FUJITSU

Fujitsu's Vision for High Performance Computing

October 26, 2004

Yuji Oinaga
Fujitsu Ltd.

CONTENTS

FUJITSU

- HPC Policy and Concept
- Current HPC Platform
  - HPC2500
  - IA-Cluster
- Major HPC customers
- Road Map
- Toward Peta-Scale Computing

FUJITSU HPC Policy

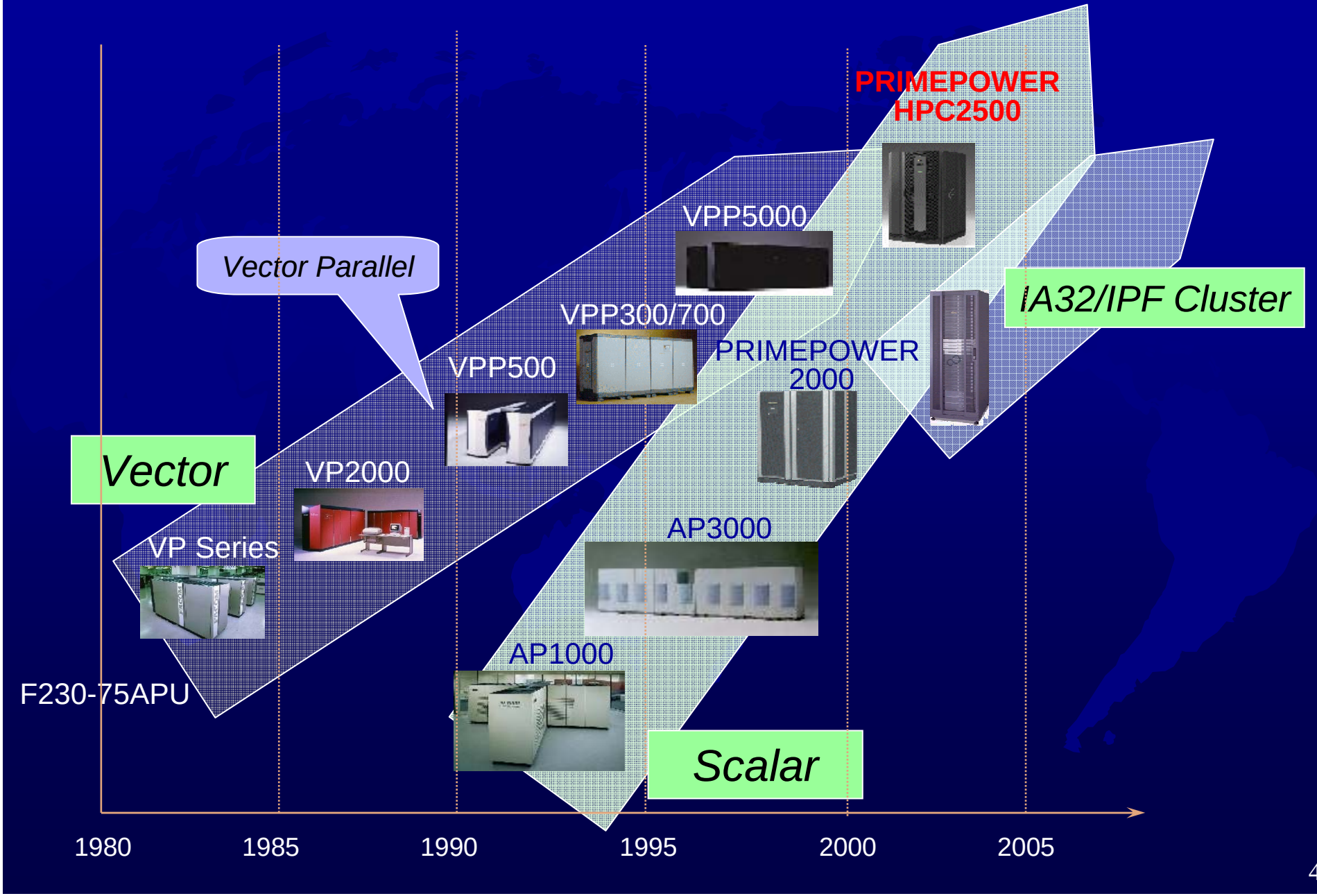
FUJITSU

~ Developing & Providing Top Level Supercomputer ~

- High Computational Power
- High Speed Interconnect
- Reliability/Availability
- Leading Edge Semiconductor Technology
- Highly Reliable, HPC Featured Operating System & Middleware
- Sophisticated Compilers and Development Tools

