

# Gelato Federation Meeting Stockholm, Sweden

October 15, 2003



## A Lustre product for HPTC

Tim Reddin (tim.reddin@hp.com)

Senior Member of Technical Staff

Steve Rowan (steven.rowan@hp.com)

Engineering Manager

High Performance Technical Computing Division

# Topics I will cover



- Why we plan a to develop a Lustre product
  - also HP's relationship with CFS & Lustre
- Target Market
- HP's value proposition
- Product Overview

# Why Lustre?



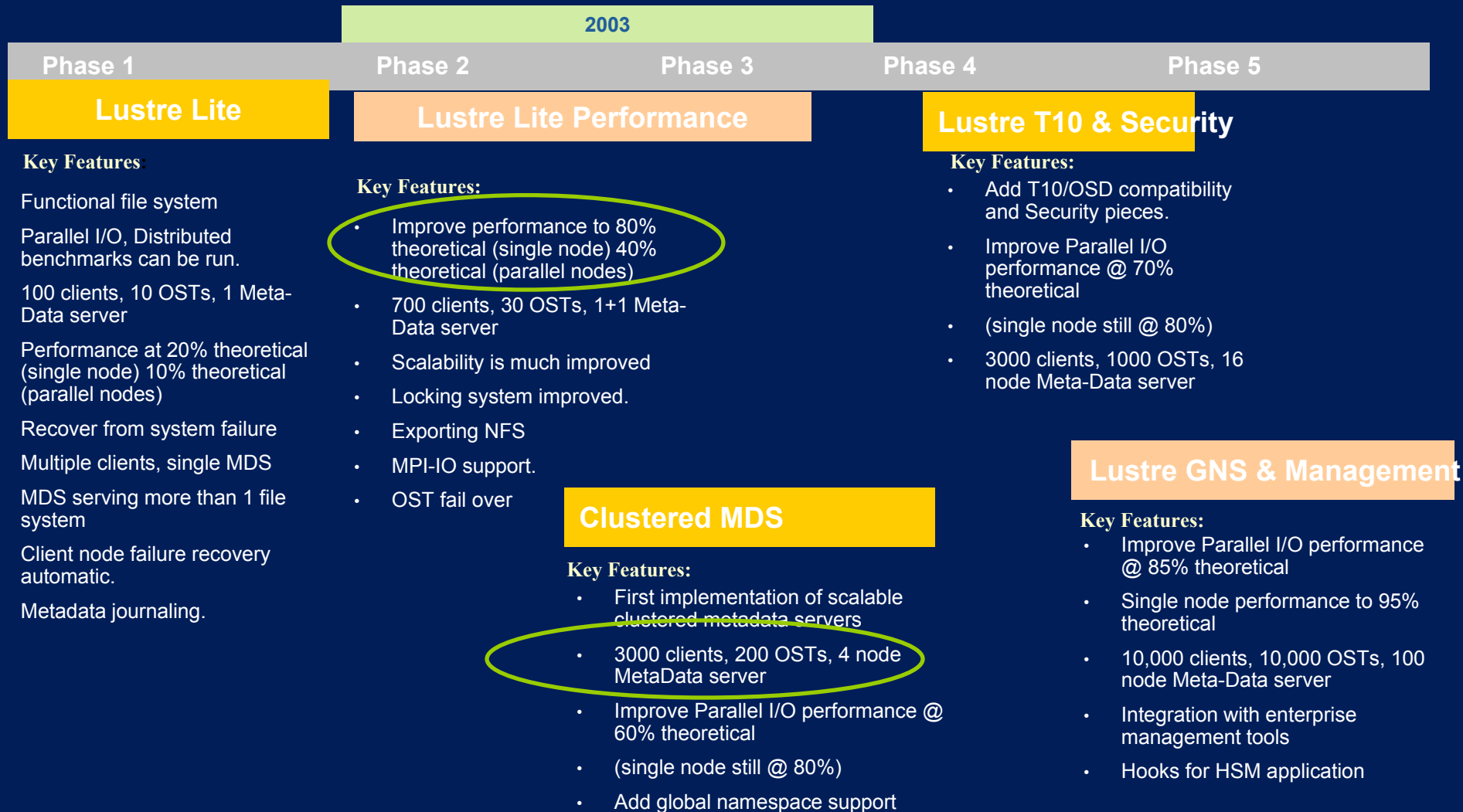
- Pre-existing history with CFS & Lustre
  - DOE PathForward contract (Hendrix)
  - HP – Compaq Merger
    - Winston Prather VP of HPTCD signs contract
  - Lustre seen as solution for Linux scale out
  - Coincides with transition from proprietary
    - AlphaServer SC / Tru64Unix
      - Migrates to
    - XC – IA64/IA32 / Linux

