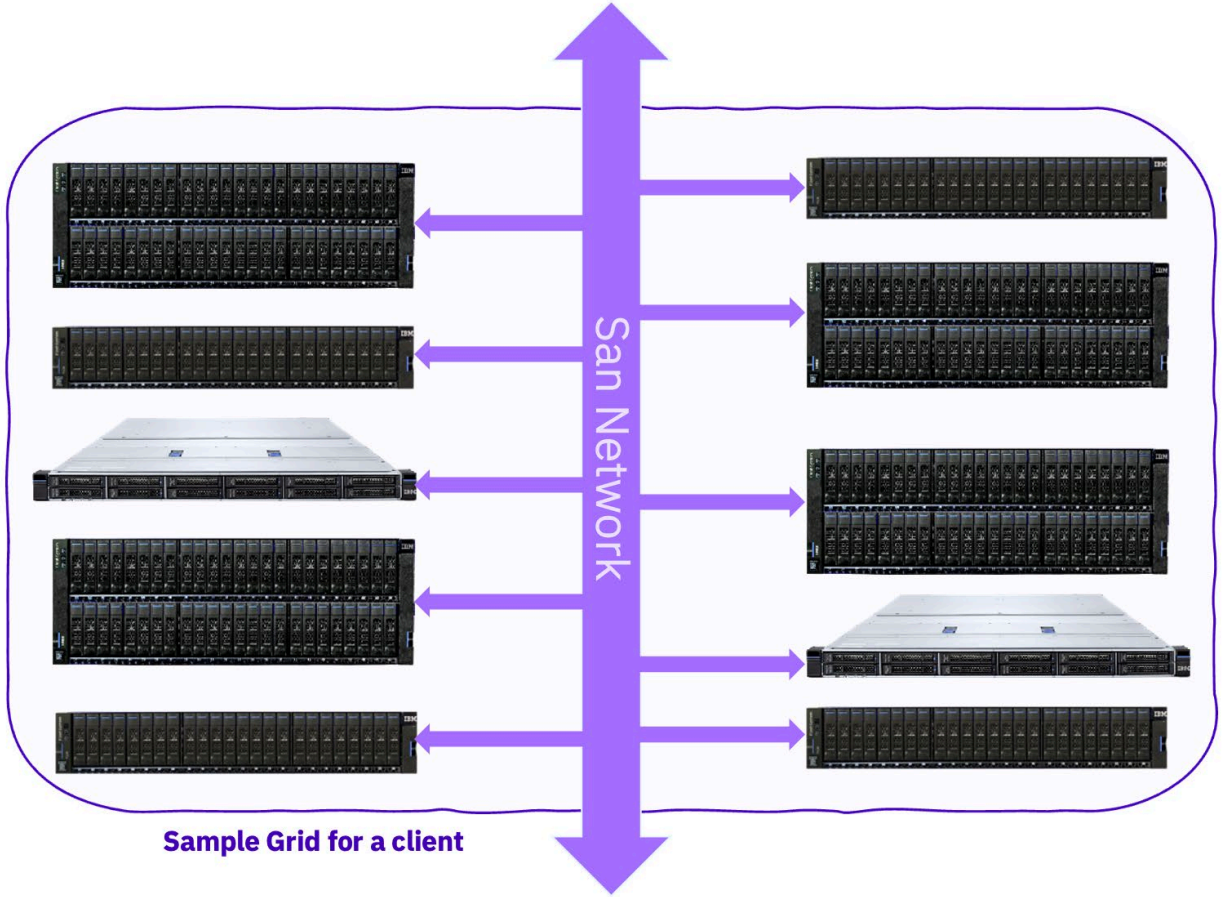
Using AI to Optimize a Grid of FlashArrays

Andrew Walls
IBM Fellow (Retired)
Owner Great Walls of Storage

Differentiating and Exciting FlashSystem Grid

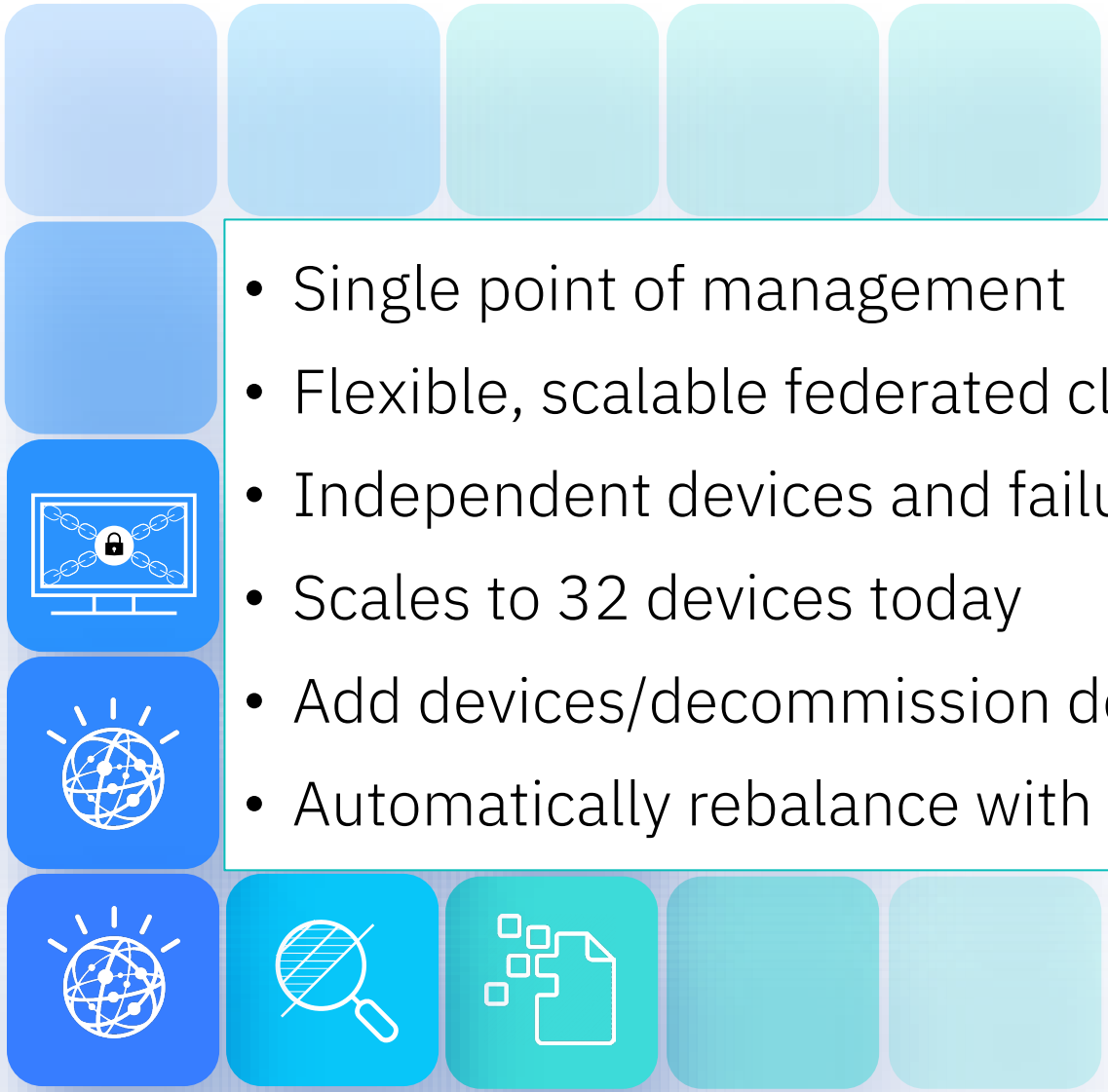
Manage your entire estate as though it was a SINGLE SYSTEM – from a single pane of glass!