History of HPC Platform

FUJITSU



Concept of Fujitsu HPC

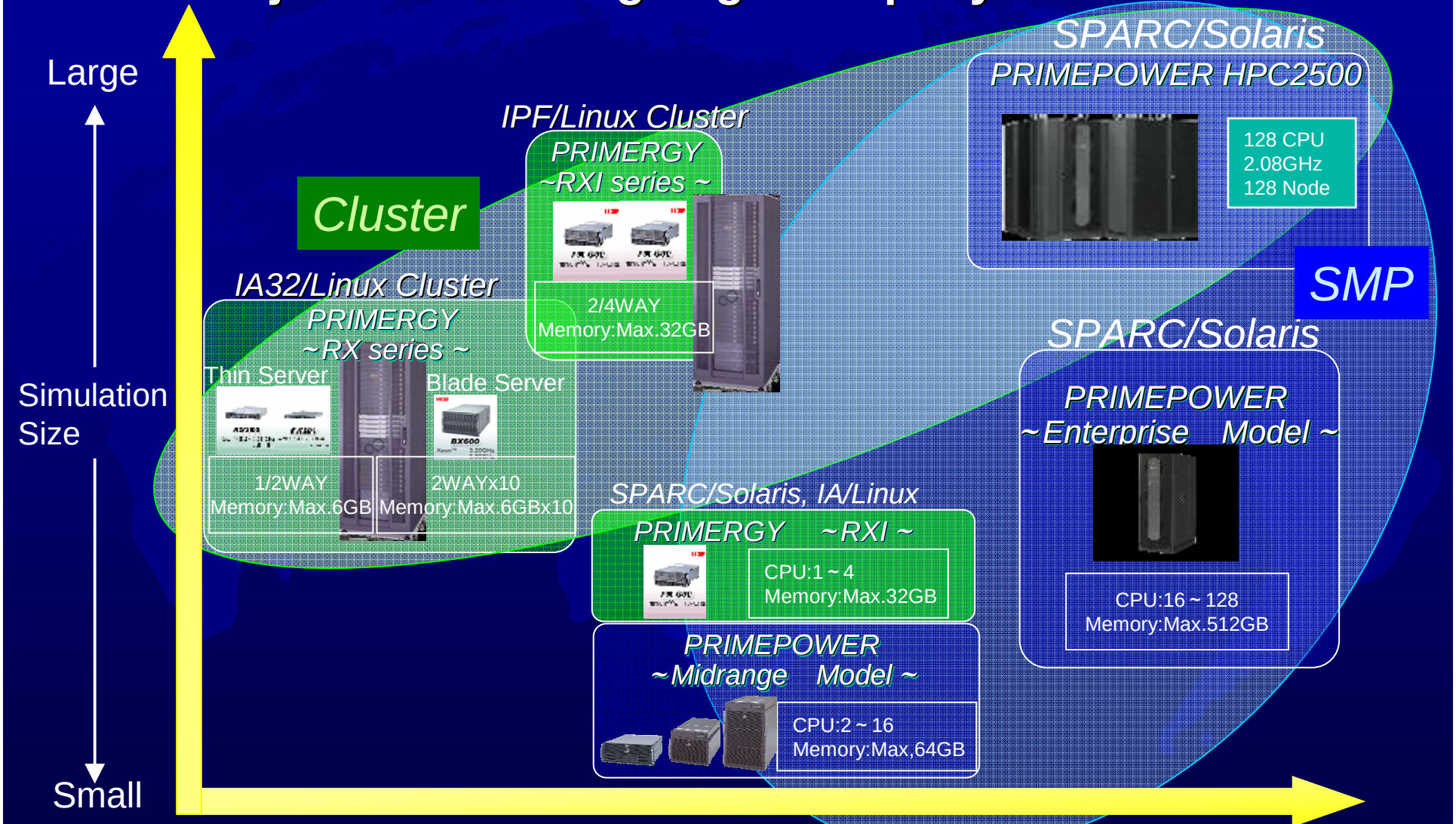
FUJITSU

- **Provide the best/fastest platform for each applications.**
  - -Scalar SMP : PRIMEPOWER, New Linux SMP
  - -IA Cluster : PRIMERGY Cluster
- **Provide the high speed interconnect.**
  - -Crossbar : High Speed Optical Interconnect
  - -Multi-Stage : Infiniband
- **Provide High Operability/Usability**
  - -Enhancement of Operability/Usability :
    - Extended Partitioning, Dynamic Reconfiguration
  - -Cluster Middleware :
    - Total Cluster Control System (Parallelnavi)
    - Shared Rapid File System (SRFS)

Fujitsu's Current HPC Platform

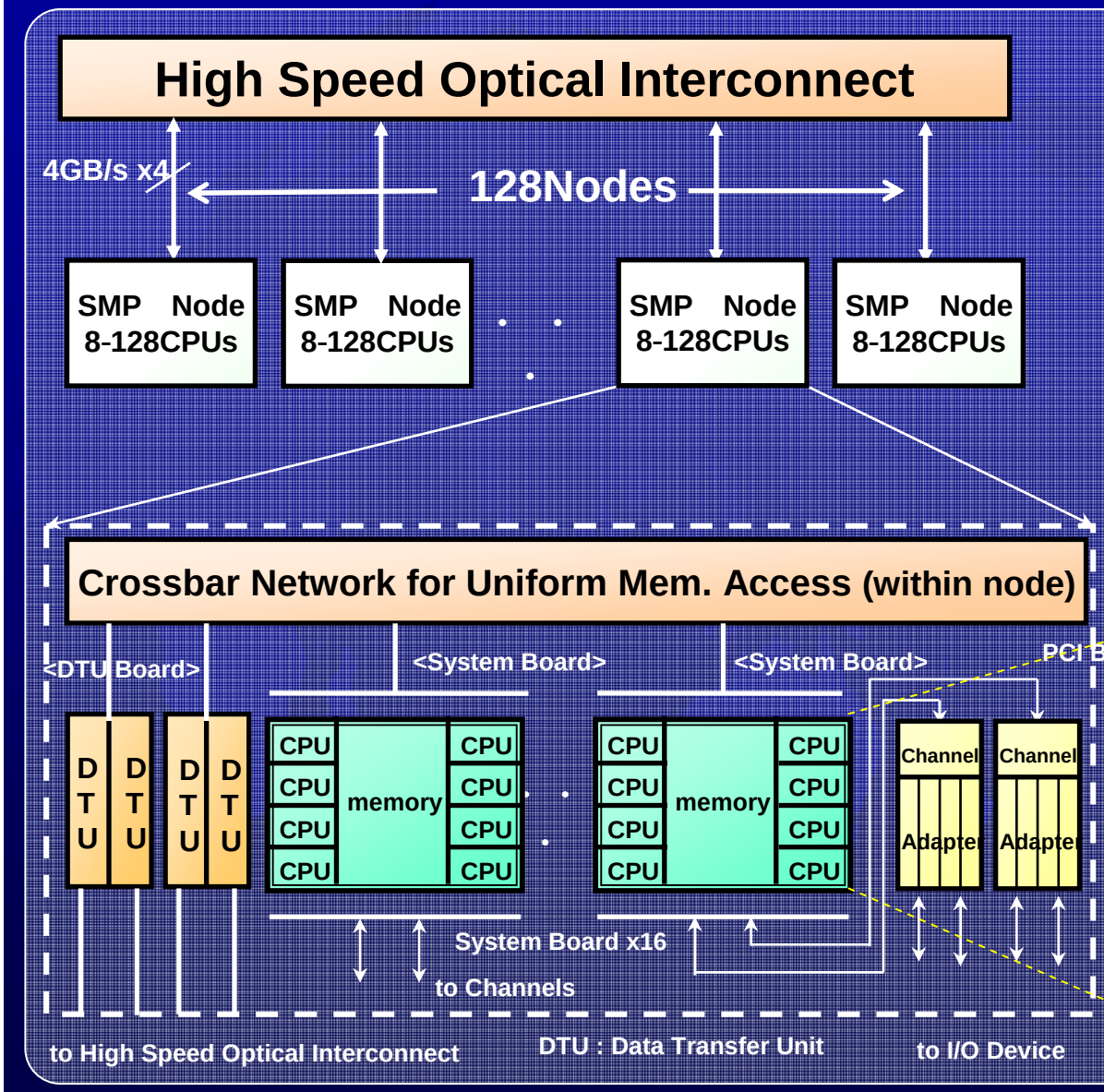
FUJITSU

~ Fujitsu is a leading edge company in HPC field ~



PRIMEPOWER HPC2500: Configuration

FUJITSU

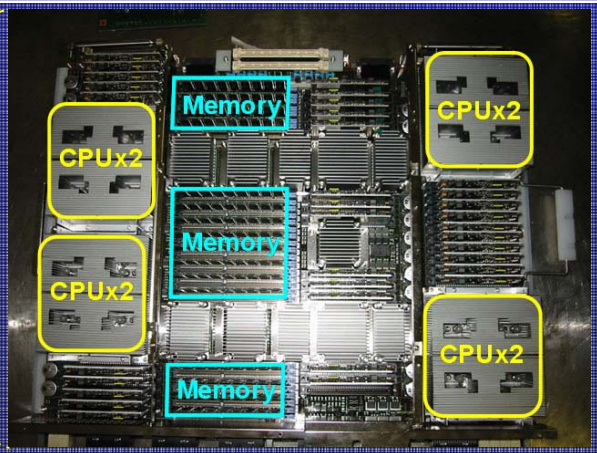


[System]

- -# of Node : Max.128
- -# of Processor : Max.16,384
- -System Performance : Max.136Tflops
- -Memory Capacity : Max. 64TB

[Node]

- -CPU Performance: 8.32Gflops(2.08GHz)
- -# of Processor : Max.128
- -Node Performance: Max.1064.96Gflops
- -Memory Capacity : Max. 512GB
- -High Speed Interconnect : 4GB/s x 2(in/out) x4



