

# Future of Storage Research in IBM

Larry Chiu ([lchiu@us.ibm.com](mailto:lchiu@us.ibm.com))  
Director, Global Storage Research  
Distinguished Engineer

# IBM Research Globally

World's largest information technology  
research organization

More than 3,000  
scientists and engineers



# Future of Data Storage – IT Industrial Trends

## Trends

New  
Apps

**Digital Innovation Explosion** – From 2018 to 2023 –with new tools / platforms, more developers, agile methods and lots of code reuse - 500 million new logical apps will be created, equal to the number built over the past 40 years.

IT Skill

**Reshaping Talent Pool Landscape** - Through 2022, the talent pool for emerging technologies will be inadequate to fill at least 30% of global demand, and effective skills development and retention will become differentiating strategies.

Security  
Privacy

**Expanding & Scaling Trust** - By 2022, 50% of servers will encrypt data at rest and in-motion; over 50% of security alerts will be handled by AI-powered automation; and 150M people will have blockchain-based digital identities.



Foundational  
Storage

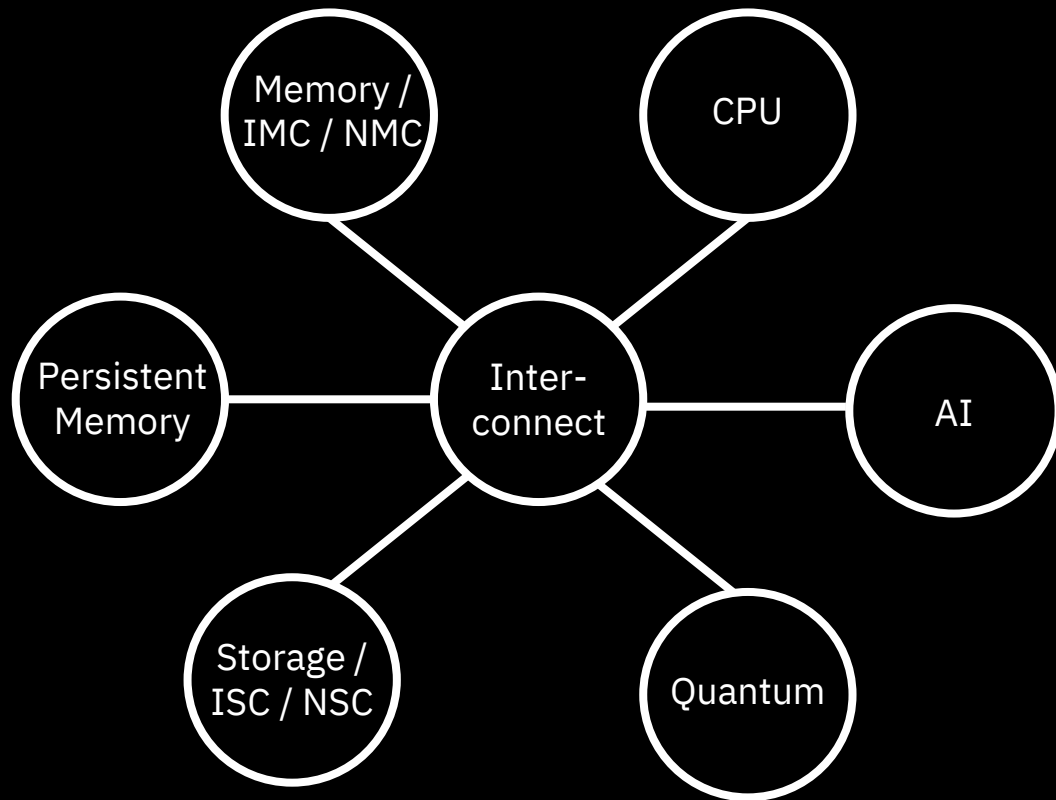
Resiliency

Security

Composable  
Architecture

# Composable Architecture

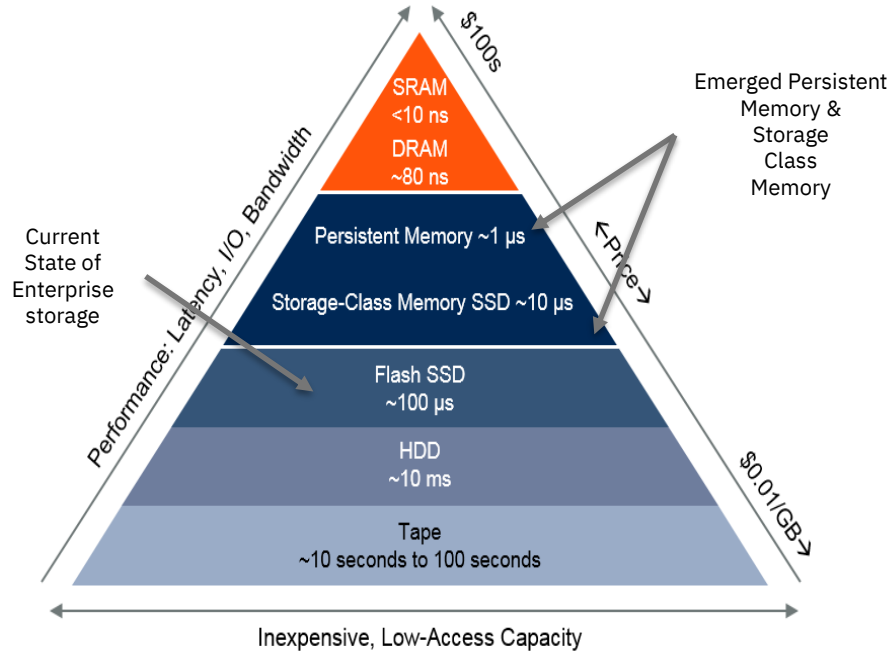
In-memory-Computing  
Near-memory-computing



In-storage-Computing  
Near-storage-computing

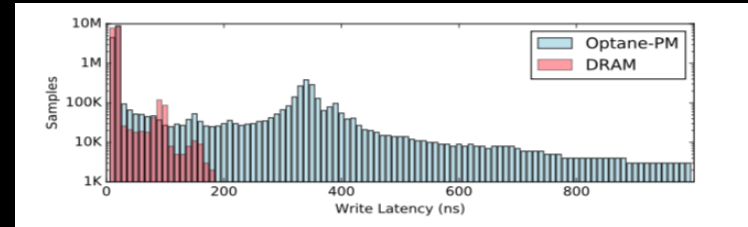
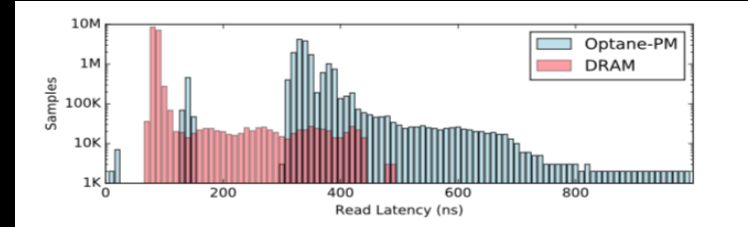
# Next Generation Persistent Memory Storage Architecture

