



# HP CCN Global File Systems

Inaugural Meeting  
March 3<sup>rd</sup> 2003

**Tim Reddin**  
**Frank O'Neill**  
**Ken Tilton**

© 2004 Hewlett-Packard Development Company, L.P.  
The information contained herein is subject to change without notice



# Agenda

- Introductions All
- HP CCN Ken Tilton
  - Forthcoming face to face meeting
- Brief History of HP & Lustre Tim Reddin
- V1.0 Product goals & description Frank O'Neill
- Beyond V1 Tim Reddin
  - Issues & Opportunities
- Discussion All
  - What do you want?

# Introductions



# HP CCN – Forthcoming Meeting

- May 26<sup>th</sup> -27<sup>th</sup> Urbana Champaign

# HP & Lustre a brief history

- **1999:** Seagate fund research at CMU
- **2000:** DOE TriLabs RFP for Scalable Global Parallel File System
  - Response from HP Labs, CFS, Intel, others
- Labs preferred CFS technology, requested HP to join as major systems vendor
- **2001:** Contract negotiated & signed by Winston Prather
- **2002:** HP- Compaq Merger
  - Winston is the new VP of HPTCD

# DOE Program and HP's Role

- Program is code named "Hendrix"
- HP is prime contractor
  - CFS and Intel sub contractors
  - CFS is main technology supplier
  - HP responsible for overall program management
  - HP engineering working on Hendrix development
  - HP responsible for verification
- Program goal is to demonstrate phased capability
  - Not to deliver product
- CFS working on other programs as well

# HP's Lustre organisation: Two teams

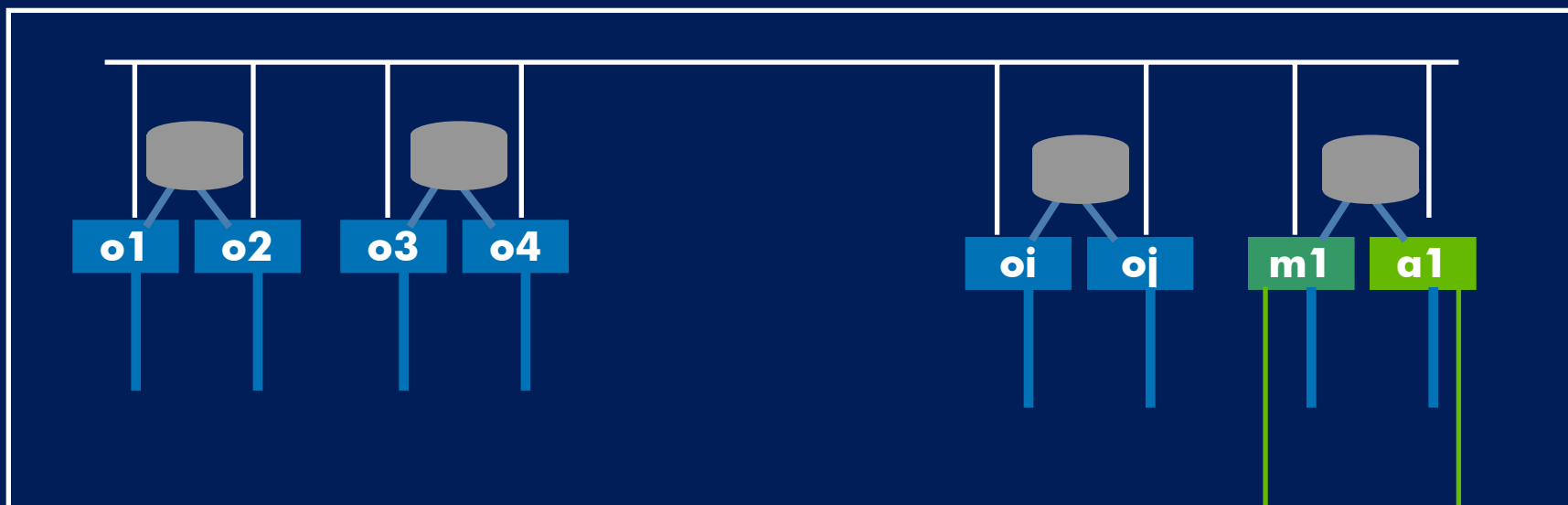
- DOE Program team
  - “Hendrix”
  - Program Management
  - Development
  - Testing
- Lustre Product team
  - “Jimi”
  - Deliver Lustre based storage solution for HPTC
  - Long term migrate back to storage organisation