# HP Lustre projects

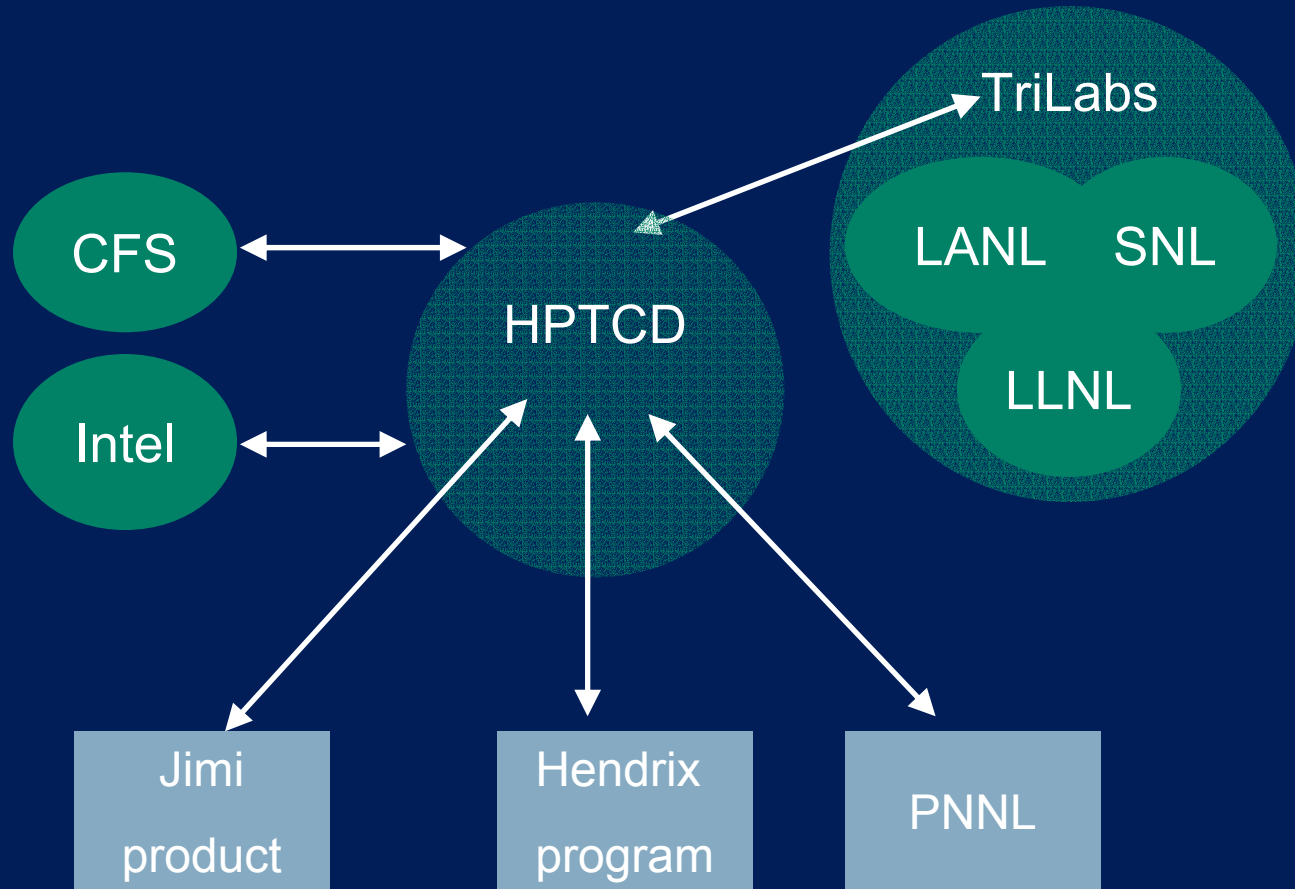


- **Hendrix:** DOE/PathForward program
  - HP is prime contractor
  - CFS Lustre technology supplier
  - Intel – performance tools
  - Phased deliverables
  
- **Jimi:** HPTCD product solution using Lustre
  - Targets next phase of Hendrix: LLP

# Lustre/Hendrix (phased development)



# Lustre/Hendrix Eco System



# Target market – HPTCD product



- Focus is the clustered super computer solution
- Integrated with target system at the level of system interconnect
  - HP's XC product
  - Integration into other clustered supercomputing solutions
  - Possible software and services solution
- All targeting the large scale customer
  - Sci/Tech
  - Life Sciences (small files)

# HPs Value Proposition



- Product Quality Lustre
  - Production level QA
  - Performance characterisation
  - Qualified HW building blocks
  - Installation & management toolset
  - Support & services
- Experienced product team
  - Concept to product experience
  - File systems
  - Performance and scalability
  - Clustered supercomputing

<b>Test Classes</b> <b>Test Levels</b>	<b>Normal input tests</b>	<b>Performance &amp; timing</b>	<b>Scaling &amp; stress tests</b>	<b>Interoper-ability</b>	<b>Fault injection tests</b>	<b>Extended period tests</b>
<b>Installation, Reinstall, Verification</b>						
<b>Configure &amp; Reconfigure</b>						
<b>Mount &amp; Unmount</b>						
<b>Usage &amp; Standards Metadata manipulation</b>						
<b>Usage &amp; Standards Locking</b>						
