IBM XIV Storage System

Enterprise Storage Reinvented

Dominique SALOMON

European Product Introduction Center (PIC) Leader

Storage Specialist Certified - IBM Montpellier

✉ Dominique.Salomon@fr.ibm.com

Agenda

- Architecture & Hardware Overview
- XIV Data distribution
- XIV Basic & Advanced functionalities
- XiV GUI

Traditional Enterprise storage solutions

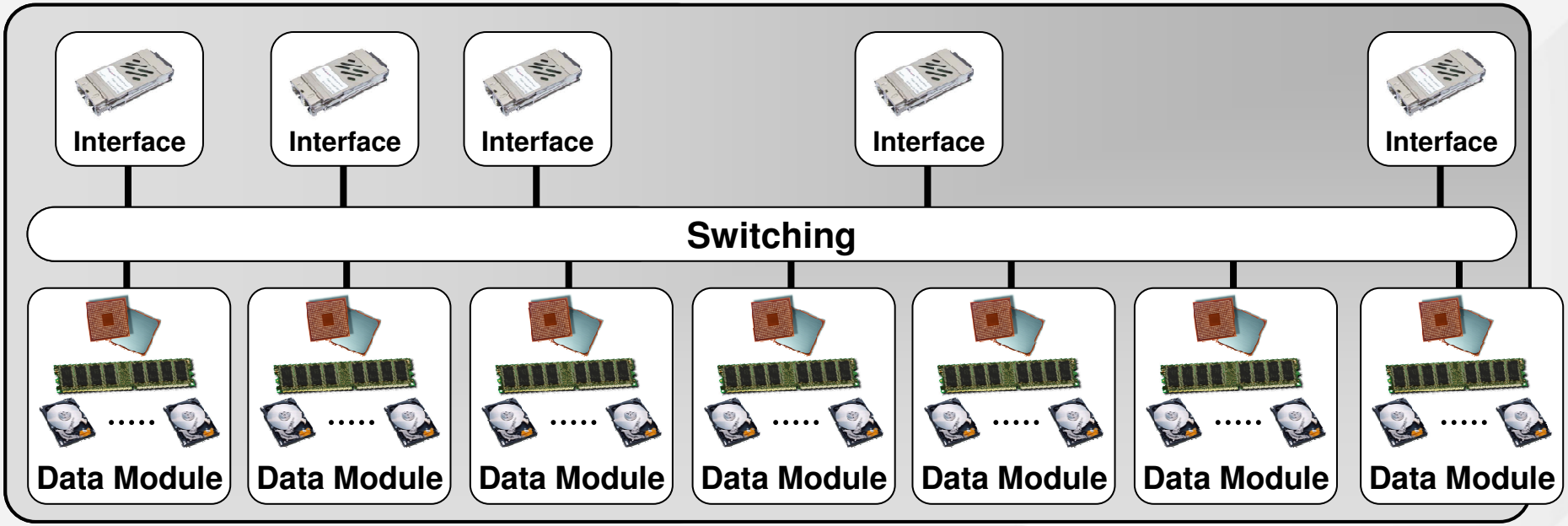
- **Traditionally storage** is improved by further optimizing the existing concept, e.g.
  - Using faster and more reliable drives
  - Adding additional cache
  - Manufacture new backplanes
  - Add new hw/sw layers for virtualization and thin provisioning
  - Forklift ´upgrades´
- This comes with a price, resulting in high cost, complex solutions and increased power consumption
- We have to look for different ways to achieve our ever growing need for larger, faster, flexible, efficient and more reliable ways to store our data
- We had to reinvent the way we look at storage

IBM XIV Storage Architecture

Design principles:

- Massive parallelism
- Granular distribution
- Off-the-shelf components
- Coupled disk, RAM and CPU
- User simplicity

Scale Out



IBM XIV Enterprise Storage Solution

- IBM XIV Storage is based on the following basic principles:
  - The entire system is one Virtual space
  - Simple storage provisioning an thin allocation
  - “Self Healing” - the failure of a component is automatically fixed with no impact on the reliability and performance of the system
  - “Self Tuning” - the provisioning and management of data should always result in the optimal use of available space
  - The speed of data access is not dependent on the speed of the drives (no disk hotspots)
  - Make use of readily available standard components
  - “GREEN” - Efficient use of resources: Power, Cooling, Space
  - Best in class TCO - Cost effective

IBM XIV Storage Distribution Algorithm

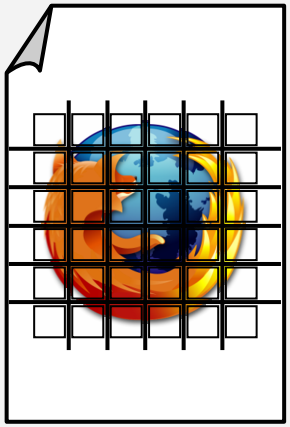
- Each volume is spread across all drives
- Data is "cut" … the disks
- XIV algorith … across **all** disks in the …

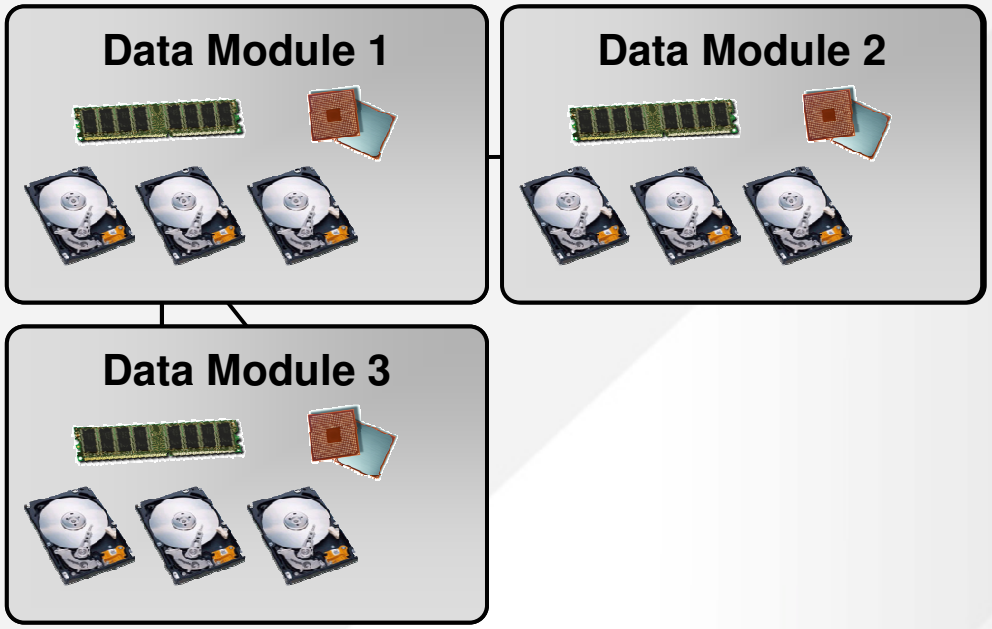
Nextra disks behave like connected vessels, as the distribution algorithm aims for constant disk equilibrium.

Thus, IBM XIV's Storage ove… disk usage could approach 100% utilization when load…

XIV Distribution Algorithm on System Changes

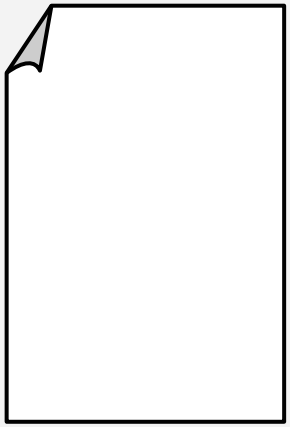
- Data distribution only changes when the system changes
  - Equilibrium is kept when new hardware is added
  - Equilibrium is kept when old hardware is removed
  - Equilibrium is kept after a hardware failure

