ALICE DCS – Technical Challenges

Peter Chochula
For ALICE DCS

Outline

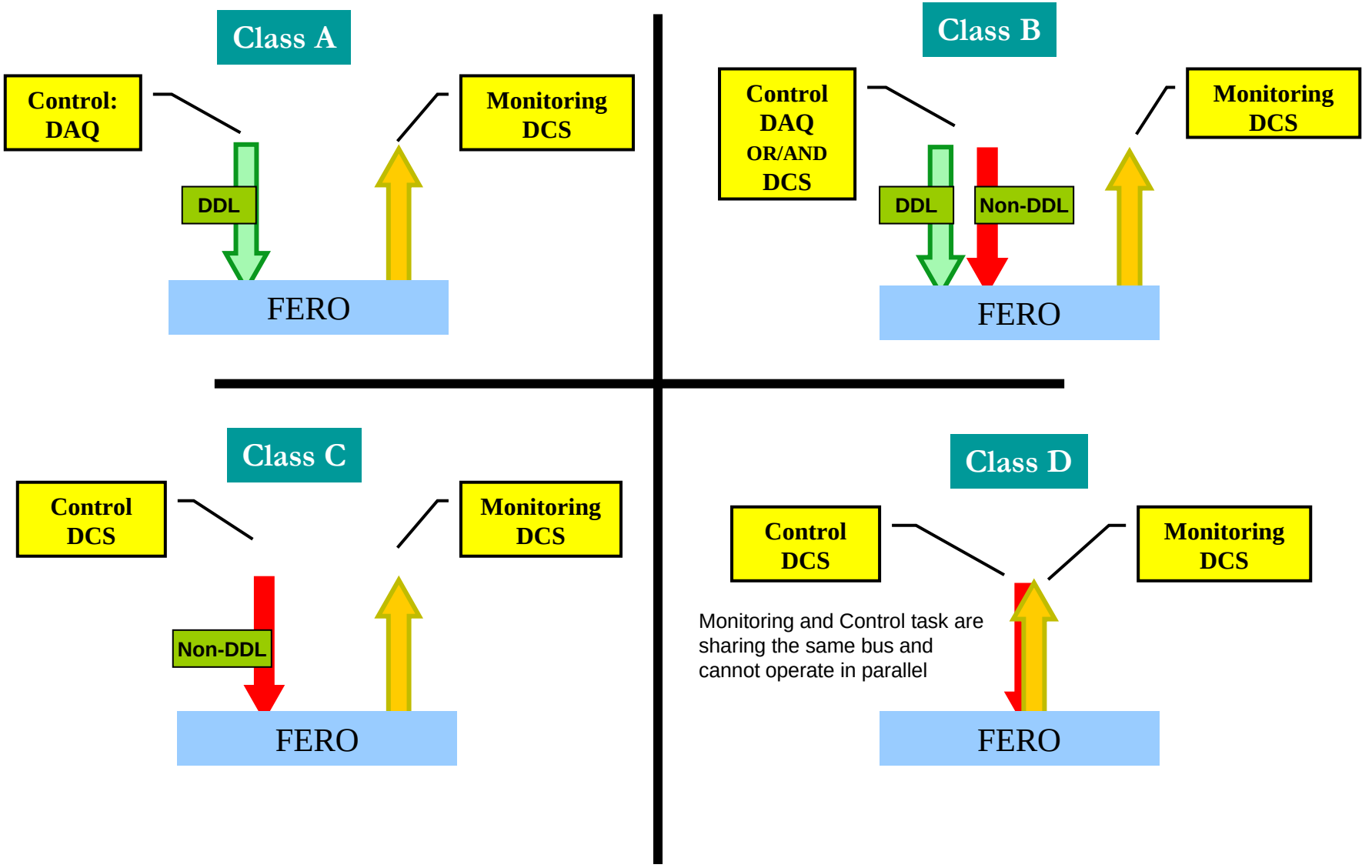
- ALICE Front-End and Readout Electronics
- DCS Performance Studies
- ALICE DCS Computing

- **Alice-specific technical challenge – the Front-end and Readout electronics (FERO)**

ALICE FrontEnd and Readout Electronics (FERO)

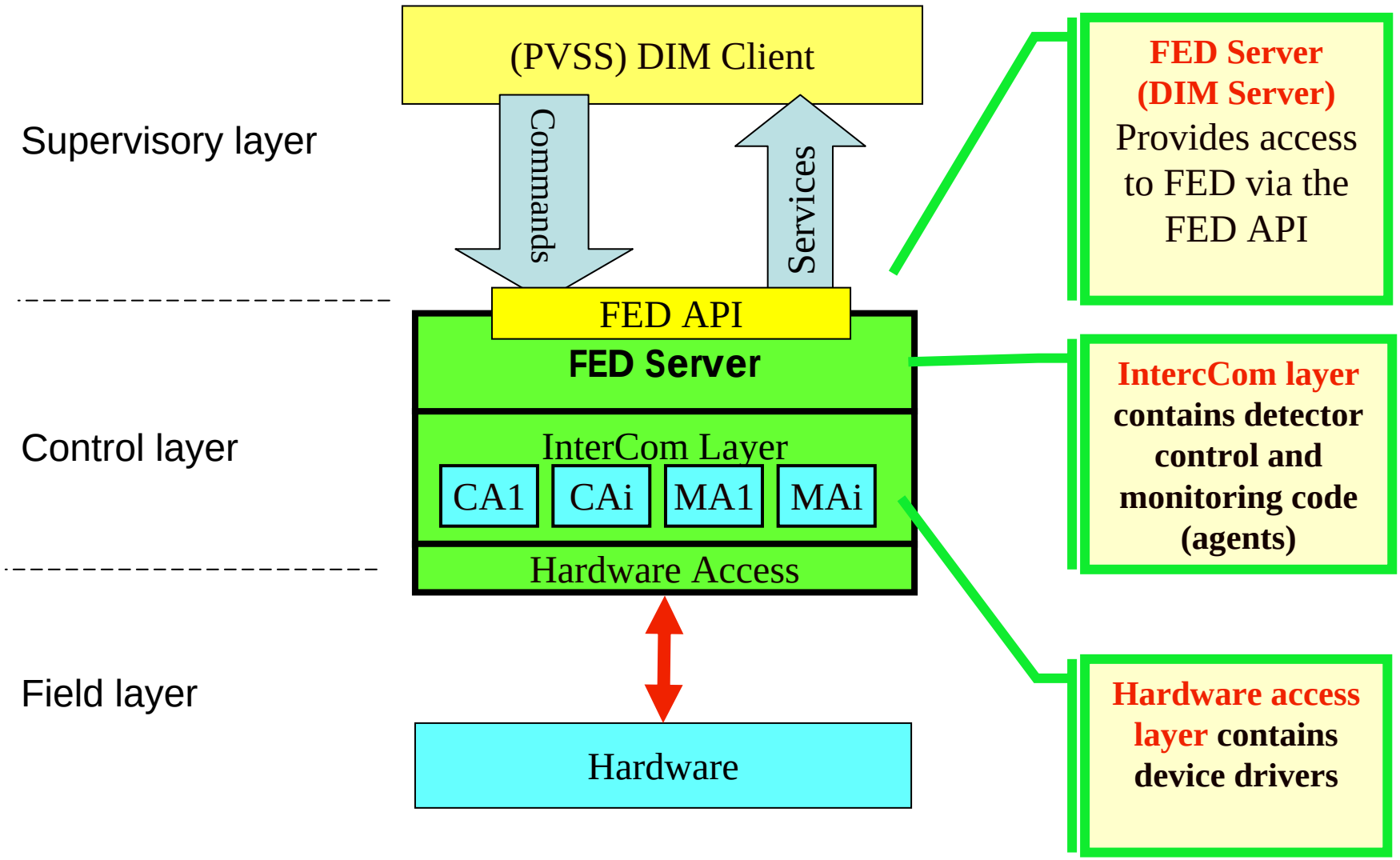
- Several architectures for FERO access are implemented in ALICE sub-detectors
  - Different buses (JTAG, CAN, DDL, Ethernet, I^2C, custom…), different operation modes
  - **DAQ** is in charge of control of architectures connected via the optical link (DDL)
  - **DCS** is in charge of controlling the rest
    - For some detectors both systems are involved

