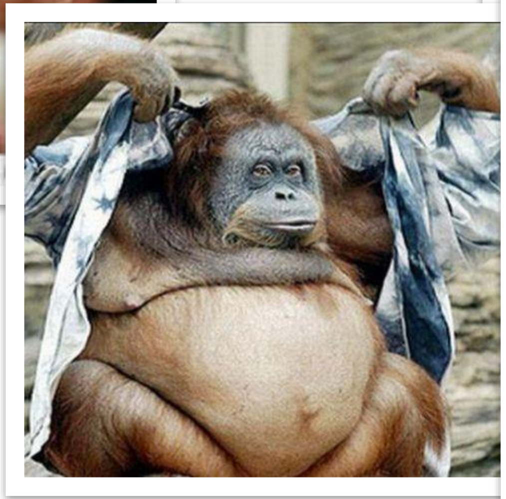
Applying RAG AI & Seq2Seq to Ransomware

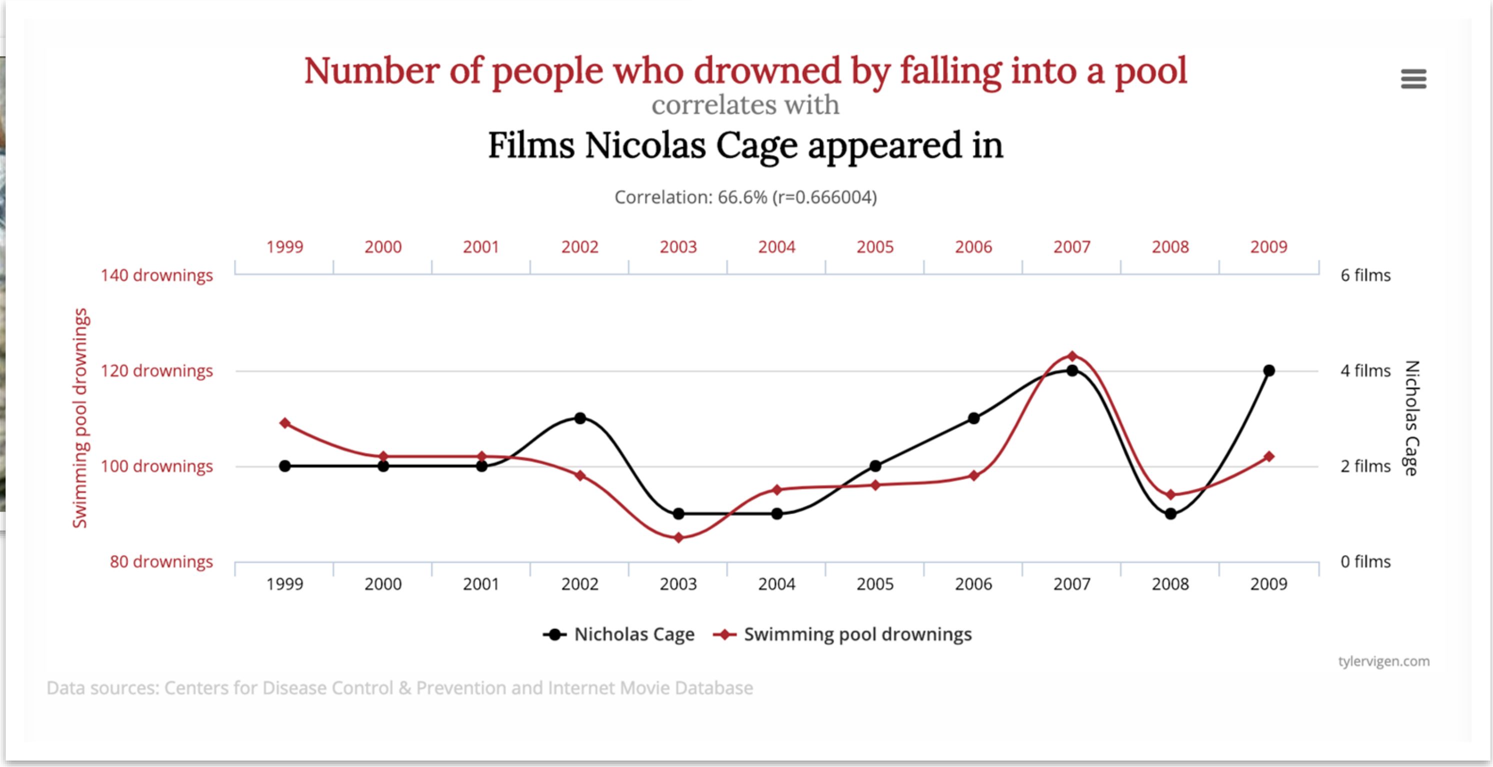
Mike Riley

Sr. Engineer, Cohesity

Slide to Satisfy Slide Request

(Yes, you can have my slides)





Cohesity: Making Backup Sexy Again
Leveraging Backup for Data Security and Management