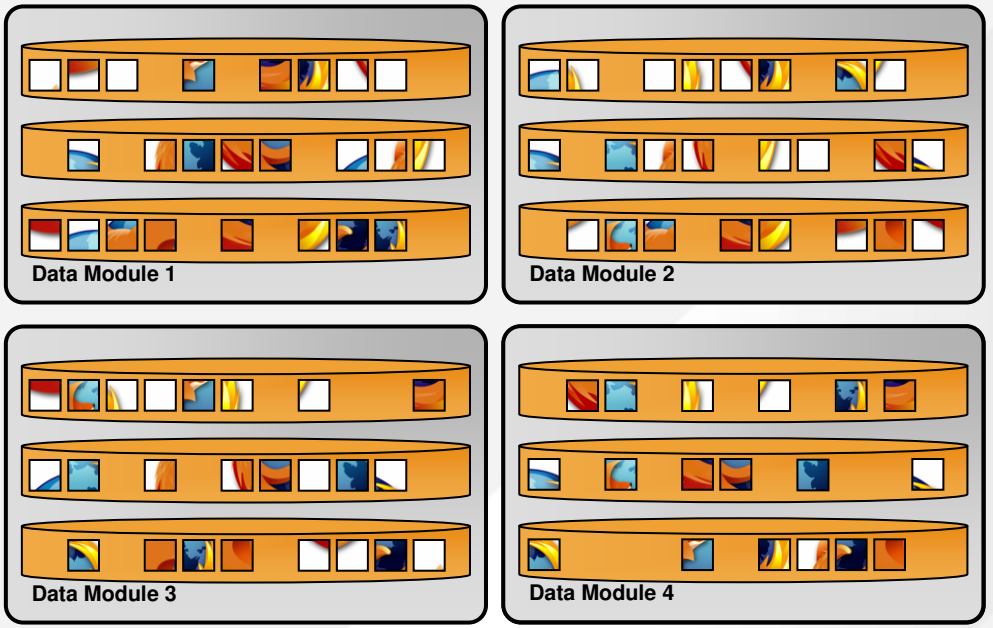




XIV Distribution Algorithm on System Changes

- Data distribution only changes when the system changes
  - Equilibrium is kept when new hardware is added
  - Equilibrium is kept when old hardware is removed
  - Equilibrium is kept after a hardware failure



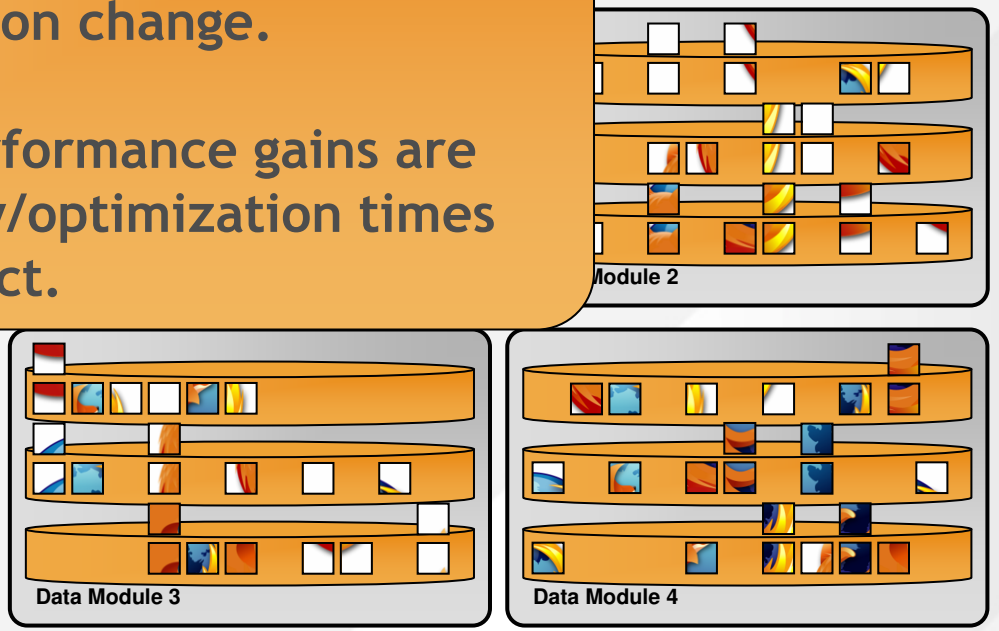


XIV Distribution Algorithm on System Changes

- Data distribution only changes when the system changes
  - Equilibri
  - Equilibr
  - Equilibr

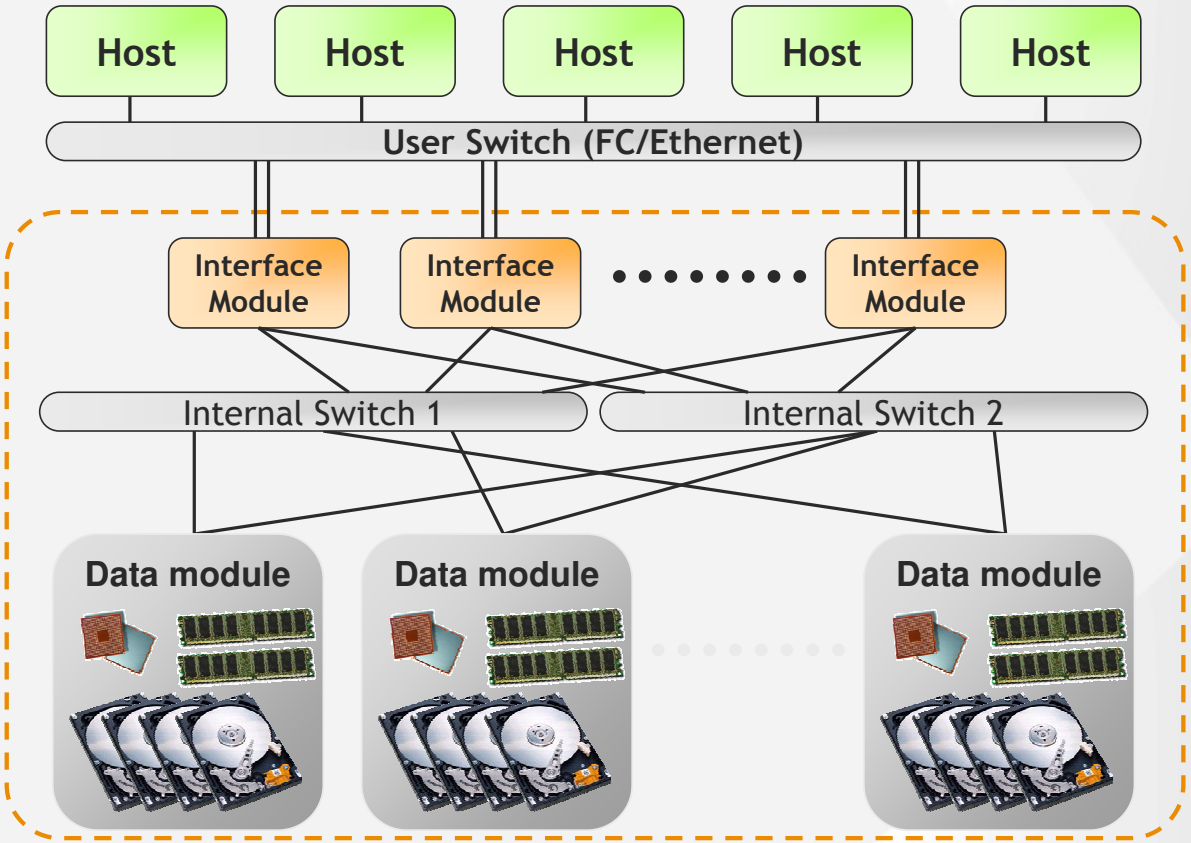
The fact that distribution is full and automatic ensures that all spindles join the effort of data re-distribution after configuration change.

Tremendous performance gains are seen in recovery/optimization times thanks to this fact.



Storage in a grid

- Efficient and Green by design
- Simplified architecture
- Each volume a spread on all the drives
- Use of large SATA Disks
- Integrated software
- Thin, smart and simple to manage
- Simple migrations
- How does it work?



- Data is redundantly spare space are spread over all the drives, with parallel access and smart caching to match the performance of high end systems
- If a drive fails, the system replicates the lost data across the other drives. System is **fully redundant in less than 30 min** with minimal performance impact.

