



# Leveraging Software-Defined Storage to Meet Today and Tomorrow's Infrastructure Demands

Unleash Your Data Center's Hidden Power

September 16, 2014

**Molly Rector**

CMO, EVP Product Management & WW Marketing

## DDN | The “Big” In Big Data



### 800%

Paypal accelerates stream processing and fraud analytics by 8x with DDN, saves \$100Ms.



### 1TB/s

The world's fastest file system, to power the US's fastest supercomputer, is powered by DDN.



### Tier 1

Tier1 CDN accelerates the world's video traffic using DDN technology to exceed customer SLAs.



# DDN | Our Vision and Mission

## Our Vision

Enable Organizations to  
Maximize the Value of  
All Information Everywhere

## Our Mission

Build the High-Scale Global  
Leader in Big Data Solutions,  
Powered by Storage, Compute  
and Analytics IP and Expertise



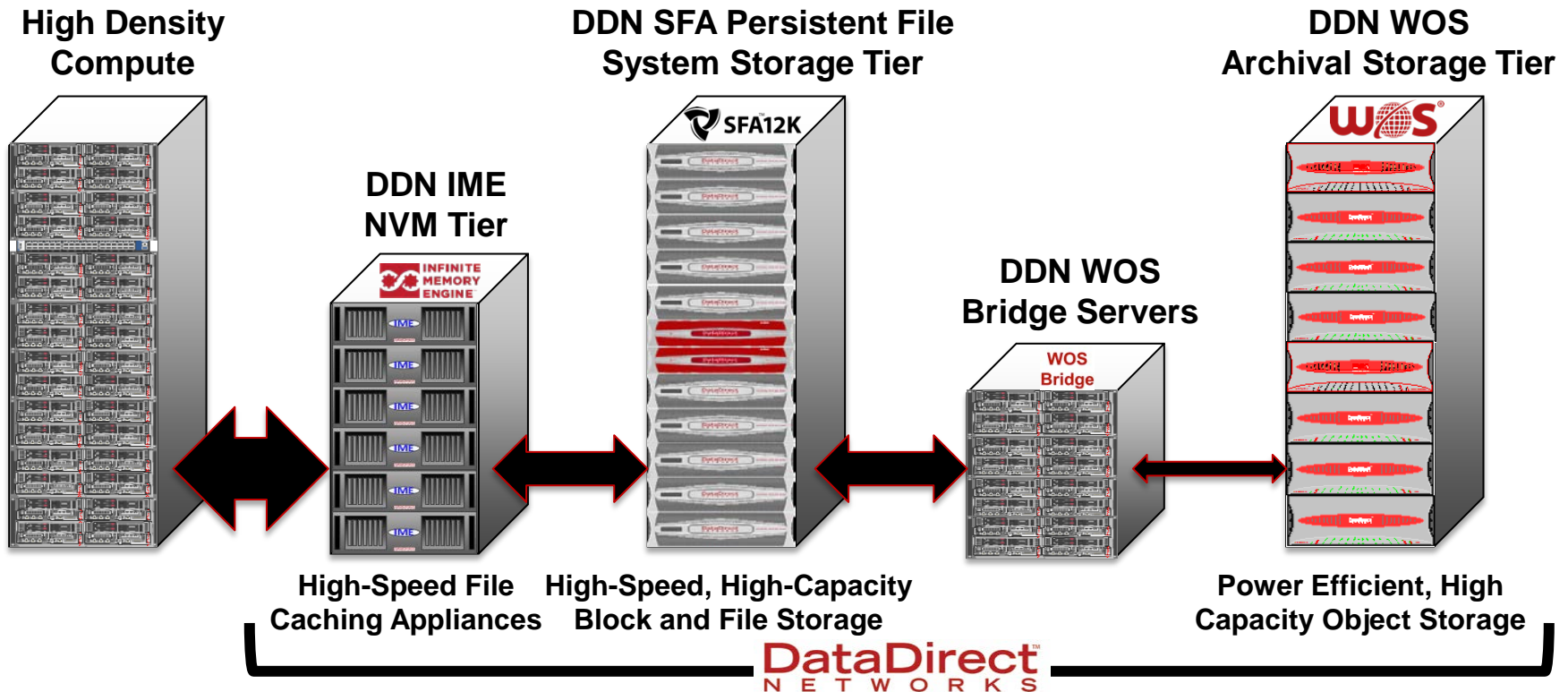
# 4 Forces Driving the Datacenter Evolution



