EMC²
where information lives®

Application Aware Intelligent Storage

Sorin Faibish – EMC Distinguished Engineer
P. Bixby, J. Forecast and J. Cardente
EMC USD Advanced Development

26th IEEE MSST2010 Symposium - Tahoe
May 3-7, 2010

Outline

EMC²
where information lives®

- Current enabling technologies
- Current State-of-the-Art in storage systems
- New trends in storage users needs
- Storage unresolved needs: User, IT
- Storage vendor’s dilemma
- Cloud or Centralized
- What users really need (and cannot ask for today)
- What is missing: intelligence
- Our new storage concept: Idea
- Proposed implementation
- New storage values for users

Current Enabling Technologies

EMC² where information lives®

- Large numbers of cores in CPUs
- Large cheaper memory caches
- Large EFD and flash drives
- Very large cheap SATA drives
- CDP for I/O journaling
- Fast Data De-duplication
- Efficient Compression Algorithms
- Flexible Data Encryption both at rest and in flight
- Optimized Policy Engines for Closed Loop Operation
- State-of-the-Art of Machine Intelligence and Learning

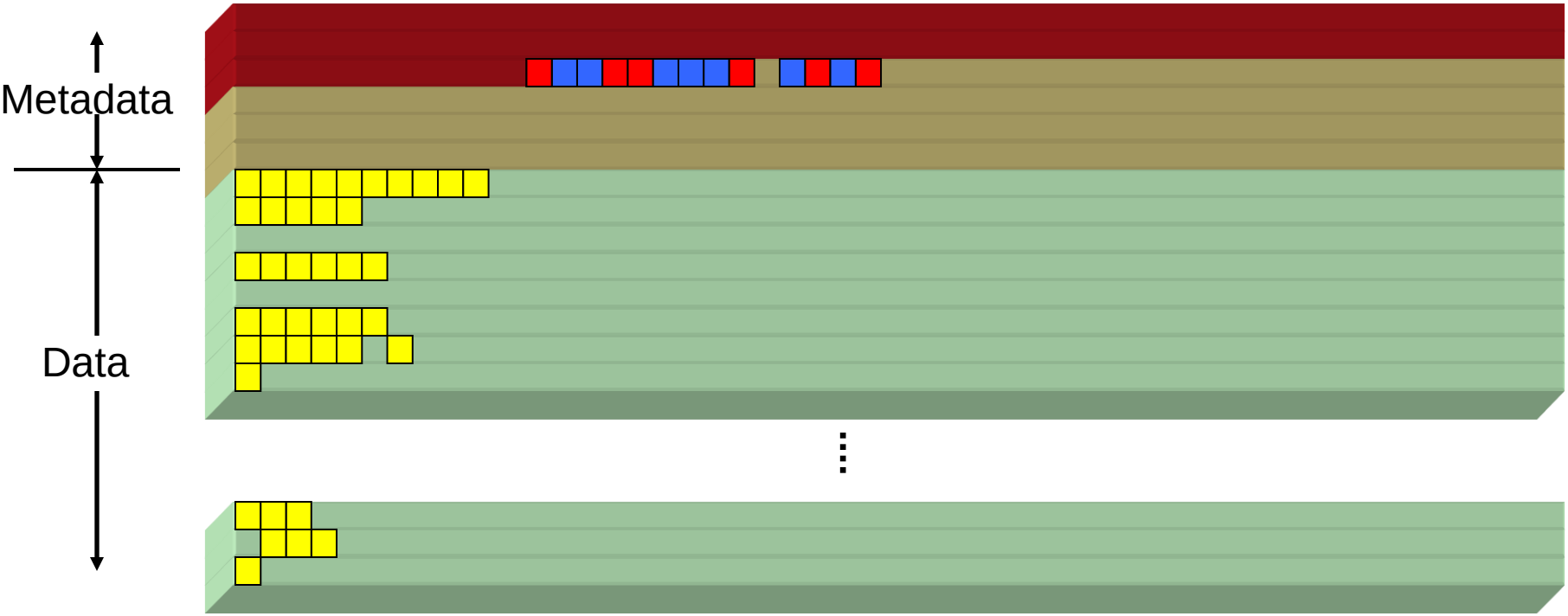
Current State-of-the-Art in Storage Systems