IBM XIV Storage System Hardware Platform

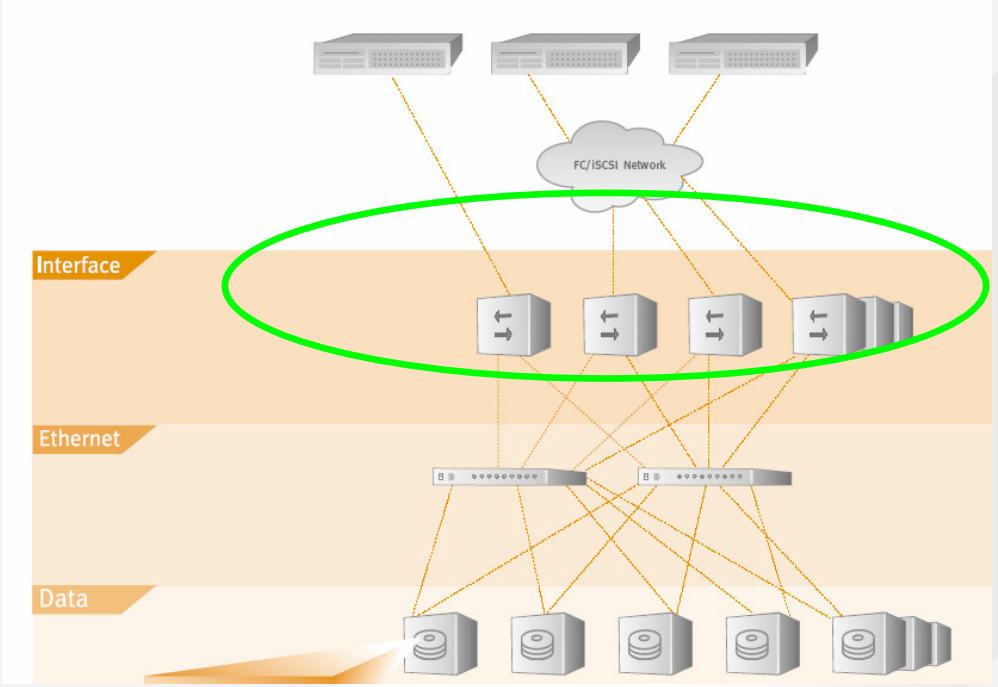
Machine Type: 2810-A14

- **180 disks** per rack
  - 15 modules per rack
    - 12 disks per 2U module
  - **1TB 7200RPM SATA** disk drives
- Single rack provides **79TB usable capacity**
- **120GB of system cache** per rack (8GB per module)
- Up to **24 4Gb FC** host ports
- **6 1Gb iSCSI** host ports
- **3 UPS** systems

IBM XIV Storage

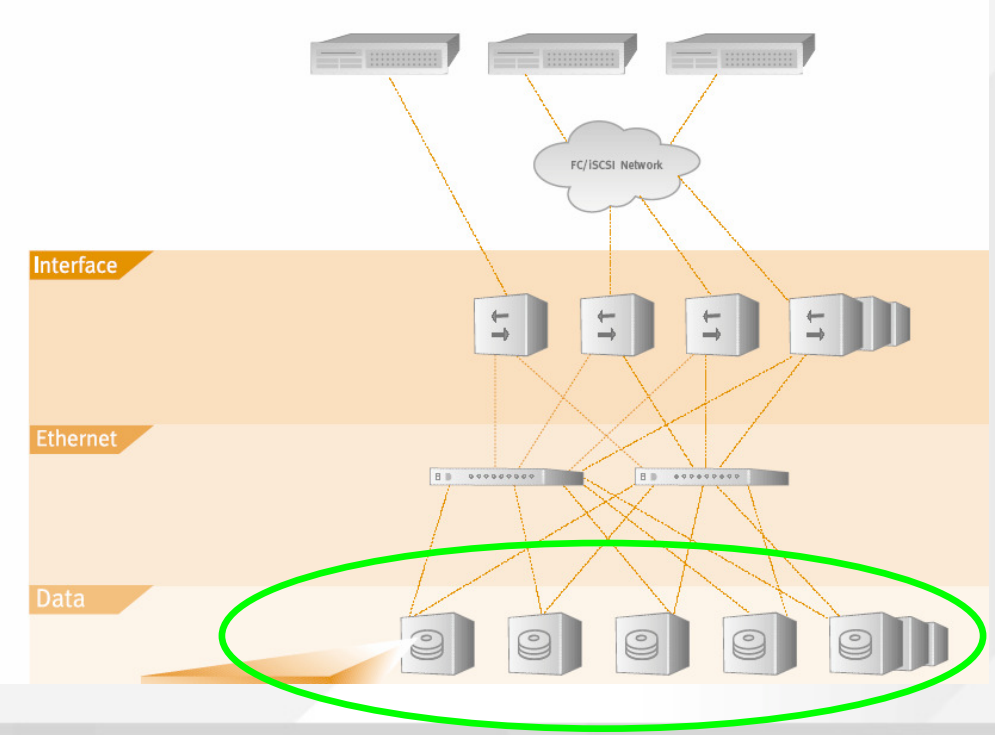
Architecture: Interface modules

- Provides iSCSI and FC connectivity to the hosts
- Characteristics:
  - Any interface can access the entire system
  - Each module works independently
  - Scalable connectivity
  - Practically unlimited connection redundancy
- Active-active multipath (load balancing)



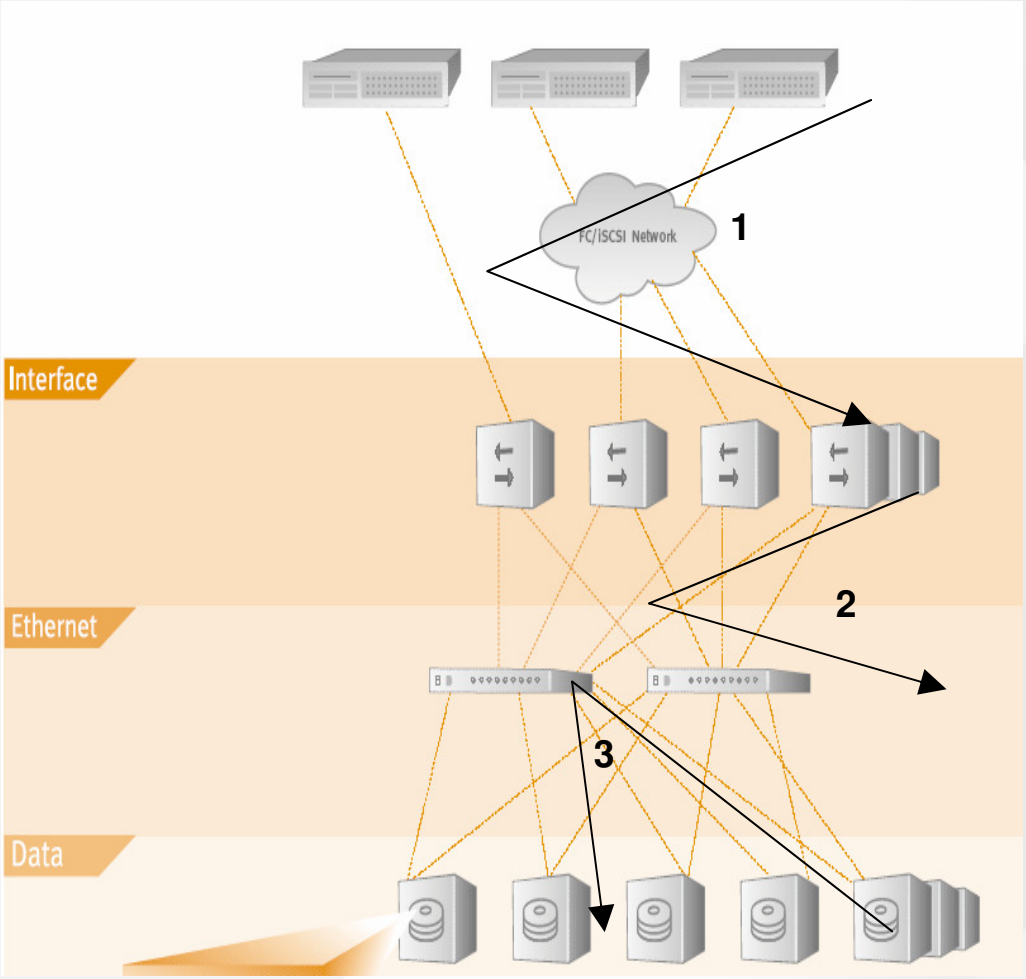
Architecture: Data modules

- Data Modules, contain SATA drives and perform the data service execution, as well as caching and snapshot functionality.
- Data is stored on disks within the module
- Scalable capacity and performance
- Dual write mechanism provides cache redundancy
- Powerful caching approach
- Very high throughput allows aggressive pre-fetching
- Powerful CPU allows for smart LRU



Architecture: Data Path

1. Host sends write to interface
2. Interface sends write to primary data module
3. Primary data module sends write to secondary data module
4. Host is acknowledged only after write is completed on both modules



XIV Storage Pools: Management of storage space

- Storage Pools are Logical
  - Storage System is partioned into Storage Pools
  - No disk drives, RAID groups or any other physical resources are reserved per pool
  - Can be of any size.
  - Same performance for all storage pools
  - Volumes can be moved between Storage Pools
  - No data movement involved… immediate results

- Storage Pools control the storage resources
  - Use for specific applications or departments
  - Physical and Virtual storage used by volumes
  - Limits physical space for clones
  - Can be resized dynamically as needed… depending on available resources

- *Storage Pools are for capacity isolation reasons, not performance*

XIV Volumes

- Volume Characteristics
  - A volume is always a part of one and only one Storage Pool
  - Volumes can be moved between pools
  - A volume may have multiple Clones
  - A volume may be part of one and only one Consistency Group
- Volumes size can be **dynamically resized**
- *User does not plan the layout of the volumes relative to physical drive resources*