Gartner 3x MQ Leader

Forbes 4x Cloud 100

FORRESTER 2x Leader

3700+ CUSTOMERS

6 OF THE FORTUNE 10

40% OF THE FORTUNE 100

85+ AVERAGE NPS SCORE

8 OF THE TOP 20 US BANKS

520+ GLOBAL FIN SERV CUSTOMERS

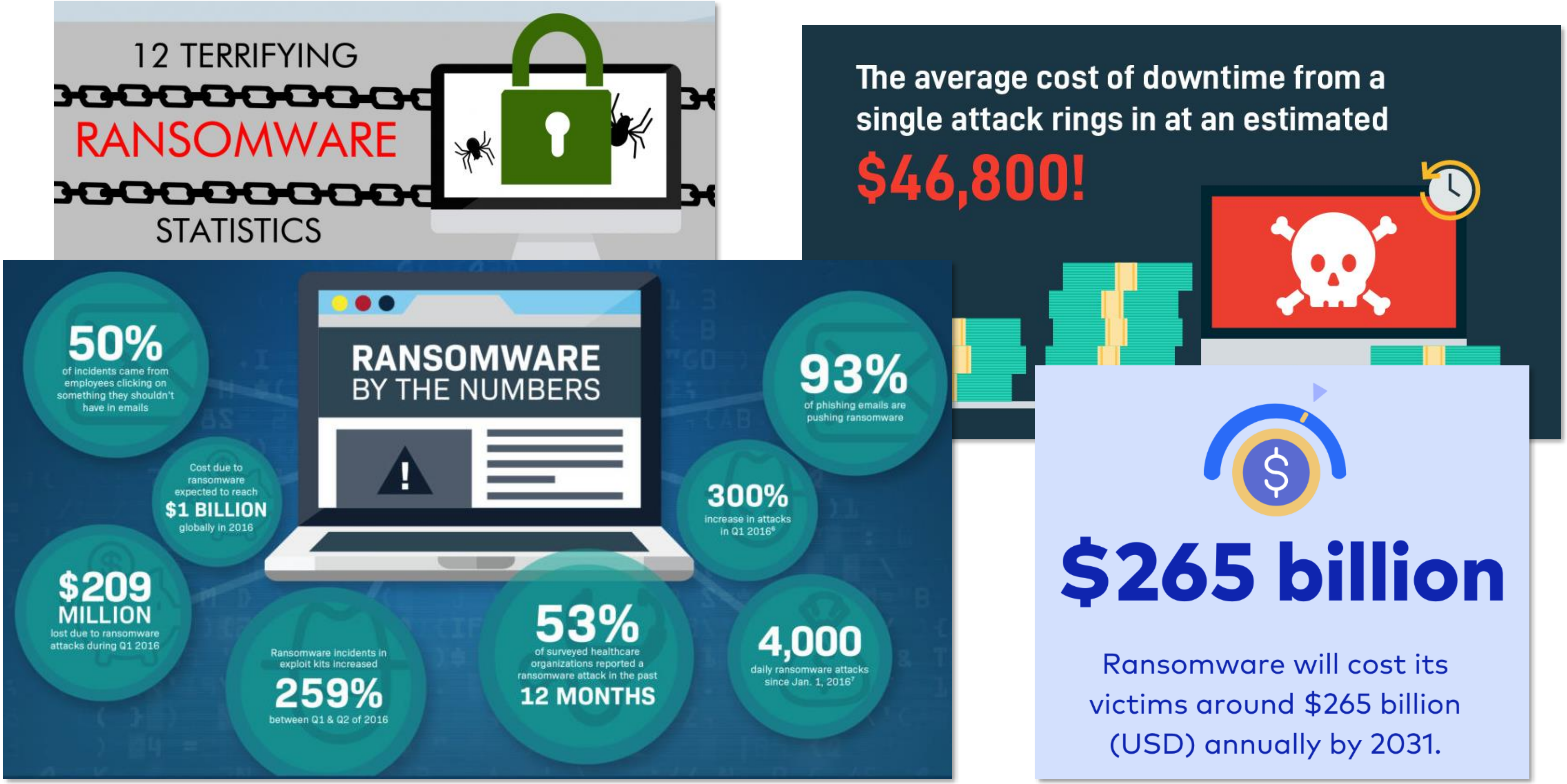
210+ FEDERAL AGENCIES

1500+ GLOBAL PARTNERS

SEQUOIA DFJ amazon

cisco Hewlett Packard Enterprise SoftBank

Obligatory “Ransomware is Scary” Slide



12 TERRIFYING
RANSOMWARE
STATISTICS

The average cost of downtime from a single attack rings in at an estimated **$46,800!**

RANSOMWARE
BY THE NUMBERS

50% of incidents came from employees clicking on something they shouldn't have in emails

Cost due to ransomware expected to reach **$1 BILLION** globally in 2016

$209 MILLION lost due to ransomware attacks during Q1 2016

Ransomware incidents in exploit kits increased **259%** between Q1 & Q2 of 2016

53% of surveyed healthcare organizations reported a ransomware attack in the past **12 MONTHS**

4,000 daily ransomware attacks since Jan. 1, 2016[7]

300% increase in attacks in Q1 2016[6]

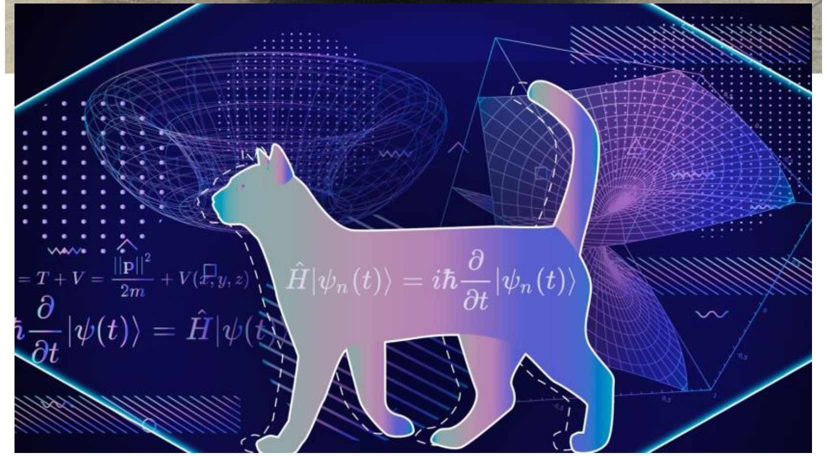
93% of phishing emails are pushing ransomware

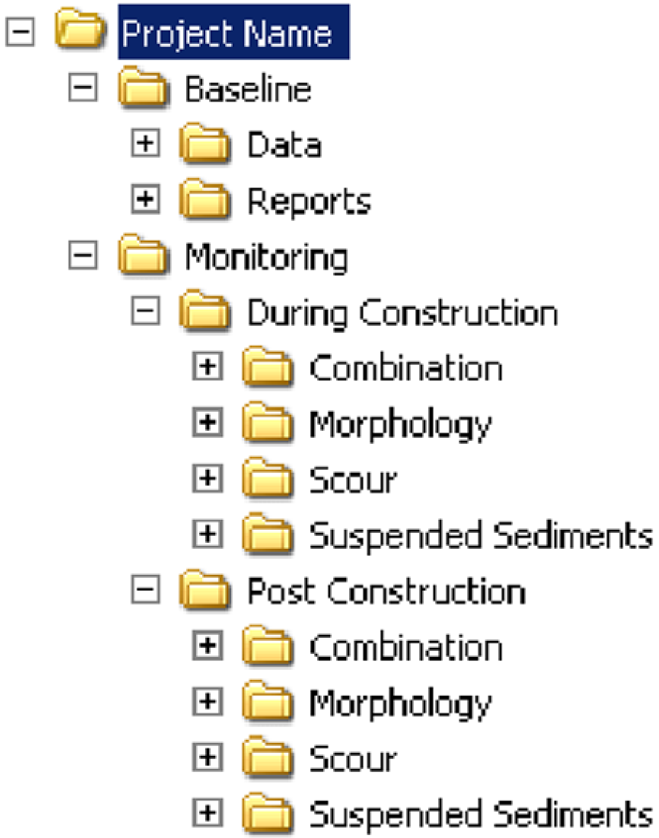
$265 billion

Ransomware will cost its victims around $265 billion (USD) annually by 2031.

What’s the one app that touches everything critical?

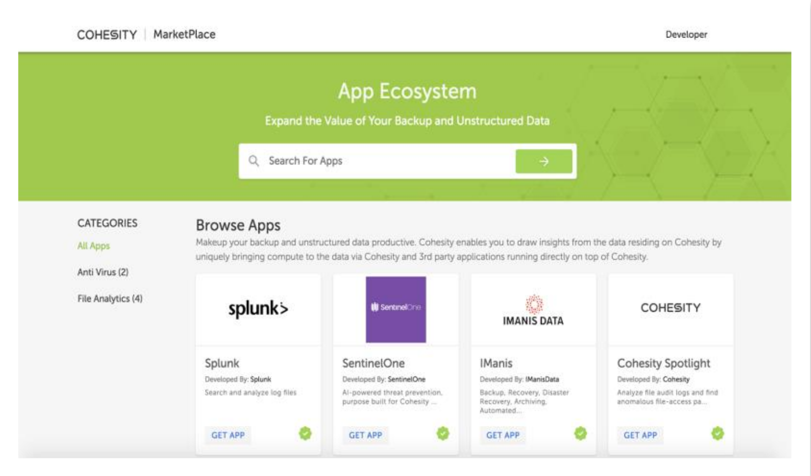
Backup: The Schroedinger's Cat of the IT Industry


$\hat{H}|\psi_n(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi_n(t)\rangle$
$= T + V = \frac{||\mathbf{p}||^2}{2m} + V(x, y, z)$
$\frac{\partial}{\partial t}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle$



- ⊟ Project Name
  - ⊟ Baseline
    - ⊞ Data
    - ⊞ Reports
  - ⊟ Monitoring
    - ⊟ During Construction
      - ⊞ Combination
      - ⊞ Morphology
      - ⊞ Scour
      - ⊞ Suspended Sediments
    - ⊟ Post Construction
      - ⊞ Combination
      - ⊞ Morphology
      - ⊞ Scour
      - ⊞ Suspended Sediments

Bring Apps to the Data!



App Ecosystem

Expand the Value of Your Backup and Unstructured Data

Browse Apps

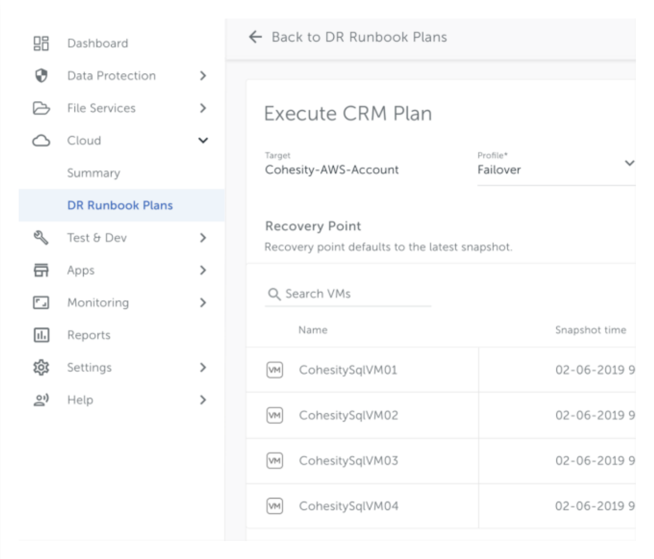
Makeup your backup and unstructured data productive. Cohesity enables you to draw insights from the data residing on Cohesity by uniquely bringing compute to the data via Cohesity and 3rd party applications running directly on top of Cohesity.