EMC² where information lives®

- **File System Technologies**: enable thin provisioning or block device FS
- **Thin provisioning**: open the door to block Storage virtualization
- **Storage virtualization**: disconnected the physical from logical location to allow re-mapping and **F**ully **A**utomated **S**torage **T**iering
- **FAST**: allow automatic data migration between storage tiers with no application awareness and motion aftermath
- **I/O Tagging**: allow data placement according to application but needs to modify applications
- **pNFS**: protocol that standardize cluster FS and massive parallel data access
- **Intelligent storage**: use machine learning to classify I/Os on the fly and place the data in best place
- **Policy engine**: reinforce policies associated to tiered storage and apply different services transparently using policy feedback

File System for Block Storage Virtualization

EMC² where information lives®



metadata (fixed)

indirect blocks

data blocks

directory blocks

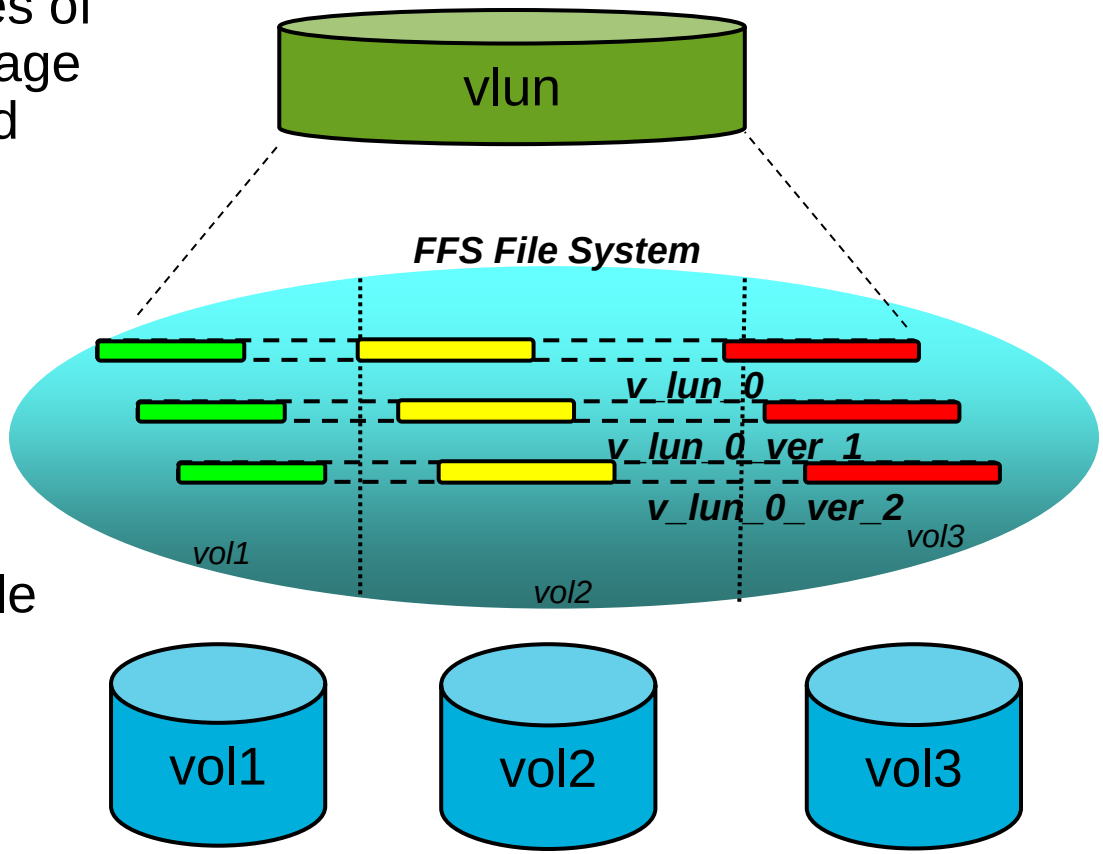
unused metadata blocks

unused data blocks

Block Virtualization (Thin Provisioning)