**The Requirement for a New IT Infrastructure is Happening Now**

- ✓ Scalable
- ✓ Secure
- ✓ Flexible
- ✓ Efficient
- ✓ Intelligent
- ✓ Cost-Effective

# End-to-End Big Data Portfolio

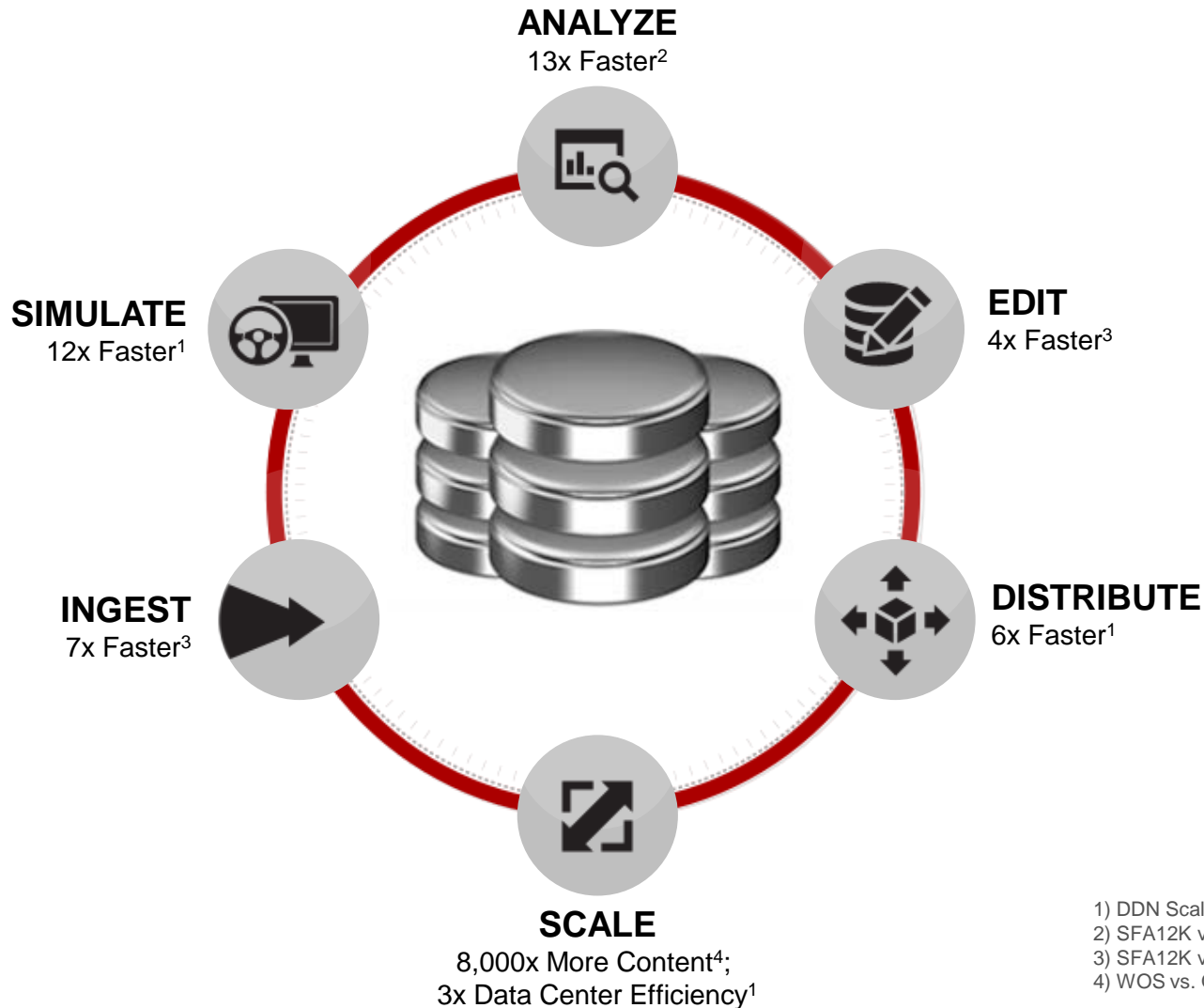


Accelerate processors,  
eliminate file system limits,  
collocate and protect NVRAM  
in compute layer

Simple, highest-efficiency  
primary storage platforms for  
high-performance scale-out file  
I/O & analytics tools

Hyperscale object storage to  
enable next-generation active  
archiving, web scale and  
collaborative cloud computing

# DDN Technology Accelerates & Scales



- 1) DDN Scalers vs. EMC Isilon
- 2) SFA12K vs. EMC VNX (SAP HANA)
- 3) SFA12K vs. NTAP e5500
- 4) WOS vs. Quantum StorNext

# Applications Are Increasingly Demanding



## Life Science/Genomics

- Drug discovery, plant human genome



## Analytics

- Time to insight



## Weather

- Finer grids & more frequent sampling



## Exascale

- x1000 needs . . .