## Performance Hierarchy: Introduction of Persistent Memory



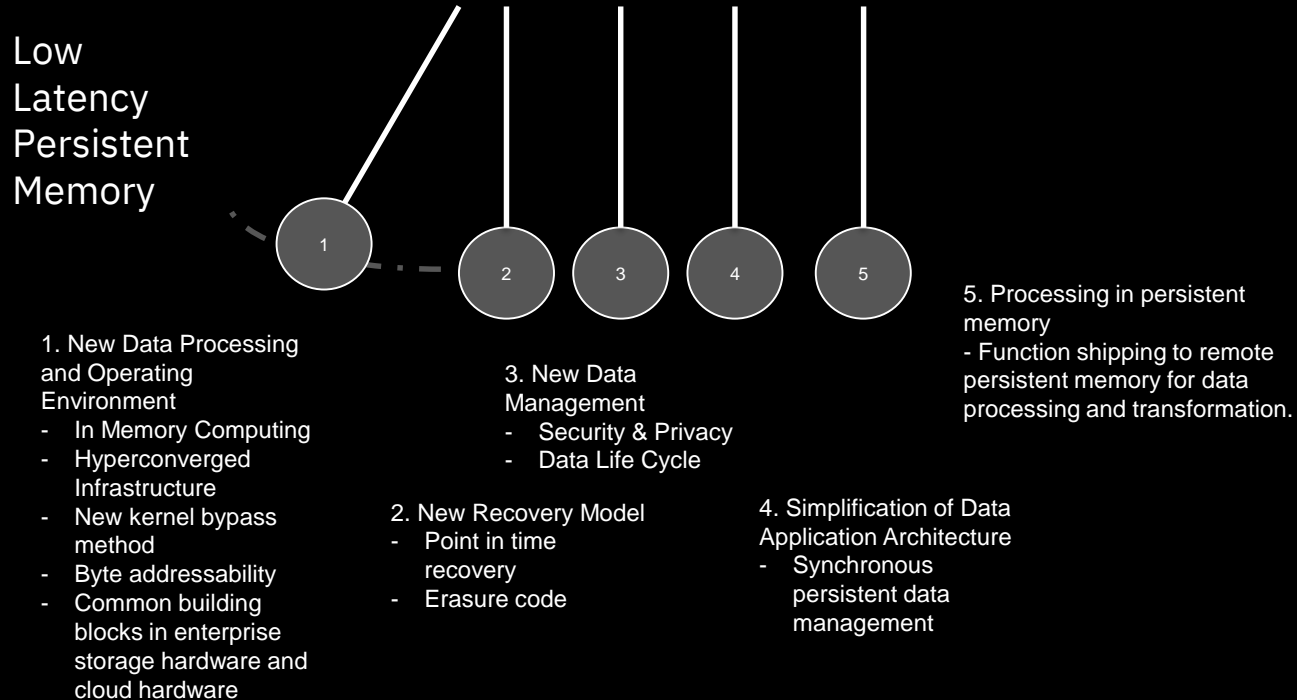
ID: 375147

© 2018 Gartner, Inc.



- Byte Addressable
- Memory Speed
- Access Bypass Kernel
- Processing in memory

# Implication of Next Generation Data Applications from Persistent Memory





# Memory and Storage? Or Accelerator?

Getting the data where the compute is?

Getting the compute where the data is?


### UPMEM PIM-DRAM big data accelerator

- UPMEM DIMMs
  - Replacing standard DIMMs
  - DDR4 R-DIMM modules
    - 8GB+128DPUs (16 PIM chips)
- UPMEM PIM-DRAM chips
  - 4Gb DDR4 2400 DRAM + 8 DPUs @500MHz
  - Single die, standard 2x nm DRAM process
- Massive additional compute & bandwidth
  - 2TB/s DRAM-DPU BW for 128GB+2048 DPUs config
- Easily programmable SDK: C-programmable

Copyright UPMEM® 2019

HOT CHIPS 31

up mem



Standard PIM Processor Unit

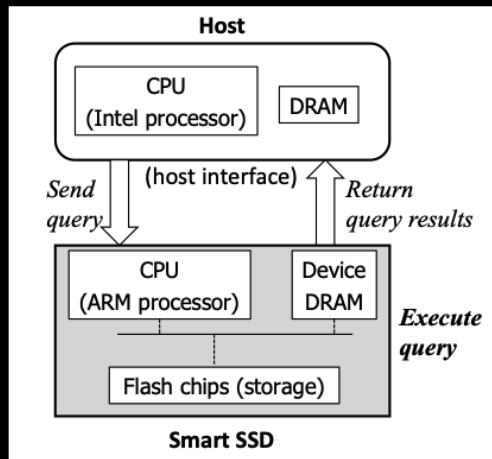
PIM server: Typically with 128GB DRAM/2048 DPUs

**Take away**

- Scalable as compatible with
- Current servers
- Unmodified DRAM process
- Programmers ;)