- **Optimize workload placement within the grid**
- **Rebalance the estate SIMPLY and EASILY from one screen**
- **Will be the best in the world at automated balancing**
- **AI will optimize placement based on many parameters like performance**
- **Will prevent out of space conditions by making moves if warnings not heeded**



Sample Grid for a client

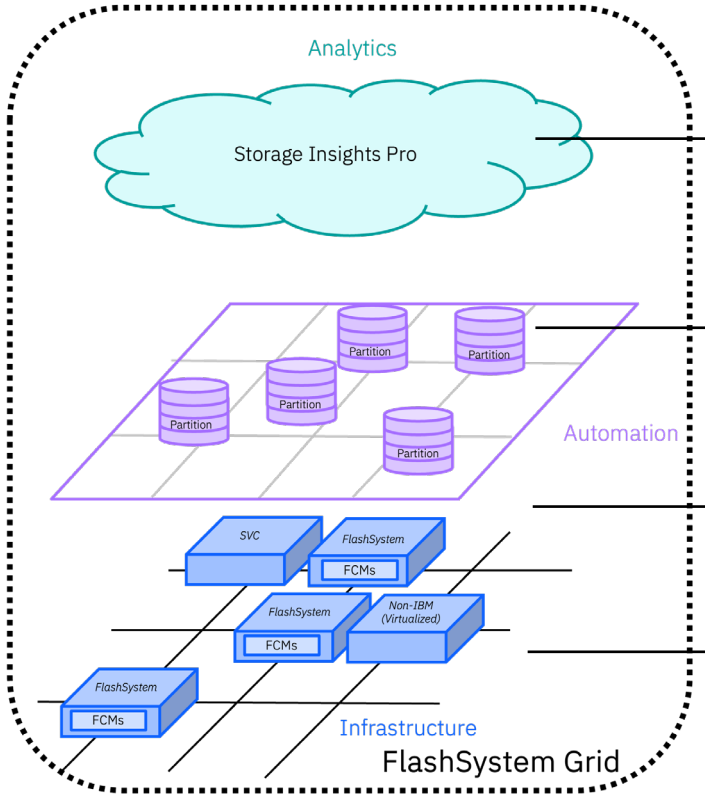
FlashSystem Grid



- Single point of management
- Flexible, scalable federated cluster
- Independent devices and failure domains
- Scales to 32 devices today
- Add devices/decommission devices …
- Automatically rebalance with nondisruptive migration

Revolutionizing Data Storage with IBM FlashSystem grid

Harness the power of virtualization and predictive analytics to anticipate storage requirements, dynamically adapting and scaling to meet the needs of a rapidly changing business.



Anticipates storage needs to optimize capacity and performance

Mobilize data non-disruptively for dynamic tiering and rebalancing

Fleet-wide management of a highly scalable platform

Virtualize FlashSystem, SVC and even non-IBM storage

FMS
the Future of Memory and Storage

Capacity Utilization will increase

Unlock more capacity for data growth or safeguarded backups. Systems are often underutilized today.

40%

Less effort

Industry-leading automation minimizes time spent on planning, provisioning and data mobility

100%

Availability

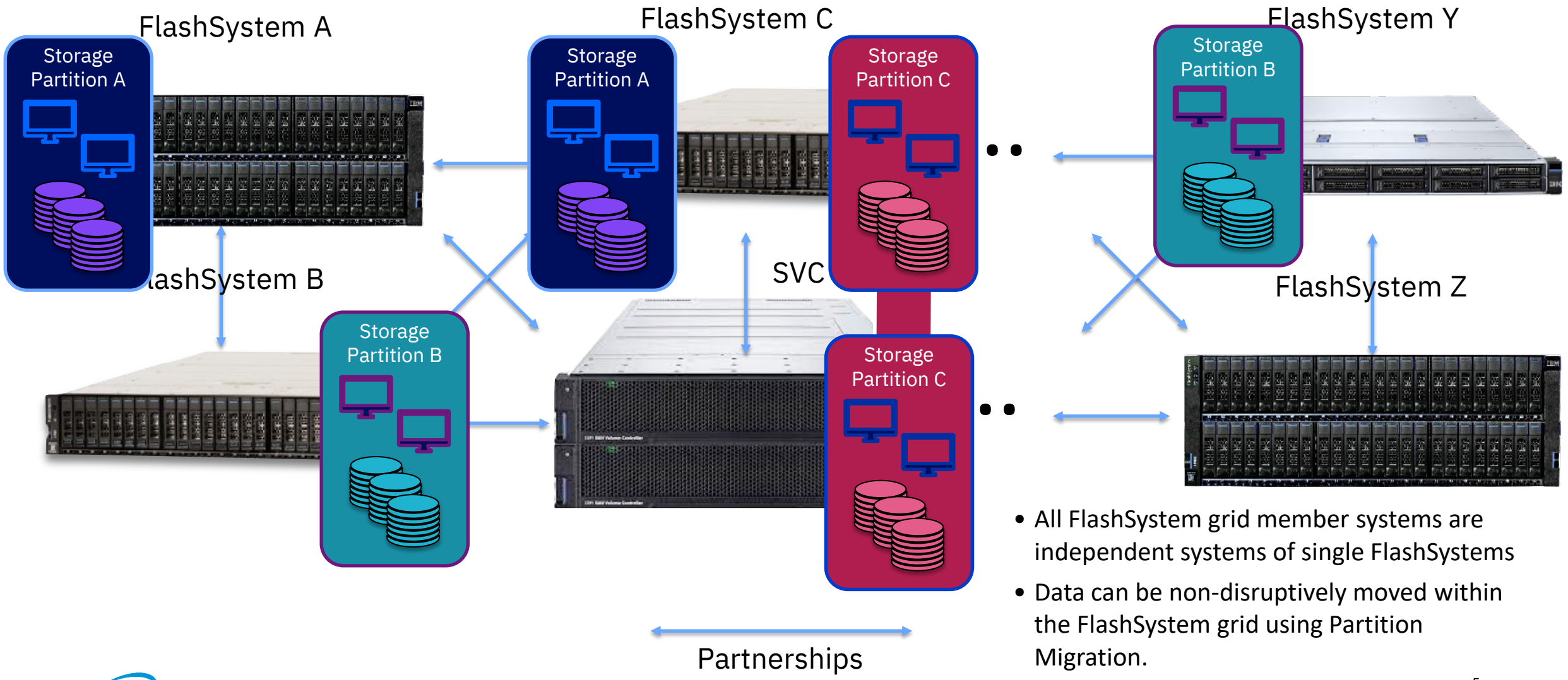
Virtualize your storage infrastructure so your data is always available, wherever you need it

Ready for the future

Easily scale capacity and performance. Keep infrastructure modern with hardware refreshes

FlashSystem Grid Vision

Non-disruptive Storage Partition mobility between resources - manually, by policy, or by SLA



- All FlashSystem grid member systems are independent systems of single FlashSystems
- Data can be non-disruptively moved within the FlashSystem grid using Partition Migration.
- The host facing configuration and the data are moved non disruptively

Simplifying Migrations for Storage and Server administrators!