## Oil & Gas

- 4D models for finer modeling



**+ Astronomy, Finance, Material Science, Manufacturing, ...**

# Consider the Impact of VMware

## Decouples Physical Server from Logical Server

### BEFORE



- IDLE CAPACITY
- INEFFICIENT SERVERS
- OVERPROVISIONED COMPUTE

### AFTER



- MULTITASK ACROSS APPS
- SHARE PROCESSOR RESOURCES
- FULLY USE SERVERS



# Software Defined Storage Opens the Door to an Equivalent Storage Case Study

- **Extreme I/O Acceleration**
- **Seamless Scalability and Provisioning**
- **Significant Data Center Efficiency**

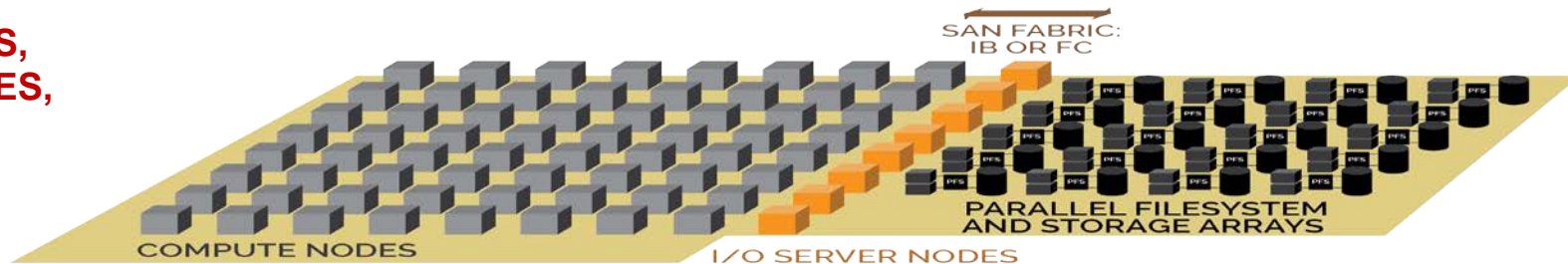
**What is Needed is Game Changing  
Technology & Economics**

# The Evolution of a Large Scale Datacenter

## BEFORE

## EVERYTHING IS OVERPROVISIONED

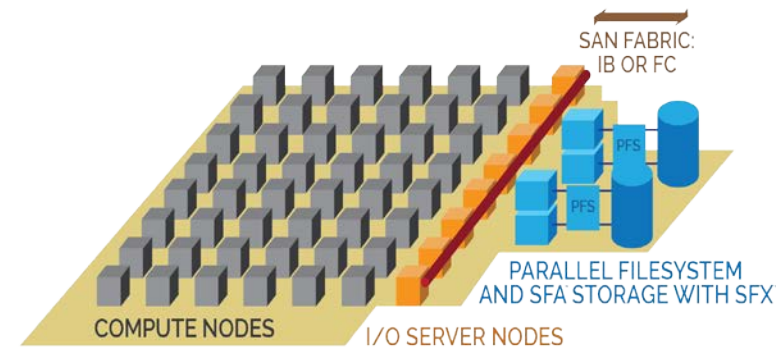
INEFFICIENT  
COMPUTE, DISKS,  
NETWORK, NODES,  
ARRAYS, ADMIN,  
HARDWARE...