<b>Usage &amp; Standards I/O</b>						
<b>Usage &amp; Standards Other</b>						
<b>Monitoring &amp; Management</b>						
<b>Failover, Failback &amp; Recovery</b>						

Architecture

IA-32 with GIG-E

IA-64 with Quadrics

IA-32 Myrinet

Test

	Normal input	Performance & timing	Scaling & stress	Interoper-ability	Fault injection	Extended period tests
Installation, Reinstall, Verification						
Configure & Reconfigure						
Mount & Unmount						
Usage & Standards						
<b>Monitoring &amp; Management</b>						
Failover, Failback & Recovery						

Test Intensity

- Component Evaluation
  - Storage Subsystems
  - HBAs
  - Server node utilisation
  - Interconnect
  - Enable the construction of balanced servers
    - Preserve investment
- Facilitate performance problem diagnosis
- System performance
  - Predictability
    - N to N, N to 1, 1 to 1, meta data churn
  - Hardware building blocks

# Performance & Scalability



- Scalable management
  - Installation
  - SW configuration
  - Booting
  - Storage configuration

# Installation and management



- Drop in and use philosophy
- Installation
  - Automated installation of MDS and OST server nodes
  - Self discovery of Fibre Channel configuration
    - Automated configuration of FC controllers and LUN creation
  - Automatic generation of Lustre “LMC” configuration
  - Functional and performance sanity checks
- Management
  - Server SW configuration
  - Health monitoring
  - Performance monitoring
  - Storage monitoring