Samples & apps available

In / Near memory Computing



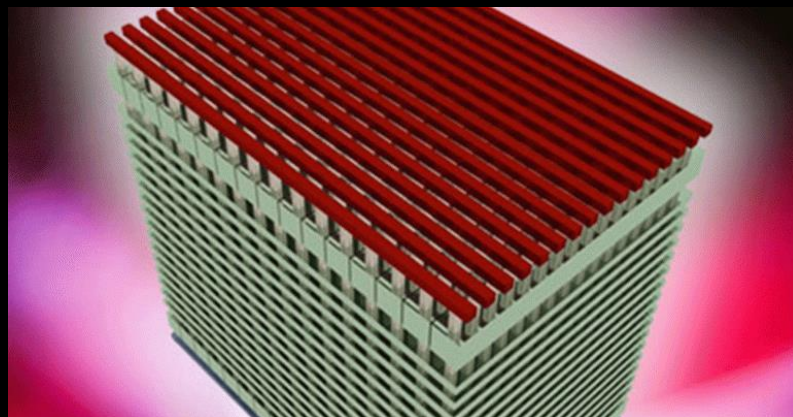
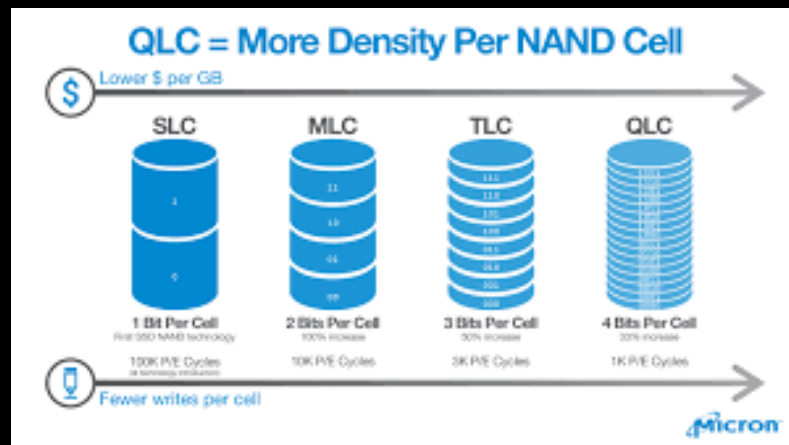
In / Near Storage Computing



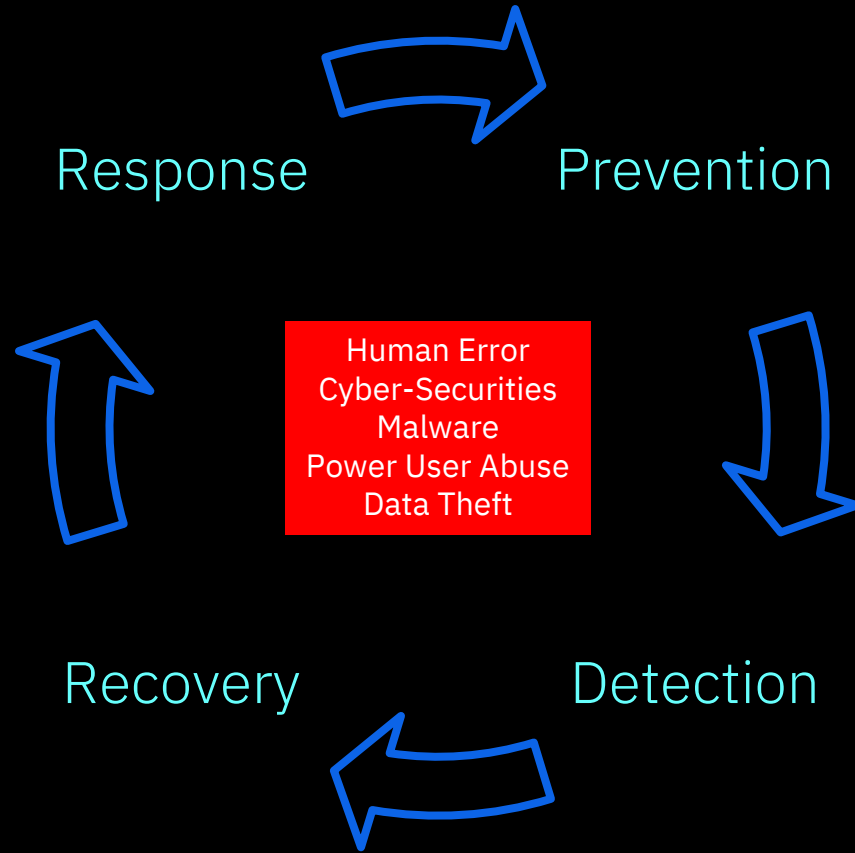
# Flash Storage Research – Self Performance Tuning



- Uses AI to predict health of blocks (Health Binning)
- Heat Segregation to put hot data on healthy blocks
- Advanced Read Level Shifting in background to avoid rereads
- Consistent performance is key
- Unique and incredibly strong ECC which never requires data to be reread



# Security



Prevention

Detection

Recovery

Response

Fine Grained Access Control

Key Virtualization

Quantum Safe Key

Secured Domain Isolation

Detect as data is accessed in wrong place, wrong person, wrong methods, wrong time