- Information on Fibre Channel hosts is gathered from the FDMI data:
- Operating system
- Version

- This can be used to determine if the operating system is known to periodically rescan targets.

- If all host operating systems in the partition will automatically rescan, no host admin intervention is required.

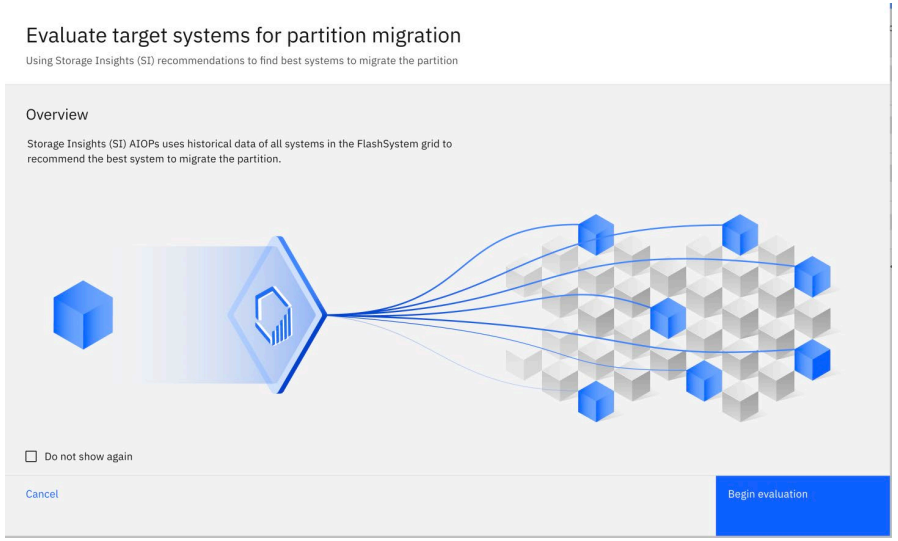
- The system automatically monitors what's happening between the storage and the server to validate that the rescan has happened:
- Have all of the expected logins opened?
- Has the host issued 'discovery' style commands to the new logins?
- Has the host started using the new logins for I/O?

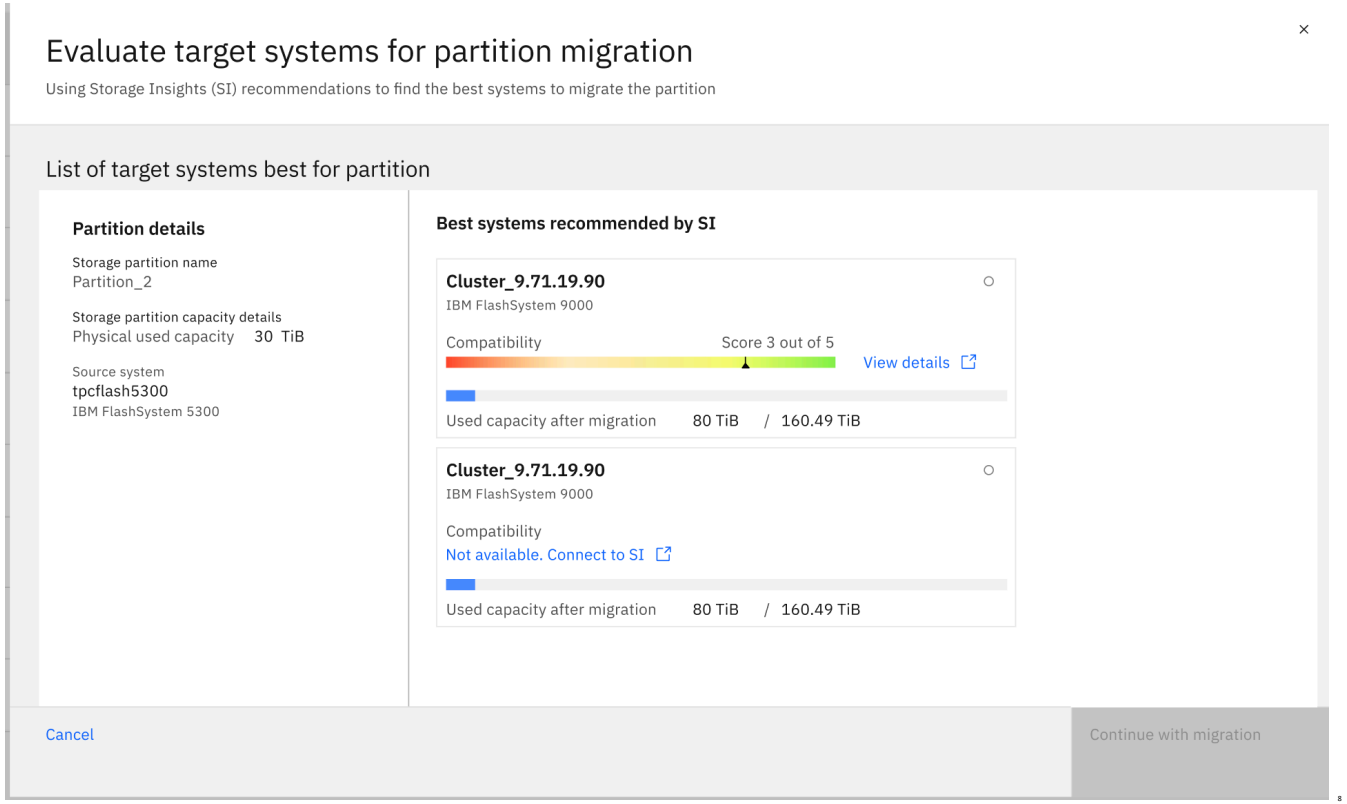
- If the above three tests pass, then it's safe to continue with the migration as the multipath driver has picked up the new paths.

- For operating systems that do not automatically rescan, the system can monitor their rescan progress to ensure it happens before continuing.