Splunk
Developed By: Splunk
Search and analyze log files

SentinelOne
Developed By: SentinelOne
AI-powered threat prevention, purpose built for Cohesity ...

IManis
Developed By: IManisData
Backup, Recovery, Disaster Recovery, Archiving, Automated...

Cohesity Spotlight
Developed By: Cohesity
Analyze file audit logs and find anomalous file-access pa...



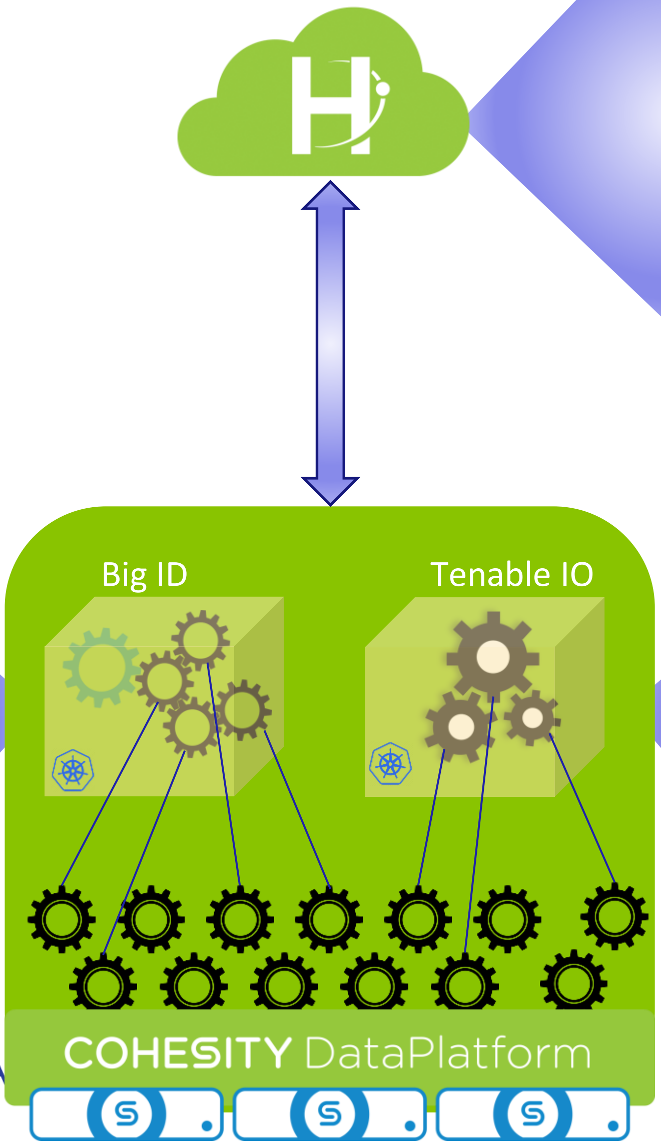
Execute CRM Plan

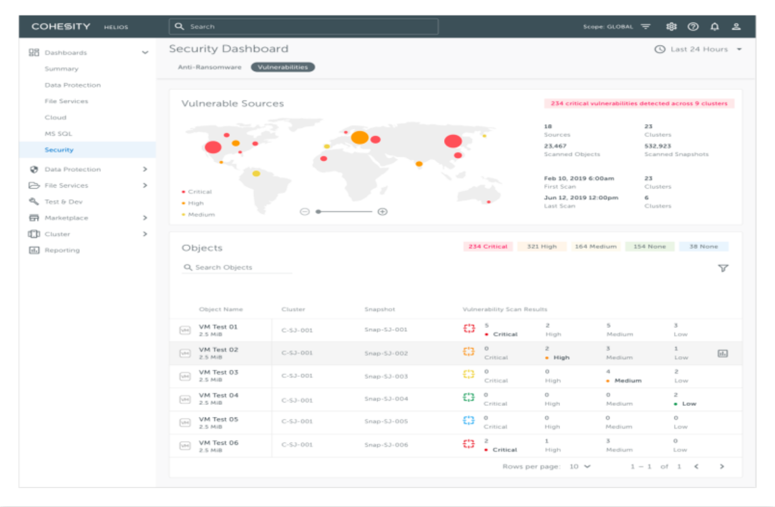
Target: Cohesity-AWS-Account
Profile*: Failover

Recovery Point

Recovery point defaults to the latest snapshot.

Name	Snapshot time
CohesitySqlVM01	02-06-2019 9
CohesitySqlVM02	02-06-2019 9
CohesitySqlVM03	02-06-2019 9
CohesitySqlVM04	02-06-2019 9





Security Dashboard

Vulnerable Sources

234 critical vulnerabilities detected across 9 clusters

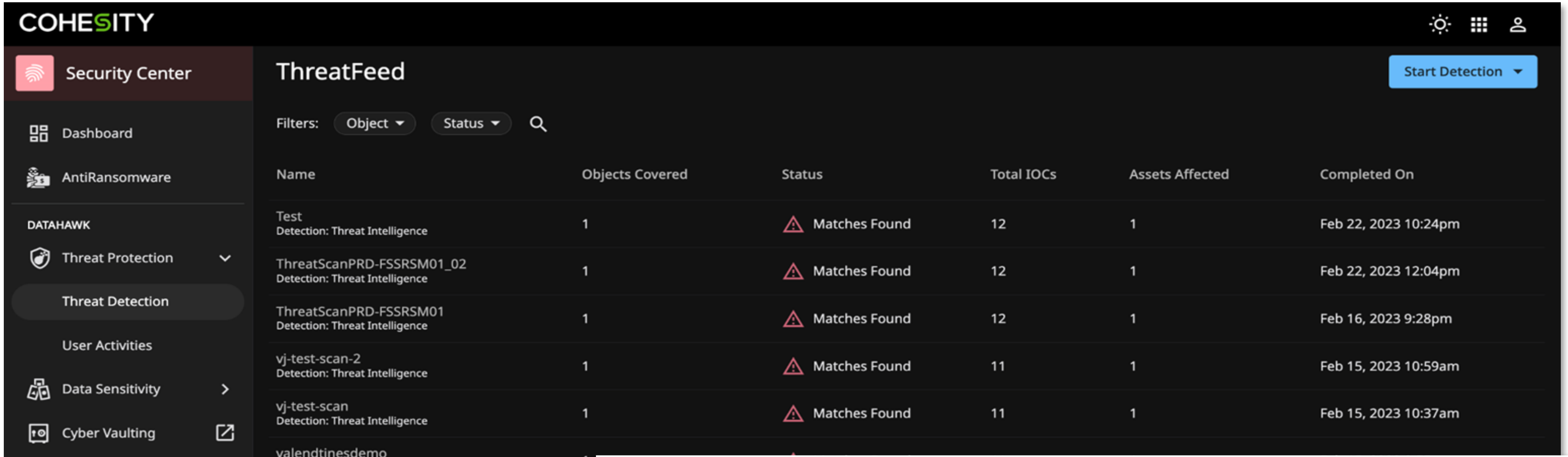
18 Sources
23 Clusters
23,467 Scanned Objects
532,923 Scanned Snapshots
Feb 10, 2019 6:00am First Scan
23 Clusters
Jun 12, 2019 12:00pm Last Scan
6 Clusters