## AFTER

## MORE SPEED TO THE APPLICATION

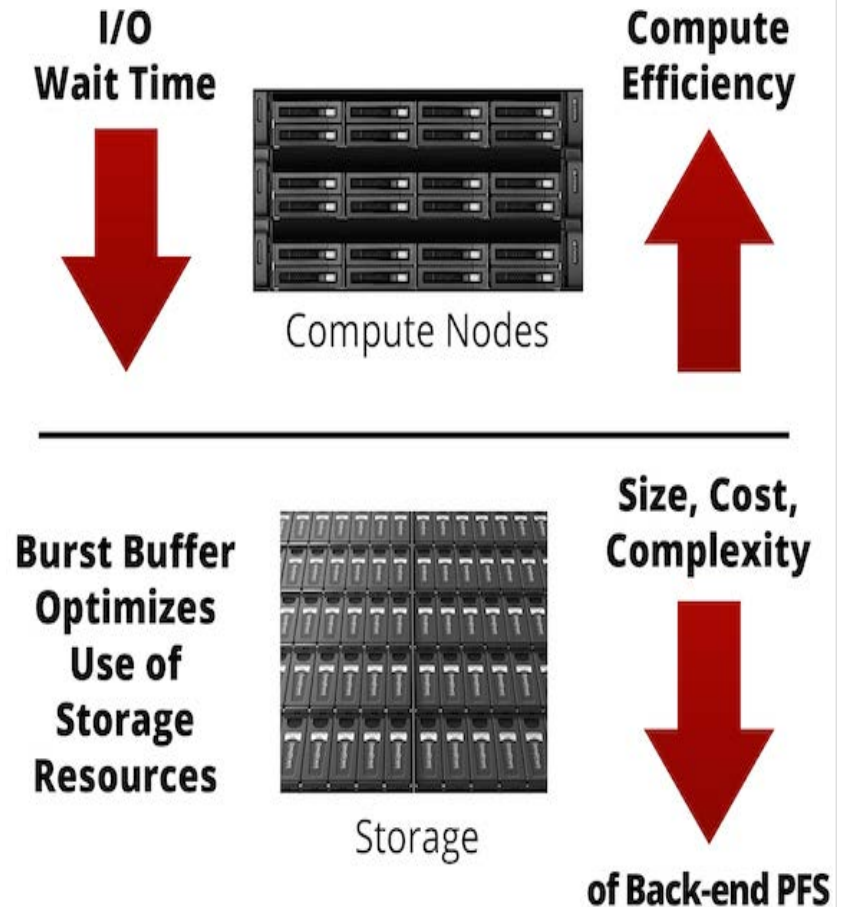
LESS COMPUTE  
NODES, DISK,  
NETWORK, FLOOR  
SPACE, POWER,  
ADMIN, LESS \$\$\$\$\$

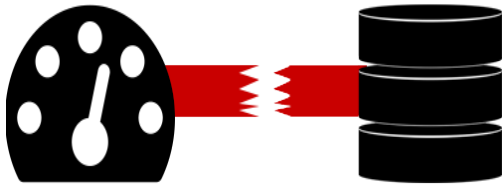


IME SSD PLACEMENT

# Software Defined: Extendable Non Volatile Memory Layer

- ➔ Provision storage performance and capacity independently
- ➔ Achieve full utilization of compute's processing capabilities
- ➔ Achieve full capacity utilization of every HDD in the storage arrays

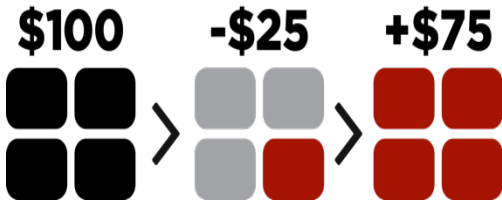




- ① **Decouples Storage Performance from Capacity**  
(SSD vs Spinning Disk)



- ② **Speeds Up Apps by Moving I/O Next to Compute**  
(Bandwidth & IOPS, Read & Write, Small & Large)



- ③ **Shrinks Cluster Idle Time With I/O Provisioning**  
(You Bought a \$100 Cluster But Are Using Only \$25, IME Gives Back the Other \$75)

**Application Acceleration**

Run More Complex Simulations  
Faster With Less Hardware

**Reduced Latency**

Optimizing Workload Performance  
to reduce time to insight and  
discovery

**Scales Memory to 100s of TB**

To Move Large Datasets Out of  
storage & into memory  
extremely fast, without storage  
latency