Migration Recommendations Using AI

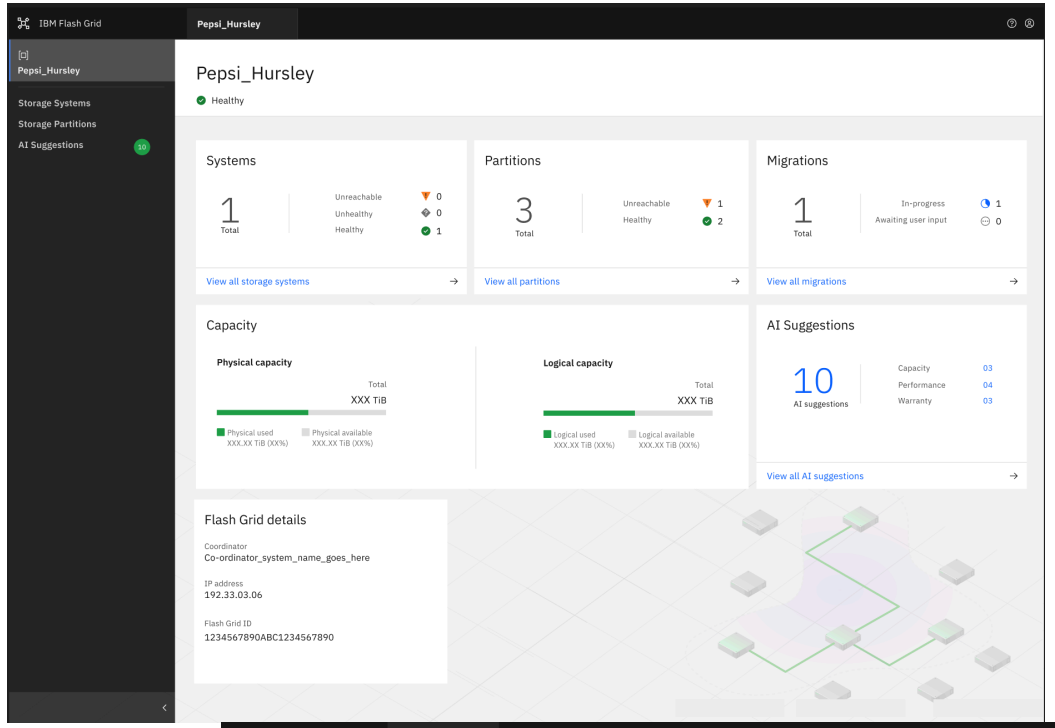
- With AIOPs capabilities, Storage Insights can help determine the best place to migrate a partition
  - Provides compatibility score based on 10s of metrics
  - Uses all historical data and generative AI to help recommend best placement
  - Done in the background, avoids blocking further actions
  - User still has control over migration action

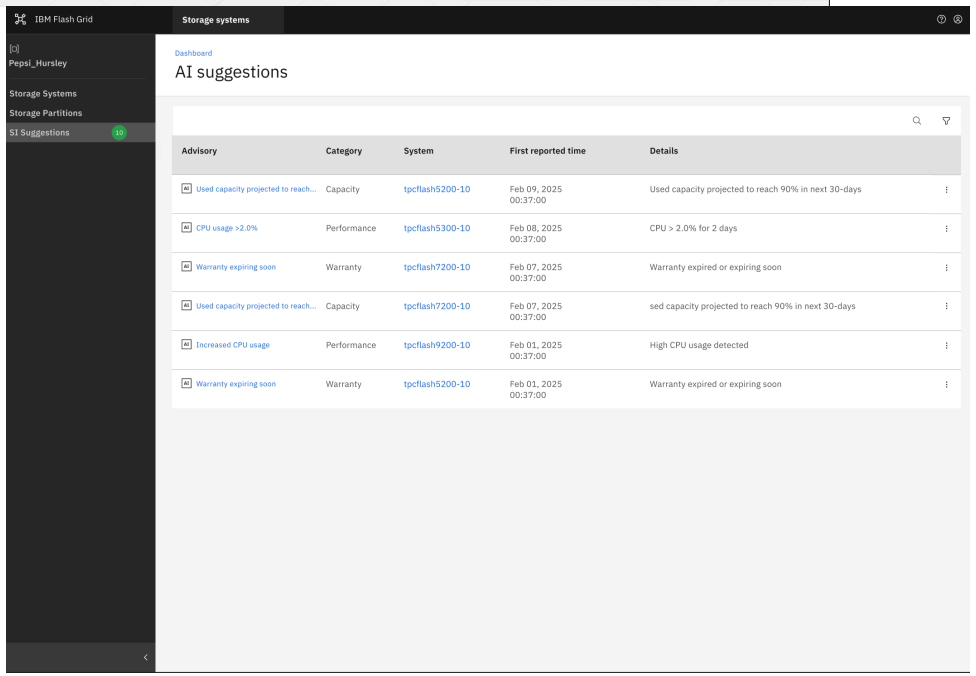




Storage Insights AI Suggestions

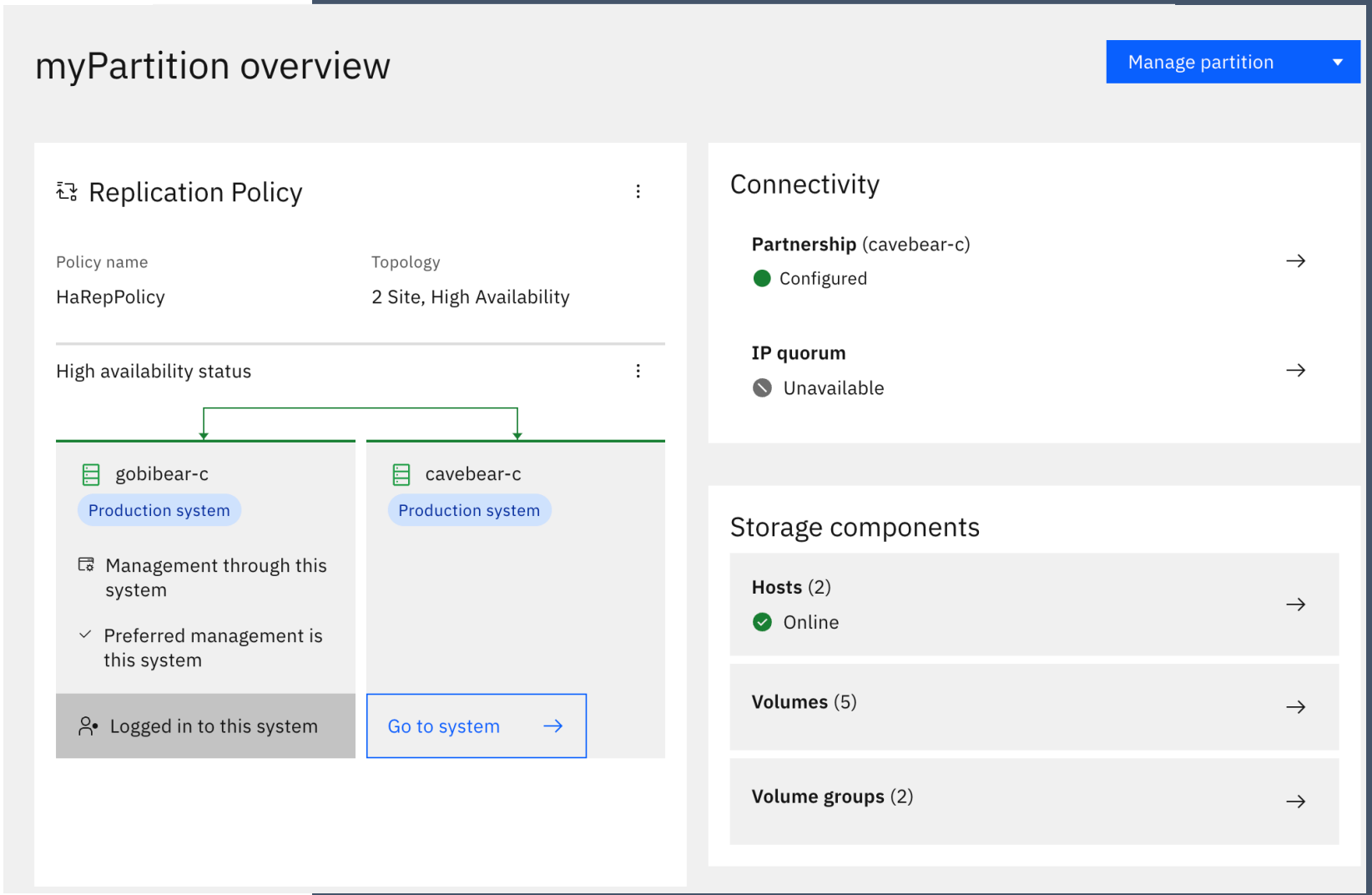
- FlashSystem grid dashboard includes a card with AI generated suggestions
  - Capacity
  - Performance
  - Warranty
- Equivalent suggestions as the AI Advisor tab in SI
  - In-line to the FlashSystem grid UI without having to go to SI
  - Provides links to provide additional information directly in SI