# Lustre based storage solution

- Goals

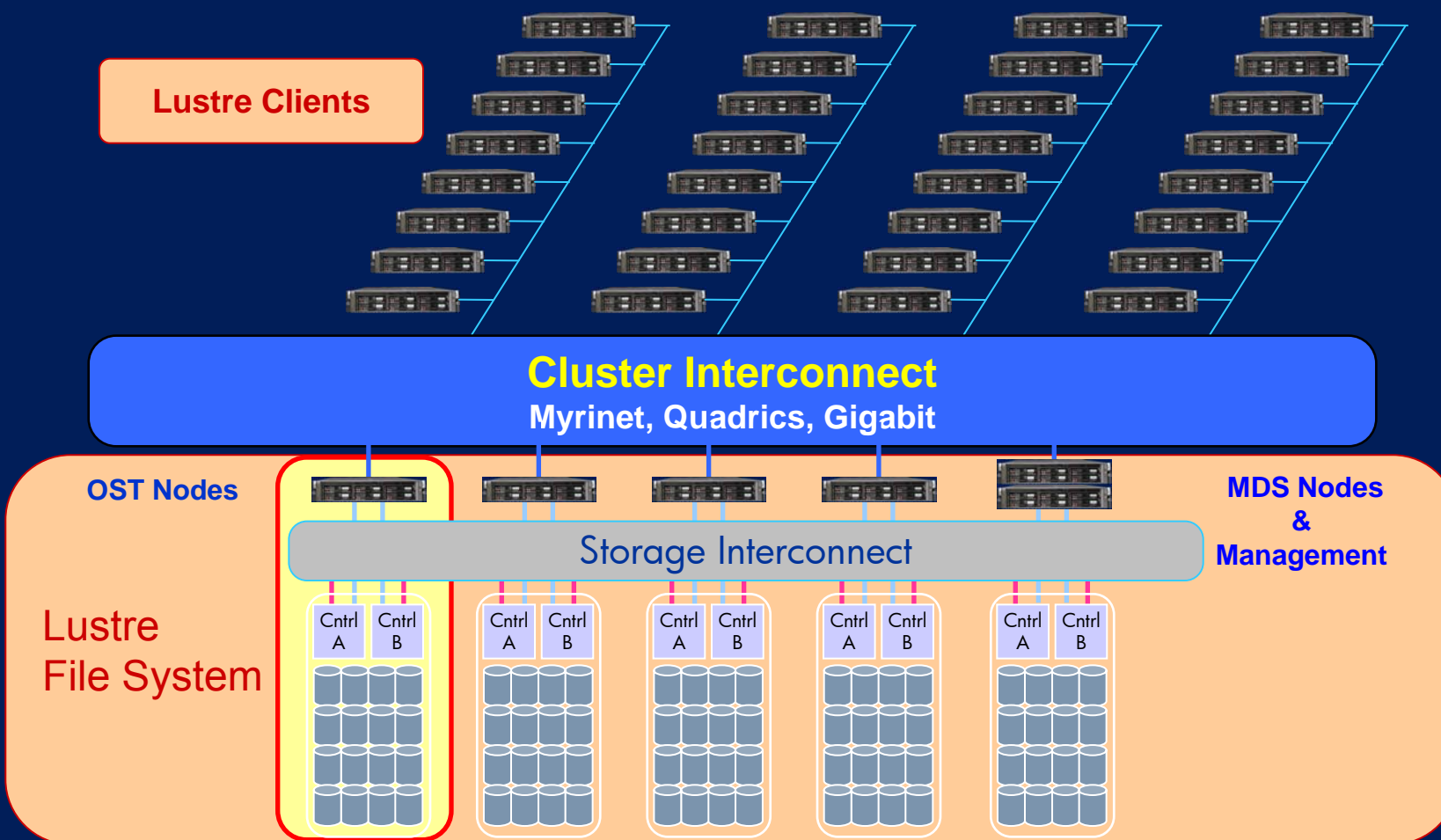
- Deliver Lustre capability using storage appliance model
- Deliver integrated HW & SW solution
- Managed appliance
- Supported, balanced product
- Drive low cost, scalable performance

# Basic Product Architecture



- Portals/ Lustre protocols
- SSH, https for external management
- IP – internal management