Machine Learning and prediction on security events

Autonomous Recovery

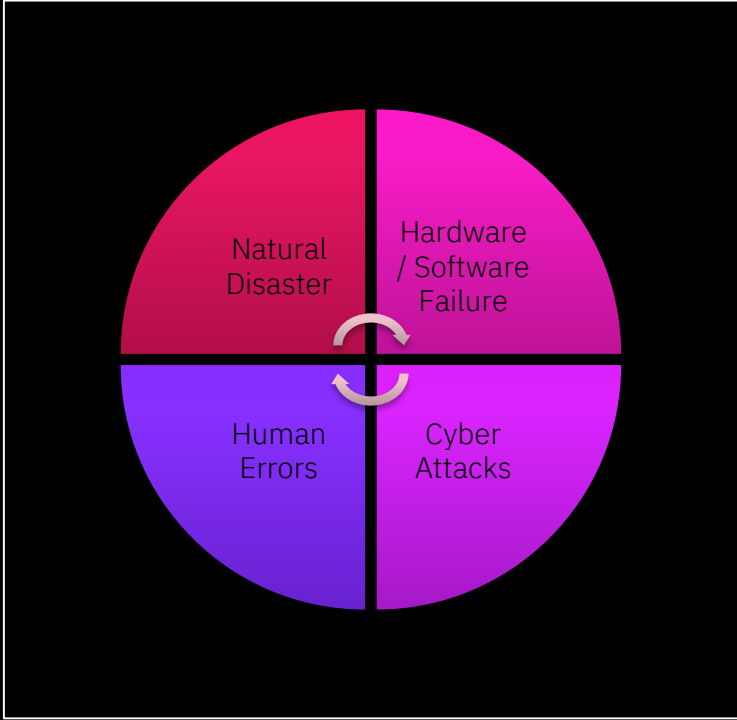
Elastic Recovery

Regulatory Reporting and Response Management