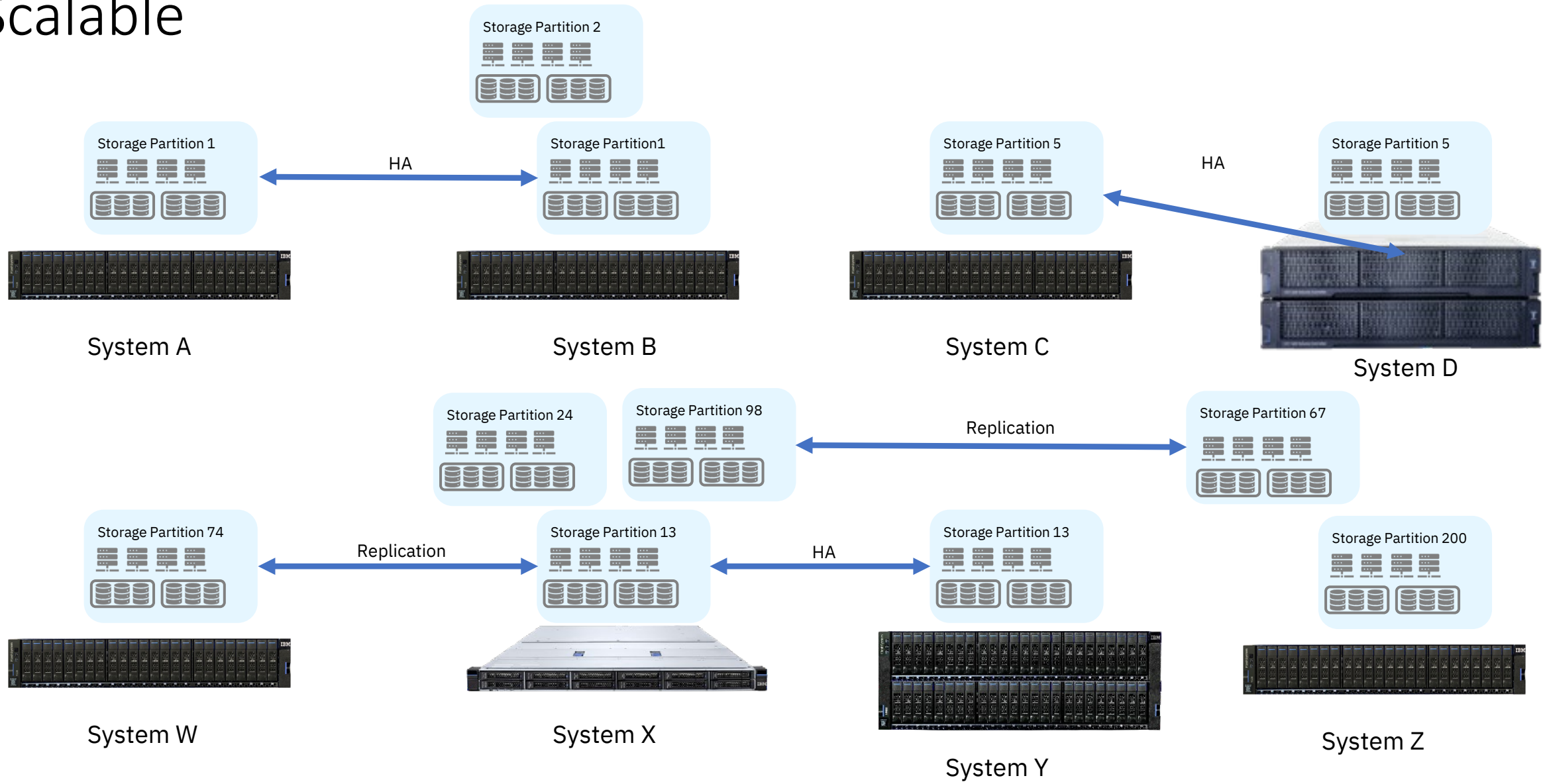


Storage Partitions

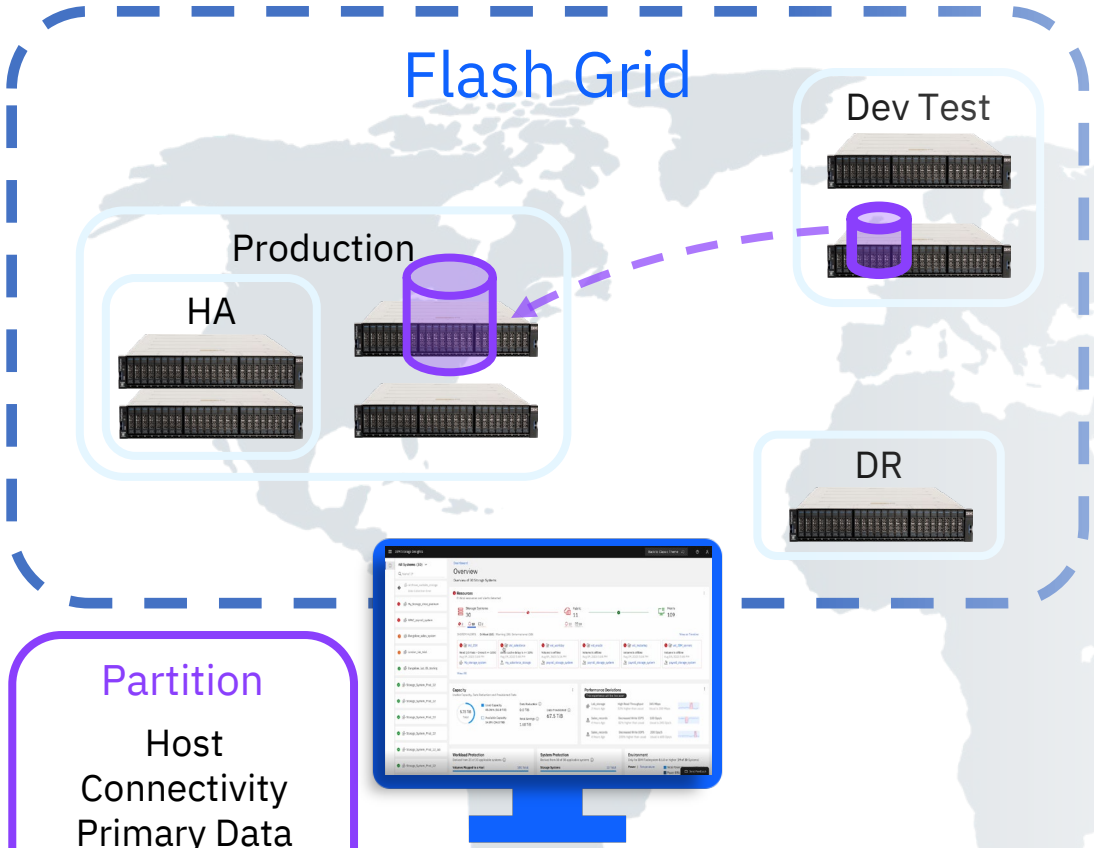
- A subsystem collects associated storage objects
  - Volume Groups, Volumes, Hosts
- Add associated protection against partitions
  - High Availability
  - Disaster Recovery
- Allows for mobility
  - Move partitions between systems
  - Partition Mobility using SI or SV
- Future Capabilities
  - Basis for Flash Grid
  - Multiple vCenters