FERO Access Architectures

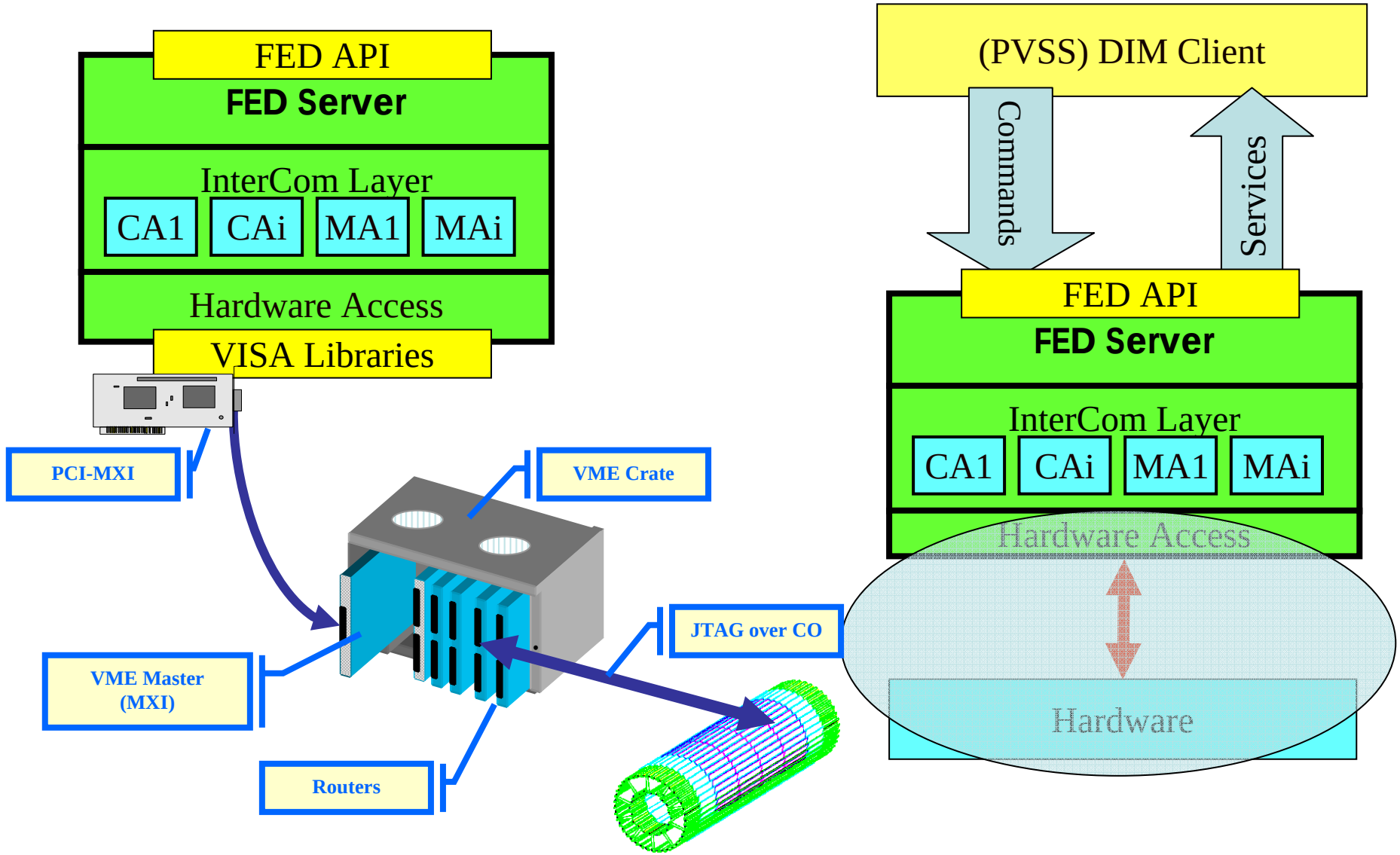


- The variety of FERO access mechanisms almost excludes implementation of common solutions
- The FrontEnd Device (FED) provides a API between PVSSII and custom architectures
  - Standard in ALICE
  - API definition (Commands, Services, Operational Guidelines) available
  - Based on DIM

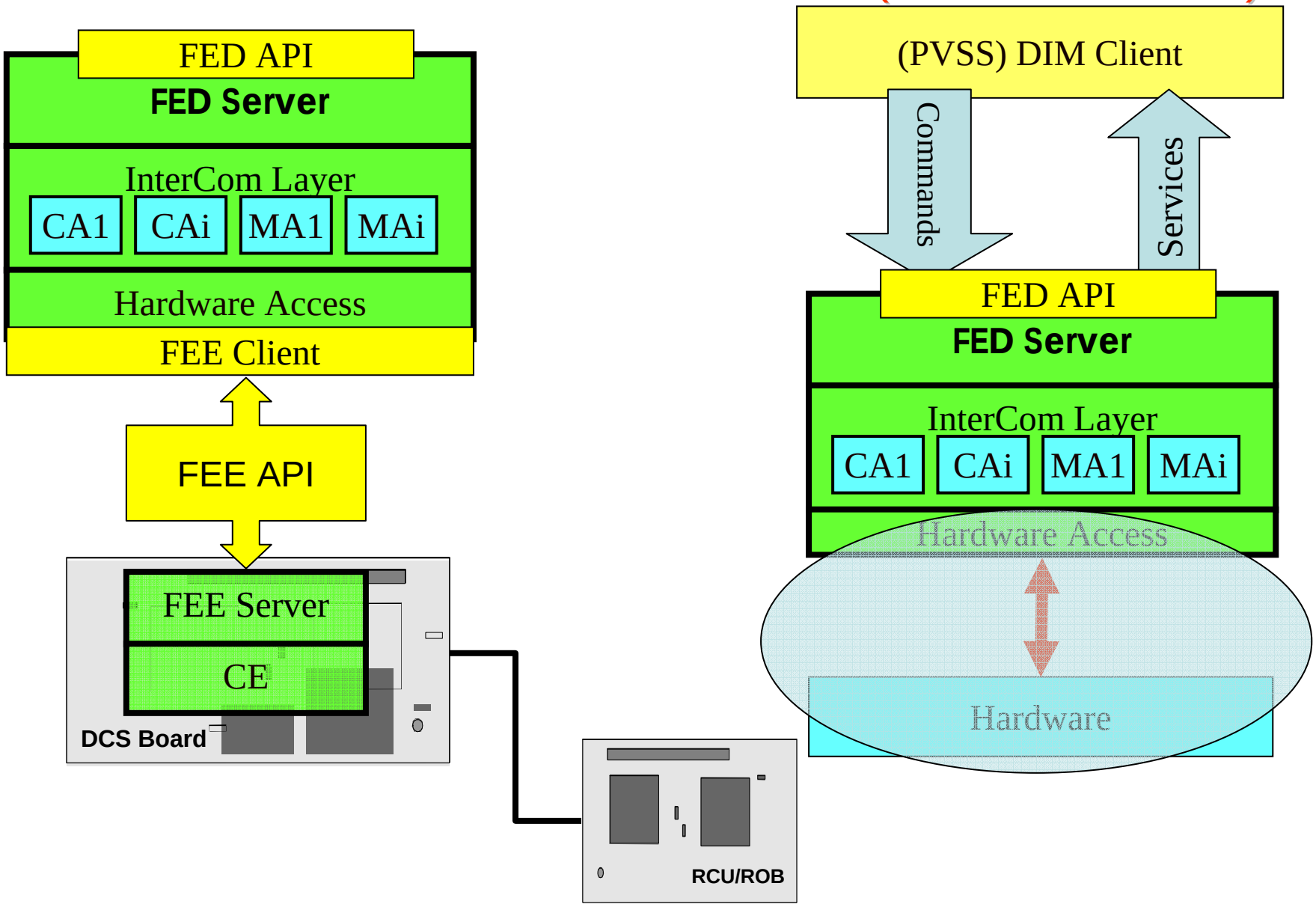
Generic Architecture of the FED Server



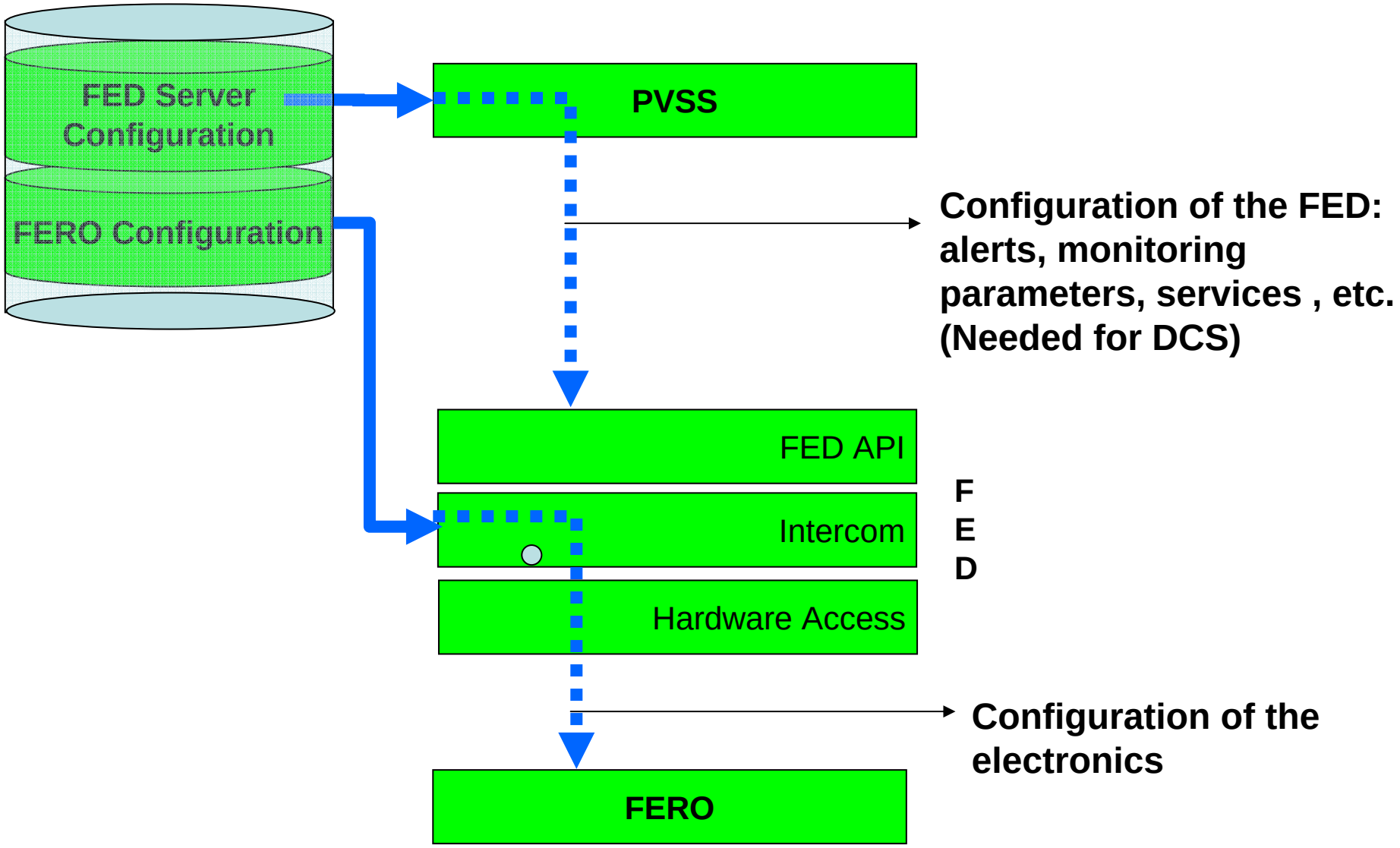
Example of “simple” Hardware Access Layer (SPD)



Yet Another Example of Hardware Access Layer (TPC/TRD/PHOS)