XIV Mapping Volumes to Hosts

- Used to control server access to LUNs
  - **LUN Maps**
    - For each host, a LUN identifies a single volume or snapshot
    - Different hosts can use the same LUN to access different volumes or snapshots.
    - Logical volumes can be added to or removed from any map dynamically
  - **Cluster Maps**
    - Grouping of several hosts together that have the same mapping to all of the hosts
    - Mapping of volumes to LUN identifiers is defined per cluster
    - Applies concurrently to all the hosts in the cluster
    - No way to define different mappings for different hosts belonging to the same cluster

XIV reliability: Scrubbing

- Verifies integrity and redundancy of stored data
- Enables early detection of errors and early recovery of redundancy
- Runs as a background process, on all disks in parallel
- Checks whether data can be read from partitions employing checksums
- 1 partition is examined every second

IBM XIV Storage: Concept of "Spare"

- Traditional approach
  - Dedicated disks used for spares
  - In many systems spares are dedicated for a RAID group
- IBM XIV Storage approach
  - Recovery time: 30 minutes for 1 TB disk (if full)
  - No dedicated spare disk, only global capacity
  - All disk are equally used
  - Minimize the risk of technician mistakes
  - Higher availability with no performance impact
- 180TB raw is 79 TB net
  - Spare space for 3 disks and one full module
  - 79 = (180 – 12 – 3)/2 – 3.5 (internal use)

IBM XIV Storage: Thin Provisioning

- Defining logical volumes bigger than physical capacity
- Installing physical capacity only **if** and **when** needed
- No space consumed when data is 0
- Pools are used to manage quota
- Results:
  - Reduced overall direct storage cost
  - Storage expenses spread over time, exploiting price reductions
  - Easier management
  - Save 20-50% of storage capacity

IBM XIV Snapshots - Virtually without Limits

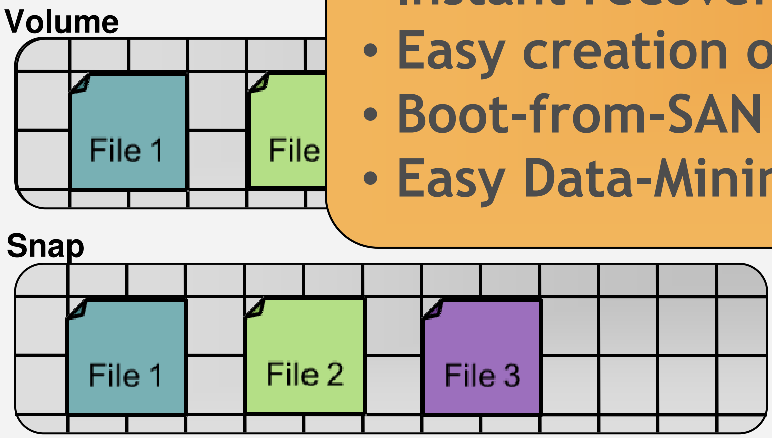
- Snapshot creation/deletion is instantaneous
- High performance WITH snapshots
- Unlimited nu
- Differential s
- **Redirect-on-**

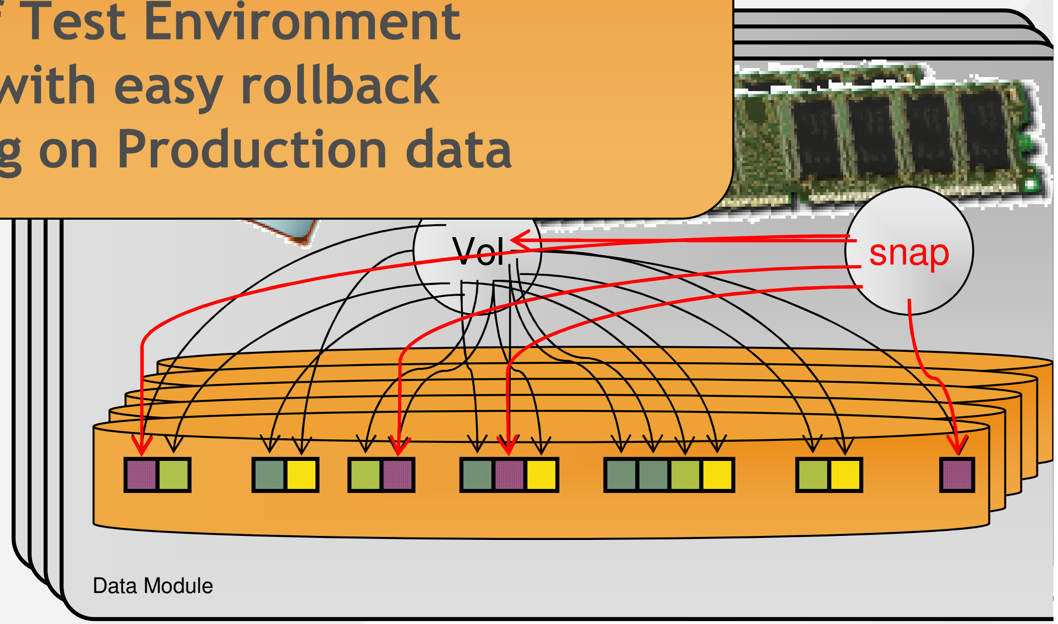
Distributed snapshots on each server. Extremely fast memory

s fast as
n volumes

High performance snapshots provide:

- Easier Physical Backup to Tape
- Instant recovery from Logical Backup
- Easy creation of Test Environment
- Boot-from-SAN with easy rollback
- Easy Data-Mining on Production data





Restore Volume from snapshot copy

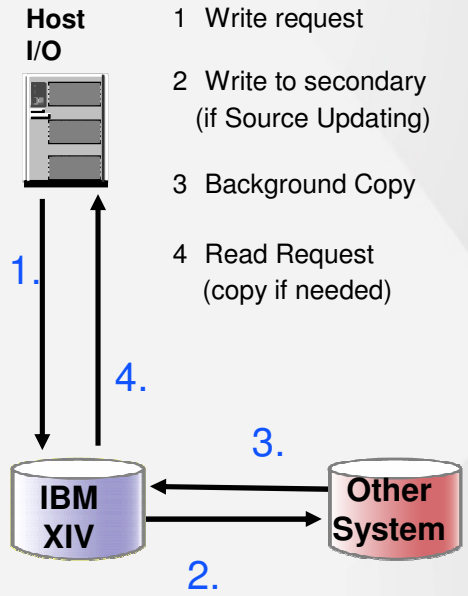
On a snapshot, each server simply points to original volume: memory only operation

XIV Volumes copy

- Copy data to another volume
  - Target can be in another Storage Pool
  - Target cannot be a Clone or a master volume
- Like a Snapshot but the target is independent (read / write)
  - Logically identical volumes
- Volume is available immediately
  - Space is allocated at creation (not differential)
  - Copy is a background activity

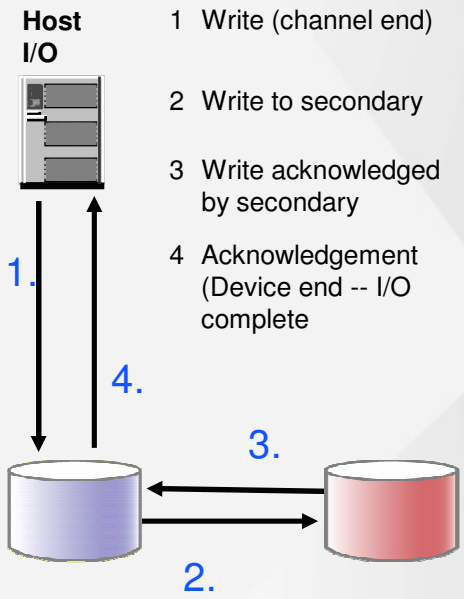
IBM XIV Storage System Data Migration

- Enables the transition to an IBM XIV Storage System
  - Host connects to the IBM XIV Storage System
  - Data is available before the copy is complete
  - Synchronizes by volume as a background process
- Hosts connect to IBM XIV Storage System as storage
  - Either FC or iSCSI
- IBM XIV Storage System connects to legacy storage as a host
  - Either FC or iSCSI
- Protocol connectivity does not have to be the same
- Thick to Thin provisioning



XIV Synchronous Remote Mirroring Support

- Synchronous remote Mirroring
  - Link Types and Distance
    - 100 km's
    - Ethernet/IP/iSCSI
    - Fibre Channel (FCP)
  - Replication to multiple target systems is allowed
  - Replication can be performed in both directions
    - Remote site can use the local site as a secondary
    - While the Local site uses the remote site as a secondary
  - Switch roles is possible