Scalable


Storage Partition 2
Storage Partition 1
HA
Storage Partition1
Storage Partition 5
HA
Storage Partition 5
System A
System B
System C
System D
Storage Partition 24
Storage Partition 98
Replication
Storage Partition 67
Storage Partition 74
Replication
Storage Partition 13
HA
Storage Partition 13
Storage Partition 200
System W
System X
System Y
System Z

AIOPs with Flash Grid



1H25: Intelligent Mobility using AI

Suggest performance/balancing optimizations

Defect overloaded systems and suggest partition moves

Optimizations based on user defined criteria

Migration/upgrade use cases

Ensure adherence to SLA (ex: always HA+DR across geos)

Target lower tiered system usage first

Optimize partition placement for power consumption

Other Key AI functions for the grid now and in the future

- **AI for Support and service**
- **Moving from migration advice to an Autonomous FlashSystem Grid**
- **Ransomware Detection and what can be done for the Grid**

AI for Support – learning from past issues

- IBM and other Storage providers invest significantly in making storage systems reliable and resilient

- YET –

- Things still go wrong – parts fail, software bugs occur, etc
- IBM practices first time data capture, so lots of data is saved and sent home when there is an issue
- We do not just send back data when there is an issue, but performance and health data is sent back constantly
- So, IBM uses all this data to train AI models

AI for Support – learning from past issues

- New Tickets are inferenced against resolved issues to see how similar issues were resolved in the past
- The same proven AI then used to spot anomalies in the data before the customer may even be aware!!
  - Performance, network issues, non- optimal configurations, etc
- It can go further and tell the user the corrective action
  - Perhaps before it becomes an issue
- Moving closer to Self Healing infrastructure

Performance Anomalies and Issues

- They can be extremely difficult for humans to debug
- They can involve the storage controller, the media, the network, the server, the application
- Or a combination of several
- The impact to the users can be significant

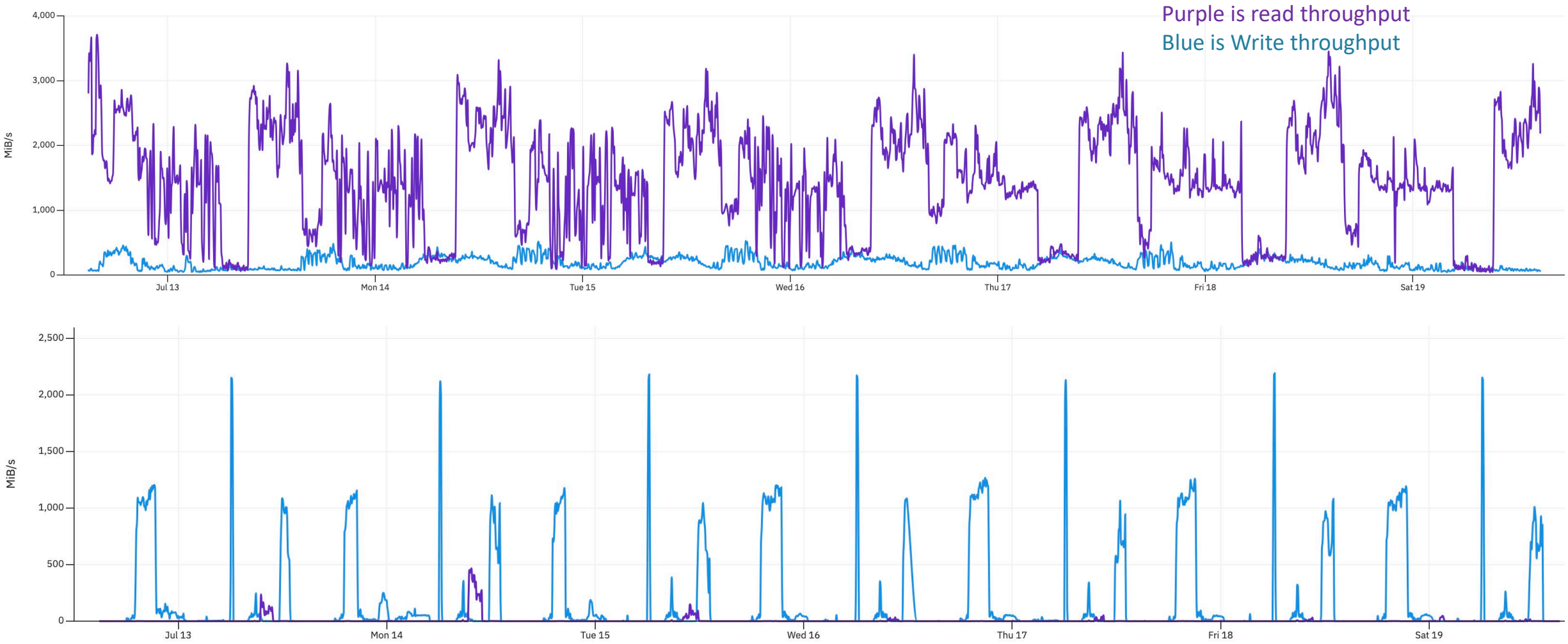
AI Detecting Performance Anomalies

- Software issues, Hardware malfunctioning, intermittent issues, design bugs can cause performance issues
  - System continues operating but response time and throughput might suffer dramatically
  - Causes dissatisfaction with users and worse – missed SLAs and financial losses.
- Storage Insights from IBM analyzes over 200 performance metrics and uses AI to provide support and client assistance in isolating the issue

But one can go further. . .

- Performance sometimes worsens over time due to overloading – More Transactions, larger files, more snapshots, etc
- Can be persistent latency increases or spikes at certain times of day.
- SI with AI can monitor the load and the entire grid and make suggestions how to optimize the different partitions
  - Capacity optimizations
  - Think about Data Reduction optimizations
  - Workloads are not constant, optimizing the grid must take into account the difference in workload at all points in the day, week, month, quarter, etc