FERO and FED server configuration



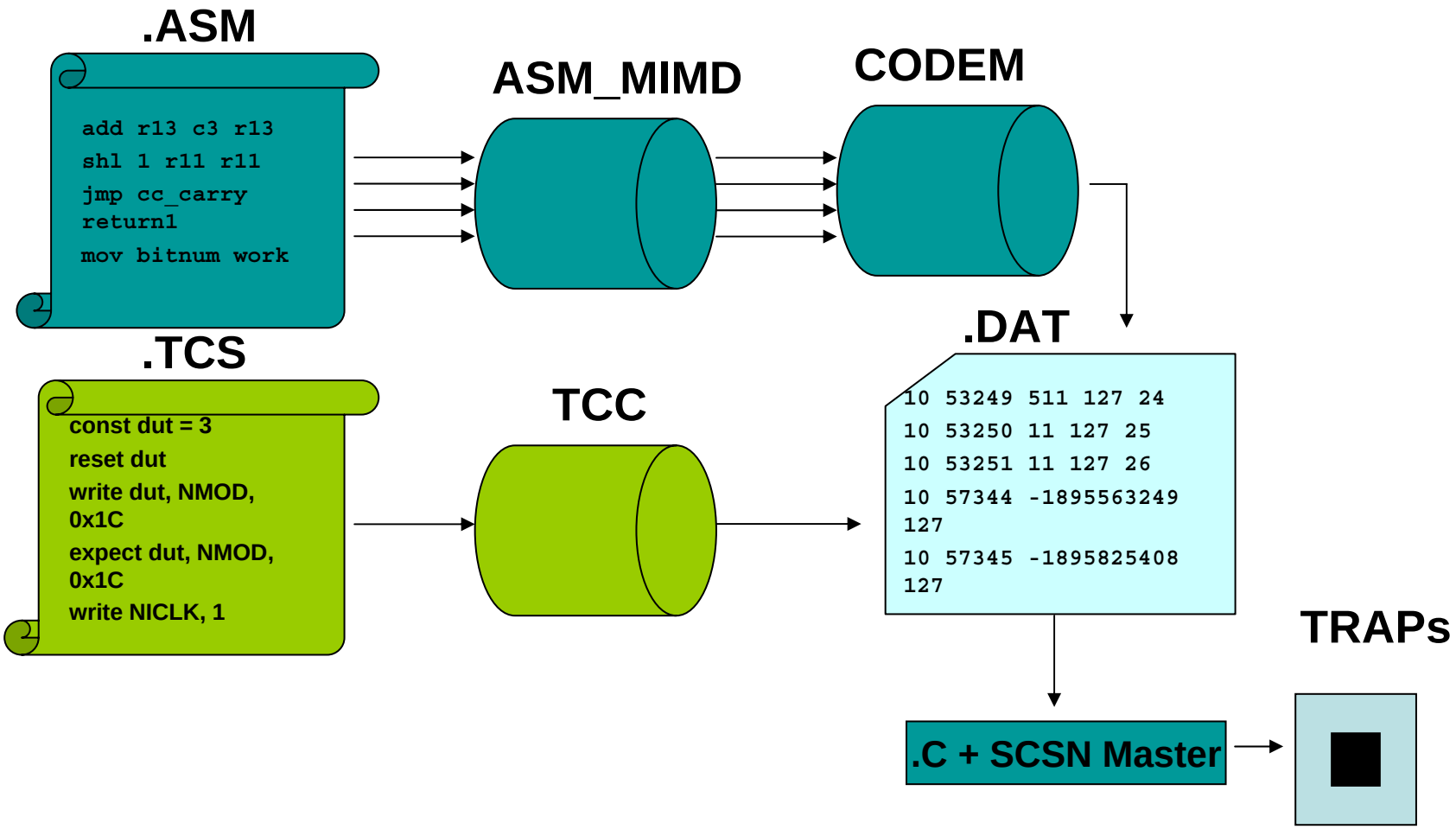
What is stored in the FERO configuration?

- The FERO Configuration contains all setting needed for the FERO operation such as
  - DAC settings
  - Thresholds
  - Mask matrices …
- Sometimes the configuration contains code for embedded processors
  - This code might be compiled on-fly by dedicated software (to avoid repetition of huge data blocks)

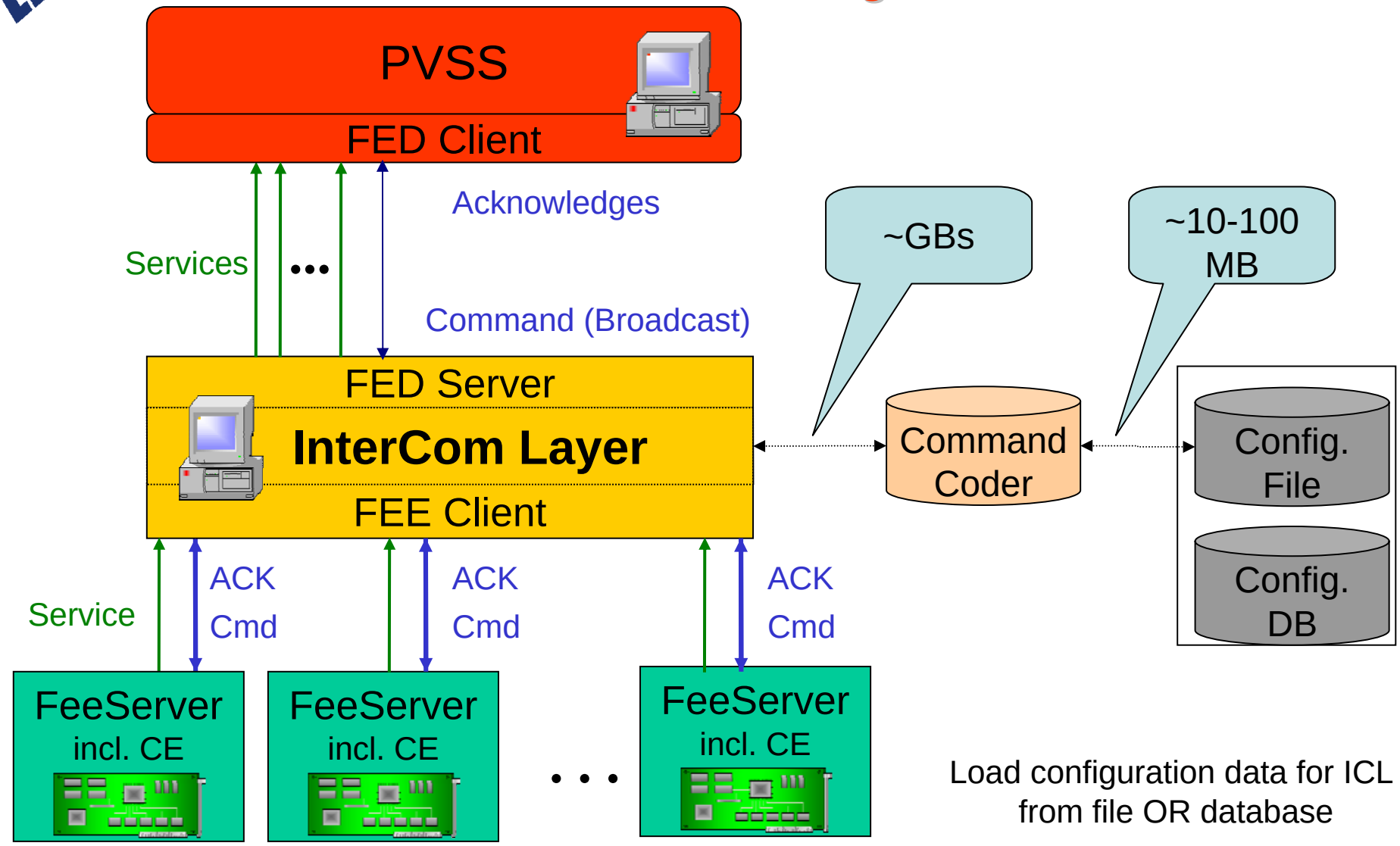
- Expected data size differs from detector to detector and ranges from few Bytes up to ~100 MB
  - the data might be compiled on fly, amount of data written to the FERO might be considerably bigger that the amount of data read from the DB

- Some parameters written to the chips are not available for the DCS monitoring
  - Data cannot be easily provided to offline

Assembling a Configuration Record



Configuration Data Flow



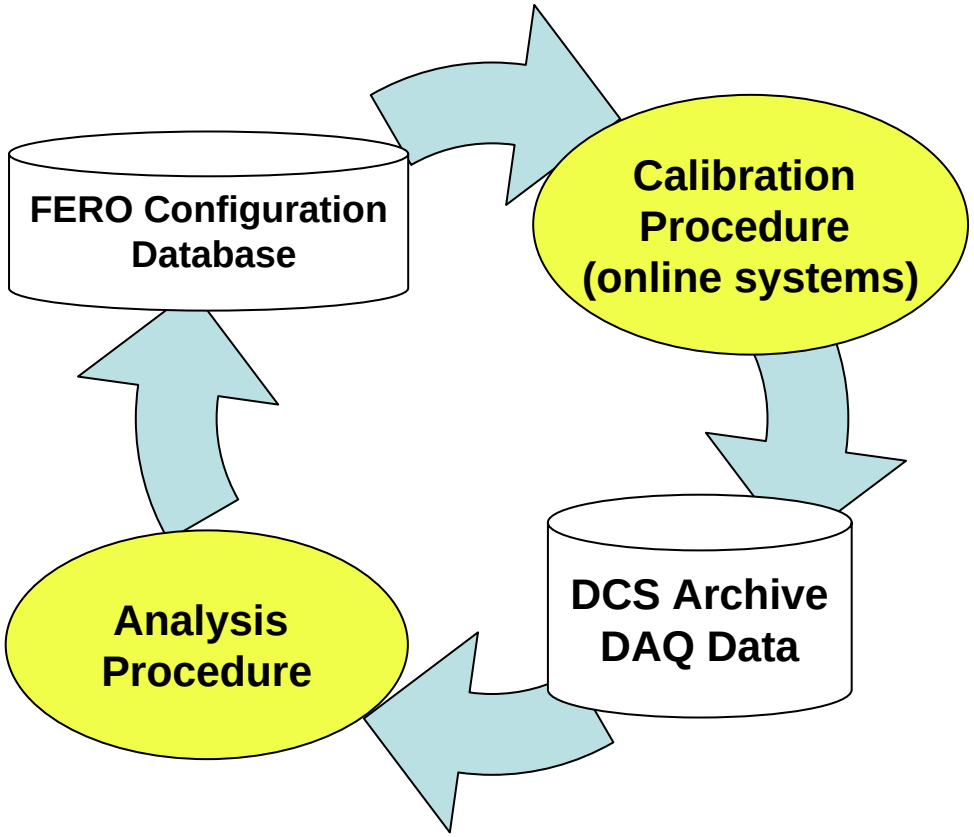
Load configuration data for ICL from file OR database