**Lower Cost**

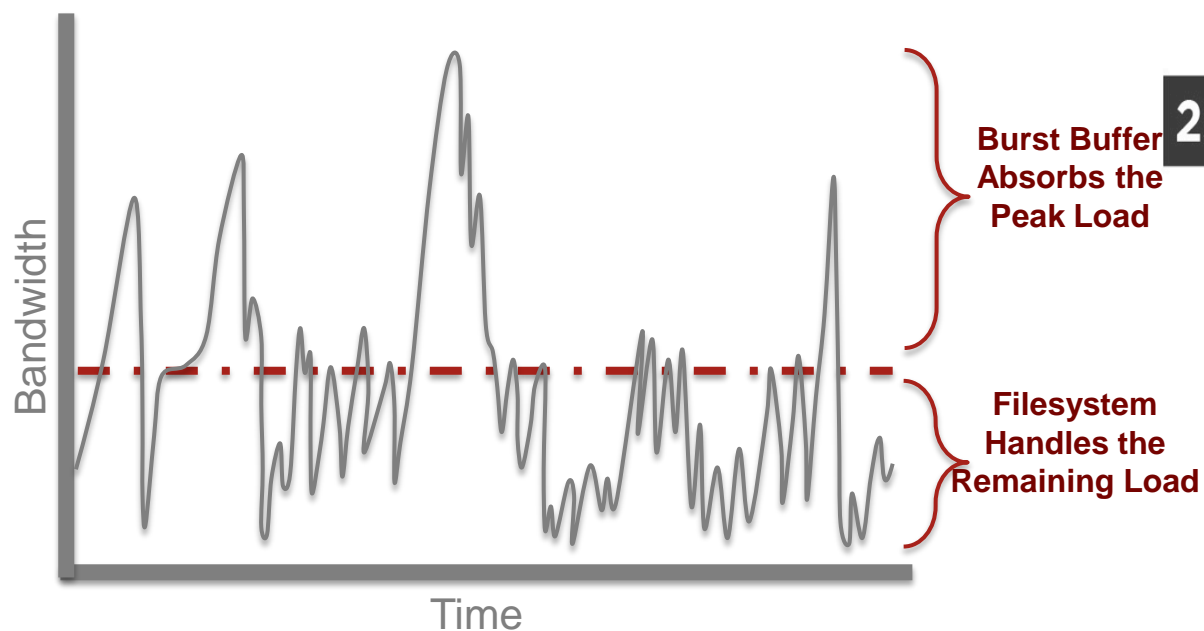
Infinite Scalability With the Highest  
Efficiency To provision I/O  
Performance with the Highest  
Efficiency

## 14

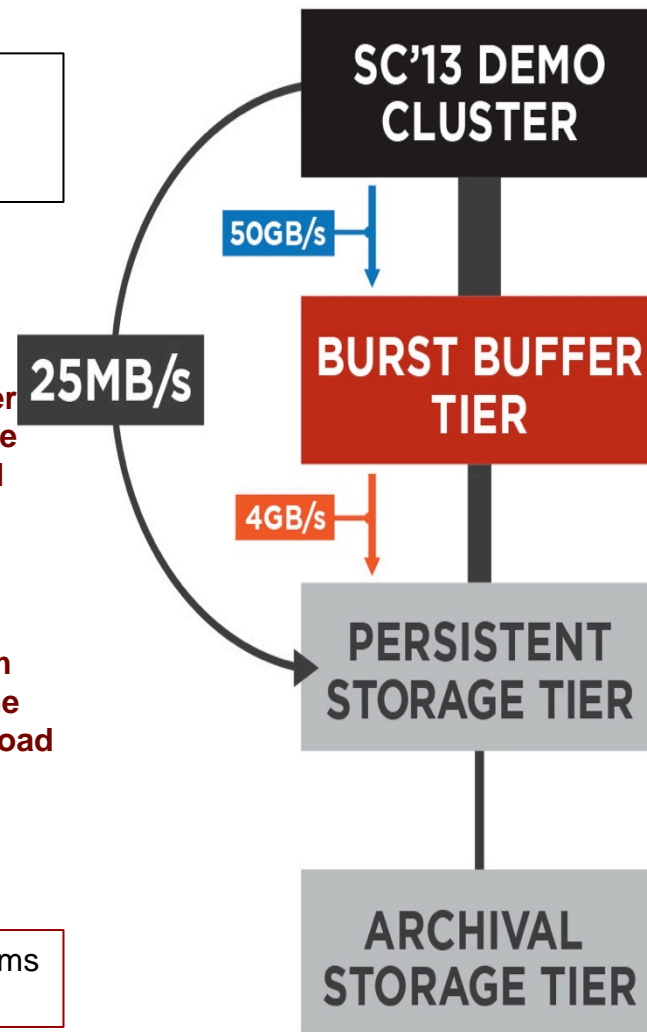
# Sample Use Case

**ANALYSIS****Argonne's LCF Production Storage System**

- 99% of the time, storage BW utilization < 33% of max
- 70% of the time, storage BW utilization < 5% of max