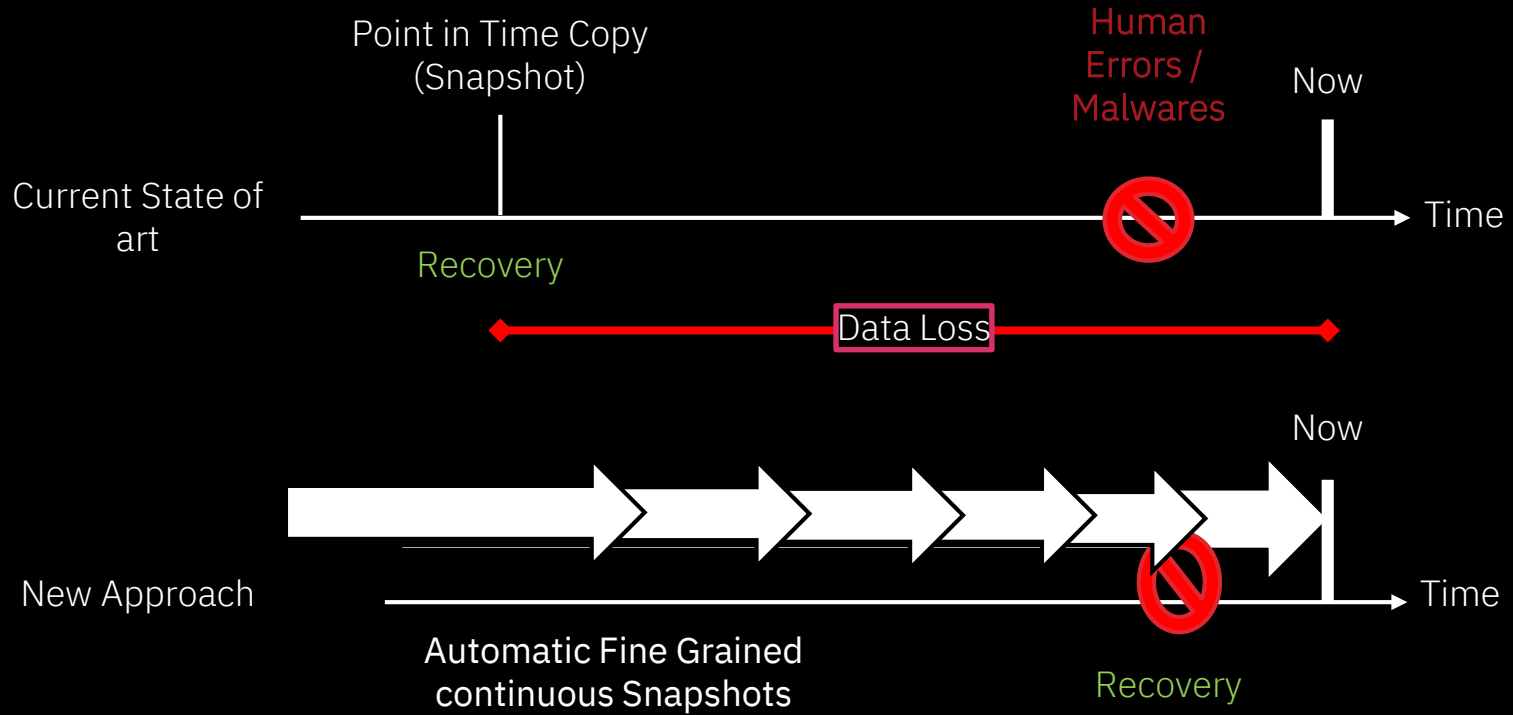
# Resiliency



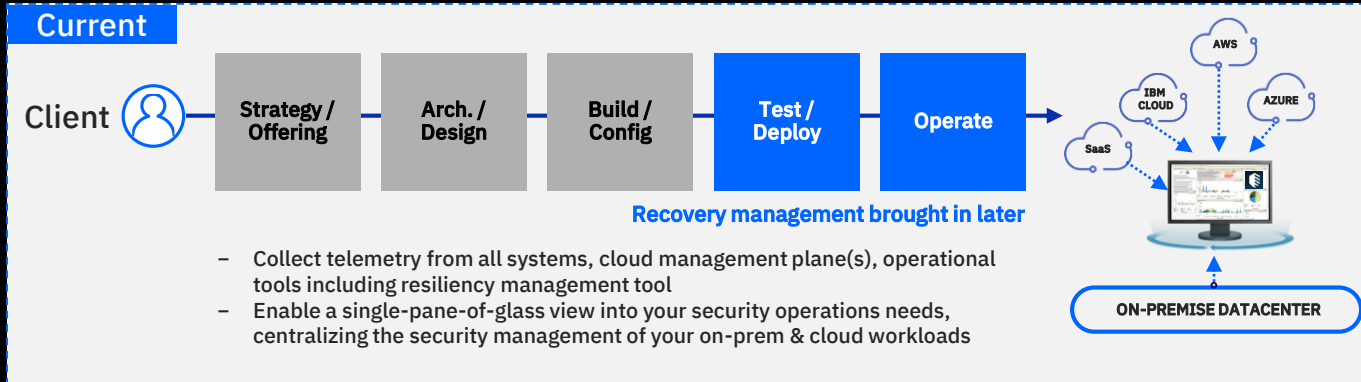
Source: World Economic Forum, The Global Risks Report 2019, 14th Edition