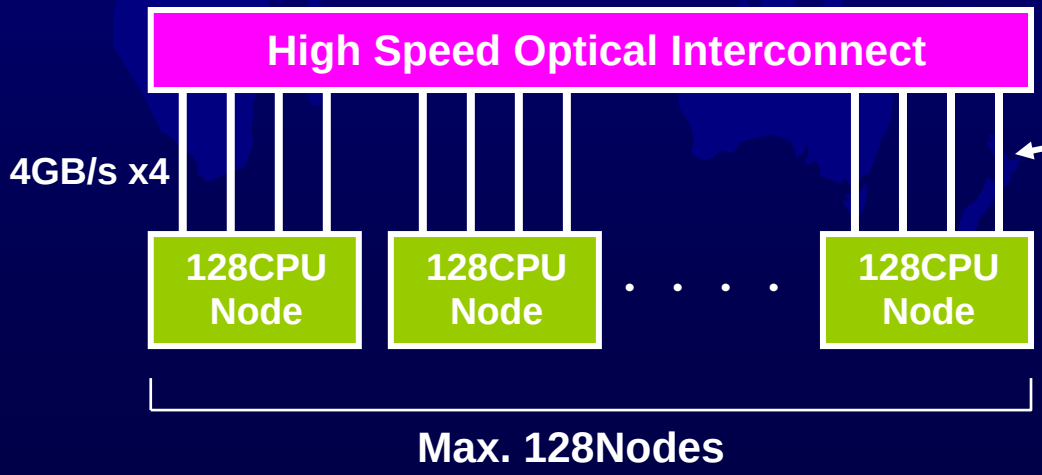
System Board

PRIMEPOWER HPC2500: Interconnect

FUJITSU

~ Parallel Optical data transfer technology for higher scalability and performance ~

- Crossbar Architecture
- Connects up to 128 nodes(16384 CPUs)
- Realizes 4GB/s x4 data throughput for each node
- Allows hundreds of node cabinets to be placed freely with 100m optical cables