**TREND**

Burst buffers demand smaller, highly robust parallel file systems that sustain very high bandwidth efficiency



## Additional Benefits



Breaks down network bottlenecks. More efficient data center operation. **Result: More Compute, Less Network Cost**



Less power consumption. Better data center density. Lower system cost. **Result: IME Software Cuts Down Hardware Costs**



Makes exascale a reality. Enables the enterprise to run HPC jobs. **Result: Gives HPC500 the Power of HPC50**

# 16 Realities of Modern Data Analysis

- Distributed data sources
  - Multiple administrative domains
  - Heterogeneous storage environments (tape, disk, cloud, web, database, ...)
- Collaborative research
  - Collaborators from multiple institutions
  - Shared data and shared workflows
  - Management of provenance, description, administrative information
- Need for reproducible analyses
  - Compare analyses
  - Modify analyses
  - Statistical ensemble across analyses

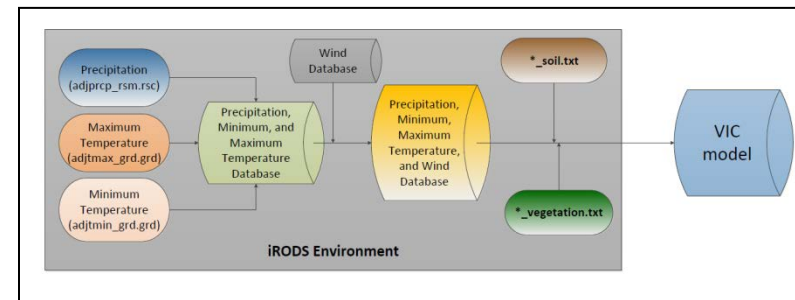
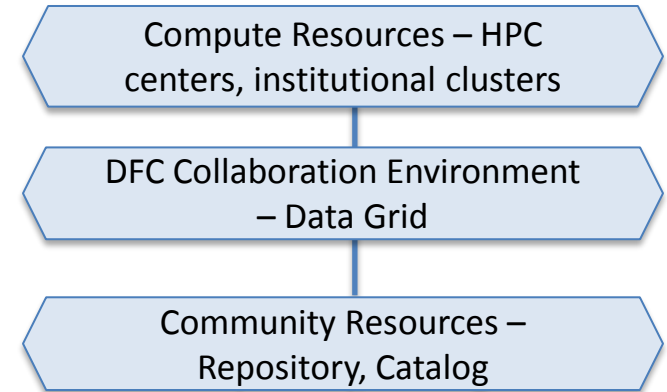


# Software Defined Storage Can be the Link

## Policy-Based, Heterogeneous Platform Data Management

# DataNet Federation Consortium Vision

- Enable collaborative research
  - Sharing of data, information, and knowledge
- Build national data cyberinfrastructure
  - Federation of existing data management systems
- Support reproducible data-driven research
  - Encapsulate knowledge in shared workflows
- Enable student participation in research
  - Policy-controlled access to “live” data



# Cloud & Service Provider Ecosystem



## Internal Cloud Customer

- Transfer EC2 & S3 workload in-house to improve security & lower costs
- Medium to large multi-site enterprise
- Provides services to internal BU's
- Lowers costs by optimizing utilization of CPU & storage
- Addresses Amazon Security concerns
- Control data location

## Managed Service Provider

- Provides hosting services for a few large customers
- Hosts at local site or third party data center
- May share some resources across multiple customers
- Extremely security conscious

## Public Cloud

- Shares resources across many customers
- Hosts at third party data centers
- Subscription pricing for CPU, storage, & network usage
- Offers lowest CAPEX, Subscription pricing
- Usage reporting for bill-back

# Integrating Policy-based Systems

- There is a strict separation between data management and storage management
- Many storage systems provide policy driven data placement which acts on data being placed but also on already resident data.
- These packages have the responsibility to keep the data intact, but it doesn't know what the data means (nor does it need to)
- Policy Managers give insight into what the data mean to the end user (descriptive, provenance, administrative metadata).
- These tools can work together to enable: automated work flows, data access management, audit trails, etc.



Thank You