# Logical Data Recovery – Human Errors or Destructive Malwares

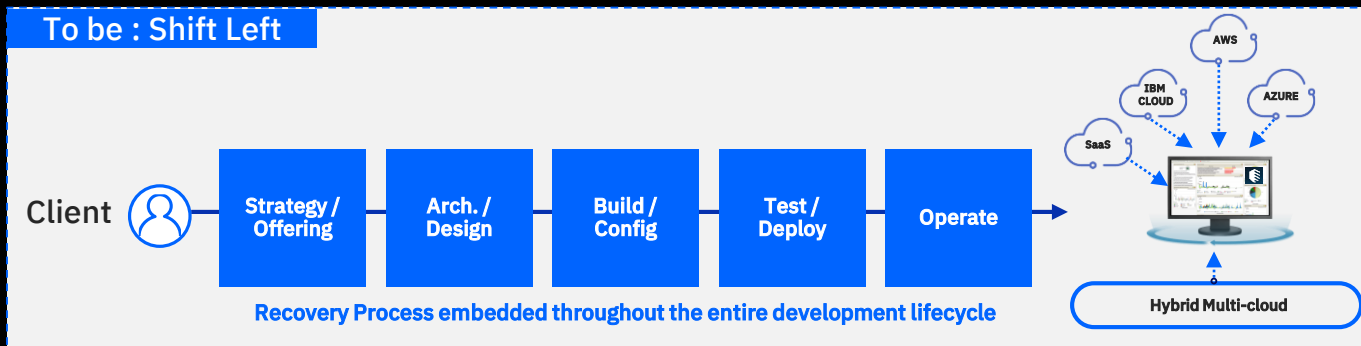


# Keeping pace with DevOps: Deploying Recovery management at cloud speed



## Automated Recovery Provisioning

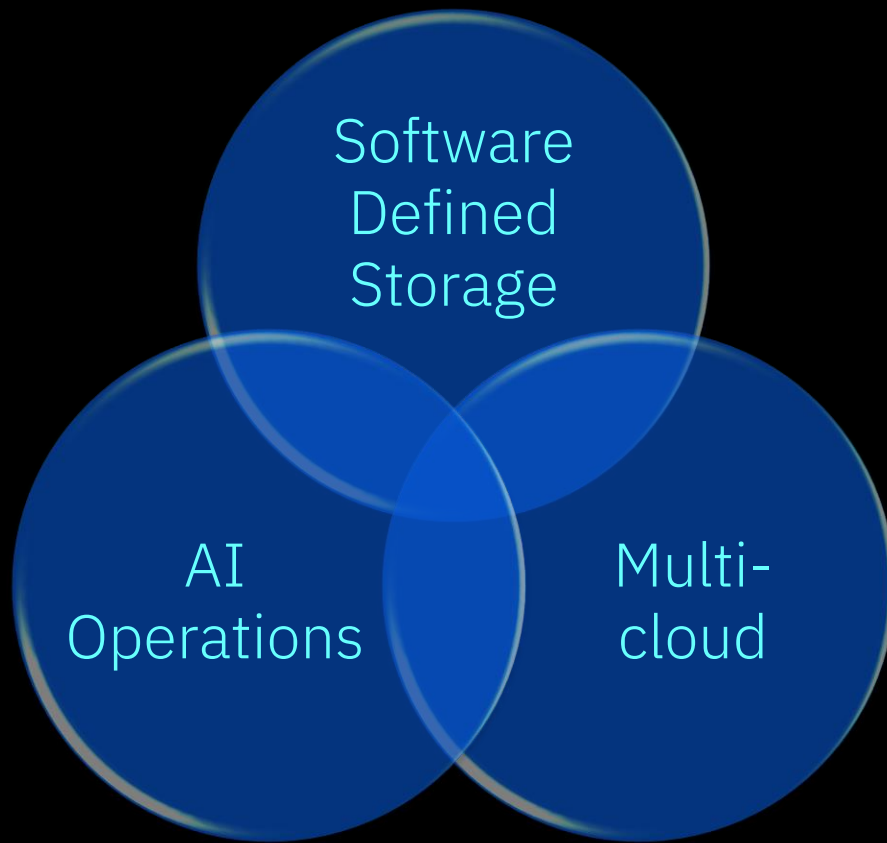
Automate provisioning of recovery technology and policies to new cloud workloads at cloud speed.



## Recovery-by-Design Approach

Intersect Recovery management provisioning with infrastructures.

# Foundational Storage





# The Hybrid Multicloud World

**85%** of companies operate a **Hybrid multicloud environment**

**91%** of public cloud adopters use **internal private cloud**

Companies average

**5**

Private and  
Public clouds

Migrate from public cloud

**Security,  
Performance,  
Cost,  
Control**

*IDC Survey*

**80%** of companies moved applications or data from Public Cloud in 2018

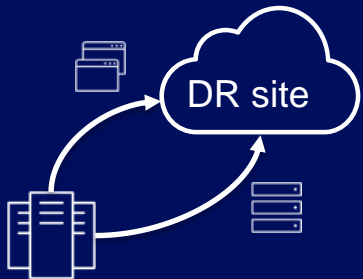
*IDC Survey*

**98%** will be **Hybrid Multicloud in three years**

# Storage choices matter because cloud use cases matter

## Business Continuity

leveraging public  
cloud resources



## DevOps Analytics

Temporary data  
copies from on-prem  
to cloud



## Workload migration

Transparent data  
migration to cloud



## Hybrid multicloud

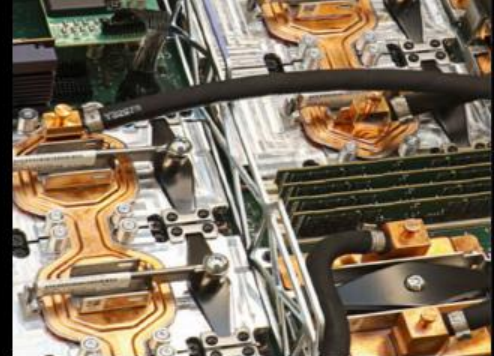
Operating  
infrastructure both on-  
prem and in public  
cloud