SATA disks to save even more power

- The power consumption of a system is the sum total of the power used by its components
  - Since there are so many of them, disks are typically the biggest users
- SATA vs. FC disks
  - SATA drives provide 2-10 times the capacity
  - Lower spin rate means each disk requires 25-30% less power
  - The result: A lot less power is used to drive each raw TB (3 to 15 times less)
- using SATA drives is not a compromise
  - The XIV architecture offers primary storage performance for all volumes
  - Perfectly adapt to any future changes in volumes and capacity

Enjoy the power-efficiency of SATA drives with a supporting architecture that handles tier-1 performance and reliability

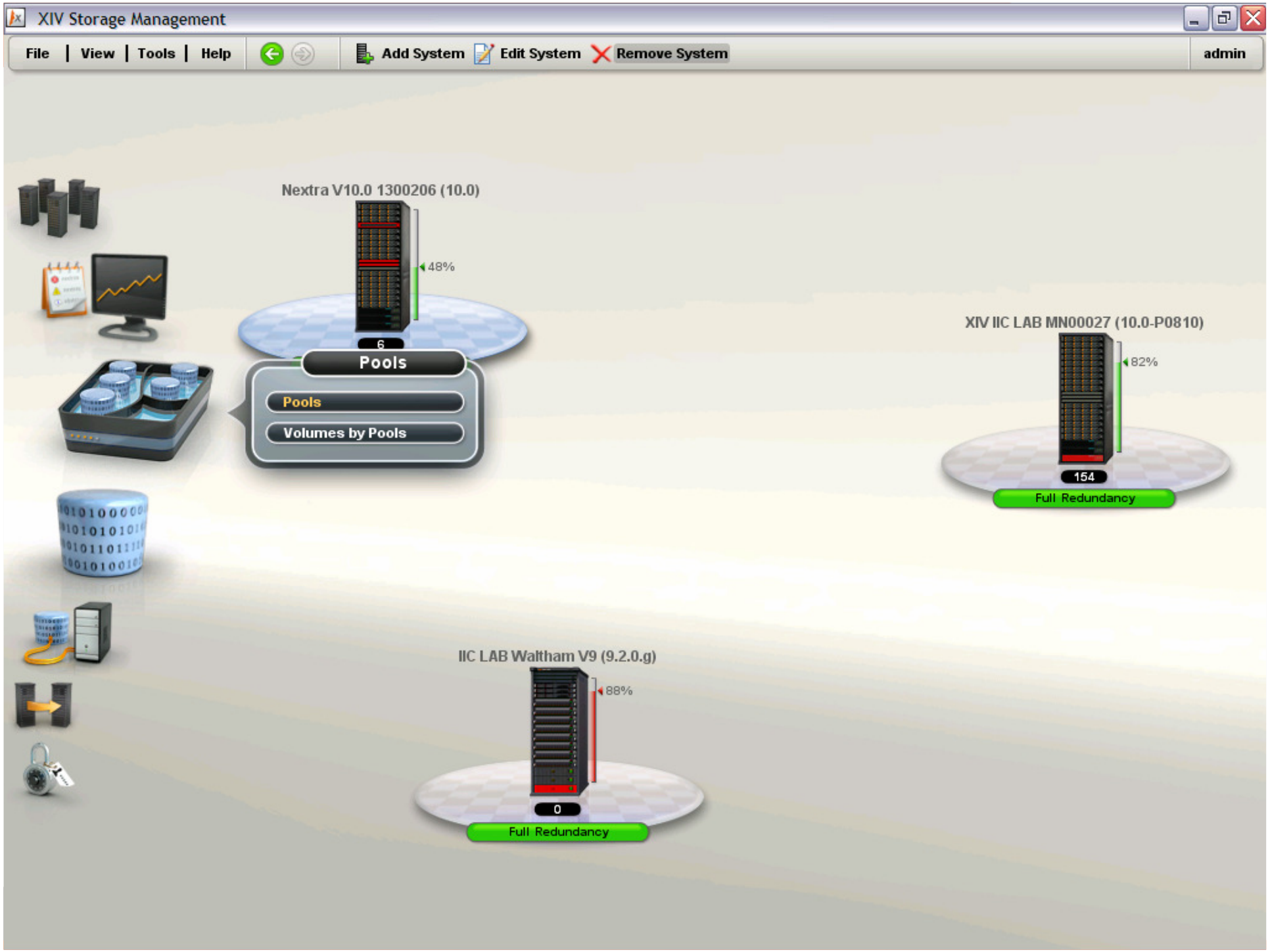
System Power Usage

- Power consumption of a system comparable to XIV is 180–380W per raw TB
  - Typically using 146GB 15K rpm disks
- Power consumption of an XIV rack is 7.7KW
  - 180TB raw capacity, 79TB net capacity
  - 42W per raw TB today
- Rack power consumption will not change much with 2TB disks
  - But capacity will double
  - Consumption per raw TB expected to drop to 21W

The new solution uses 4 to 9 times less power for the same (or better) performance and reliability levels

IBM XIV Storage Simple and Intuitive Management

- Intuitive GUI (Java based) with Script Generator
- No dedicated management station
- Command Line Interface (CLI)
- XML over SSL
- Event management (SNMP)
- Complete Event Logging
- Events notification via email, SNMP and SMS
- Role based management:
  - Storage Admin
  - Application Admin
  - Operator


XIV Storage Management
File | View | Tools | Help
Add System
Edit System
Remove System
admin
Nextra V10.0 1300206 (10.0)
48%
6
Pools
Pools
Volumes by Pools
XIV IIC LAB MN00027 (10.0-P0810)
82%
154
Full Redundancy
IIC LAB Waltham V9 (9.2.0.g)
88%
0
Full Redundancy

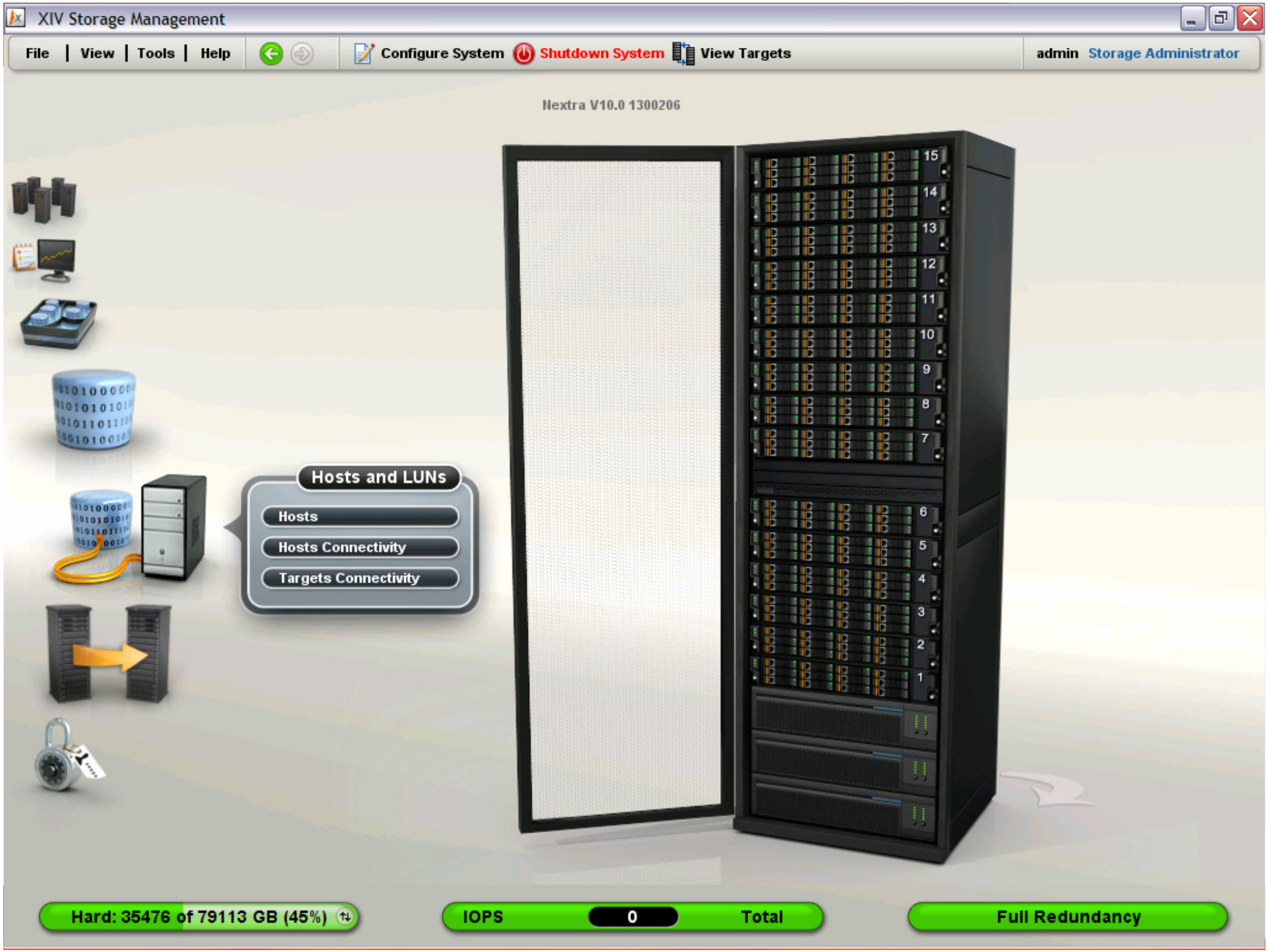

XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy


XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
Patch Panel
FC Port: 2
(Data Module: 6)
WWPN:
5001738000CE0161
User Enabled: yes
Rate (Current): 4
Rate (Configured): Auto
Role: Target
State: Online
Status: OK
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy


XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
Monitor
System
Events
Statistics
IP Interfaces - iSCSI
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy


XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
Pools
Pools
Volumes by Pools
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy

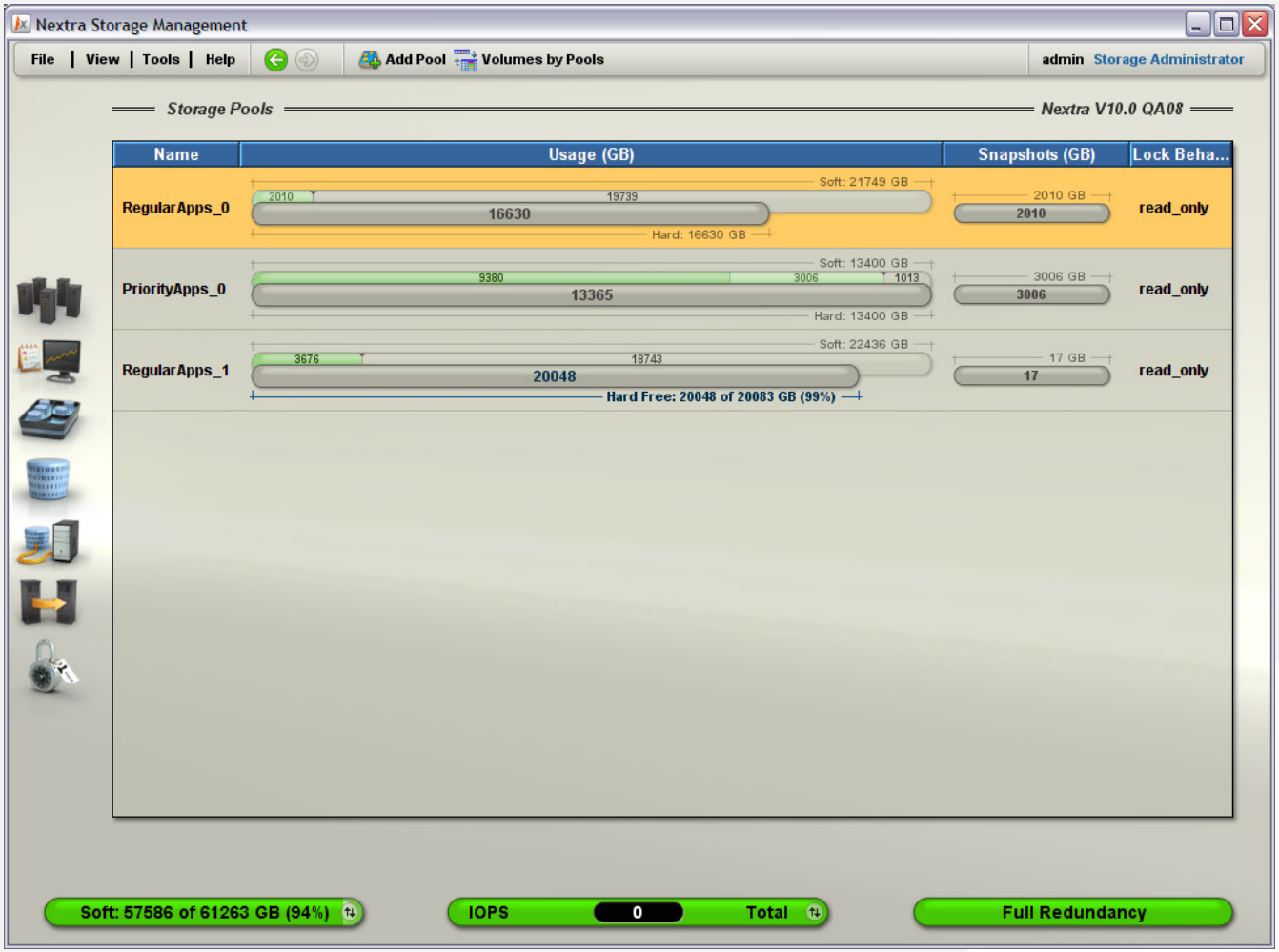

XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
Volumes
Volumes & Snapshots
Snapshots Tree
Volumes by Pools
Consistency Groups
Snapshots Group Tree
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy


XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
Hosts and LUNs
Hosts
Hosts Connectivity
Targets Connectivity
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy

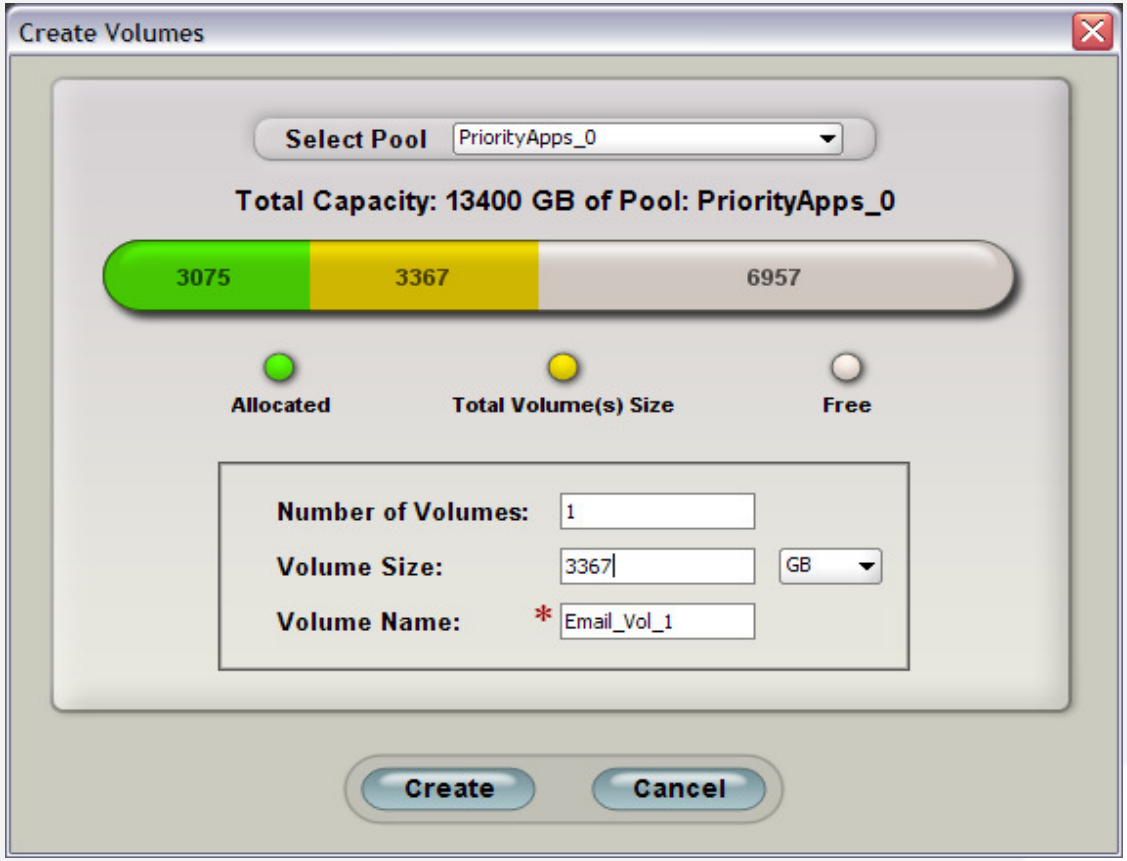

XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
Remote
Targets Connectivity
Remote Mirroring
Data Migration
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy


XIV Storage Management
File | View | Tools | Help
Configure System
Shutdown System
View Targets
admin Storage Administrator
Nextra V10.0 1300206
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Access
Users
Users Groups
Access Control
Hard: 35476 of 79113 GB (45%)
IOPS
0
Total
Full Redundancy

IBM XIV Storage Pools

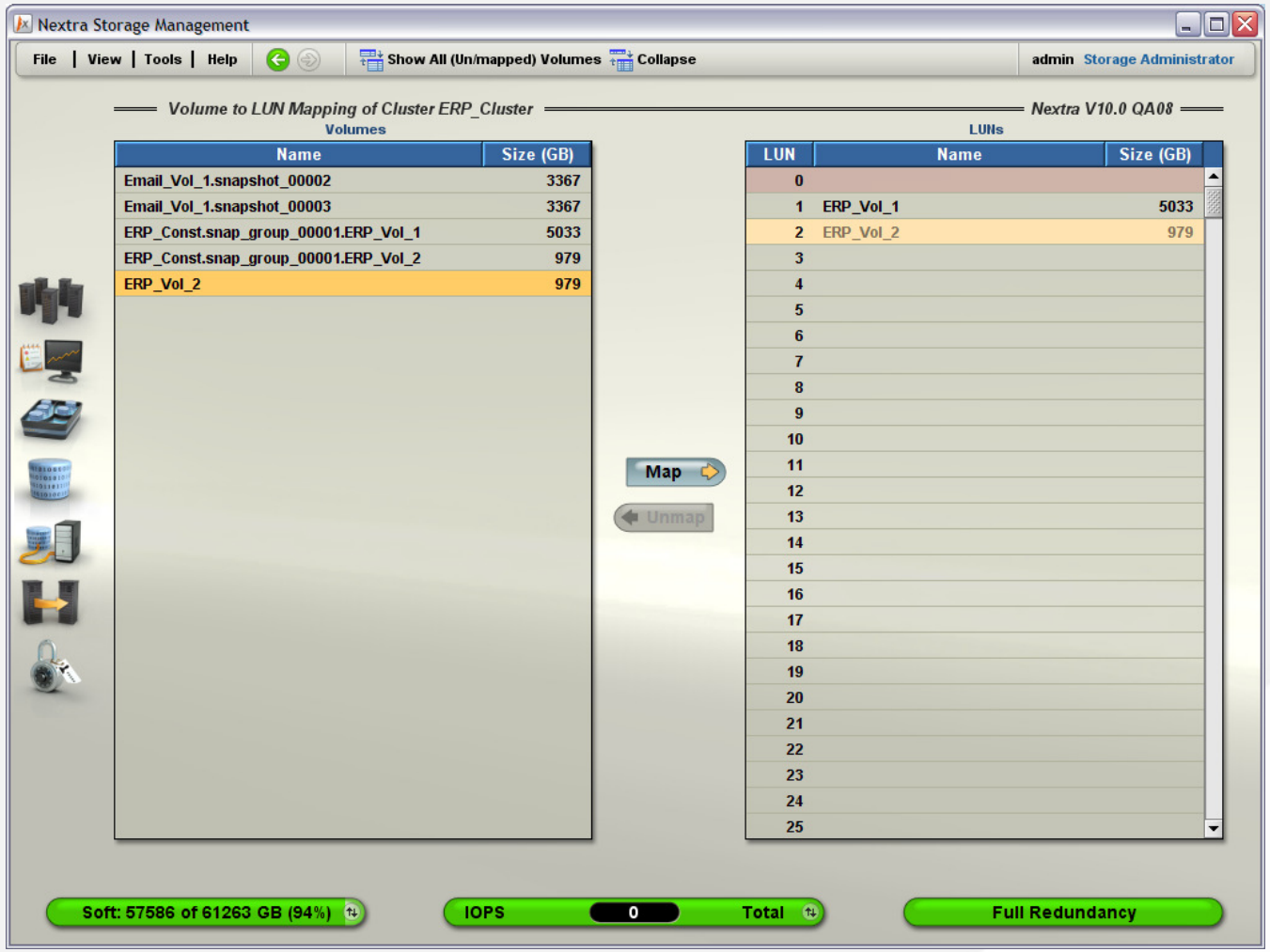


IBM XIV Storage Simple Intuitive Management example: Creating a Volume



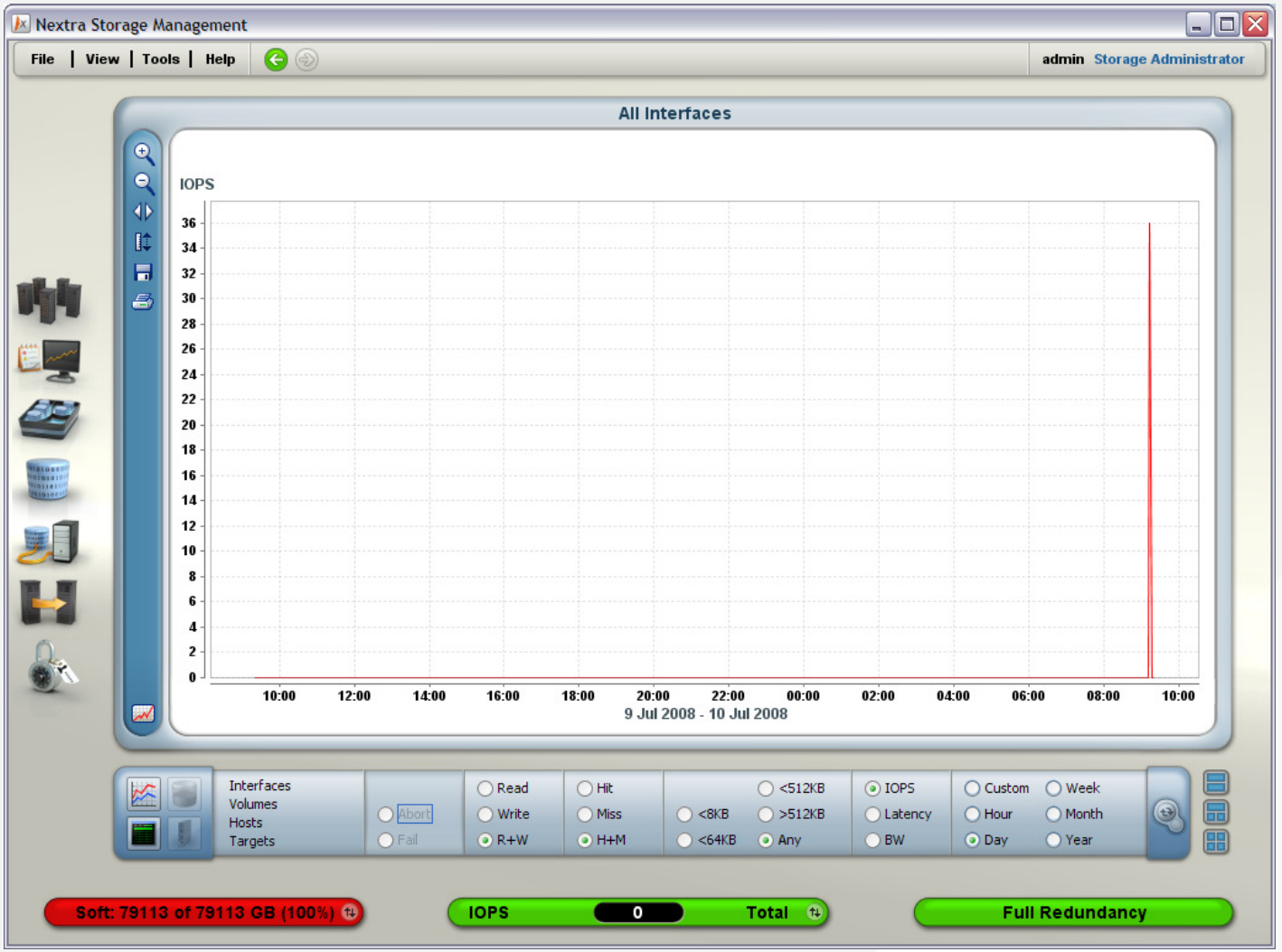
- Used capacity is always known !