# Lustre Cluster





# Jimi V1.0 Core Features & Functions

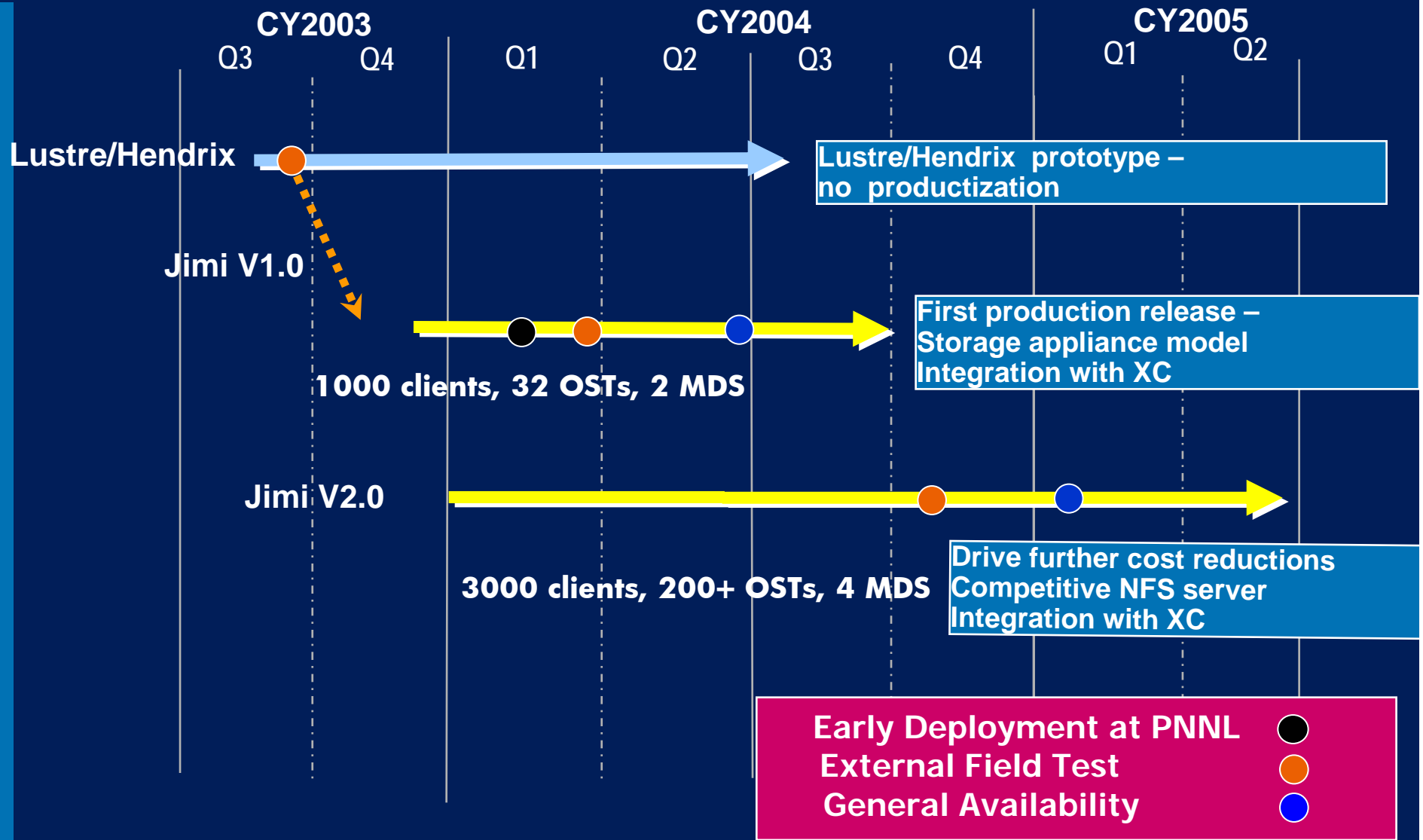
Functionality	Jimi V1.0
Linux Kernel	2.4.20 2.4.21
Clients Supported	1000
OSTs Supported	32
MDSs Supported	2
<b>File System</b>	
File data coherency	●
Meta data coherency	●
<b>Networking &amp; Interconnects</b>	
NFS export	●
TCP networks	●
Gigabit Ethernet	●
Elan3 (restricted to XC V1.0 accounts)	●
Elan4	●
Myrinet XP, 2XP	●

Functionality	Jimi V1.0
<b>High Availability</b>	
No single point of failure	●
Disaster recovery tool (lfsck)	●
<b>Performance &amp; Scalability</b>	
I/O optimizations for large files	●
MetaData scalability	●
<b>Management</b>	
OST addition/removal	●
LDAP Management	●
SNMP	●
<b>Standards</b>	
POSIX compliant	●
<b>Security</b>	
POSIX security compliant	●

# HP Value Add

- Base software: Lustre 1.2.x
- Possible support of additional Lustre options :
  - Myrinet, LFSCK, NFS performance, POSIX locking.
- Solution Management
  - Storage discovery and configuration
  - Lustre configuration
  - Software administration of OST/MDS nodes
  - Hardware monitoring
- Qualified Building Blocks
- Standard HP product infrastructure
  - Documentation, training, support