# Summit & Sierra by the numbers



**2.5 TB/sec**  
single stream IOR



Together, more than  
**44,000 NVIDIA GPUs**

**>400 PB**  
of IBM Storage



**2.6 Million** 32K files  
created/sec

**50K creates/sec**  
per shared directory

**#1 & #2 most powerful  
supercomputers  
built for AI**

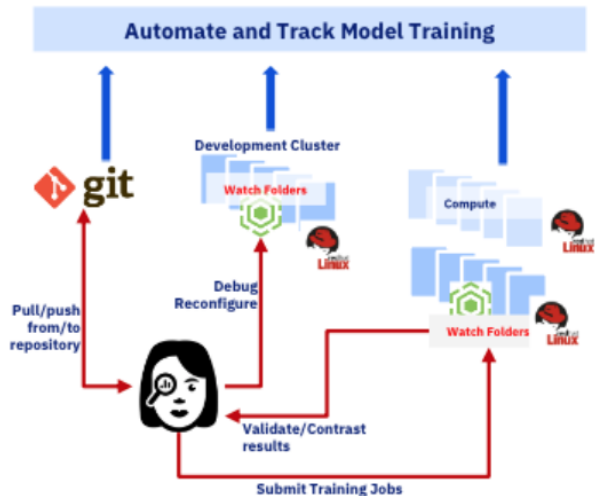


**Single Node 16 GB/sec**  
sequential read/write

**1 TB/sec**  
1MB sequential  
read/write



# IBM Research - Data Curation and Provenance Management



## Track facets of analytical pipelines

- Source, Intermediate data, final result
- Model, script, algorithm changes

## Track transparently

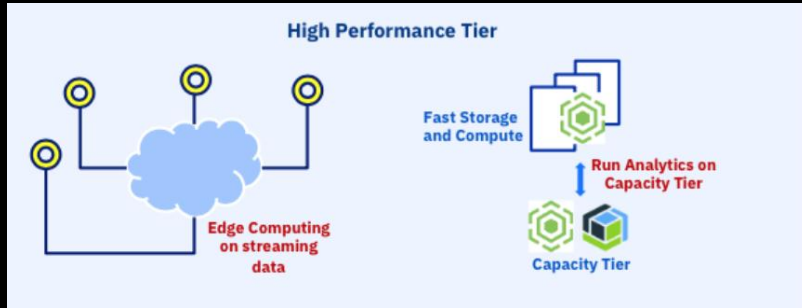
- Spectrum Scale Watch Folders To track data creation and transformation
- RedHat auditd To track processes
- Git (for source code change tracking) To track changes in code

## Track, contrast, compare iterations

- Anomaly detection
- Result contrast and forensics

**Cumulatively, make analytics reproducible**

# IBM Research – Data Affinity for AI and Analytics

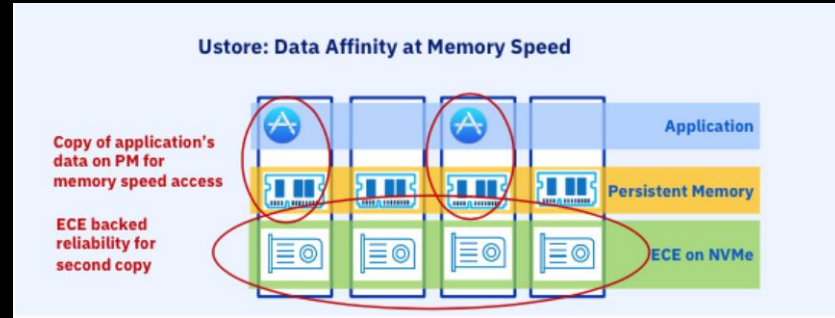


## Affinity in a diverse Storage Topology

- Co-locate data, compute, analytics
- Stream data and data at rest (3<sup>rd</sup> party repository)
- Exploit Fast Storage (local attached storage)

## Use Cases

- Edge Computing
- Analytics on Streaming Data



## Improve Data Locality in a distributed storage environment

- Application directed storage managed data locality
- Access data at memory / persistent memory speed

## Use Cases

- Storage platform for Hybrid Transaction and Analytics Platform (HTAP)



Foundational  
Storage

Resiliency

Security

Composable  
Architecture

IBM®