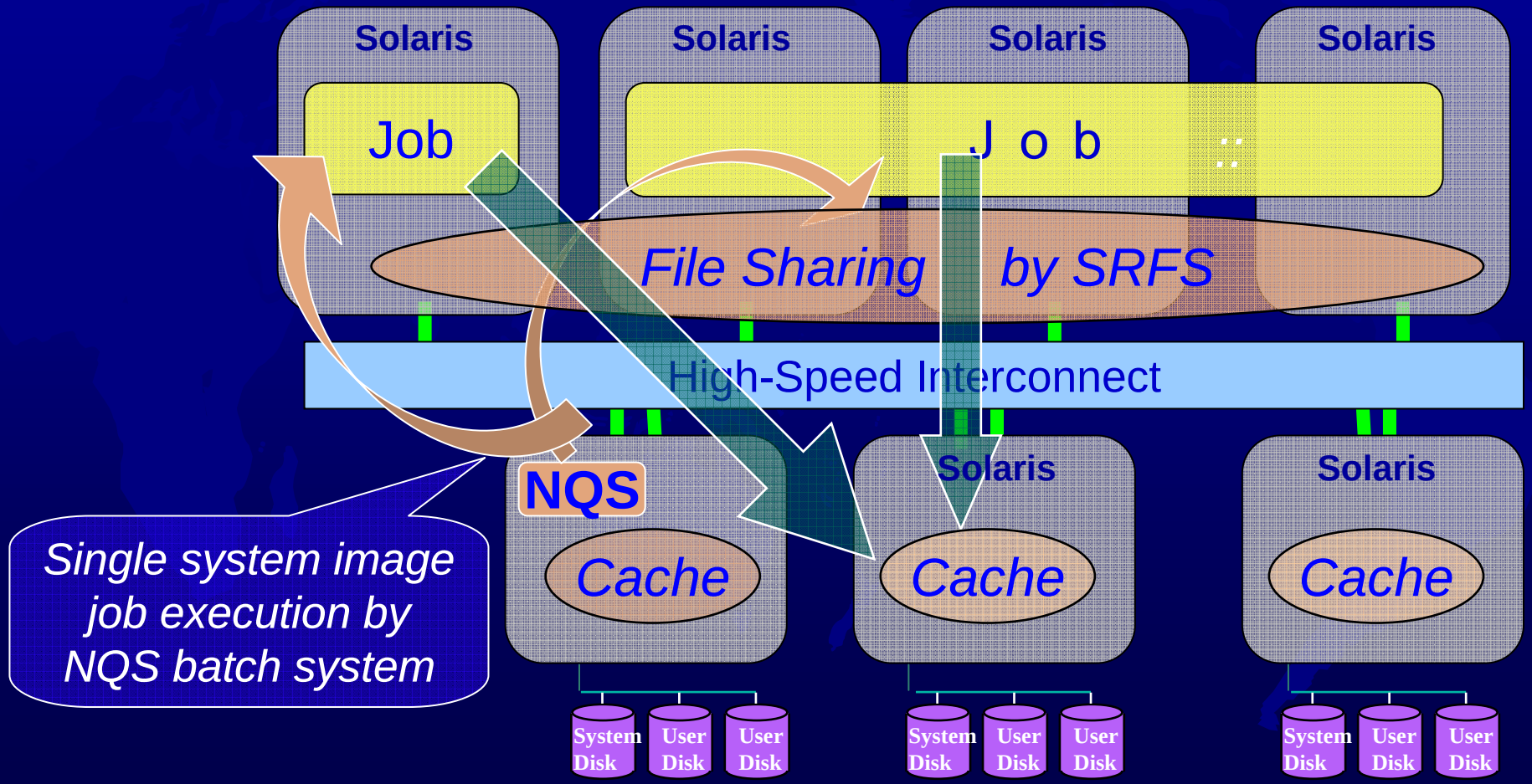
PAROLI Module

PAROLI:PARallel Optical LInk

ParallelNavi : Cluster Middleware

FUJITSU

Single system image for Job execution / File access

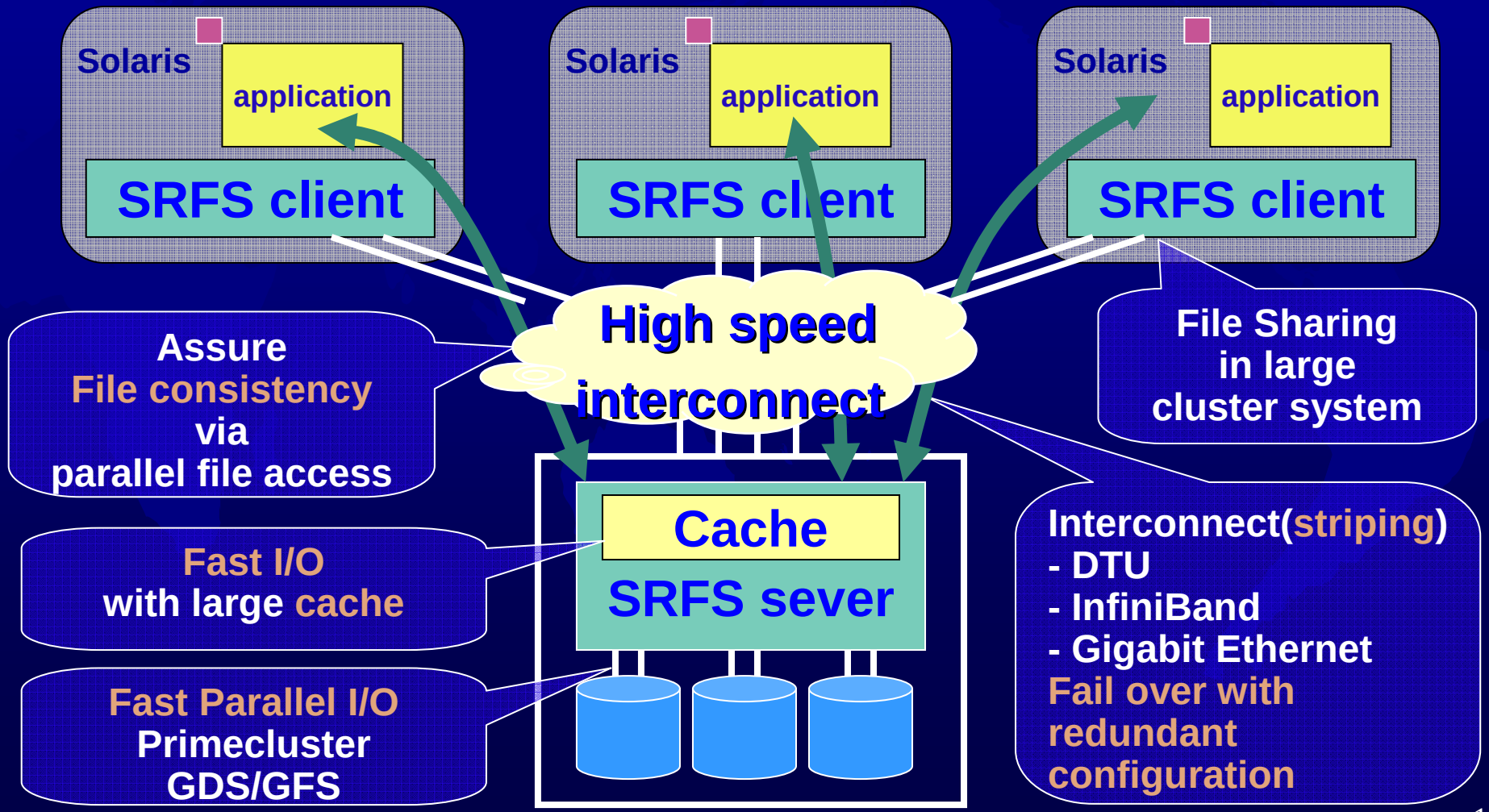


SRFS(Shared Rapid File System)

FUJITSU

File Sharing in large cluster system

⇨ ***High performance/Single system image file access***

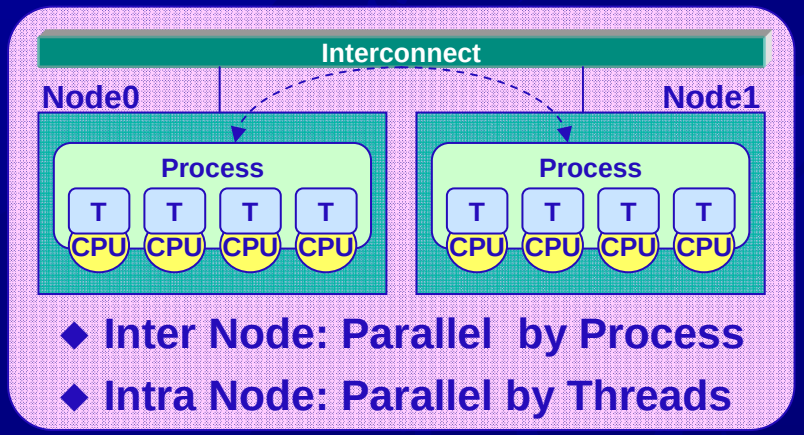


Language System

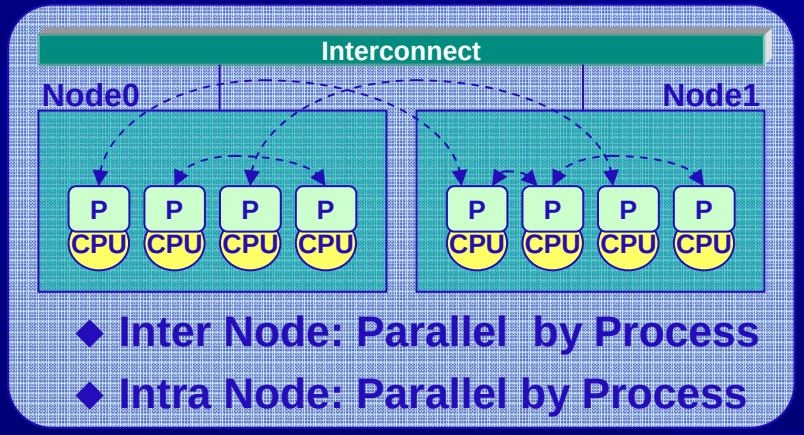
FUJITSU

- Programming Model on Cluster System

[Hybrid Model]



[Flat Model]



P : Process T : Thread Inter-Process Communication

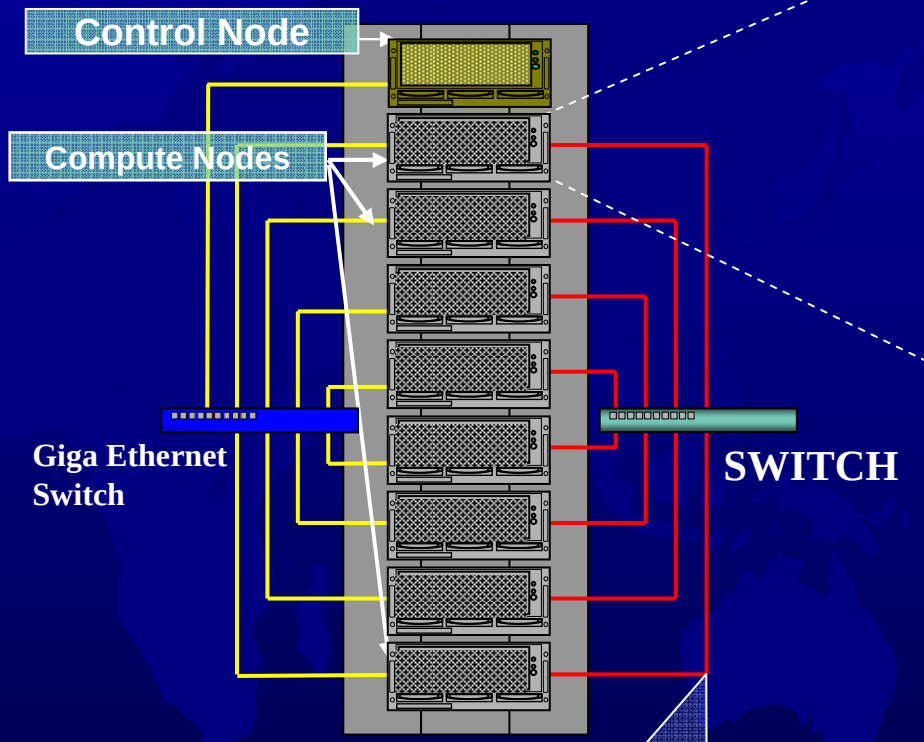
- Fujitsu Language System for Cluster

Programming Model/ Process Parallel Method		Inter-Node	Intra-Node
Hybrid Model	Data Parallel	HPF, XPF (Fujitsu)	OpenMP Compiler Auto Parallelization
	Message Passing	MPI	
Flat Model	Data Parallel	HPF, XPF (Fujitsu)	
	Message Passing	MPI	