Objects

234 Critical | 321 High | 164 Medium | 154 None | 38 None

Object Name	Cluster	Snapshot	Vulnerability Scan Results			
VM Test 01 2.5 MiB	C-SJ-001	Snap-SJ-001	5 Critical	2 High	5 Medium	3 Low
VM Test 02 2.5 MiB	C-SJ-001	Snap-SJ-002	0 Critical	2 High	3 Medium	1 Low
VM Test 03 2.5 MiB	C-SJ-001	Snap-SJ-003	0 Critical	0 High	4 Medium	2 Low
VM Test 04 2.5 MiB	C-SJ-001	Snap-SJ-004	0 Critical	0 High	0 Medium	2 Low
VM Test 05 2.5 MiB	C-SJ-001	Snap-SJ-005	0 Critical	0 High	0 Medium	0 Low
VM Test 06 2.5 MiB	C-SJ-001	Snap-SJ-006	2 Critical	1 High	3 Medium	0 Low

Bring Apps to the Data: Threat Detection App

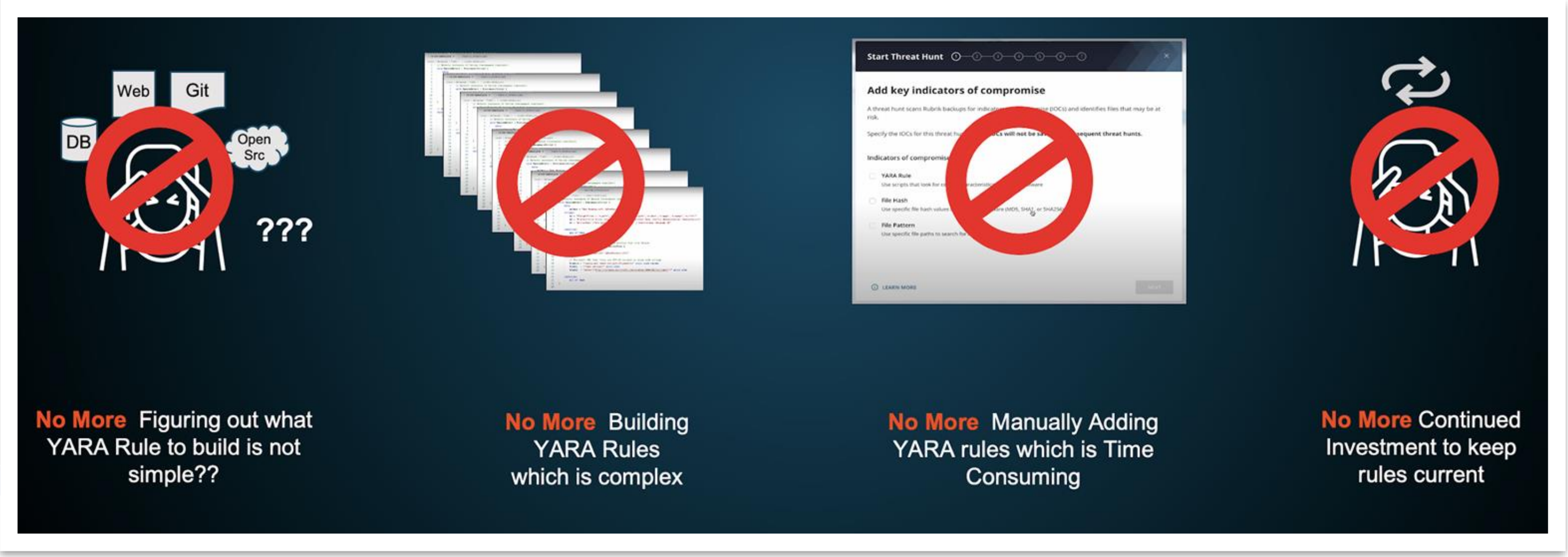


ThreatFeed

Name	Objects Covered	Status	Total IOCs	Assets Affected	Completed On
Test Detection: Threat Intelligence	1	⚠ Matches Found	12	1	Feb 22, 2023 10:24pm
ThreatScanPRD-FSSRSM01_02 Detection: Threat Intelligence	1	⚠ Matches Found	12	1	Feb 22, 2023 12:04pm
ThreatScanPRD-FSSRSM01 Detection: Threat Intelligence	1	⚠ Matches Found	12	1	Feb 16, 2023 9:28pm
vj-test-scan-2 Detection: Threat Intelligence	1	⚠ Matches Found	11	1	Feb 15, 2023 10:59am
vj-test-scan Detection: Threat Intelligence	1	⚠ Matches Found	11	1	Feb 15, 2023 10:37am
valendtinesdemo Detection: Threat Intelligence	1				
Scan_1 Detection: Threat Intelligence	1				

FUN FACT?

Let's say it take 3 mins to Research, create & upload 1 yara rule. To run 100,000 rules it will take 208 days working 24x7



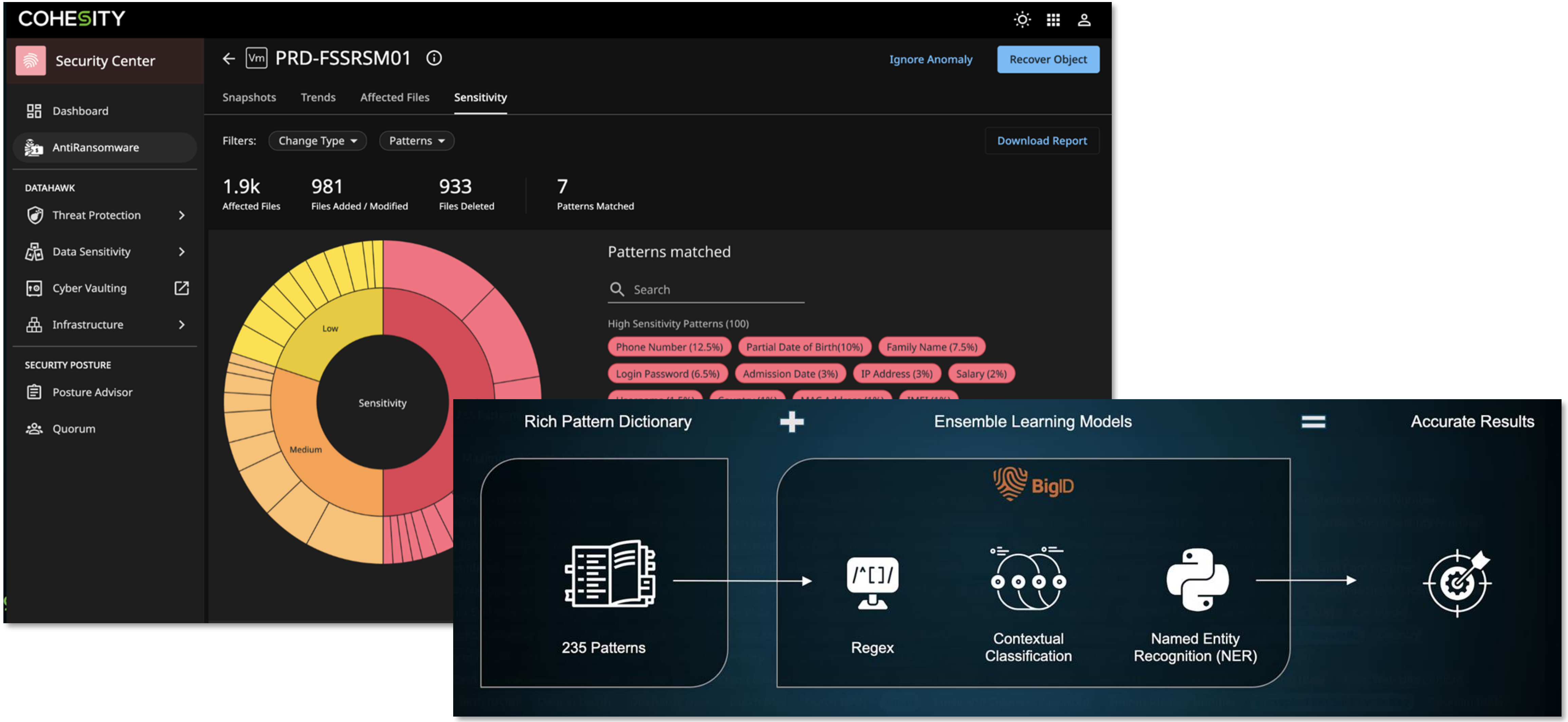
No More Figuring out what YARA Rule to build is not simple??