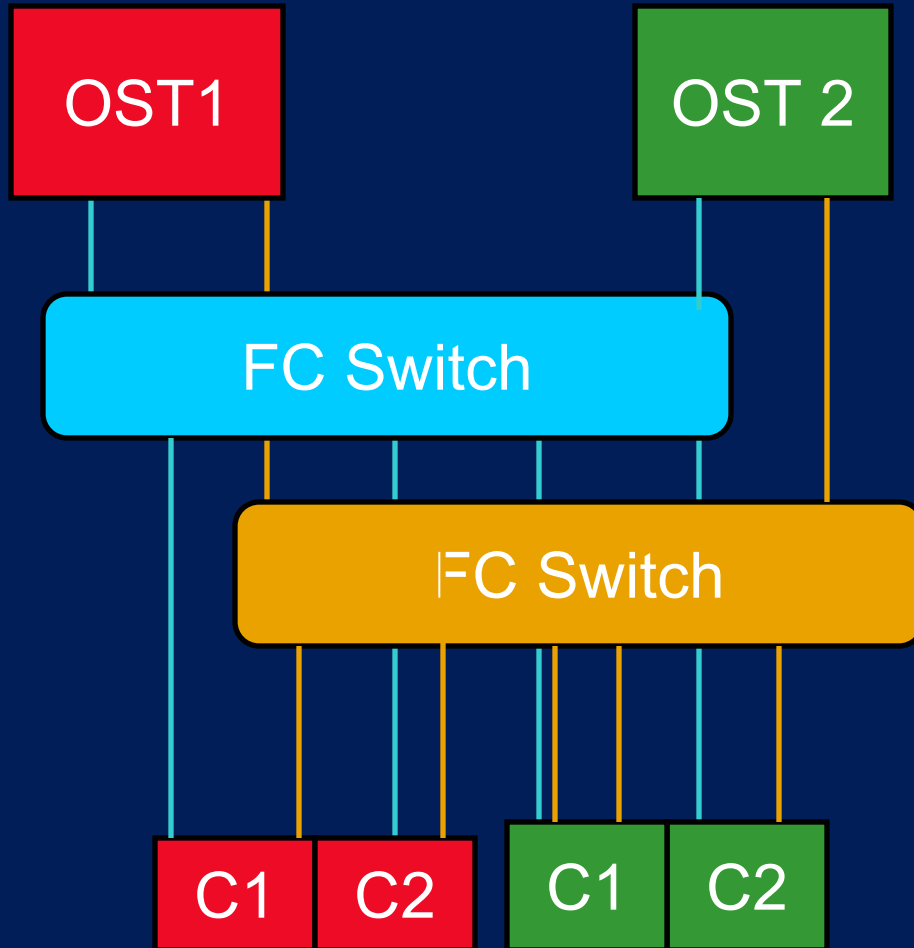
- Standard product support infrastructure
  - Remedial call center support
    - Lustre Engineering
      - LOSL (HP's Linux org)
      - Cluster File Systems
  - On-site hardware support
- Professional services
  - Installation
  - Training
  - Consultancy / knowledge transfer

# Product Overview



- Basic components:
  - 32 and 64 bit server nodes
    - Client – server interoperability
  - Interconnect
    - Elan, Myrinet & GigEther
  - EVA Storage
  - 2Gb Fibre infrastructure
  - HW Management