# Lustre POR Roadmap



# Beyond V1

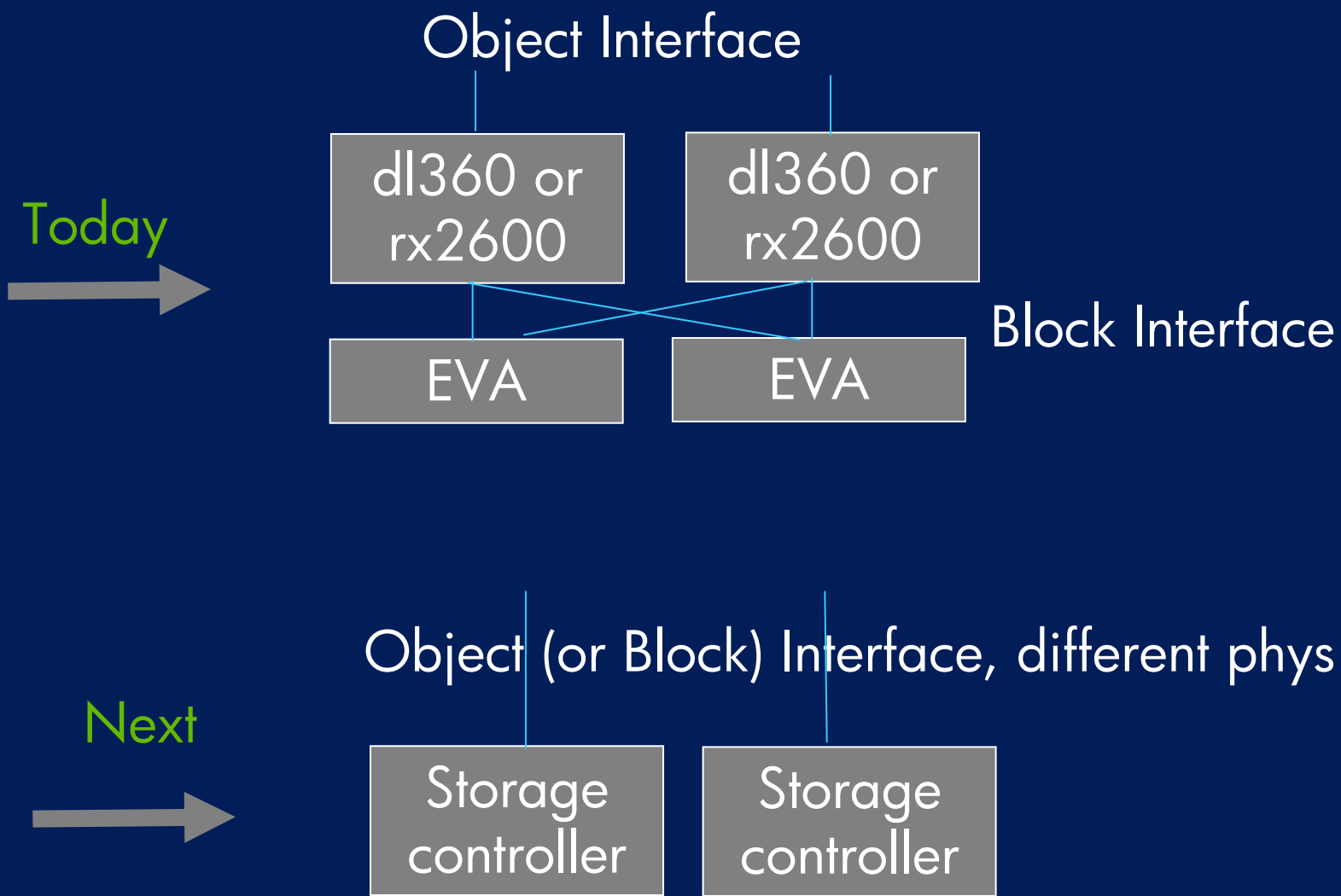
- Goals

- Integrated Lustre Storage node
  - e.g. OST appliance
- XC Integration
- Feature enhancement, e.g.
  - Bandwidth on Demand/ Quality of Service
  - Lustre for general purpose file system use
  - Good NFS support
  - HSM
- What are your priorities?

# Integrated Lustre storage node

- Two trends in storage controller hardware
  - 1. Becoming more powerful
    - Faster CPUs, buses etc
    - External connectivity
      - PCI-X, PCI-Express
      - Supports iSCSI, but also opportunity for object interfaces
  - 2. Possible use of standard low cost servers as storage controllers
    - Fuelled by low cost, faster commodity processors
    - Standard interconnects
      - 10 GigE, RDMA, Infiniband
      - iSCSI
    - Makes it possible to consider riding the commodity curve
      - Instead of expensive NRE

# Reduce \$/TB by eliminating HW



# Which form of Integration is best?

- Lustre/Controller
  - Reduce cost
  - Improve performance
  - Drive OST appliance model & increase acceptance
    - Need more capacity, throughput?
    - Plug another OST onto your network
- OST/Cluster
  - Integrates OST node into cluster fabric
  - One management domain
  - Use OST nodes for computes for Grand Challenge jobs
  - (Restricts OST node type)

# Discussion

- What are your expectations for the HP CCN?
- Topics for future meetings
- Meeting Frequency?



i n v e n t