EMC²
where information lives®

- Automatic mapping of pieces of information to the ideal storage devices in a transparent and durable manner
- Virtualizes Storage Logical Units (LUNs) as files in a File System



Fully Automated Storage Tiering- FAST

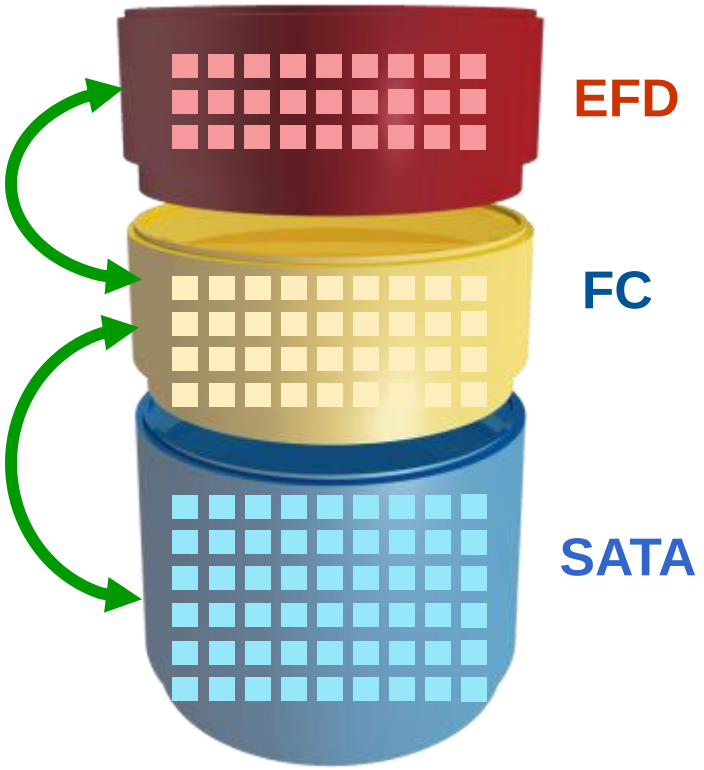
EMC² where information lives®

Automates movement and placement of data based on changing needs

Enables the use of the latest drive technologies

Optimizes both performance and cost

Tiered Storage



I/O Tagging

EMC² where information lives®

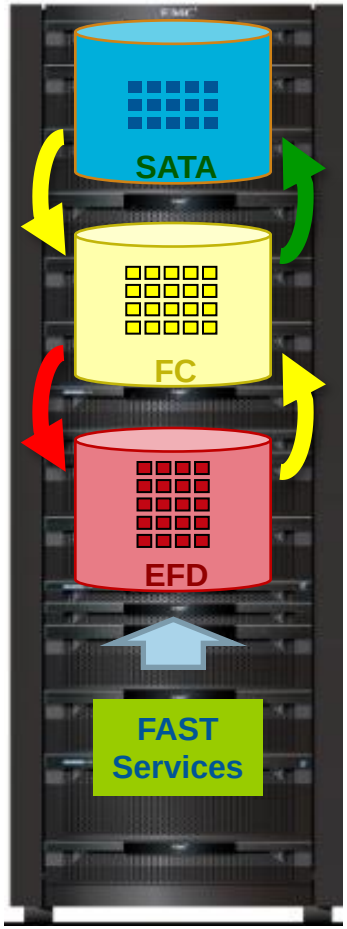
Allowing applications to influence data placement