No More Building YARA Rules which is complex

No More Manually Adding YARA rules which is Time Consuming

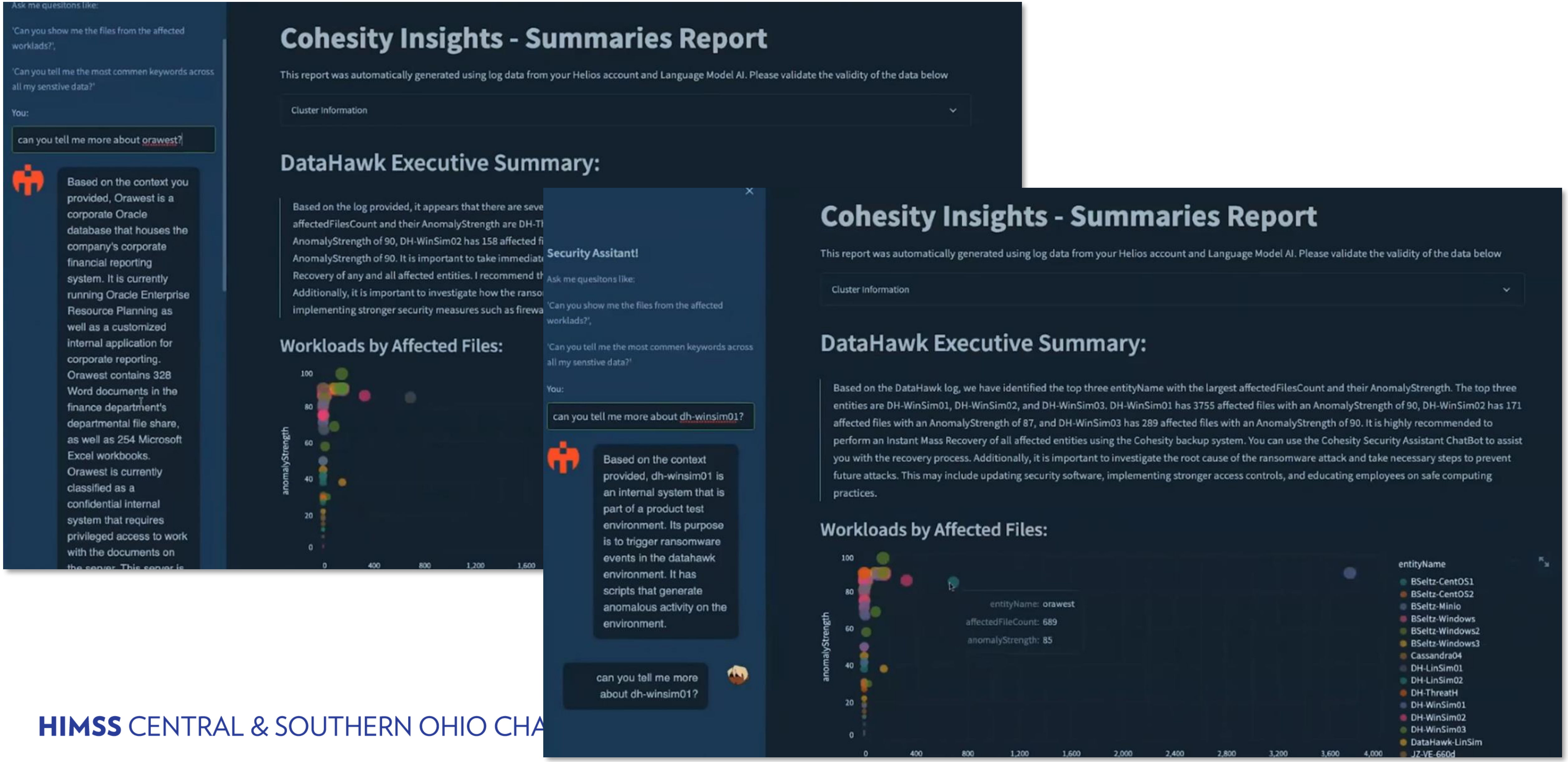
No More Continued Investment to keep rules current

Bring Apps to the Data: Data Classification



HIMSS CENTRAL & SOUTHERN OHIO CHAPTER

Bring Apps to the Data: AI (RAG + Seq2Seq)



Ask me quesitons like:

'Can you show me the files from the affected worklads?',

'Can you tell me the most commen keywords across all my senstive data?'

You:

can you tell me more about orawest?

Based on the context you provided, Orawest is a corporate Oracle database that houses the company's corporate financial reporting system. It is currently running Oracle Enterprise Resource Planning as well as a customized internal application for corporate reporting. Orawest contains 328 Word documents in the finance department's departmental file share, as well as 254 Microsoft Excel workbooks. Orawest is currently classified as a confidential internal system that requires privileged access to work with the documents on the server. This server is

Cohesity Insights - Summaries Report

This report was automatically generated using log data from your Helios account and Language Model AI. Please validate the validity of the data below

Cluster Information

DataHawk Executive Summary:

Based on the log provided, it appears that there are seve
affectedFilesCount and their AnomalyStrength are DH-Tl
AnomalyStrength of 90, DH-WinSim02 has 158 affected fi
AnomalyStrength of 90. It is important to take immediat
Recovery of any and all affected entities. I recommend th
Additionally, it is important to investigate how the ranso
implementing stronger security measures such as firewa

Workloads by Affected Files:

Security Assitant!

Ask me quesitons like:

'Can you show me the files from the affected worklads?',

'Can you tell me the most commen keywords across all my senstive data?'

You:

can you tell me more about dh-winsim01?

Based on the context provided, dh-winsim01 is an internal system that is part of a product test environment. Its purpose is to trigger ransomware events in the datahawk environment. It has scripts that generate anomalous activity on the environment.

can you tell me more about dh-winsim01?

Cohesity Insights - Summaries Report

This report was automatically generated using log data from your Helios account and Language Model AI. Please validate the validity of the data below

Cluster Information

DataHawk Executive Summary:

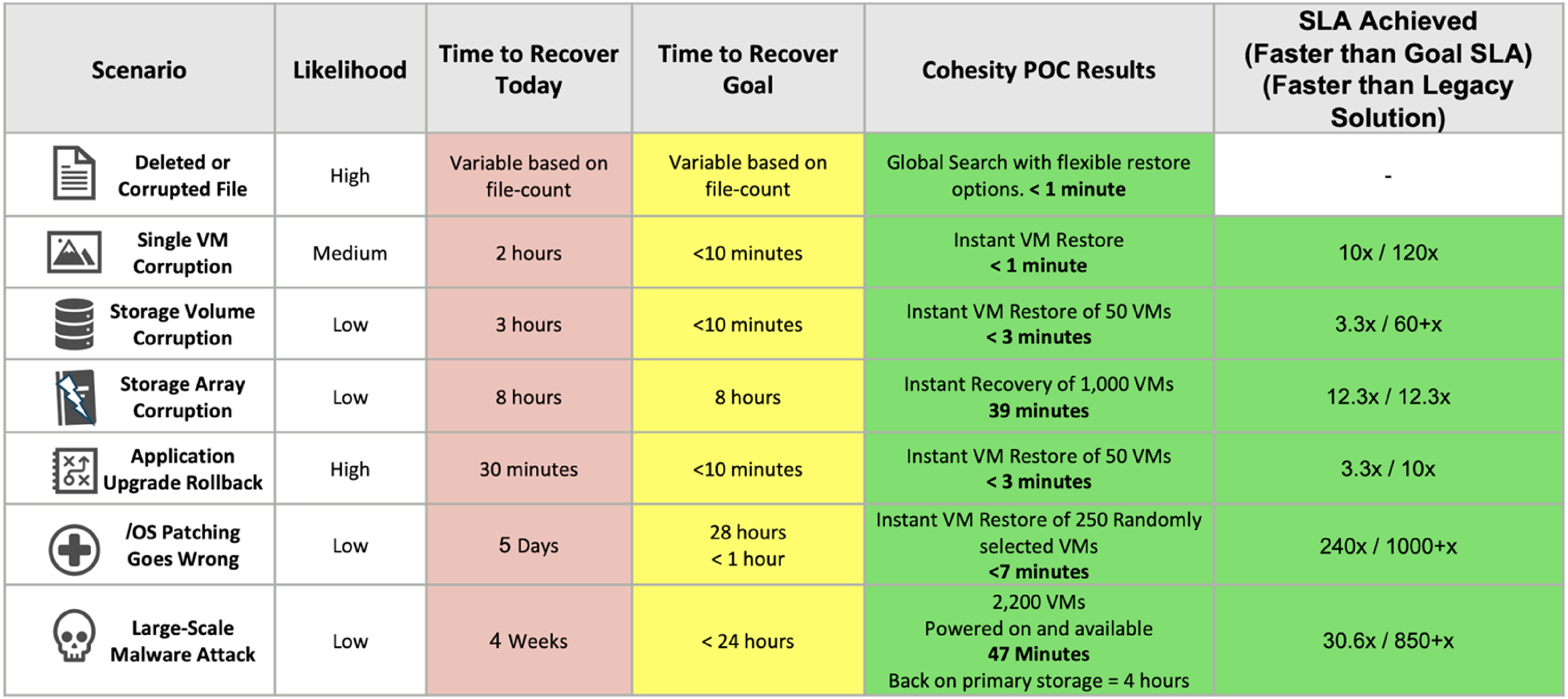
Based on the DataHawk log, we have identified the top three entityName with the largest affectedFilesCount and their AnomalyStrength. The top three entities are DH-WinSim01, DH-WinSim02, and DH-WinSim03. DH-WinSim01 has 3755 affected files with an AnomalyStrength of 90, DH-WinSim02 has 171 affected files with an AnomalyStrength of 87, and DH-WinSim03 has 289 affected files with an AnomalyStrength of 90. It is highly recommended to perform an Instant Mass Recovery of all affected entities using the Cohesity backup system. You can use the Cohesity Security Assistant ChatBot to assist you with the recovery process. Additionally, it is important to investigate the root cause of the ransomware attack and take necessary steps to prevent future attacks. This may include updating security software, implementing stronger access controls, and educating employees on safe computing practices.

Workloads by Affected Files:

entityName: orawest
affectedFileCount: 689
anomalyStrength: 85

Backups are Worthless; Restores are Priceless

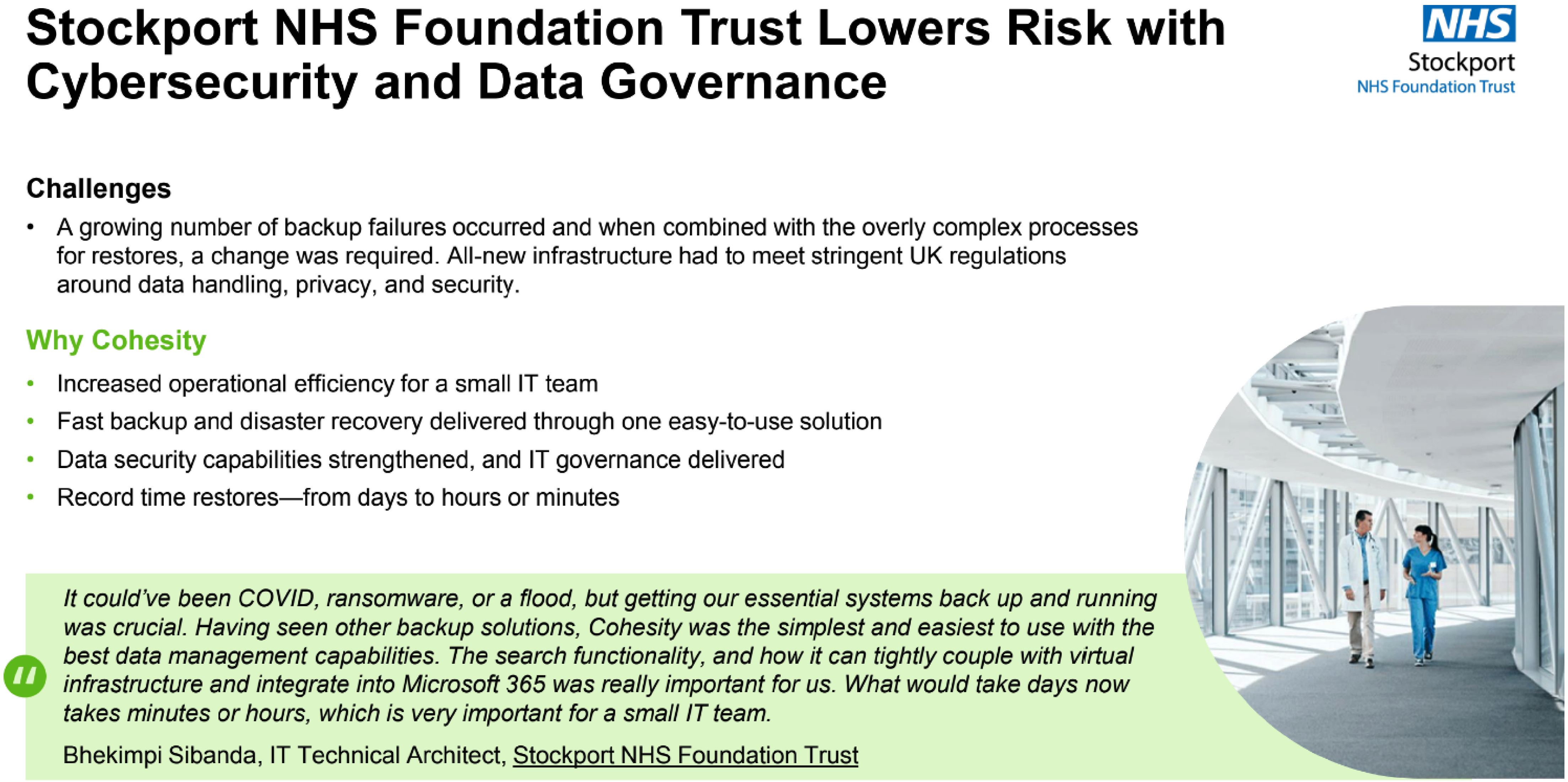
Leverage AI/ML to compress triage and accelerate restores



Scenario	Likelihood	Time to Recover Today	Time to Recover Goal	Cohesity POC Results	SLA Achieved (Faster than Goal SLA) (Faster than Legacy Solution)
Deleted or Corrupted File	High	Variable based on file-count	Variable based on file-count	Global Search with flexible restore options. **< 1 minute**	-
Single VM Corruption	Medium	2 hours	<10 minutes	Instant VM Restore **< 1 minute**	10x / 120x
Storage Volume Corruption	Low	3 hours	<10 minutes	Instant VM Restore of 50 VMs **< 3 minutes**	3.3x / 60+x
Storage Array Corruption	Low	8 hours	8 hours	Instant Recovery of 1,000 VMs **39 minutes**	12.3x / 12.3x
Application Upgrade Rollback	High	30 minutes	<10 minutes	Instant VM Restore of 50 VMs **< 3 minutes**	3.3x / 10x
/OS Patching Goes Wrong	Low	5 Days	28 hours < 1 hour	Instant VM Restore of 250 Randomly selected VMs **<7 minutes**	240x / 1000+x
Large-Scale Malware Attack	Low	4 Weeks	< 24 hours	2,200 VMs Powered on and available **47 Minutes** Back on primary storage = 4 hours	30.6x / 850+x

Case Studies

Case Study 1: Stockport NHS Foundation



Stockport NHS Foundation Trust Lowers Risk with Cybersecurity and Data Governance