IA-Cluster: Configuration

FUJITSU

●System Configuration

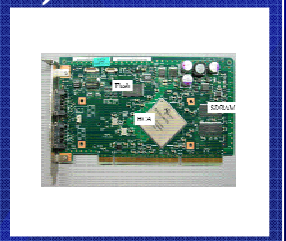


● Compute Node

- FUJITSU PRIMERGY (1U)
  Xeon(~3.2GHz): 2CPU
- PRIMERGY BX600
  Max. 10 blades in a 7U chassis
  Xeon(~3.2GHz): 2CPU
- PRIMERGY RXI300/RXI600
  Itanium2(1.5GHz): 2/4CPU

● Compute Network

- InfiniBand (8Gbps x 2 ports)
  HCA Card: Fujitsu's own development
  133MHz, 64bit PCI-X bus supported



- Myrinet2000 (2+2Gbps)
- Gigabit Ethernet

● Software

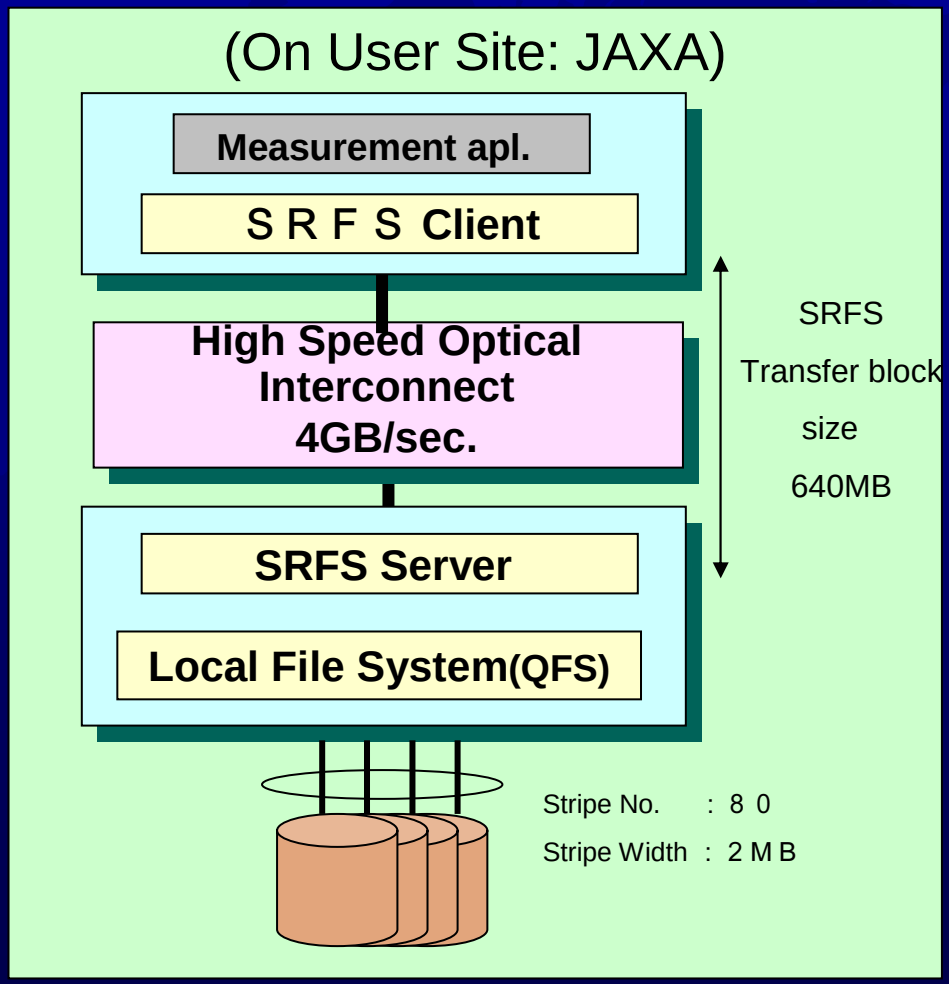
-OS : Linux
-Middleware:
SCore/Beowulf

*SCore is PC Cluster Control System developed by PCCC(PC Cluster Consortium in Japan).

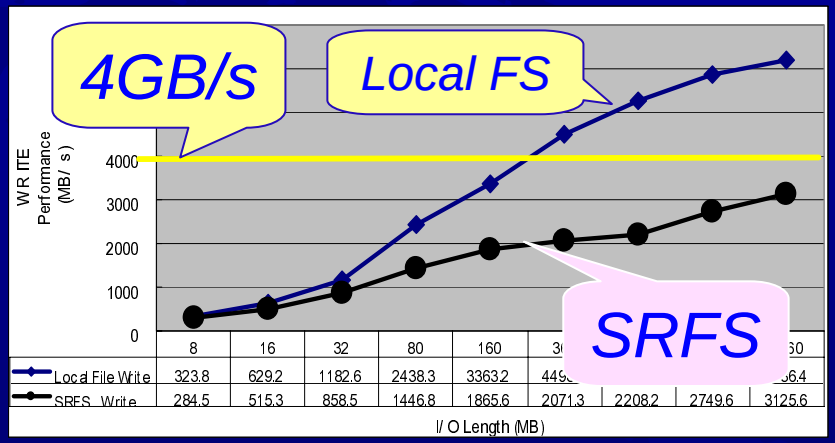
Performance Example: SRFS

FUJITSU

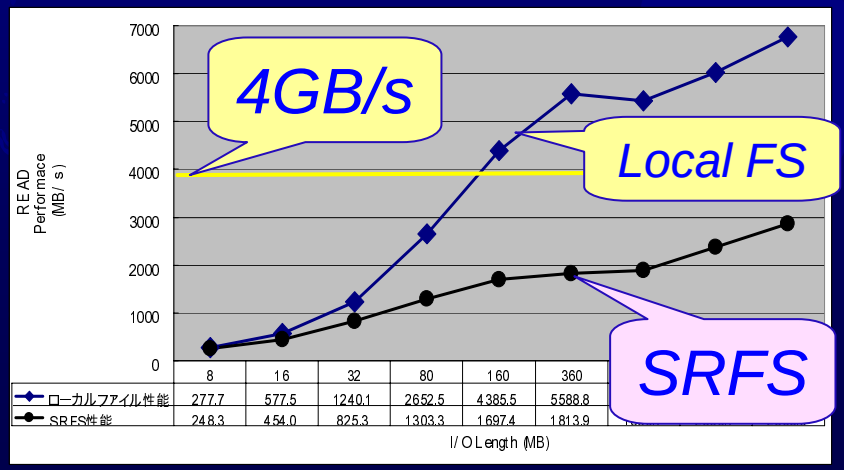
- SRFS performance on HPC2500 -



- Write -



- Read -



Achieved high performance (Over 3GB/s) with optical interconnect

JAXA : Large SMP System

FUJITSU

- 9.3 Tflops peak performance, 5.406Tflop/s with Linpack
  - 56 compute nodes, 32 CPUs per node at 5.2Gflop/s per CPU
  - Fujitsu PRIMEPOWER HPC2500 * 14 (18 in total)
- 3.6 TB of main memory
- 57 TB total disk space, 620 TB tape library