Application
Understands data classes and their importance

Data Access

Data QoS Hints

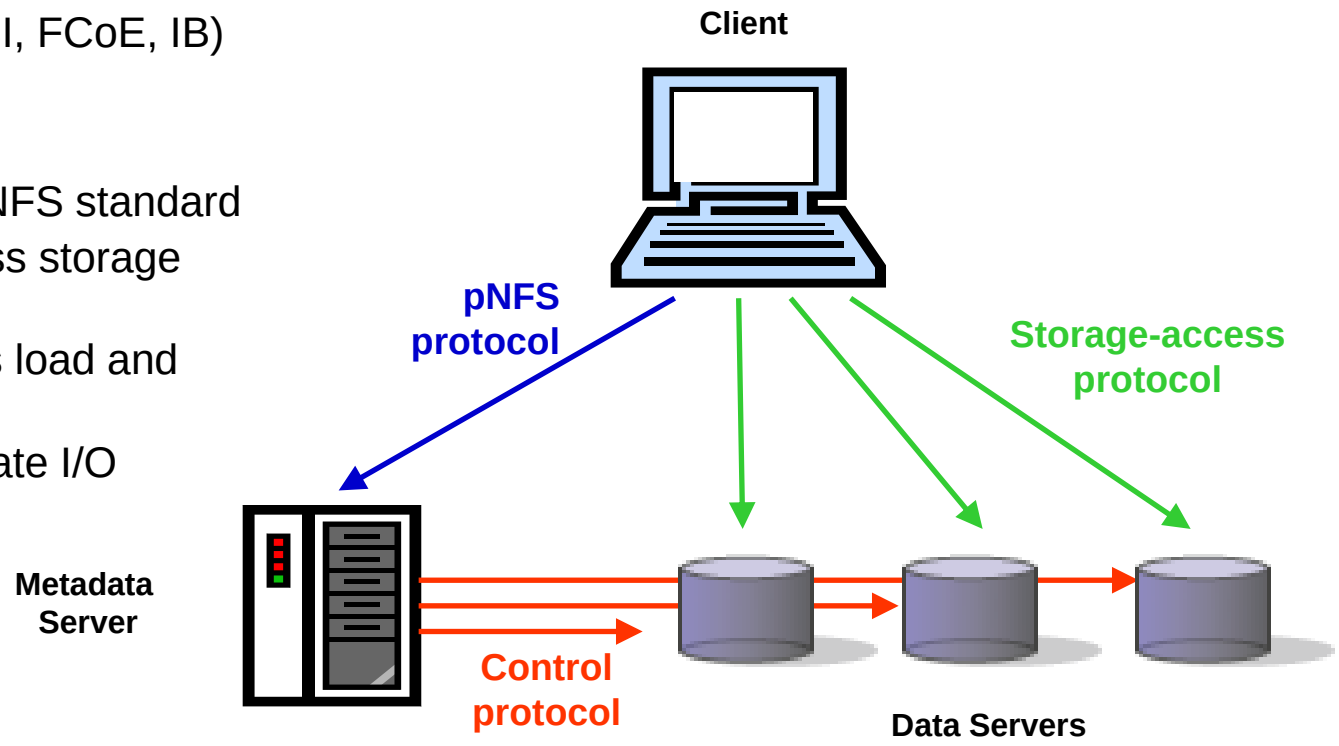
Application provides QoS hints to storage system.