Status of FERO Developments

- FERO access implementation is advancing well for 4 detectors: SPD, TPC, TRD and PHOS
  - SPD full slice is being commissioned now (ACC participation)
  - TPC and TRD tested the chain from intercom layer to devices, PVSS integration in progress, database tests started
  - PHOS aims for full chain in June (test beam)

- Main worry: manpower is missing in detector teams.
- ACC is providing help, but soon we will loose the key person (SK)

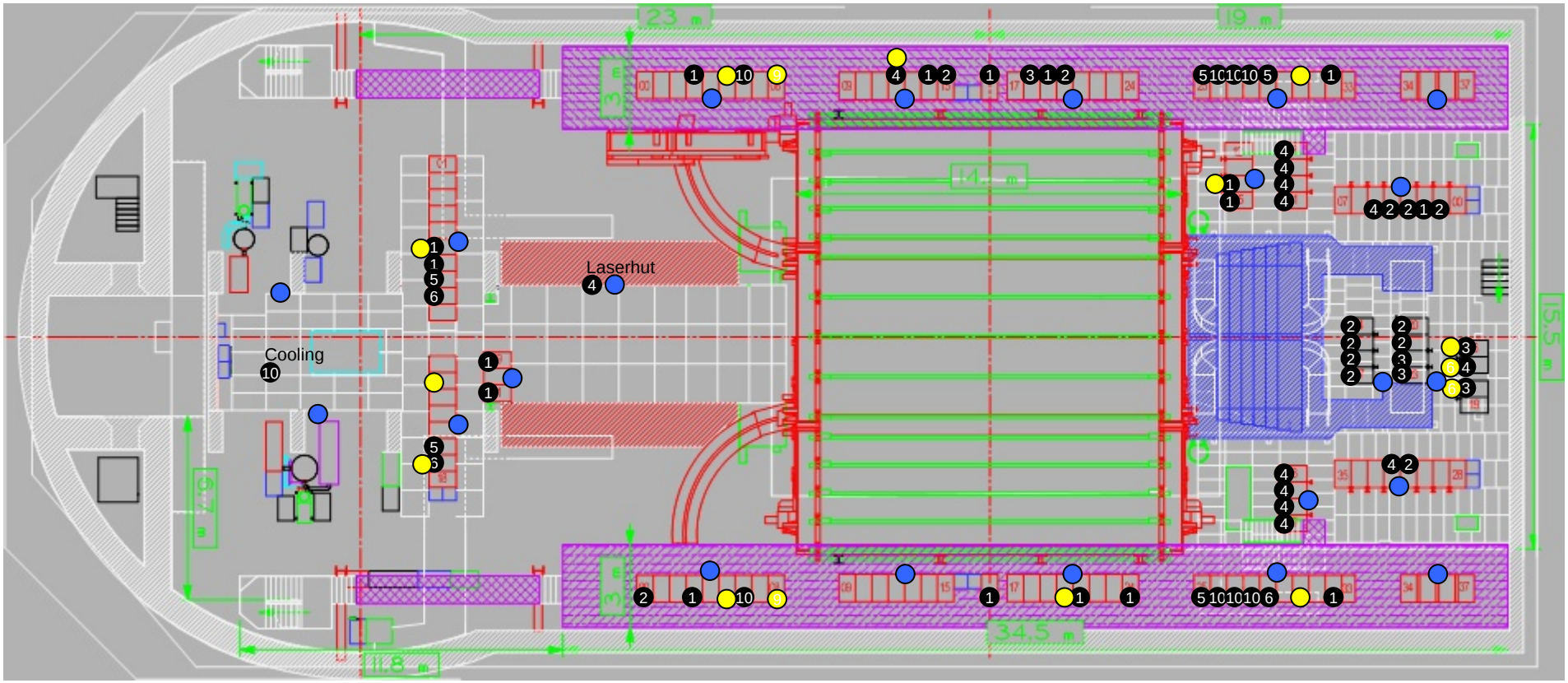
Obtaining the configuration data



- Calibration data is a result of a complex chain of steps
- Some detectors can execute several calibration procedures
- Offline and all Online systems are involved
  - We are facing insufficient information flow between different experts within the detector group
  - Difficult coordination, regular workshops of involved systems launched

The FERO and the DCS Network

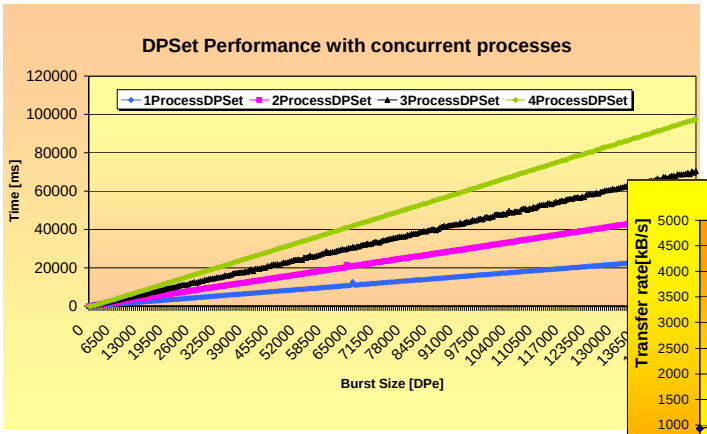
- Some FERO components rely on hardware controlled over Ethernet installed in magnetic field
  - Customized Ethernet interfaces require installation of network switches close to the detector (in the cavern)
  - No IT support for those switches
  - Hardware is tested, but long-term stability remains a question

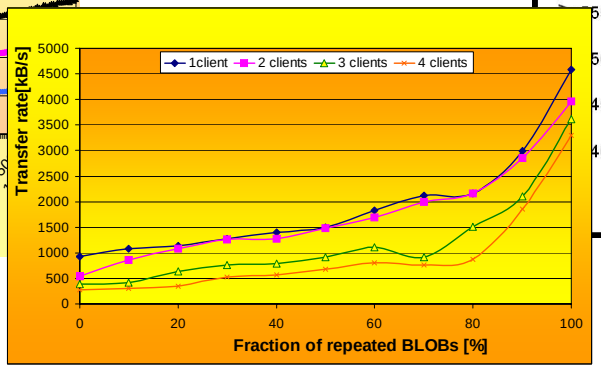


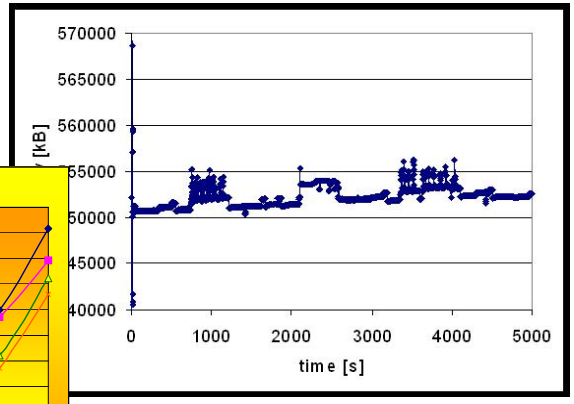
- DCS network in UX -- *[power supplies, VME crates etc.]*
  - 248 ports, only in racks
- DCS network in UX -- *["ALICE" switches for DCS boards, RCU]*
  - 41 Gbit uplinks
- GPN network -- *[e.g. for commissioning/debugging]*
  - Wireless (enough to cover whole cavern)
  - ~50 ports 'strategically distributed across rack areas'
  - Exact locations being defined now

- **The DCS Performance Tests and Related Challenges**

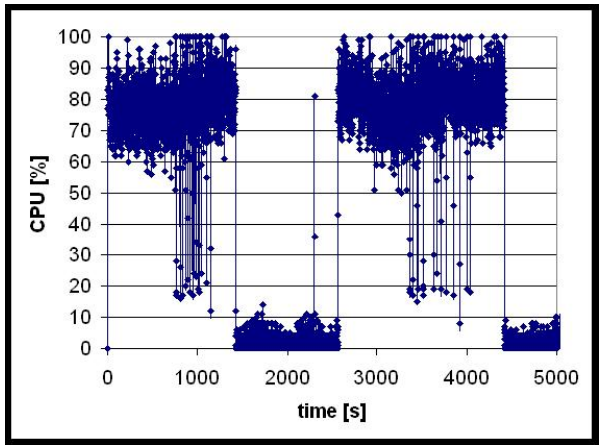
DCS Performance Tests

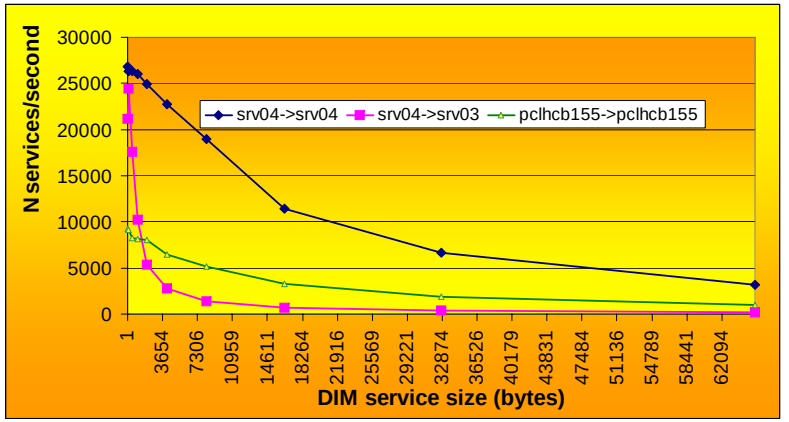






Results of performance tests presented at ALICE DCS Review 2005
http://alicedcs.web.cern.ch/AliceDCS





Summary of test Campaign

- Tests covered all aspects of the DCS
- Results of several tests were collected and evaluated:
  - JCOP tests
  - JCOP tests with ALICE contribution
  - ALICE test campaign
  - Results provided by colleagues from other experiments (special thanks to Clara)
- No major problems discovered, each PVSS system can digest its load
- Distribution of PVSS systems provides very flexible tool for performance tuning, BUT:
  - All systems will meet in a single point – the ORACLE configuration and archive. This is our major performance concern.