Disk
57TB FC RAID

Compute nodes
32way SMP×56

Network

Tape library
620TB

I/O node
128way SMP

Service/Login nodes
64way SMP×6

Kyoto University: HPC2500 System

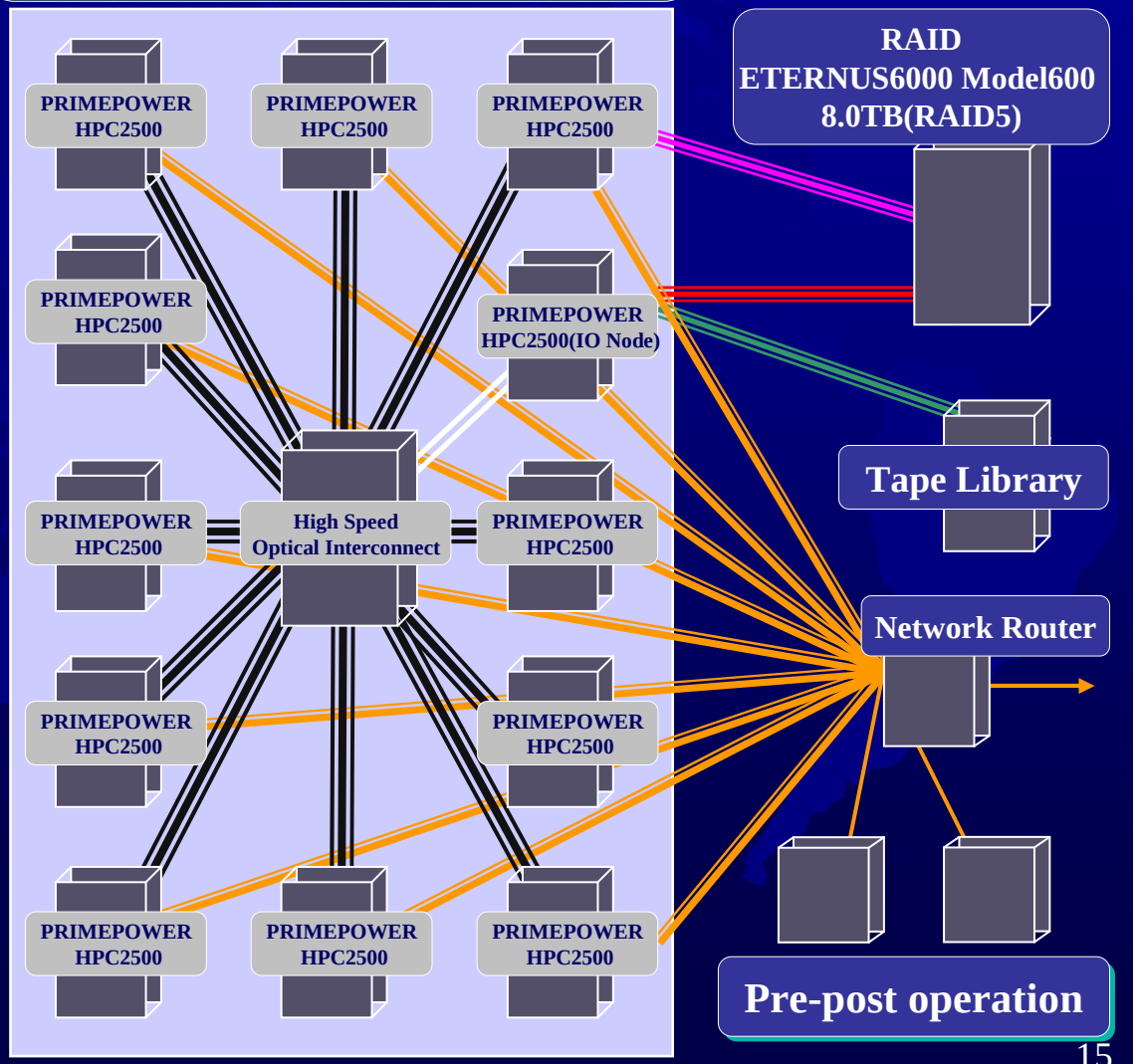
FUJITSU

Supercomputer
PRIMEPOWER HPC2500
9.185Tflops, Memory:5.75TB

[Configuration]

[PRIMEPOWER HPC2500]
- 128CPU/Node 11Cabinets
  - ➡ Compute Nodes
- 64CPU/Node 1Cabinet
  - ➡ I/O Node

[HPC2500 at Kyoto University]



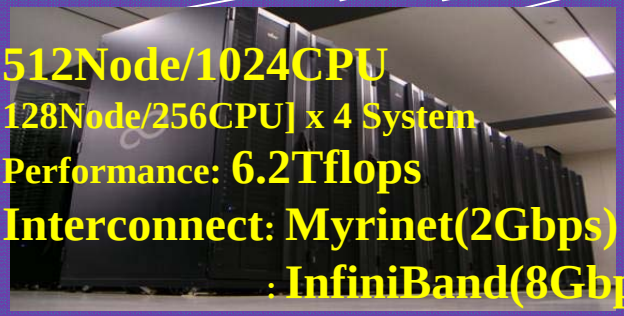
RIKEN: IA-Cluster System

FUJITSU

IA Cluster 1

512Node/1024CPU (CPU:Xeon 3.06GHz)
Performance : 6.2Tflops
Interconnect: InfiniBand(8Gbps)

Total: 1024Node/2048CPU
Total Performance :12.5Tflops



512Node/1024CPU
128Node/256CPU] x 4 System
Performance: 6.2Tflops
Interconnect: Myrinet(2Gbps)
: InfiniBand(8Gbps)

IA Cluster 2

Gigabit Ethernet Network

Firewall/VPN Device

HPSS System

Archive System

Large Memory System

20TB

File Server

Front-end Server

SAN
(Storage Area Network)

Hardware Road Map

FUJITSU

SPARC /Solaris PRIMEPOWER

HPC2500

SPARC64 V
1.3GHz/$2MB

SPARC64 V
2.08GHz/$4MB

Post-HPC2500

SPARC64 VI
UltraSPARC

IPF(IA64) /Linux PRIMERGY

IPF System
[2or4cpu]