An Example of how much workloads change during the week

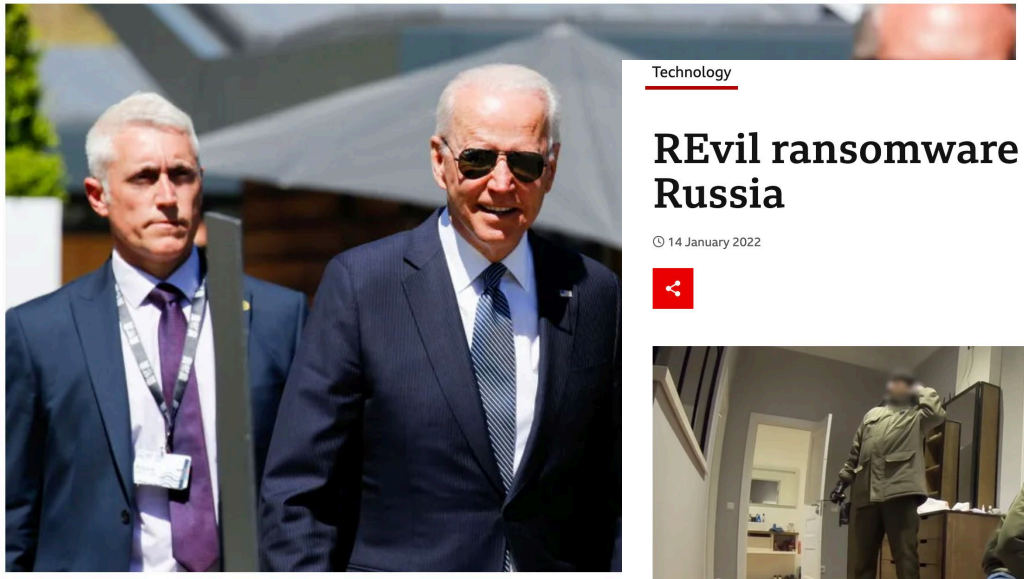


Cyber Security for the Grid

Biden, Putin conduct diplomatic dance over hypothetical hacker exchange

The two world leaders are scheduled to meet on Wednesday amid a spate of ransomware attacks.

BY JEFF STONE • JUNE 14, 2021

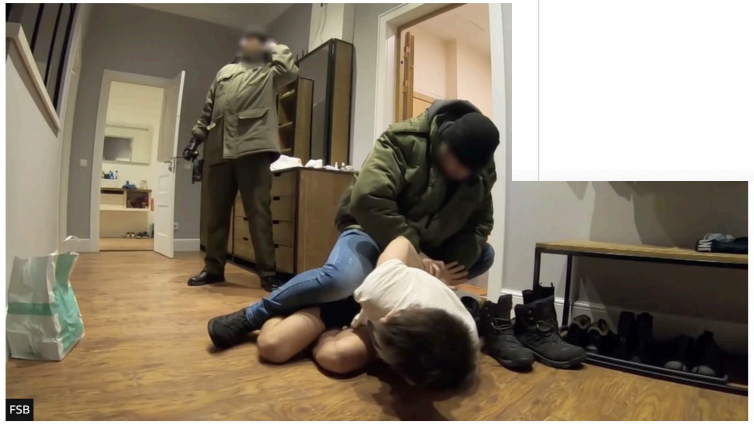


U.S. President Joe Biden arrives for a plenary session during G7 summit in Carbis Bay on Ju
Noble - WPA Pool/Getty Images)

Technology

REvil ransomware gang a
Russia

14 January 2022



The FSB has released video footage of the arrests

Authorities in Russia say they have dismantled the ransomware crime group REvil and charged several of its members.

The United States had offered a reward of up to $10m (£7.3m) for information leading to the gang members, following ransomware attacks.

Russia's intelligence bureau FSB said the group had "ceased to exist".

Cybercrime

This article is more than **1 year old**

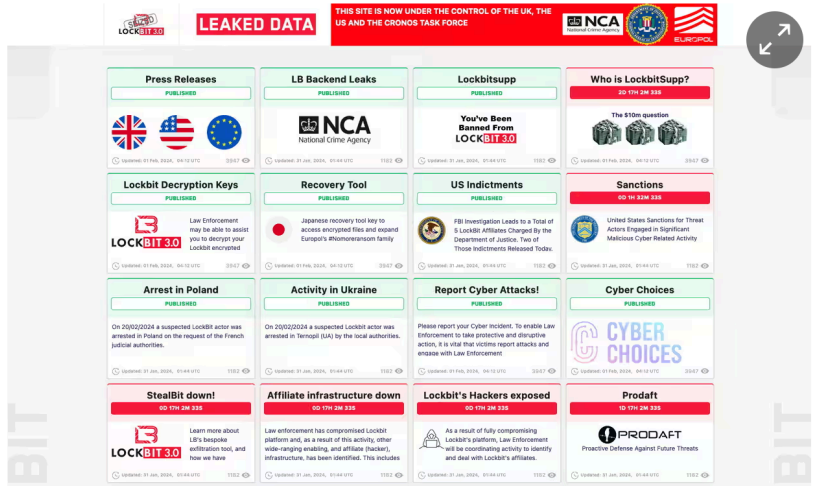
Russia-based LockBit ransomware hackers attempt comeback

Gang sets up new site on dark web and releases rambling statement explaining how it was infiltrated by law enforcement agencies

Dan Milmo *Global business editor*

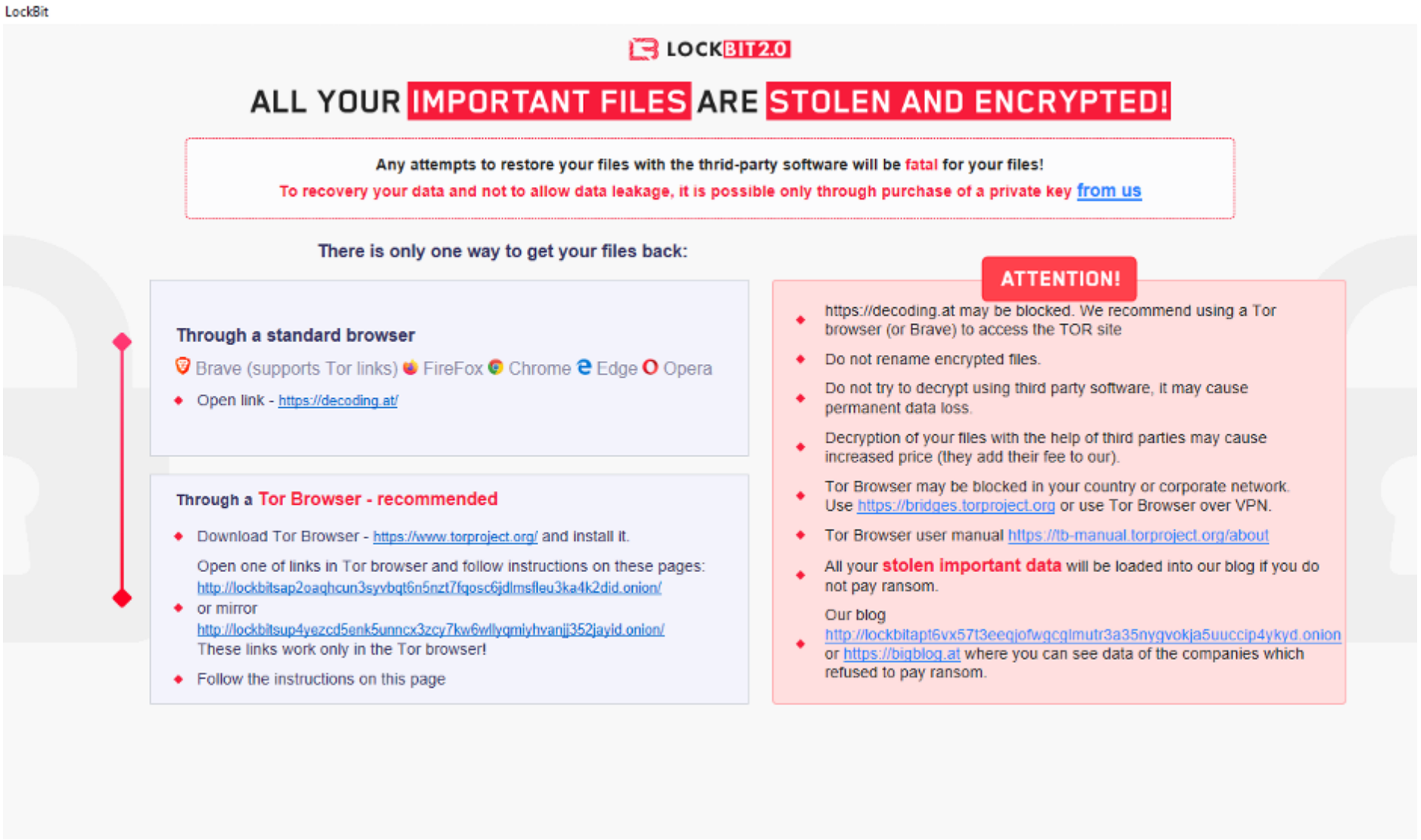
Mon 26 Feb 2024 11.34 EST

Share



A screenshot taken last week shows the dark web site after law enforcement agencies took control of it. Photograph: Reuters

