Workshop KEK - CC-IN2P3

KEK new Grid system

Koichi Murakami, KEK/CRC

Outline

2

- Computer system on KEK/CRC
  - Central Information System
  - Belle Computing System
  - Upgrade plans

- Strategy for the migration of Grid system

- List of items to be configured towards the new Grid system

Computer System on KEK/CRC

3

- "Central Computing System" as a part of "Central Information System"
  - CIS provides general information services on KEK.
    - mail system, web system, Indico, etc.
  - CCS provides work servers, batch servers, and storage system.
  - LCG system is operated.

- B Factory Computer System
  - dedicated to analysis of the Belle experiment
  - including Grid testing servers

- Supercomputer System
  - Hybrid system of a vector machine (Hitachi SR11000) a scalar system (IBM Blue Gene)
  - There are also data grid activities on Lattice QCD

Central Information System

4
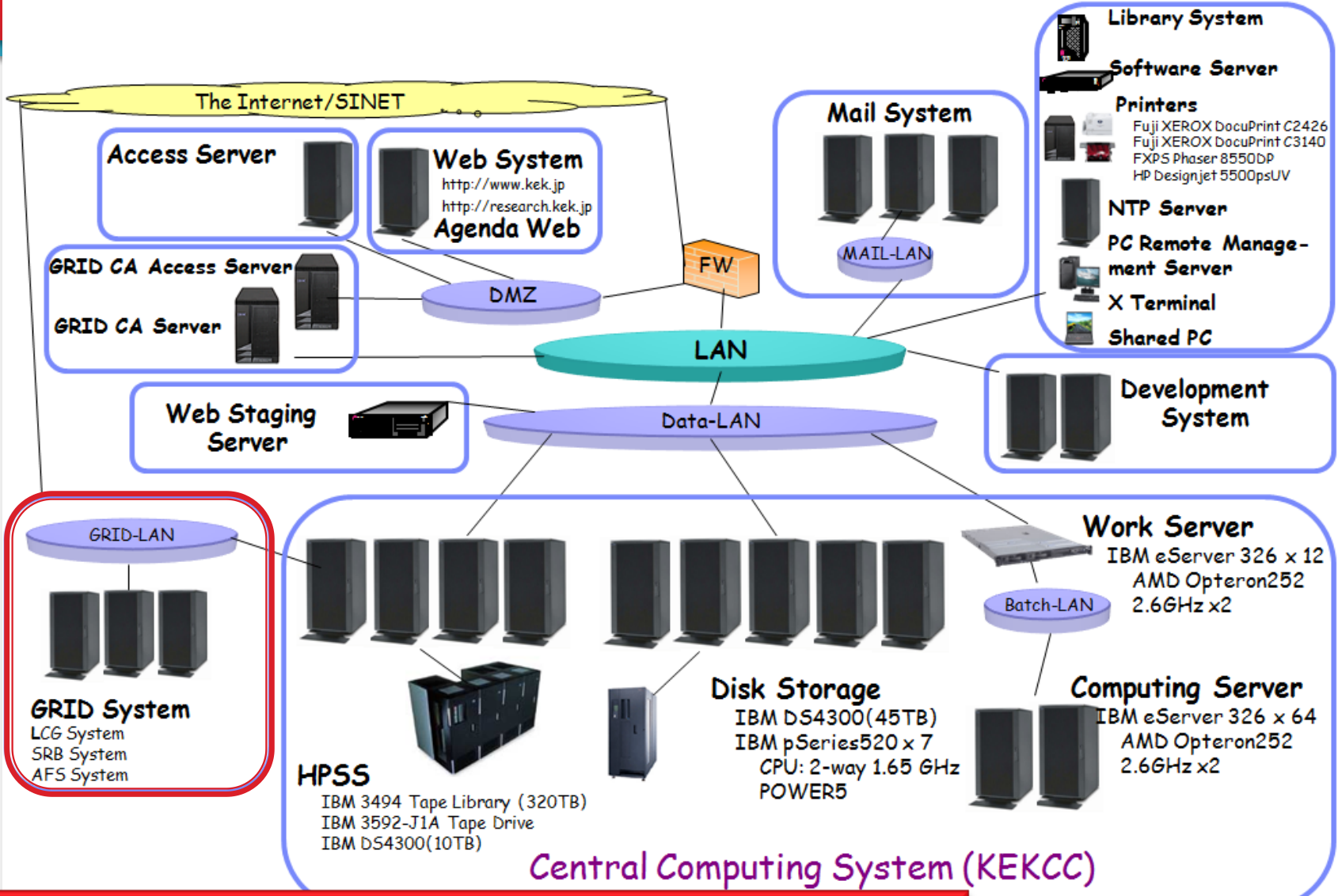
KEK-2 is operated as a part of Central Information System

Central Computing System (KEKCC)

5

- Computing Servers
  - Linux (RedHat ES4) cluster
  - 76 nodes (AMD Opteron 252 x 2), 4GB memory/node

- Disk System
  - 45TB disk storage

- HPSS Data Storage
  - HPSS tape library
  - maximum capacity of 320TB

- Grid System
  - LCG system
    - KEK-2 is operated
    - 24 WNs, 2TB disk storage
 - SRB system, AFS servers

System Replacement of CI System

6

- The current Central Information System is rental system of 3 years duration.
  - 2006/Feb. – 2009/Jan.