Dealing with Detector Performance

- DCS configured according to performance needs
  - Number of sub-systems per PVSS
  - Number of PC's per sub-system

- Critical issues
  - Switch-on of many channels
  - Configuration of many channels

Start-up Time

- Switch-on of many channels:
  - Test: Total switch-on for 180 CAEN HV channels: 7 sec
  - SDD: 520 Caen HV channels: ~15 sec
  - TRD: 1080 = 180 Iseg channels * 6 via DCS board: 7 + ? Sec
  - TOF: 3600 = 180 Caen channels * 20 fanout: 7 sec

  NOTE: **to be compared with ramping times of minutes !!**

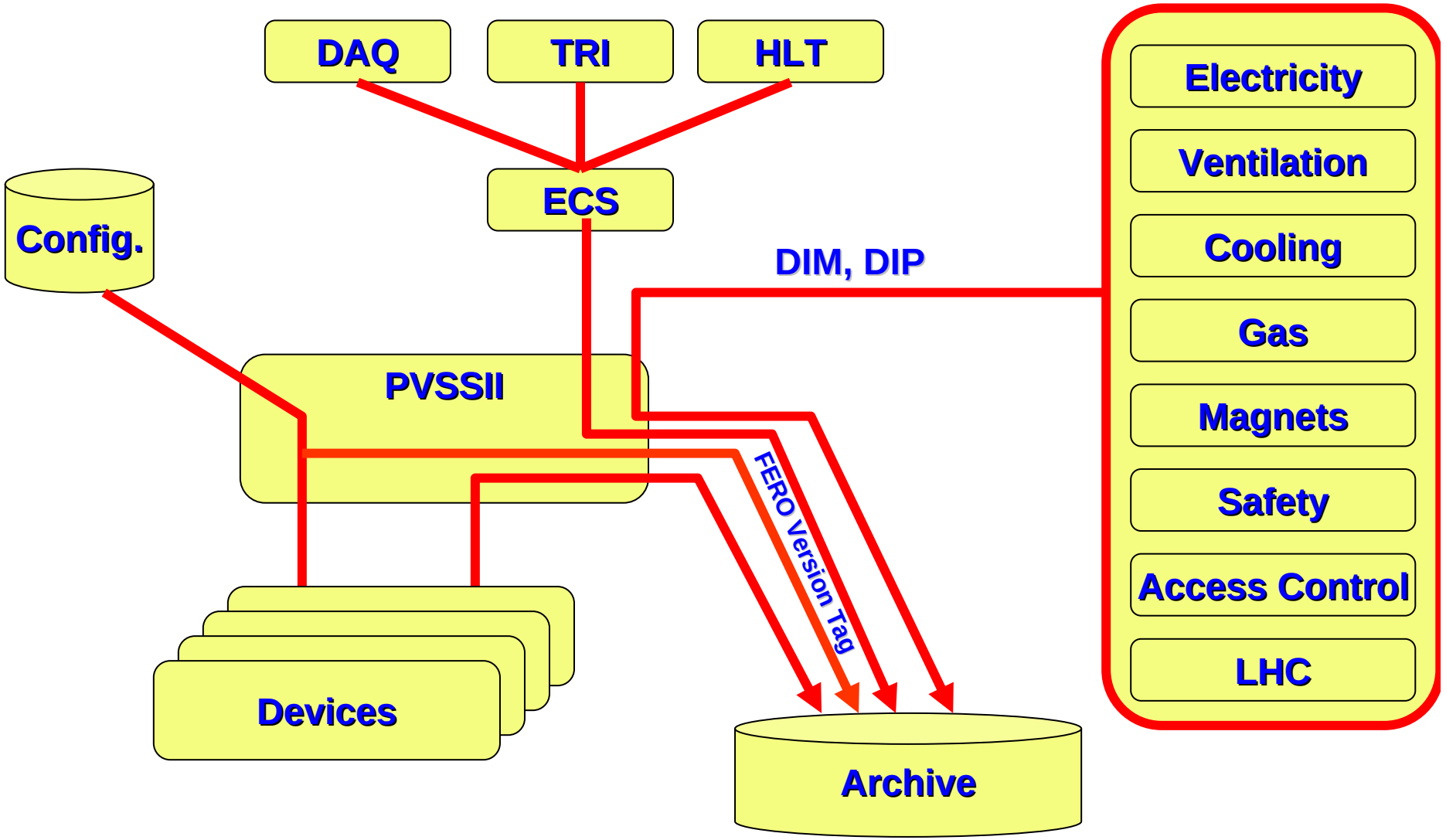
- Configuration: **normally done outside physics time !!**
  - Test: Configuration of a full Caen crate 192 ch: 20 sec
  - SDD: 520 Caen HV channels: <54 sec
  - Test: DB retrieval of FEE 150MB BLOB's: 15 – 50 sec
  - SPD: 3 sec
  - TPC: 10*10kB/DCS board: 25 sec
  - TRD: 10*10KB/DCS board: 50 sec
  - if required: Oracle tuning and Caching will improve

Performance: alert avalanches”

- Tests have shown that PVSS copes with
  - an alert avalanche of at least 10 000 alerts per PVSS system
    - ~ 60 PVSS systems in ALICE: 600 000 alerts acceptable
  - a sustained alert rate of ~200 alerts/sec per PVSS system
    - ~ 60 PVSS systems in ALICE: 12 000 alerts per second acceptable
  - all alerts from a full CAEN crate displayed within 2 sec
    - Max 6 crates on one PVSS system: all alerts displayed within 12 sec

- Many means to limit alert avalanches
  - Scattering of PVSS systems
  - Correct configuration of
    - Alert limits for each channel
    - Summary alerts & filtering
  - Verified by ACC at installation time

The DCS Archival



What needs to be archived?

- For offline use we need to archive at least HV, LV and some (many) FrontEnd parameters
- This data will be produced by ~60 machines

- Number of archived parameters:
  - LV: 3100 channels
  - HV: 20835 channels
  - FERO: ~20000 parameters

- In addition we need to archive the information provided by services, DCS states, environment, crate status, …

RDB Archival Status and Tests

- The final release of the RDB-based archival mechanism is repeatedly delayed
- At present we participate in tests of the latest release
  - setup procedure still requires expert knowledge, cannot be recommended in this stage to detector teams (it is a sort of beta version)
  - worrying problems on the server side:
    - High CPU utilization (which limits the number of clients to be handled by a single server to ~5)
      - Server overload causes loss of data
    - Big data volumes created at the database servers (mainly redo logs) resulting in unacceptable database size
- We consider today the RDB archival as still *not ready for production*

Implications of missing archival mechanism

- Some detectors already started the preinstallation and data needs to be archived
  - This data will be needed also in the future
  - we need a mechanism for transporting the data produced today into the final archive to be implemented tomorrow
- We cannot provide recommendations to detector teams for archive setup. The only solution is to use the present file-based archival and parametrize the whole project again in the future

Implications of Archival Performance

- As shown, Alice has ~40000 channels to be archived
  - Number of corresponding DPEs to be archived is higher
- ~60 computers will provide the data for archival
- The database server(s) must cope with the situation when all channels change at the same time (e.g. ramp-up)

- If the situation does not improve, we need to plan for 6-12 database servers

Implications of Archived Data Size

- As we do not know the final data volumes which will be created on the server by the archival mechanism, we cannot refine our specifications
  - For example, present mechanism creates ~2GB of data per hour for a client archiving 5000DPEs/s (tests done with 4 clients each archiving 5000DPE/s). This is about 900% overhead compared to raw information produced by the machines
- Unclear situation concerning the archival complicates developments of DCS-OFFLINE interface (see later)

Final Archival Implementation

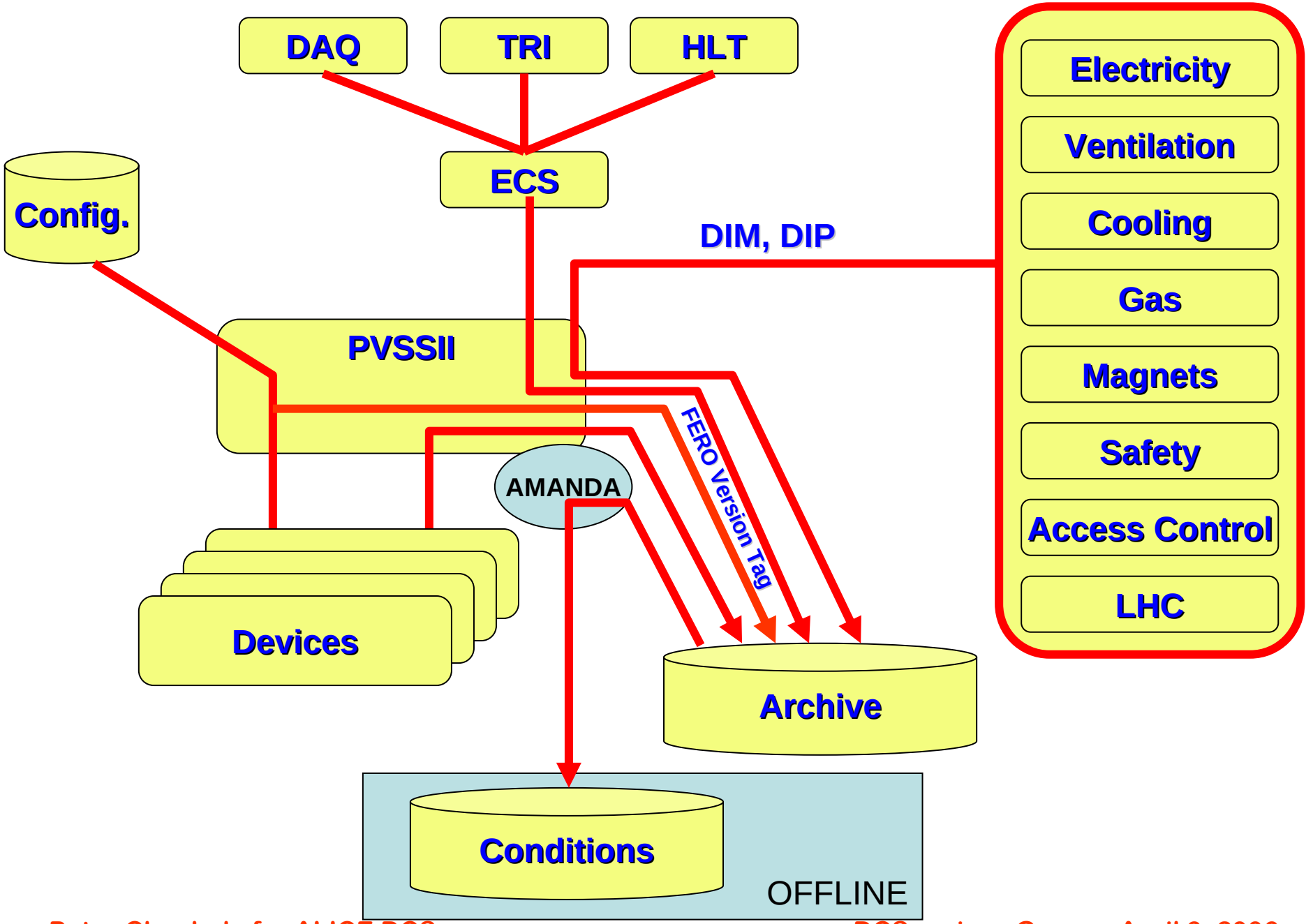
- There is no caching mechanism which will cover the period when the connectivity to DB server is lost
  - Implication: we will need to run a local database server in P2 which is not compliant with the IT policy on DB support
    - For example, the DAQ can run ~20 hours in standalone mode. The DCS must be able to cover at least this period with fully working archival