- This is Lockbit ransomware note
- Look familiar?
- All ransomware strains borrow from each other



FCM4 - Ransomware Threat Detection

LBA Addressing and Sequencing Patterns

Statistical Analysis *(Median, Distribution etc)*

Encrypted payload detection

Compression Statistics

Changes in Read / Write Throughput

Just a few of the statistics. . . .

Ransomware Threat Detection - Solution Overview

SIEM/SOAR
e.g. IBM QRadar, Splunk etc

Take SOAR actions, i.e. isolation

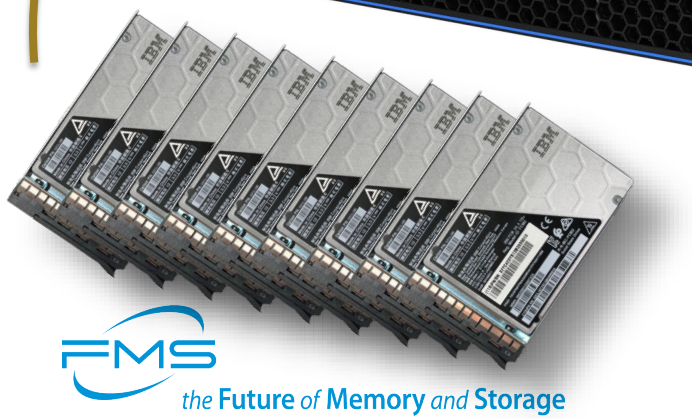
e.g. Trigger ad-hoc **Safeguarded Copy**

Granular Data Analytics & Volume Stats *(Summarised every 2s)*

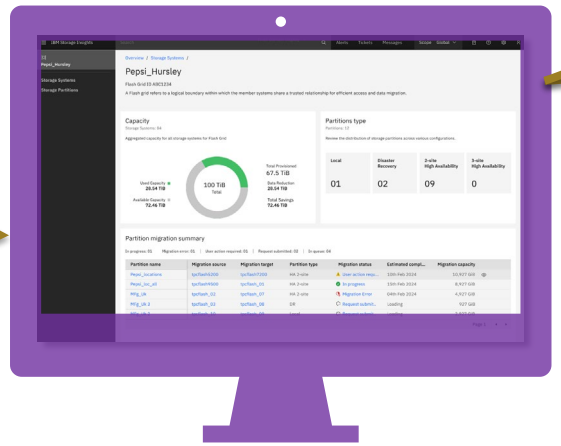
IBM Storage Virtualize Software
(FlashSystem firmware 8.6.3+)

Trends/Summary

Patchable AI Inference Engine

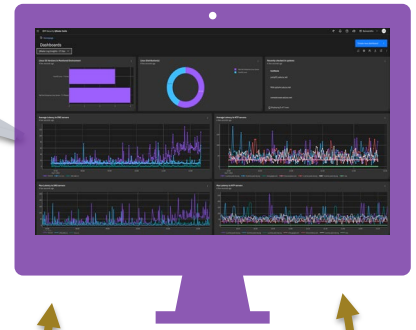


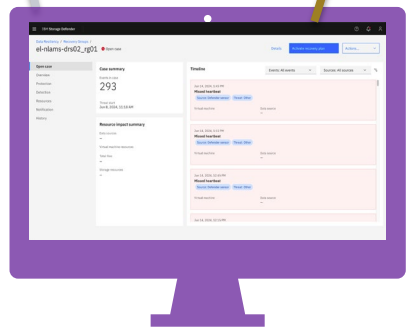
Real-time Alerts *(Typically under 1 minute)*



IBM Storage Insights Pro

Forward Details (e.g. Alert type, affected volume)



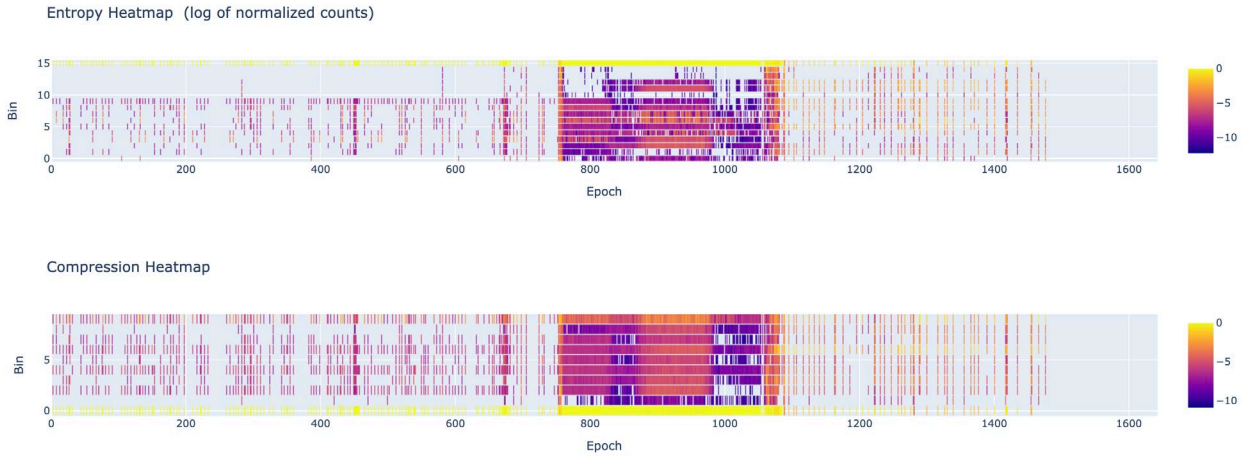


IBM Storage Defender DRS
Data Resilience Service

Lockfile Ransomware

- Encryption IO activity easily distinguished by our hardware accelerated byte-by-byte analysis
- Intermittent encryption is still

- Part of our original set of strains used for training
- Even when mixed with a MSSQL workload, as shown here, activity is easily distinguishable by the FCM statistics



Where this is heading with the FlashSystem Grid

- We have made tremendous progress with False positives and false negatives
- Imagine having accelerators in the grid and run additional models if there is an alert or if number of single classifier events increase
- Imagine sampling volumes on to the accelerator in other parts of the grid
- Imagine training on the grid with the client workloads
- Can help to deal with the increasing sophistication of cyber criminals and protect the FlashSystem grid

Other Items which can be done with AI to the Grid

- Sustainability and Power Efficiency optimization
- Virtual Storage Administrator
- Data Reduction optimization