- System replacement ("*upgrade*") is scheduled on 2009/Feb.
  - again, rental for 3 years
  - We are collecting requirements for system specifications from experimental groups.
    - HEP exp., Neutron exp. at J-PARC, ATLAS, ILC, Acc., Theory

- Estimated amount of computing resources
  - ~2,000 cores of CPU
    - LSF as batch system
  - ~4PB storage (~500TB as disk)
    - HPSS or Storage Tek

System Structure of B Computer System

7
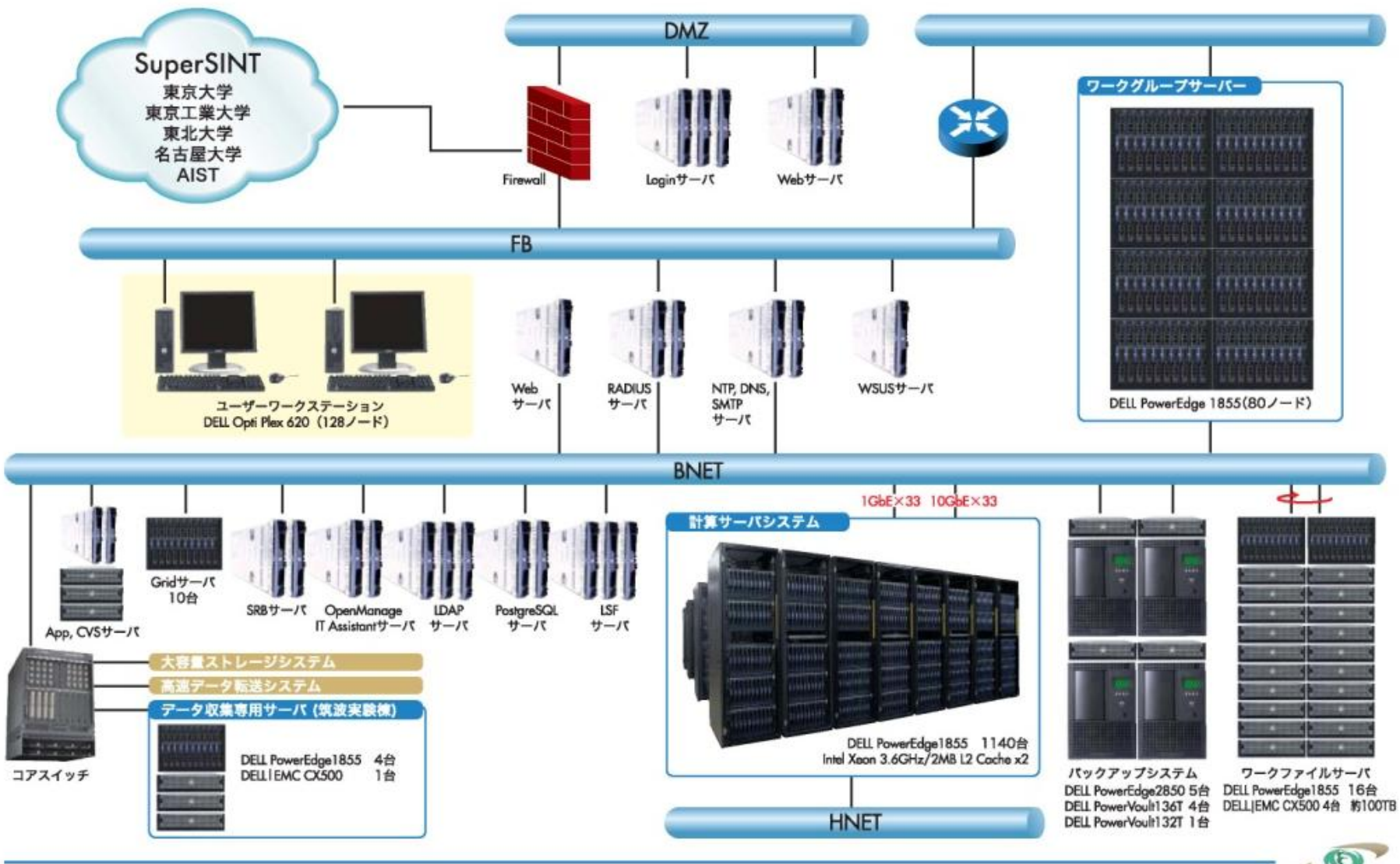
Overview of B Computer Resouces

8

	Capability	# of racks	Electrical Power Consumption (kW)
Tape	3.5 PB	21	6
Disk	1 PB	13	60
CPU	2280 CPU	23	475
Login node	80 node 80 TB	8	70
DAQ	100 TB	3	22