What are the next steps?

- We need to set a deadline on the archival solution
  - This date cannot be moved beyond May 31st
- If the RDB archival does not qualify for production, the only solution is file-based archival
  - Implication: we did not foresee the extra disk space on the DCS computers. If we have to order the extra disks, it must be done now. (The extra cost involved is ~9000CHF, ordering starts now)

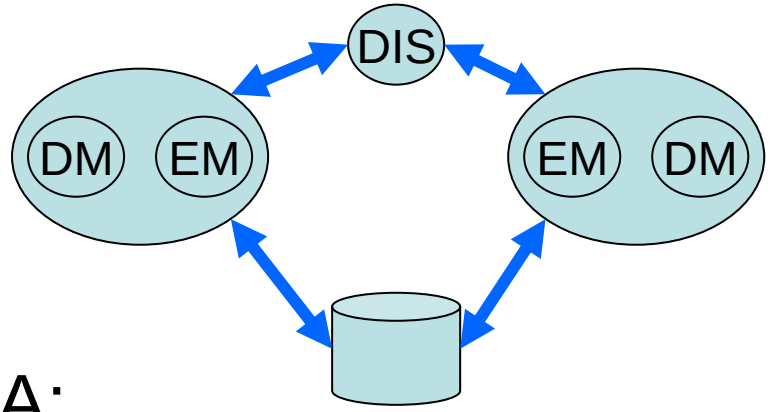
Conditions Data

- ALICE offline is not using COOL
  - DCS needs to provide an API to its data
    - RDB Archive structure is not yet settled down
    - File-based archival is using proprietary format with missing API
- DCS and Offline teams developed AMANDA
  - PVSS API manager
  - Data exchange protocol (over TCP/IP)


DAQ
TRI
HLT
ECS
Config.
DIM, DIP
Electricity
Ventilation
Cooling
Gas
Magnets
Safety
Access Control
LHC
PVSSII
AMANDA
FERO Version Tag
Devices
Archive
Conditions
OFFLINE

PVSS Architecture: Implications on Amanda

- The RDB manager is not threaded safe, one request can be handled at a time
  - Amanda needs therefore to queue the requests, which causes severe performance limitations
  - Even in distributed system, the RDB manager can retrieve data only from it's own archive
  - If data from remote system is needed, its managers will be involved as well



- Implications on AMANDA:
  - Extra load to PVSS systems is added
  - We will need to run at least 1 Amanda per detector

DCS Conditions

- The RDB archival can solve performance problems which we see in Amanda
  - Data can be directly retrieved from the database, no need to involve PVSSI API
- Developments need some time, but can be started only after the situation is clear

- BUT:
  - Conditions data is needed now (TPC commissioning, SPD commissioning, upcoming data challenge …)

- **The DCS software: installation, maintenance**

Software Developments, Installation and Maintenance

- Rules and guidelines discussed in DCS workshops
- Procedures need to be tested and refined
- Software installation procedure:
  - Basic checks done in the lab
  - Software uploaded to production network via the application gateway
  - Configuration, tuning and tests by DCS team and detector experts
- Worry: very often the full tests cannot be performed in advance because the hardware will not be available
  - The associated risk is, that detectors rely on software developments on the production network
- Management of the software installation is a challenge

Software Versions

- The DCS is a big distributed system based on components provided by many parties
  - Policy on software version is inevitable for successful integration
- We are following the FW and PVSS developments
- List of recommended software versions for ALICE DCS is released, all detectors are requested to keep up to date

Version freezing

- Problem: we need to freeze the versions at some point
  - Some detectors are already pre-commissioning now and will not be able/willing to touch their software during the tests
  - Some detectors will be installed in June 06 and will not be fully accessible until April 07
  - Some detectors finished their DCS developments using the existing FW components. The upgradea are painful and require manpower
- *We are aware that the new developments are important and we rely on the new features*
- However, at some point we need to freeze the developments
  - We can make an internal decision in ALICE and compromise on the functionality in favor of a working system, but
  - we need to assure that the recommended components will be supported at least during the next year

Example: PVSSII 3.5 Concerns and Worries

- 3.5 should be ready by end of June
- Compatibility with older releases should be assured by gateway functionality between 3.0 and 3.5, but
  - Can we really profit from it? How much will the new software depend on Qt (and therefore will not be running on 3.0)?
- new version of compiler for Windows is still not yet decided
  - We are running FED servers on the same machines as PVSSII and we are forcing our colleagues to use the compilers compatible with PVSS (to avoid problems with libraries): All FED Servers need to be recompiled !
- What will be the policy for parallel support of 3.5 and 3.0 (e.g. libraries, framework)?
- What are the final deadlines?
  - If the 3.5 is really released in June, it will need some testing. When will be the release date for sub-detectors?
  - How do we react if ETM delays the release?
- PVSS 3.5 will contain many useful features, but some important improvements will not be implemented: changes in alert handling
- If we accept 3.5, it should be really the last version valid for the startup

- **The DCS computing:**
  - **organization**
  - **hardware**
  - **Management and supervision**
  - **Remote access**

- DCS computers fall into one of three categories:
  - Worker nodes (WN) performing the DCS tasks
  - Operator nodes (ON) running the UI
  - Backend servers – providing services for the whole DCS (fileservers, remote access servers, database servers…)
- First batch of DCS computers is being delivered now
  - ON for all detectors
  - WN for DCS infracstructure
- Second (so far the last) batch is being ordered now
- Total number of DCS computers is ~100

DCS Computers

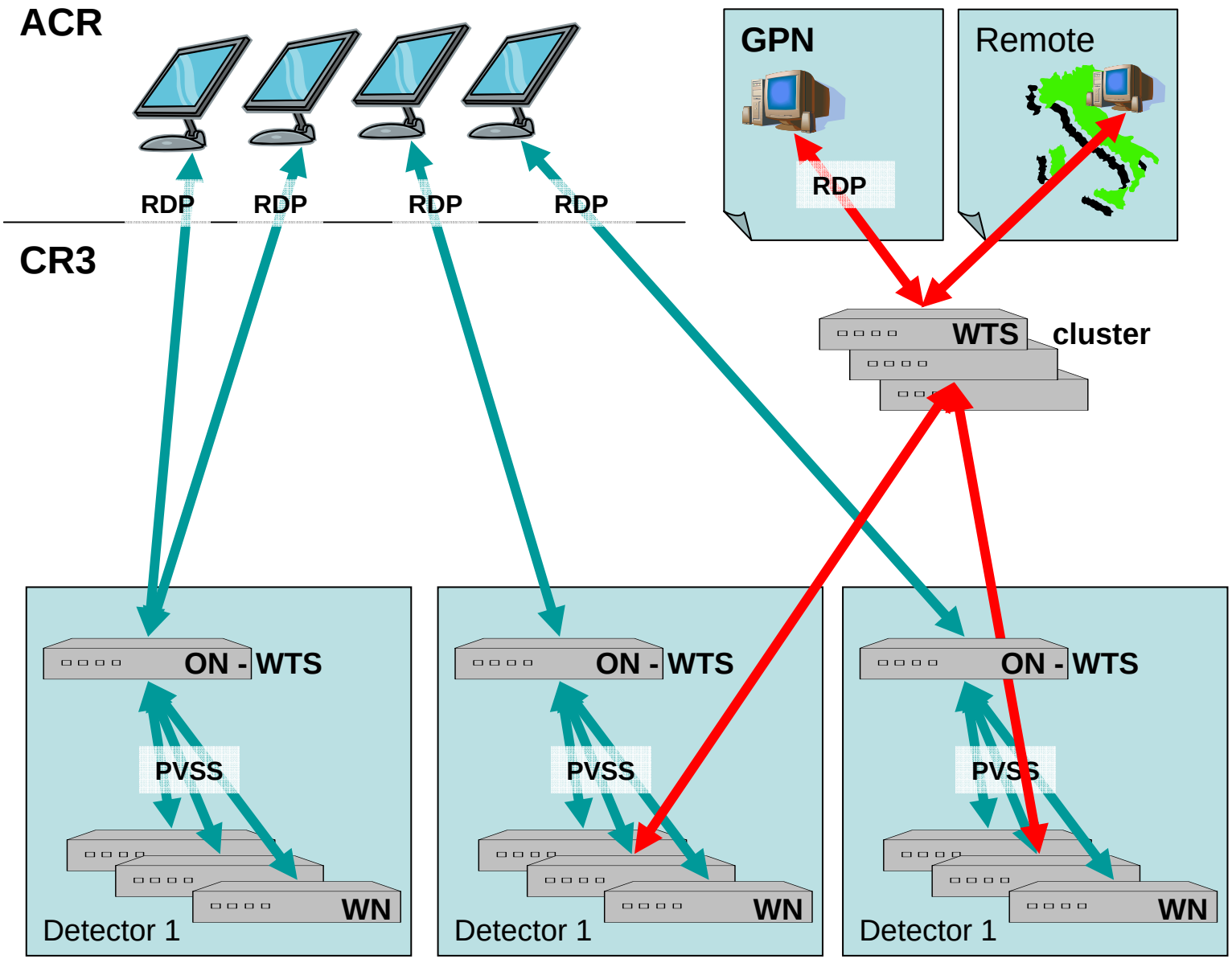
- All machines are based on Intel server boards equipped with dual core CPUs
- Special emphasis was given to HW compatibility tests (the duration of the test cycle involving the ordering of prototypes, tests, tendering and purchasing procedure is comparable with the mainboard production lifetime)
- main technical problem are the 2U PCI risers