NHS
Stockport
NHS Foundation Trust

Challenges

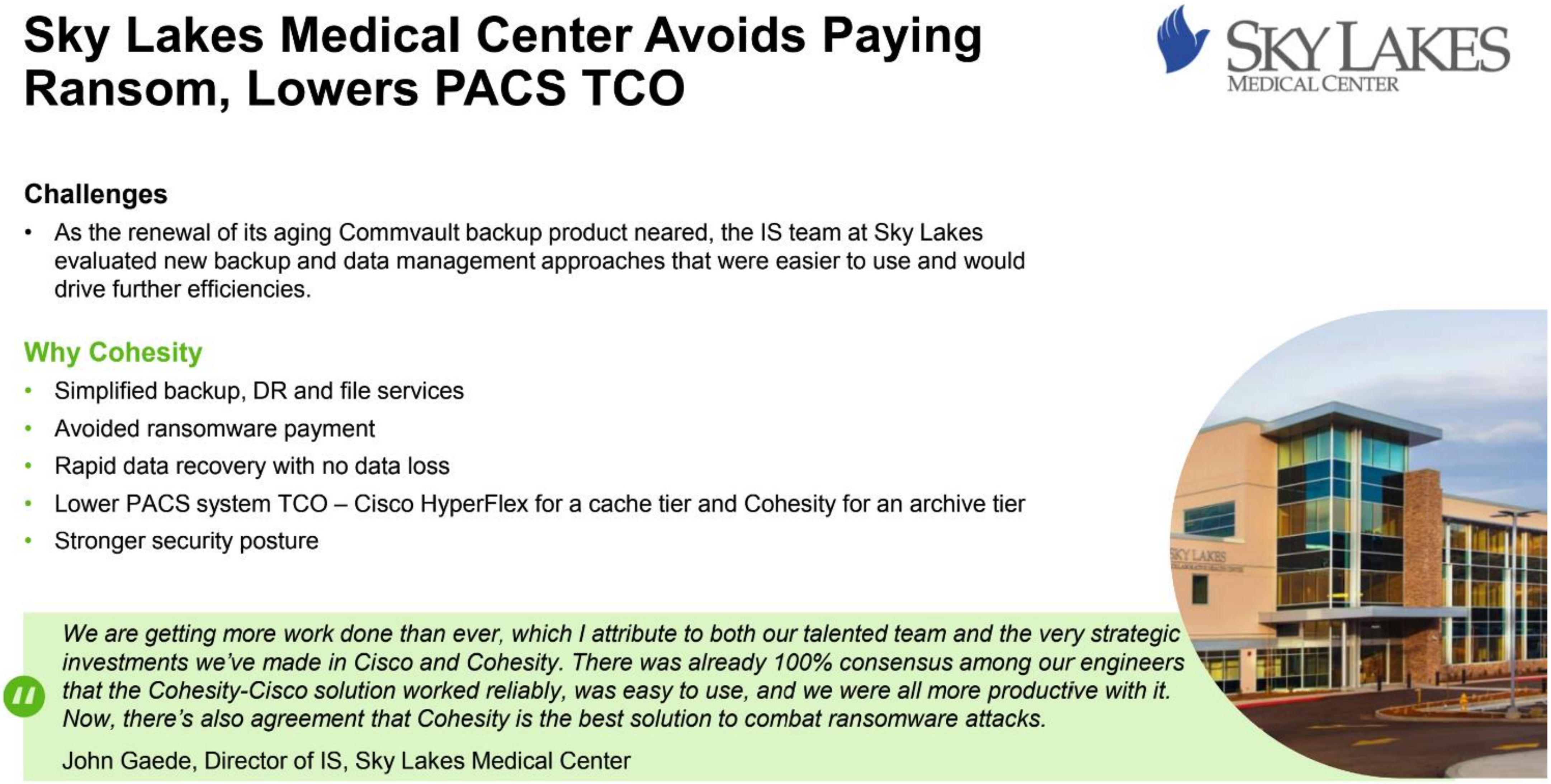
- A growing number of backup failures occurred and when combined with the overly complex processes for restores, a change was required. All-new infrastructure had to meet stringent UK regulations around data handling, privacy, and security.

Why Cohesity

- Increased operational efficiency for a small IT team
- Fast backup and disaster recovery delivered through one easy-to-use solution
- Data security capabilities strengthened, and IT governance delivered
- Record time restores—from days to hours or minutes

> *It could've been COVID, ransomware, or a flood, but getting our essential systems back up and running was crucial. Having seen other backup solutions, Cohesity was the simplest and easiest to use with the best data management capabilities. The search functionality, and how it can tightly couple with virtual infrastructure and integrate into Microsoft 365 was really important for us. What would take days now takes minutes or hours, which is very important for a small IT team.*
>
> Bhekimpi Sibanda, IT Technical Architect, Stockport NHS Foundation Trust

Case Study 2: Sky Lakes Medical Center



Sky Lakes Medical Center Avoids Paying Ransom, Lowers PACS TCO

Challenges

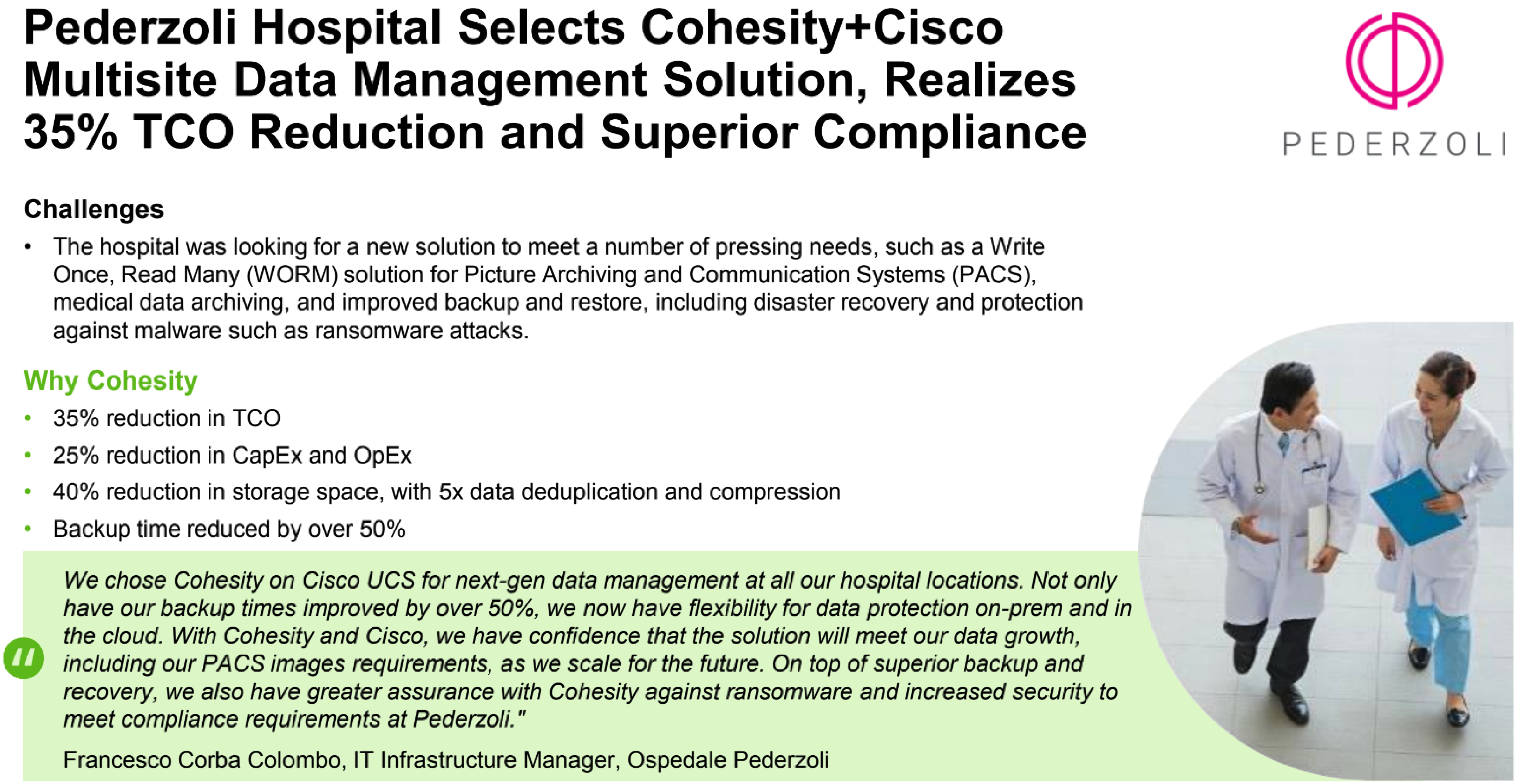
- As the renewal of its aging Commvault backup product neared, the IS team at Sky Lakes evaluated new backup and data management approaches that were easier to use and would drive further efficiencies.