633 kW in total

Computing Server (CS)

9

1 Enclosure	=	10 nodes, 7U space
1 rack	=	50 nodes
25 racks	=	4 arrays

- CS＋WG servers(80)
    - = 1208 nodes = 2416CPU
    - = 45,662 SPEC CINT 2k rate = **5,114 kSI2K base**
    - = 8.7THz
- DELL Power Edge 1855
    - Xeon3.6GHz x2, memory 1GB
- Linux
    - CentOS/CS, REL/WGS

Belle Storage System – disk -

10
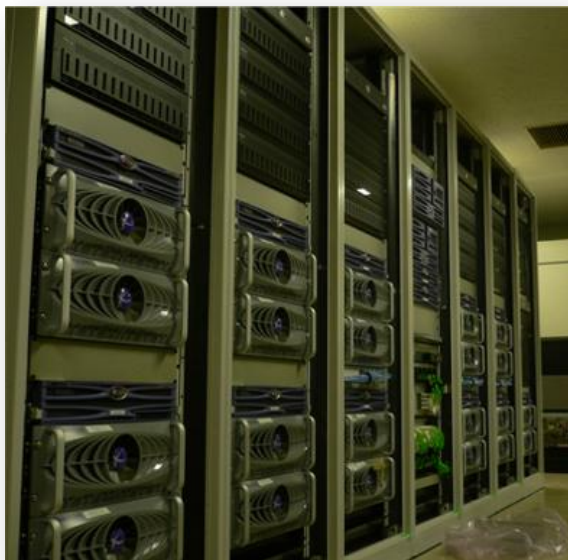
Array Master 15dr/3U/8TB

SATABeast 42dr/4U/21TB

- 1PB, on 42 File Servers
  - Nexan + ADTeX + SystemWks
- SATAII 500G drive ×~2k
  - ~1.8 f/day
- HSM = 370TB
  non HSM (no backup) = 630TB

Bell Storage System – tape -

11

- Tape Library
  - 3.5PB + 60 drives
     + 13 servers
  - SAIT 500GB/volume
  - 30MB/s drive
  - SONY SAIT-PetaSite

- Work File Server Backup
  - 90TB + 12 drives
     + 3 servers
  - LTO3 400GB/volume
  - NetVault

Upgrade plan for B Computer System

12

rental for 6 years duration

JFY April - March	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Computing Server	WS				WS+PC					PC			?		
Throughput SPECint rate2K	100				1,250					42,500 ➡ 68,000					
Speed (kSI2K) SPECint2K base	11				140					**4,760 ➡ 7,600**					
Disk (TB)	4				9					**1,000 ➡ 1,500**					
Tape (TB)	160				620					**3,500**					
Work Group Server # of hosts	3+(9)				11					80+16FS					
User Workstation # of hosts	25WS+68X				23WS+100PC					128PC					

Status of Grid Deployment

13

- Current Grid environments are deployed as a "*fragment*" of these computer system
  - R&D for how to utilize Grid environment, trying distributed analysis using Grid technology
  - mainly focused on applications other than LHC activities

- Gradual integration of Grid environment
  - Co-operation with the current (non-Grid) system is mandatory.
  - accumulating knowledge for Grid operation

Concerns about Grid deployment

14

- Support for both *non-Grid* and *Grid* users

- System configuration of such "*hybrid*" system is a key issue.
  - resource management of batch servers
    - job submission using LSF
  - providing LFC access using SRB-DSI/HPSS-DSI with SE
  - using virtualization technology (VM) for more robust and performant operation

- Interoperability between NAREGI and LCG
  - choice of GRID middleware

- There might be one possibility of migrating two computer system in the next next system replacement (2012) as a further prospect?
  - more efficient resource management in different aspects