- Additional worry: the 5V PCI disappeared on the PIV server boards
- The 3V version has a limited number of available ports/computer (replaced by PCI-E) and will probably also disappear very soon
  - Solution: probably USB
- The selected computer models solve our problems at least for the ALICE startup phase

Accessing the DCS

- Remote access to the DCS is based on the Windows Terminal Service (WTS)
- Access from the ALICE control room (ACR)
  - Consoles will display the UI from the Operator Nodes
- Access from outside
  - Dedicated Windows terminal servers

Remote access to the DCS network



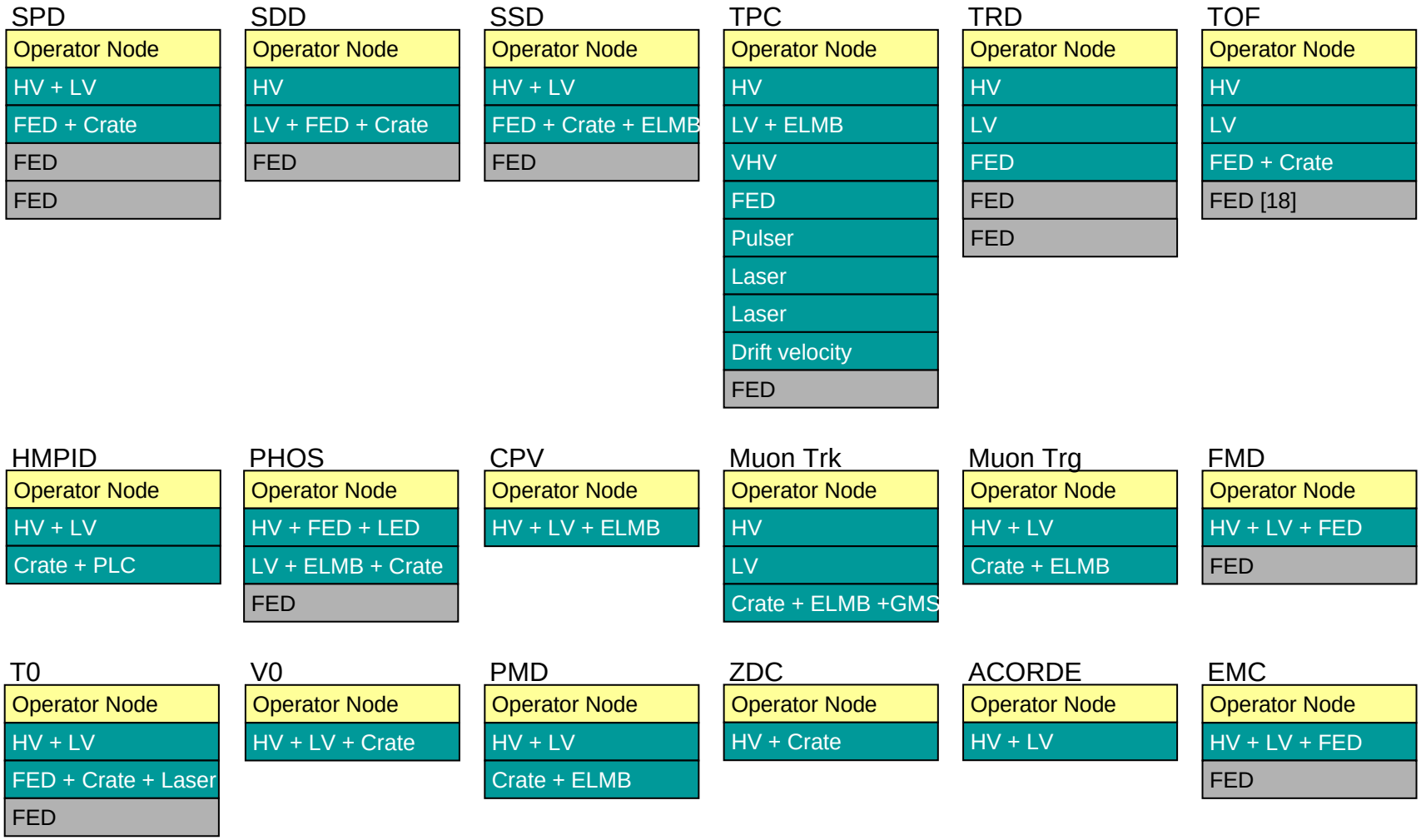
WTS Performance

- **WTS Performance was studied – no problems observed**
  - **Master project generated 50000 datapoints and updated 3000 /s.**
  - **Remote client displayed 50 values at a time**

Terminal Server		
#clients	Average CPU load [%]	Mem [kB]
60	11.2	2788719
55	11.0	2781282
45	13.8	2790181
35	12.0	2672998
25	9.7	2067242
15	7.2	1448779
5	4.2	934763
0	4.9	666914

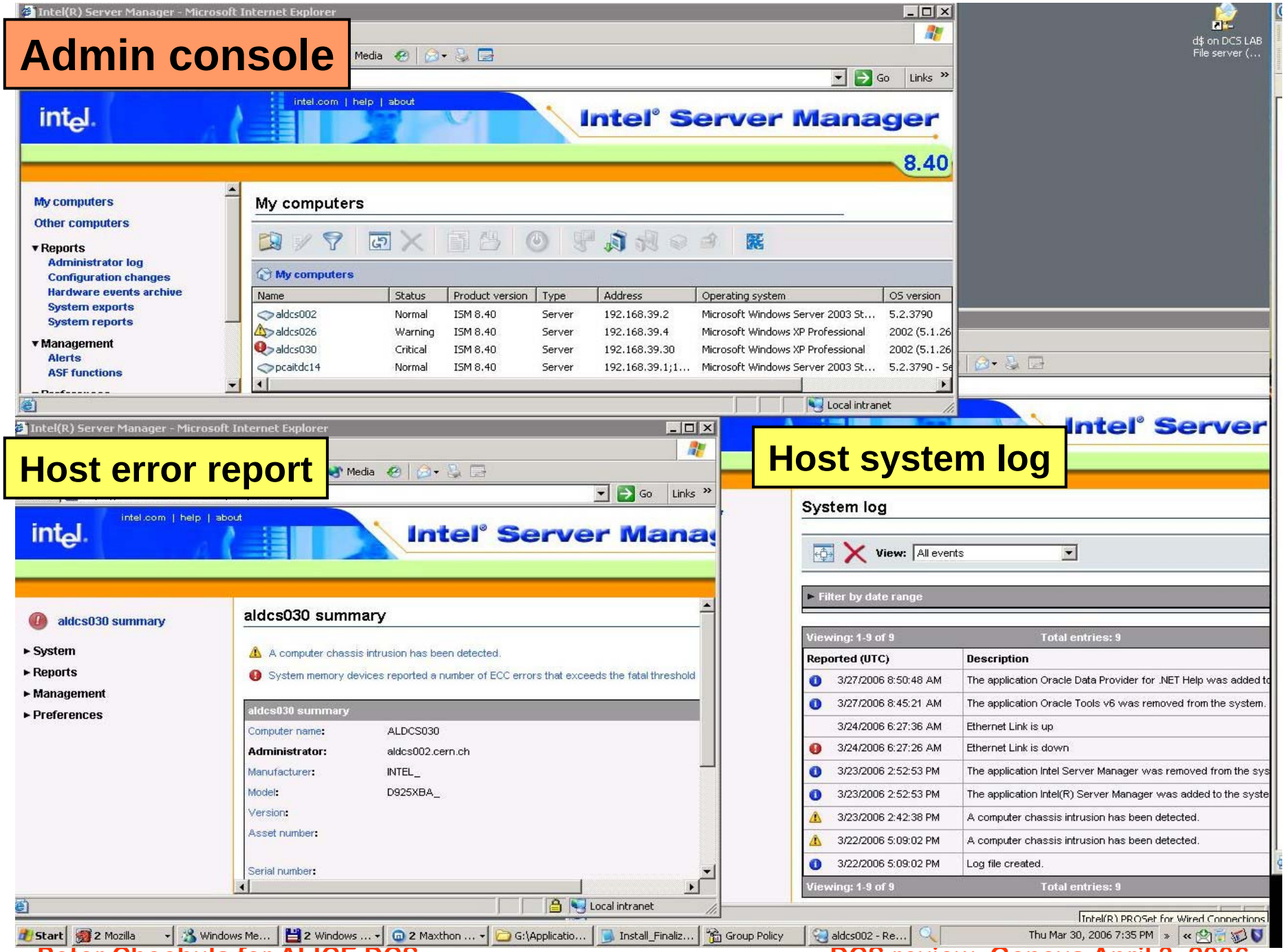
Workstation running the DCS project		
#clients	Average CPU load [%]	Mem [kB]
60	85.1	579666
55	86.6	579829
45	84.9	579690
35	81.3	579405
25	80.9	579384
15	81.4	579463
5	83.0	580003
0	83.7	579691