Itanium (R) 2 1.5GHz ~

Montecito 2Core

New Linux SMP

Montevale ~

IA32/Linux PRIMERGY

Perf-Blade
2or4cpu/Node

Xeon ~3.2GHz

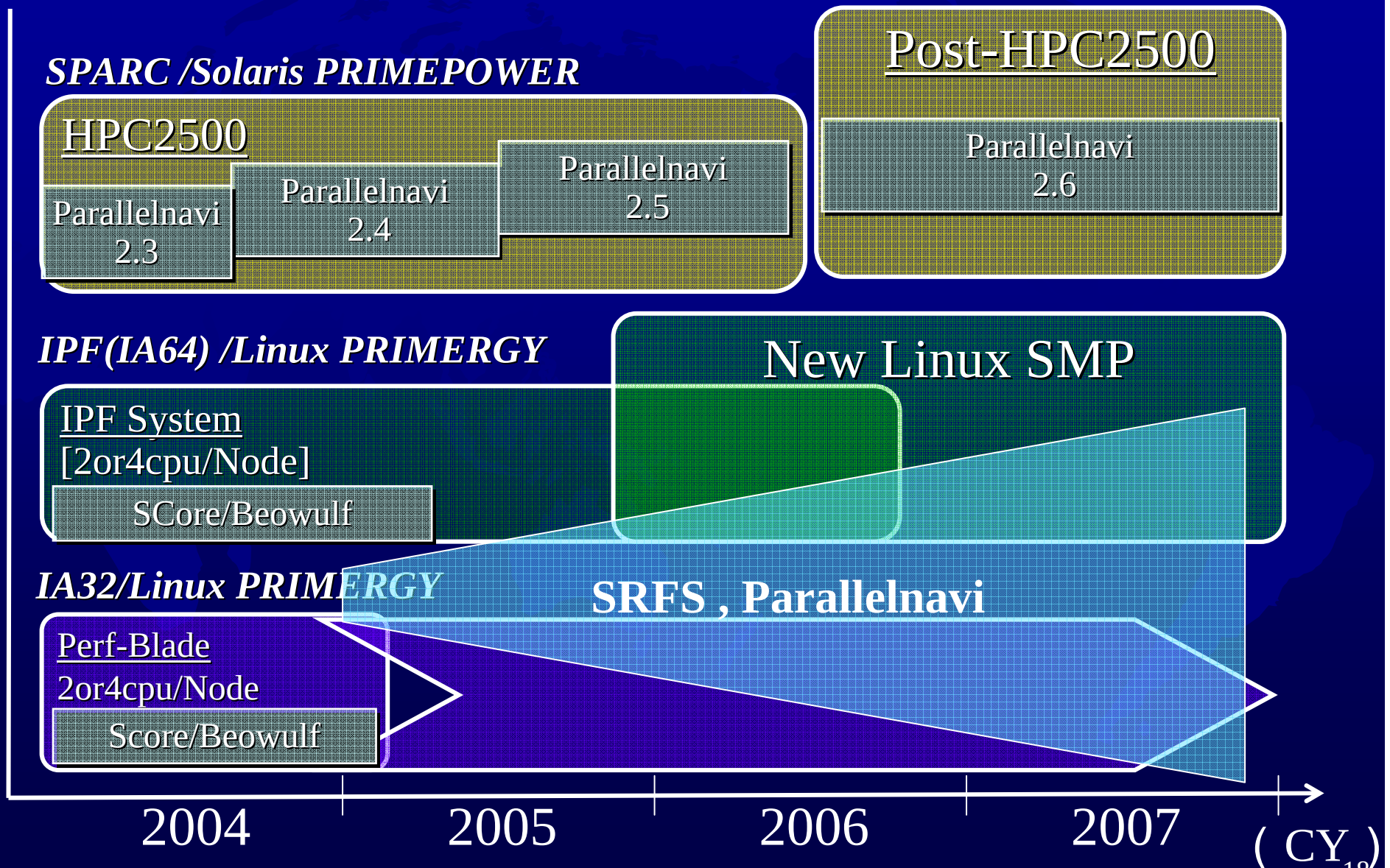
EM64T

Enhancement according to
Intel's Road Map

2004 2005 2006 2007 (CY)