Storage system uses hints to tune data placement.

pNFS

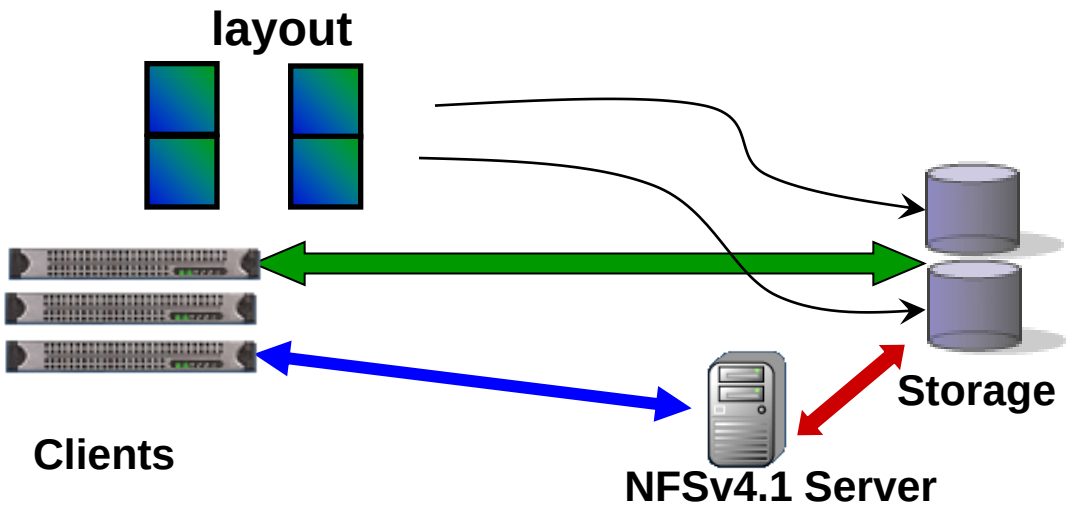
- pNFS protocol
  - standardized: NFSv4.1
- Storage-access protocol
  - files (NFSv4.1)
  - blocks (FC, iSCSI, FCoE, IB)
  - objects (OSD2)
- Control protocol
  - Outside of the pNFS standard
- Distributes data across storage cluster
- Eliminates or reduces load and capacity balancing
- And yes: can accelerate I/O



pNFS Multiple Layouts

EMC² where information lives®

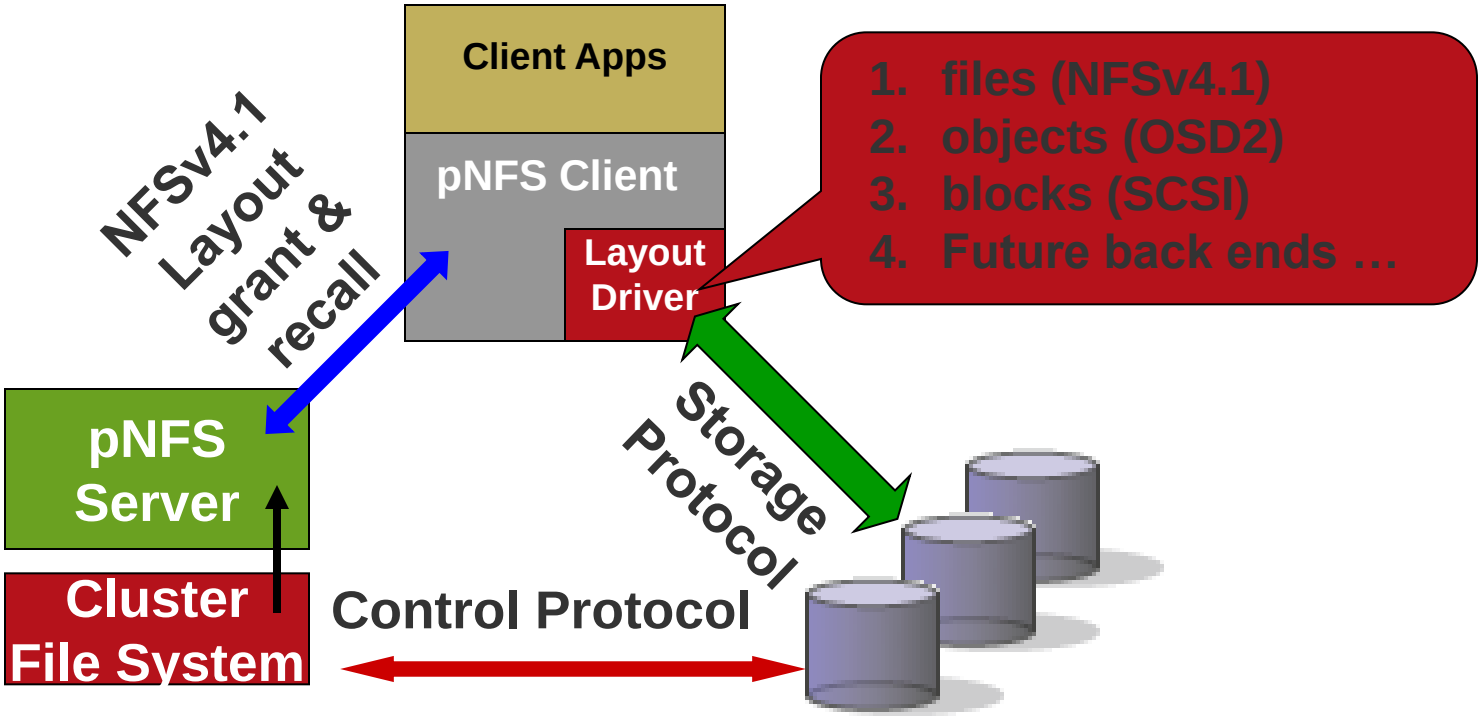
- Client gets a *layout* from the NFSv4.1 server
- The layout maps the file onto storage devices and addresses
- The client uses the layout to perform direct I/O to storage
- At any time the server can recall the layout
- Client commits changes and returns the layout when it's done
- pNFS is optional, the client can always use regular NFSv4.1 I/O`