Why Cohesity

- Simplified backup, DR and file services
- Avoided ransomware payment
- Rapid data recovery with no data loss
- Lower PACS system TCO – Cisco HyperFlex for a cache tier and Cohesity for an archive tier
- Stronger security posture

> *We are getting more work done than ever, which I attribute to both our talented team and the very strategic investments we've made in Cisco and Cohesity. There was already 100% consensus among our engineers that the Cohesity-Cisco solution worked reliably, was easy to use, and we were all more productive with it. Now, there's also agreement that Cohesity is the best solution to combat ransomware attacks.*
>
> John Gaede, Director of IS, Sky Lakes Medical Center

Case Study 3: Pederzoli Hospital



Pederzoli Hospital Selects Cohesity+Cisco Multisite Data Management Solution, Realizes 35% TCO Reduction and Superior Compliance

PEDERZOLI

Challenges

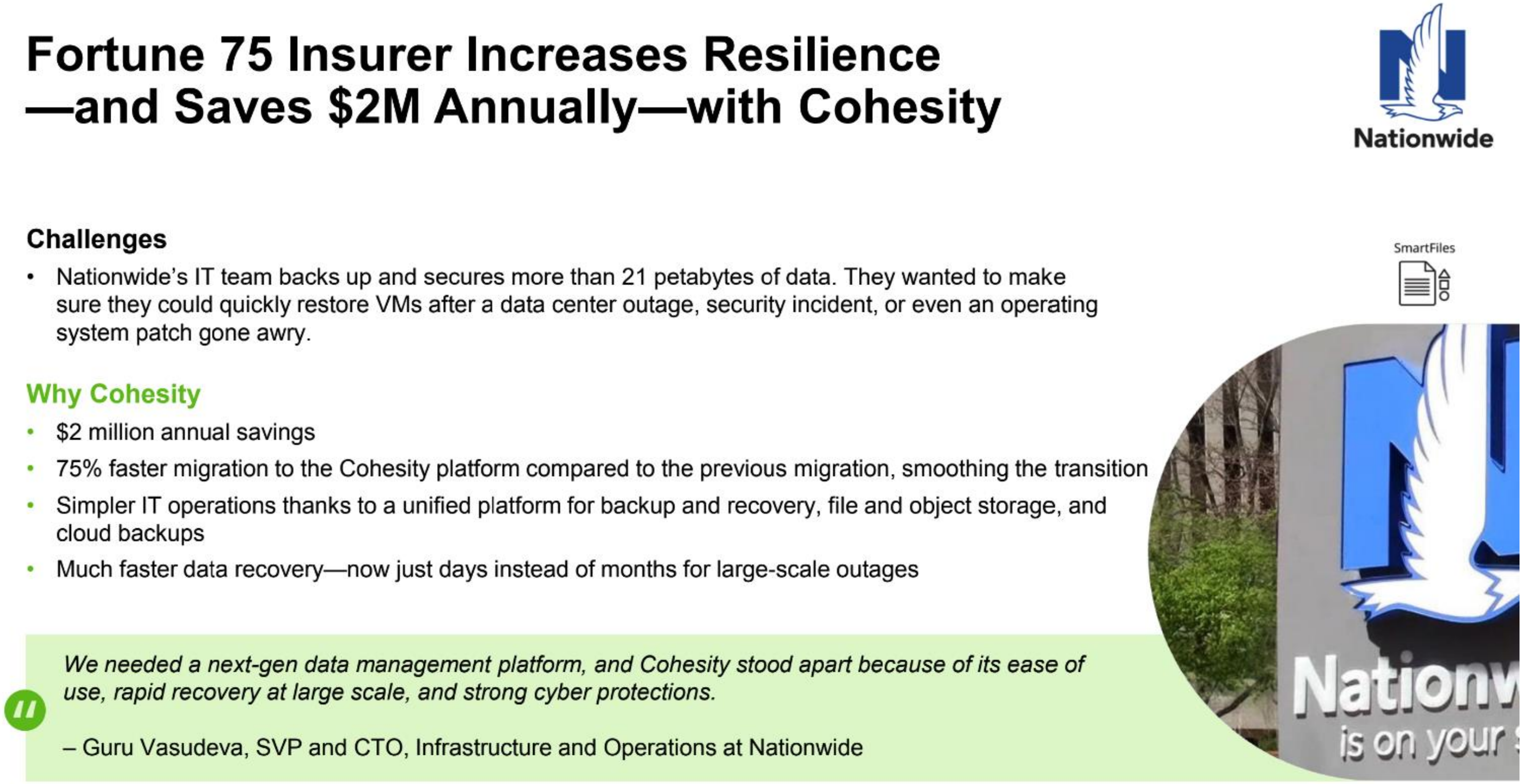
- The hospital was looking for a new solution to meet a number of pressing needs, such as a Write Once, Read Many (WORM) solution for Picture Archiving and Communication Systems (PACS), medical data archiving, and improved backup and restore, including disaster recovery and protection against malware such as ransomware attacks.

Why Cohesity

- 35% reduction in TCO
- 25% reduction in CapEx and OpEx
- 40% reduction in storage space, with 5x data deduplication and compression
- Backup time reduced by over 50%

> “ *We chose Cohesity on Cisco UCS for next-gen data management at all our hospital locations. Not only have our backup times improved by over 50%, we now have flexibility for data protection on-prem and in the cloud. With Cohesity and Cisco, we have confidence that the solution will meet our data growth, including our PACS images requirements, as we scale for the future. On top of superior backup and recovery, we also have greater assurance with Cohesity against ransomware and increased security to meet compliance requirements at Pederzoli."*
>
> Francesco Corba Colombo, IT Infrastructure Manager, Ospedale Pederzoli

Case Study 4: Nationwide Insurance


Nationwide
SmartFiles
Nationw
is on your

Fortune 75 Insurer Increases Resilience —and Saves $2M Annually—with Cohesity

Challenges

- Nationwide’s IT team backs up and secures more than 21 petabytes of data. They wanted to make sure they could quickly restore VMs after a data center outage, security incident, or even an operating system patch gone awry.

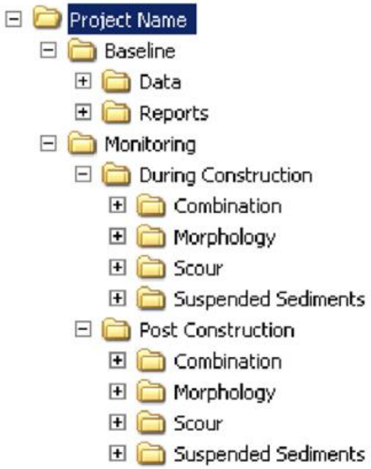
Why Cohesity

- $2 million annual savings
- 75% faster migration to the Cohesity platform compared to the previous migration, smoothing the transition
- Simpler IT operations thanks to a unified platform for backup and recovery, file and object storage, and cloud backups
- Much faster data recovery—now just days instead of months for large-scale outages

> *We needed a next-gen data management platform, and Cohesity stood apart because of its ease of use, rapid recovery at large scale, and strong cyber protections.*
>
> – Guru Vasudeva, SVP and CTO, Infrastructure and Operations at Nationwide

COHESITY

$-\frac{\hbar^2}{2m}\nabla^2\psi + V(\boldsymbol{x})\psi = E\psi$



- Project Name
  - Baseline
    - Data
    - Reports
  - Monitoring
    - During Construction
      - Combination
      - Morphology
      - Scour
      - Suspended Sediments
    - Post Construction
      - Combination
      - Morphology
      - Scour
      - Suspended Sediments