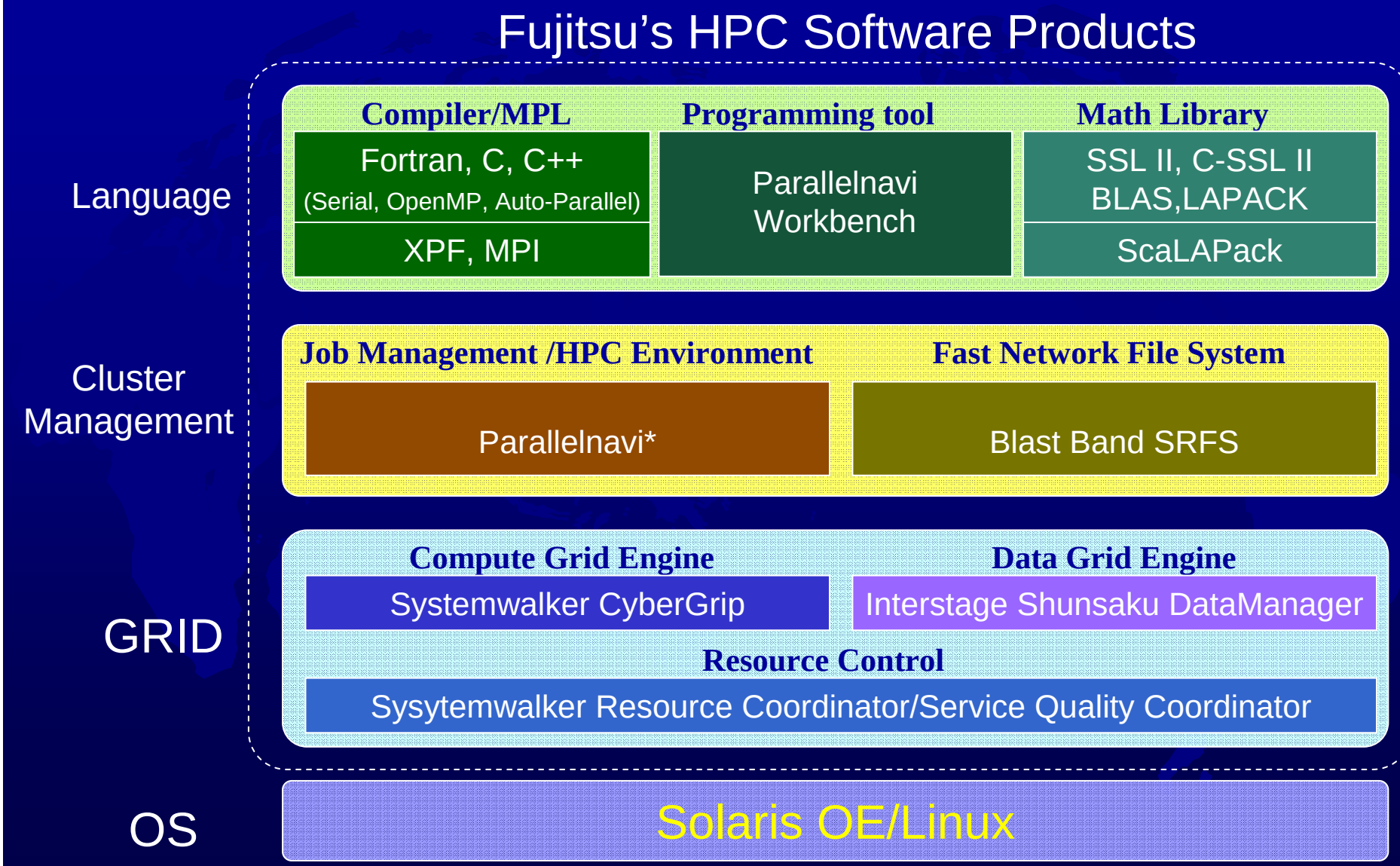
Software Road Map

FUJITSU



Fujitsu's HPC Software Structure

FUJITSU

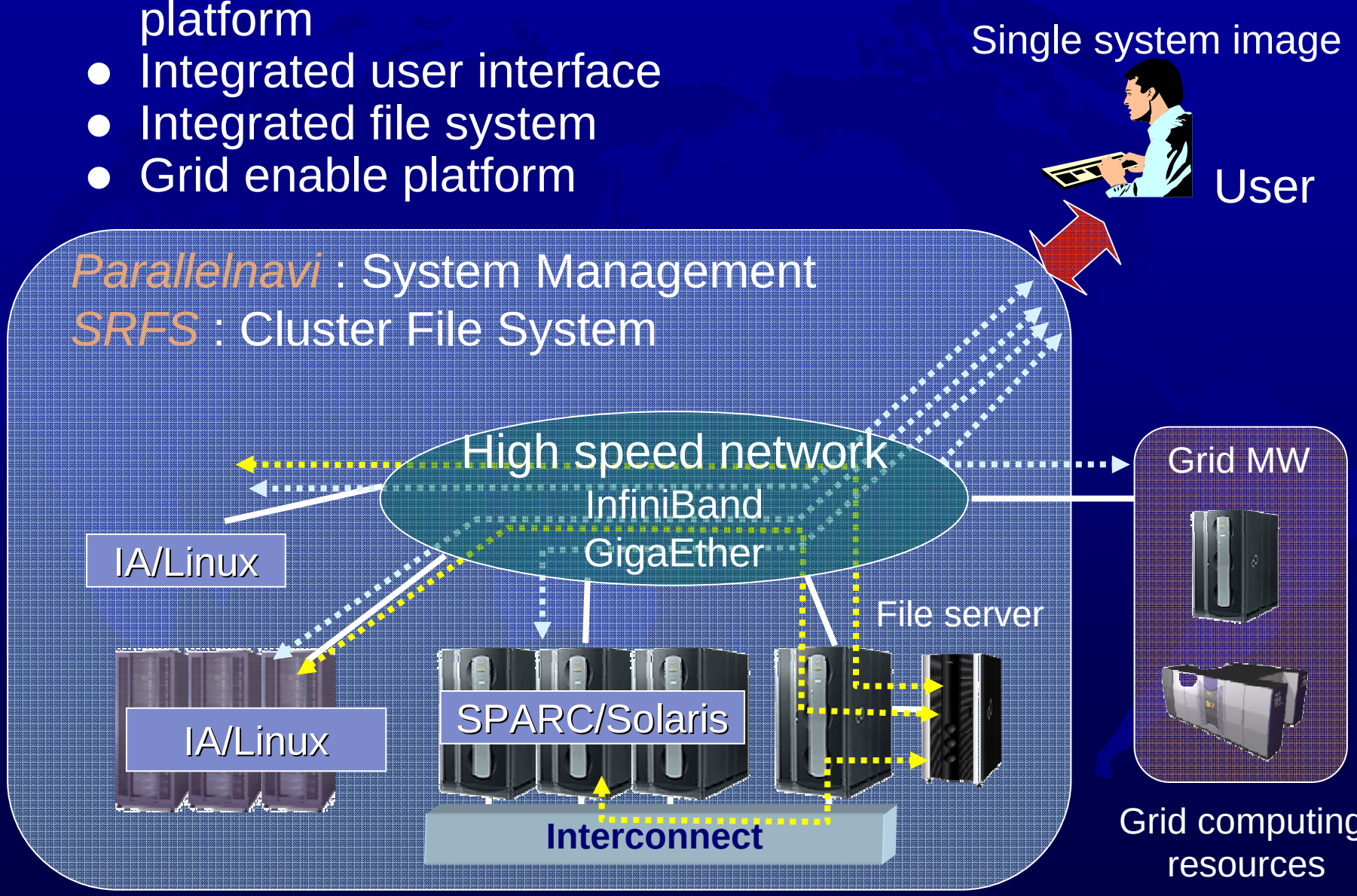


* Parallelnavi also includes language product

Integrated HPC Environment

FUJITSU

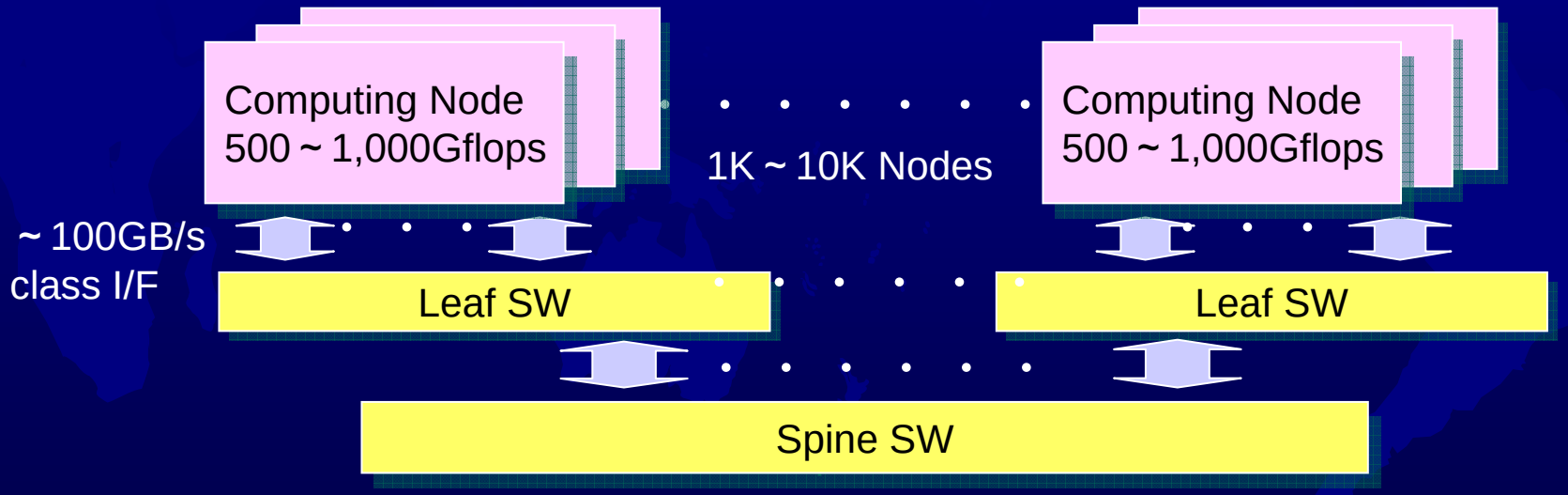
- Support heterogeneous platform
- Integrated user interface
- Integrated file system
- Grid enable platform



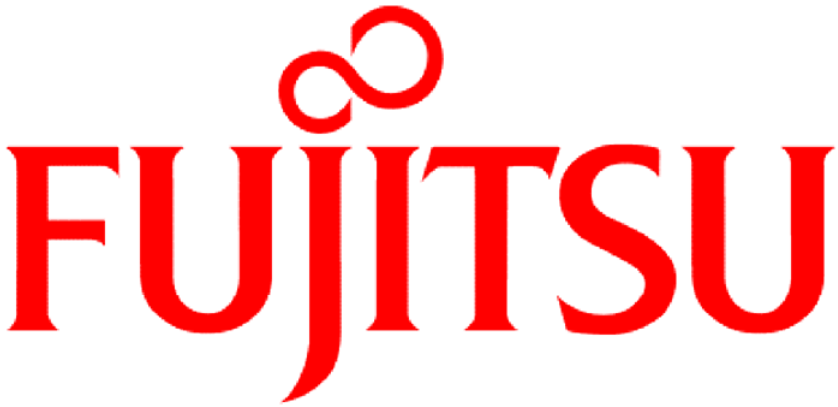
Toward Peta-Scale Computing(2010)

FUJITSU

- Achieving practical Peta-Scale computing, Fujitsu's goals are 3Pflops at peak and 1Pflops at sustain.
- We need high-performance CPU, low power consumption, faster memory access and interconnect.
- And also, high reliable, intelligent and more than 10K node connection capability, and cluster software with highly parallel system management capability.



Fujitsu organized 'Peta-Scale Computing Research Center' on October 2004.



THE POSSIBILITIES ARE INFINITE