Linux pNFS Clients

EMC² where information lives®

- Transparent to applications
- Common client for different storage back ends
- Fewer support issues for storage vendors
- Normalizes access to clustered file systems



NFSv4.1 – OpenSource Status

EMC² where information lives®

- Two OpenSource Implementations
  - OpenSolaris and Linux (file, osd and block)
- OpenSolaris Client and Server
  - Support only file-based layout
  - Support for multi-device striping already present (NFSv4.1 + pNFS)
  - “Simple Policy Engine” for policy-driven layouts also in the gate
- Linux Client and Server
  - Support files (NFSv4.1)
  - Support currently also blocks (SCSI), objects (OSD T10)
  - Client consists of generic pNFS client and “plug ins” for “layout drivers
- Windows NFSv4.1 Client from CITI - NEW

Latest Storage Trends

EMC² where information lives®

- Storage became a commodity component in any computing environment
- Storage arrays are pushed down the stack of value offered to data centers
- Intelligent storage arrays have large caches as well as high compute power
- Compute power and cache is used already for low intelligent tasks:
  - Caching IOs for speed before flush to disk
  - Protecting data, RAID computations
  - Pre-fetch IOs for increased read performance (highest level of intelligence)
  - Thin provisioning
  - Replication, mirroring

HPC Applications Storage Needs

EMC² where information lives®

- Application users require suitable performance for access the storage system and do not care of the storage type
- Users are not necessarily aware of the energy consumption or cost of the storage service
- Users need ease of use of the storage without having to know storage configuration or price tag and provisioning
- Users need to be sure the data they use is protected and do not care what protection technology is used
- Users need the data to be stored in secure storage and do not care of the way data is encrypted and managed
- Only thing users need to see is a cloud storage that satisfies all the above needs and not worry how are achieved

IT Department Needs

EMC² where information lives®

- IT department are forced to deliver the storage needs to application users at the lowest possible cost to the business
- IT needs to supply the right amount of storage such that applications are not disturbed by shortage
- IT needs to supply users with storage with right performance characteristics at lowest possible cost
- IT deploy tiered storage to achieve the SLAs committed to the business at expense of human resources
- IT must ensure full data protection and disaster recovery crucial to the business at expense of complexity
- IT must ensure maximum data security and safety specific to each part of the business and application

Storage Vendors Dilemma

EMC² where information lives®

- To address the cost/performance storage vendors use tiered storage but leave the IT the management complexity
- To address energy savings storage vendors power off disks but leave the IT to ensure SLAs to users
- Storage arrays use tiered storage to achieve the SLAs committed to the business at expense of IT resources
- To ensure full data protection and disaster recovery crucial to the business at expense of IT management complexity

How can storage vendors fulfill all these IT needs without increasing the IT head count and expenses and reduce both at the lowest possible $/GB.