IBM XIV Storage: Volume to LUN Mapping

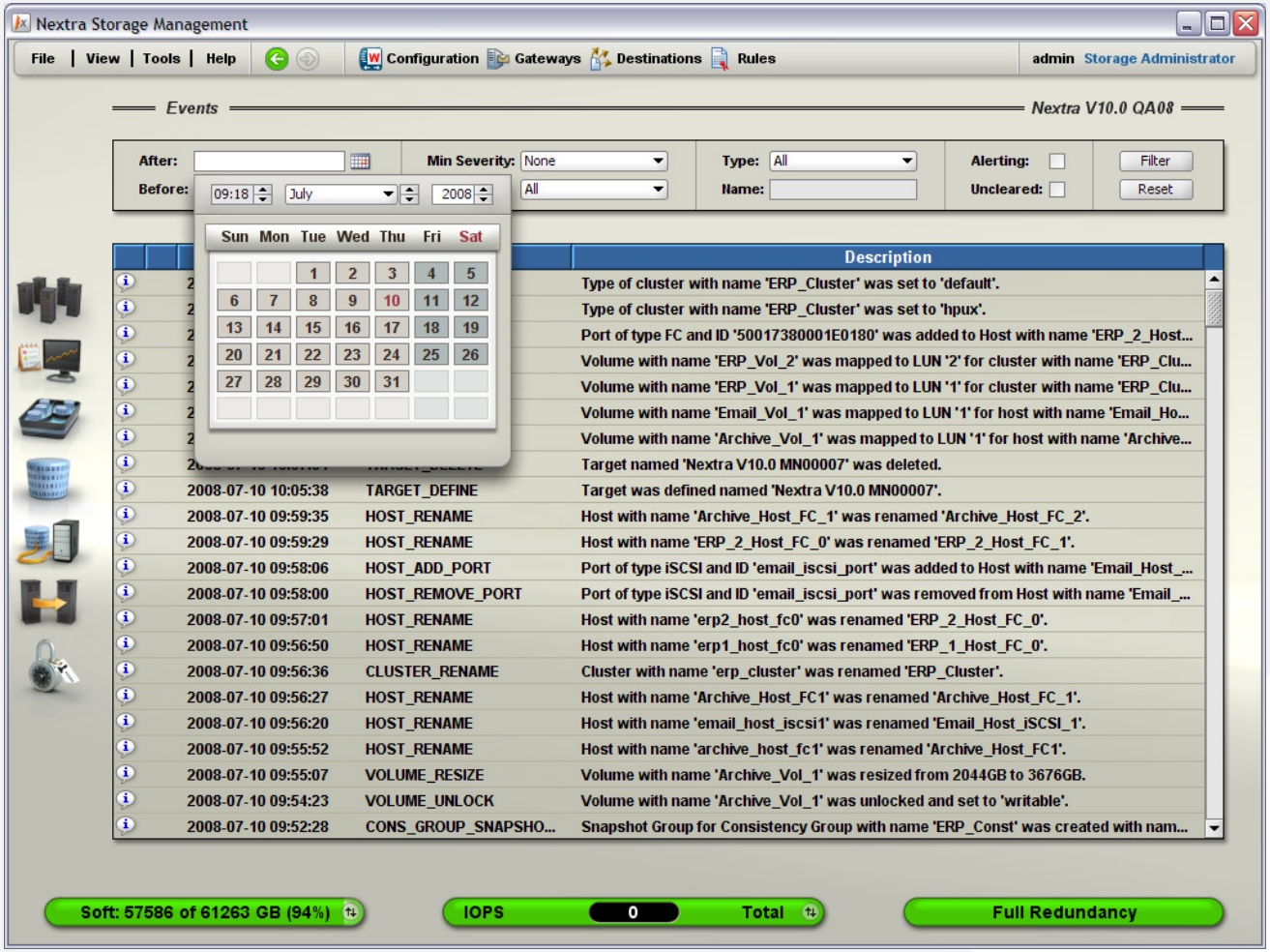


IBM XIV Storage Consistency Groups

IBM XIV Storage: Monitoring



IBM XIV Storage: Events Log



Conclusion

The Bottom Line: Real-World Benefits

- **Reliability**
  - Revolutionary self healing takes minutes, not hours
- **Functionality**
  - Thin provisioning and replication built into the architecture
- **Power and Space**
  - Minimize power, cooling and floor-space with SATA drives
- **Performance**
  - Massive parallelism, spindle utilization, self-healing and cache effectiveness boost performance dramatically
- **Manageability**
  - Simple, easy management; a logical volume has only two parameters: name and size
- **Cost**
  - Off-the-shelf components
  - No charge for software features (Snap, DR, Management)

Resources

- Internal IBM XiV saleskit (updated August 15)
  - http://w3-03.ibm.com/sales/support/ShowDoc.wss?docid=P314034Y41267Q84&infotype=SK&infosubtype=S0&node=doctype,S0|doctype,SKT|brands,B5000|geography,AMR&appname=CC_CFSS
- Partnerworld saleskit (updated August 15):
  - http://www.ibm.com/partnerworld/wps/servlet/ContentHandler/ssIBMXIVsk.skit
- Project office can help in finding XIV resources or contacts:
  - Send note to ASKXIV/Raleigh/IBM or askxiv@us.ibm.com
- XIV external web site:
  - www.xivstorage.com
- Interoperability and ISV support, will be available in the System Storage Interoperation Center (SSIC)
  - http://www-03.ibm.com/systems/support/storage/config/ssic/displayesssearchwithoutjs.wss

Thank You

Disclaimer

This publication is provided “AS IS.” IBM product information is subject to change without notice.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This information could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or programs(s) at any time without notice. The information provided in this document is distributed “AS IS” without any warranty, either express or implied. IBM EXPRESSLY DISCLAIMS any warranties of merchantability, fitness for a particular purpose OR INFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided.

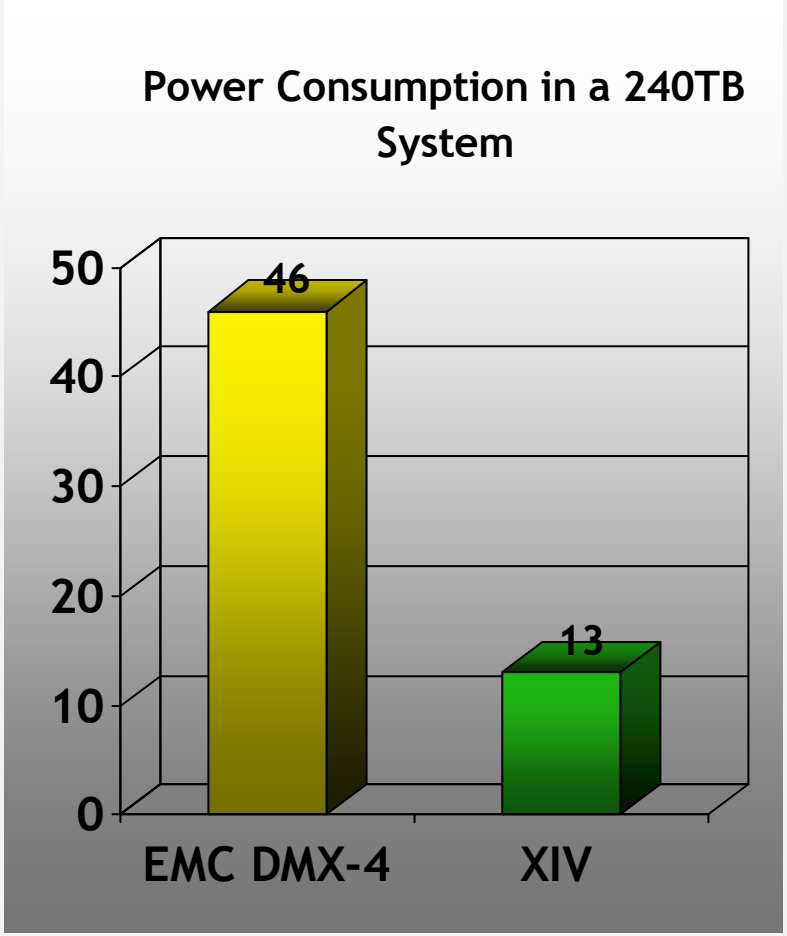
Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products.

IBM makes no representations or warranties, expressed or implied, regarding non-IBM products and services, including those designated as Server Proven.

IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of the specific Statement of Direction.

XIV Green IT: Dramatically Lower Power Consumption



- Approximately **1/4 of the power consumption per TB**, when compared to an equivalent tier-1 system
- XIV with 1TB drives; EMC DMX-4 with 146GB drives
- Future - 2TB drives will double power efficiency

Source: http://www.emc.com/collateral/hardware/specification-sheet/c1166-dmx4-ss.pdf

IBM XIV Storage System Unique Architecture

- **Virtualized grid storage**:
  - Massive parallelism
  - Data distribution across all drives
  - Data fully mirrored
  - No RAID groups to manage
- **Automatic load balancing**
  - Consistent performance
  - No manual intervention
- **Thin Provisioning** (over allocation)
- **High Performance**, flexible **Snapshots**
- **Remote Replication**
- **Intuitive Graphical User Interface (GUI)**
- Built in Data migration

IBM XIV Storage