Technical stuffs toward new Grid system
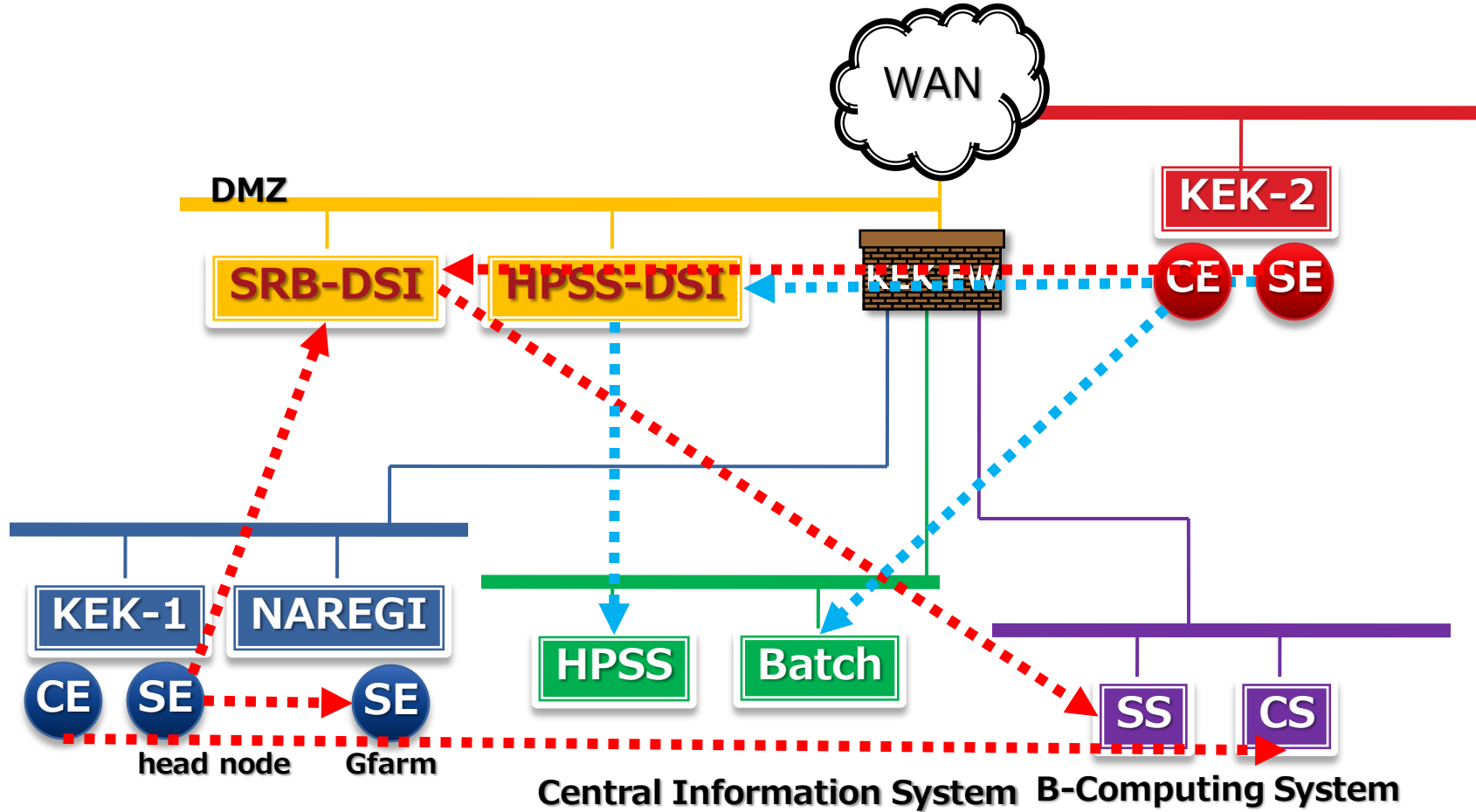
15

- Job submission via LSF on CE/WN
  - sharing resources with locally submitted jobs and grid jobs
  - optimization of setting parameters for practical operation

- SE integration
  - LFC accesses to a large scale of storage system (SRB/HPSS)
  - We can access to both SRB-DSI and HPSS-DSI via GridFTP, but not use LFC.
  - HSM access by using SRB-DSI and HPSS-DSI on SRM

- More robust and flexible operation
  - system configuration /operation using VMs

- Networking/Security
  - We were suffered from restrictions due to a domestic security policy!
  - ((need for *bureaucratic?* (not technical) global security assurance mechanism.))

Integration of different components

16

WAN

DMZ

SRB-DSI   HPSS-DSI   KEK FW   CE SE

KEK-2

KEK-1   NAREGI

HPSS   Batch

CE SE   SE

head node   Gfarm

SS   CS

Central Information System   B-Computing System

Summary

- Computer system on KEK/CRC
  - Central Information System
  - Belle Computing System
  - Grid environments are deployed as a "*fragment*" of these computer system
- Central Computing System will be replaced on 2009/Feb.
  - Migration of Grid environment
  - Support for both *non-Grid* and *Grid* access
- Technical items and plans towards new Grid system