Cloud Storage or Storage Arrays?

EMC² where information lives®

- Storage arrays have much more powerful processing units then needed for simple tasks
- Servers use JBOD but need to consume some of the CPU on I/O so the 2 worlds will meet in the "cloud"?
- **Cloud Computing**: bring the storage to the compute but need multiple replicas: efficient analytics compute
- **Storage Arrays**: bring the compute to the storage but need data migration: Computing is optimized for I/O
- **Dilemma**: which is cheaper? Cloud storage or Central Storage who scales better?
- **Is pNFS the answer?**: pNFS and parallel computing address both worlds in a distributed way.

What Users Really Want

EMC² where information lives®

- A storage system able to accommodate any type of application workloads transparently
- A storage that will deliver optimal performance to all the applications concurrently maximizing the value of all the storage resources
- A storage that will send feedback to the IT with recommendations for additional resources required to fulfill all the SLAs of all applications
- A storage that will automatically rebalance storage resources based on changes in application mix
- A storage that will be able to utilize optimally new resources added to it

In short: a highly artificially intelligent storage system

New Storage Functionality

EMC² where information lives®

- There is additional computation power that can be used for higher intelligent tasks like:
  - Recording/indexing and cataloging all the I/Os landing in the cache
    - Input I/Os from SAN to cache; writes
    - Output I/Os from disk to cache; reads
  - Use cataloging data to learn IO patterns and behaviors and match with knowledge of host applications
  - Adaptively locate the data blocks on different storage tiers based on their characteristics
  - Grouping I/Os based on spatial and temporal locality criteria and correlation
  - Learn behavior of applications using model based learning and identify applications specific workloads
  - New functionality can be used with any type of storage: block, file or object
  - Improve power management using retention of I/Os and placement on optimal power tier

Higher Intelligence Storage Services

EMC² where information lives®

- The additional computation power can be used to deliver highly intelligent services:
  - Perform security scrutiny and selective encryption based on policies and learned patterns
  - Adaptively control the number of copies based on learned access patterns
  - Block and file de-dupe compression or decompression of data blocks on the fly (while transit in the cache)
  - Adaptive multi-tiered data location on a global scale according to policies and learned behaviors
  - Block replication based on learned access patterns; in and out,
  - Automatic features detection and indexing images,
  - Sounds indexing in MP3 audio files

Benefits to Storage Users

EMC² where information lives®

- Allow users to add storage resources based on the available budget and get maximum performance from given resources
- Report back to users shortage of resources required to achieve the SLOs of all applications and make recommendations for additional resources to achieve the required SLOs
- Report back to users changes in application behavior generating missed SLOs
- Allow deliver different SLOs for different applications: lower SLOs for test applications than production applications sharing resources
- Always use the minimal energy to achieve all applications SLOs with minimal storage resources provisioning

Deliver highest price/performance storage value at any time

Additional Value to Users

EMC² where information lives®

- Create temporary copies of I/Os based on hot spot detections
- Encrypt/decrypt data blocks based on similarity of location
- De-dupe and compress I/Os on the fly and place them in faster storage based on access frequency (number of copies)
- Add new services (pre-stored VM service images) as new needs appear
- Store DB of less frequently used VM service providing images
- Can be implemented as a cache appliance in front of JBOD tiered storage as a cache appliance
- Can gradually introduce to cloud storage small number of services in the beginning and add services to existing systems


EMC²
where information lives®