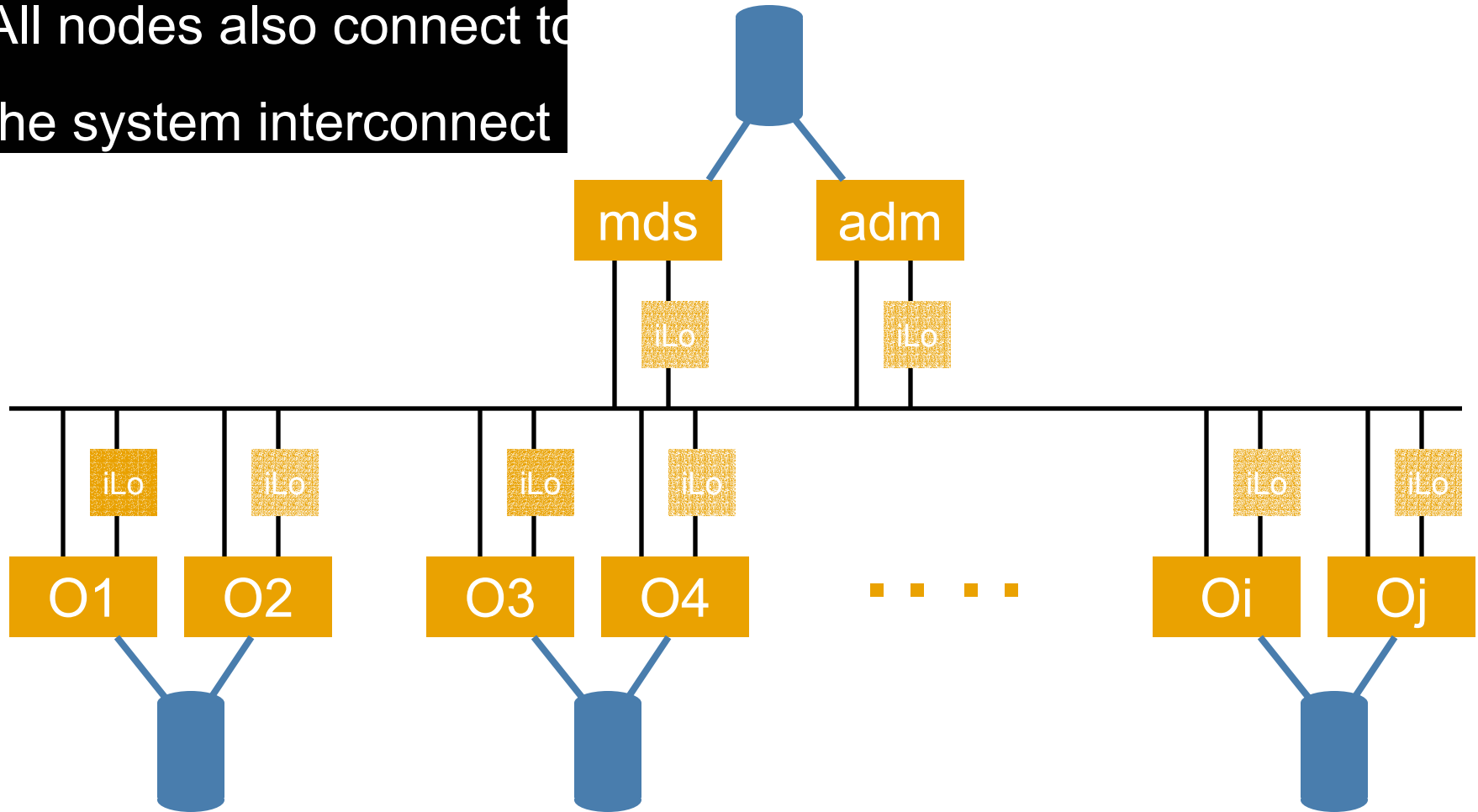
# OST Server pair



# Inside the Jimi server



All nodes also connect to the system interconnect



- OST server nodes
  - Pair wise connected
  - Clumanager for failover
  - Peer storage visibility
    - Pre-configured building blocks
  - Management processor
    - iLo or MP
  - Minimum OS configuration
- MDS nodes
  - Pair wise with admin node
- Admin node
  - Configuration and management utilities
  - SNMP

# Summary



- Initial product goal for Q204
- Focus on qualification and performance characterisation for V1.0
- Installation and management capabilities
- Full product
  - Support
  - Service
  - Training



**i n v e n t**