DCS Computers – Services and Back-End not Included

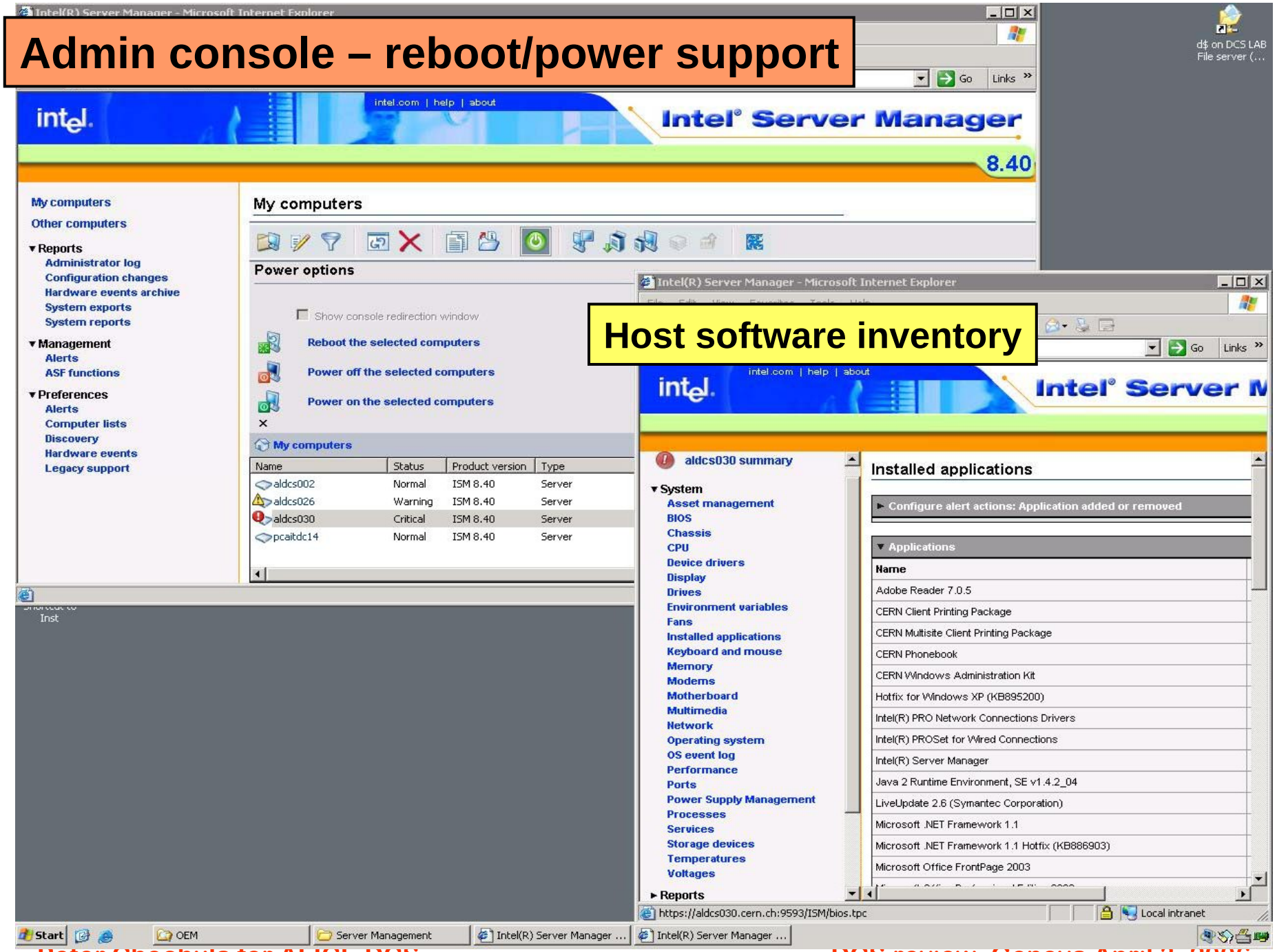


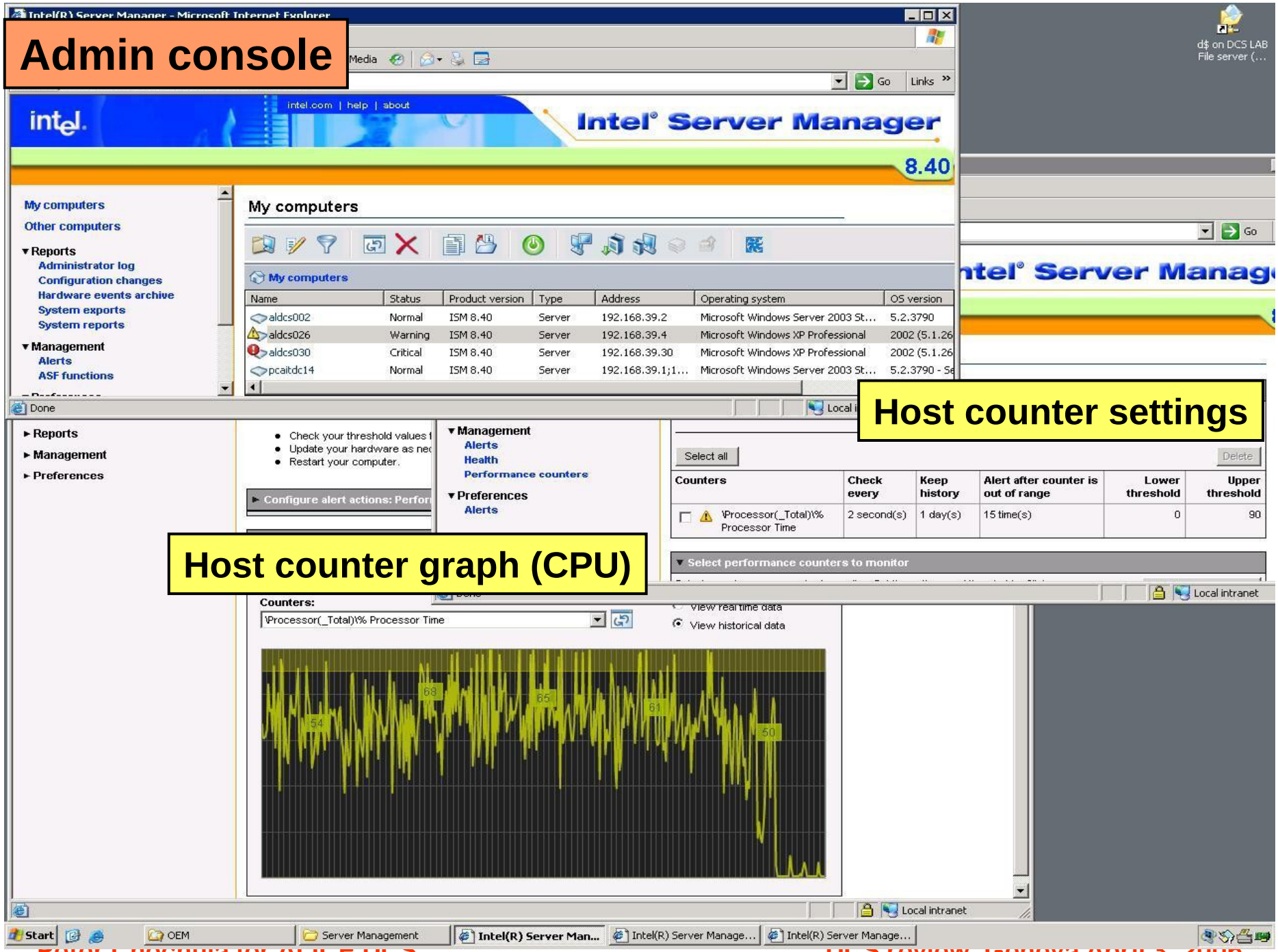
Monitoring the D CS farms – Intel Server Management (ISM)

- ISM selected as the temporary monitoring and supervision tools
  - IPMI-based
  - Windows/Linux monitoring agent
  - Out-of-band monitoring (supported mainboards)
    - We are using Intel server boards everywhere in the DCS, but this might change n the future
  - Admin console (subnet monitoring), web GUI
  - Alerts, logs, counters, graphs
  - Software monitoring (logs changes)


Admin console
Intel(R) Server Manager - Microsoft Internet Explorer
intel.com | help | about
Intel® Server Manager
8.40
My computers
Other computers
Reports
Administrator log
Configuration changes
Hardware events archive
System exports
System reports
Management
Alerts
ASF functions
My computers
Name
Status
Product version
Type
Address
Operating system
OS version
aldcs002 Normal ISM 8.40 Server 192.168.39.2 Microsoft Windows Server 2003 St... 5.2.3790
aldcs026 Warning ISM 8.40 Server 192.168.39.4 Microsoft Windows XP Professional 2002 (5.1.26
aldcs030 Critical ISM 8.40 Server 192.168.39.30 Microsoft Windows XP Professional 2002 (5.1.26
pcaitdc14 Normal ISM 8.40 Server 192.168.39.1;1... Microsoft Windows Server 2003 St... 5.2.3790 - Se
Local intranet
Host error report
Intel(R) Server Manager - Microsoft Internet Explorer
aldcs030 summary
System
Reports
Management
Preferences
A computer chassis intrusion has been detected.
System memory devices reported a number of ECC errors that exceeds the fatal threshold
Computer name: ALDCS030
Administrator: aldcs002.cern.ch
Manufacturer: INTEL_
Model: D925XBA_
Version:
Asset number:
Serial number:
Host system log
System log
View: All events
Filter by date range
Viewing: 1-9 of 9
Total entries: 9
Reported (UTC)
Description
3/27/2006 8:50:48 AM The application Oracle Data Provider for .NET Help was added to
3/27/2006 8:45:21 AM The application Oracle Tools v6 was removed from the system.
3/24/2006 6:27:36 AM Ethernet Link is up
3/24/2006 6:27:26 AM Ethernet Link is down
3/23/2006 2:52:53 PM The application Intel Server Manager was removed from the sys
3/23/2006 2:52:53 PM The application Intel(R) Server Manager was added to the syste
3/23/2006 2:42:38 PM A computer chassis intrusion has been detected.
3/22/2006 5:09:02 PM A computer chassis intrusion has been detected.
3/22/2006 5:09:02 PM Log file created.
Start
Thu Mar 30, 2006 7:35 PM

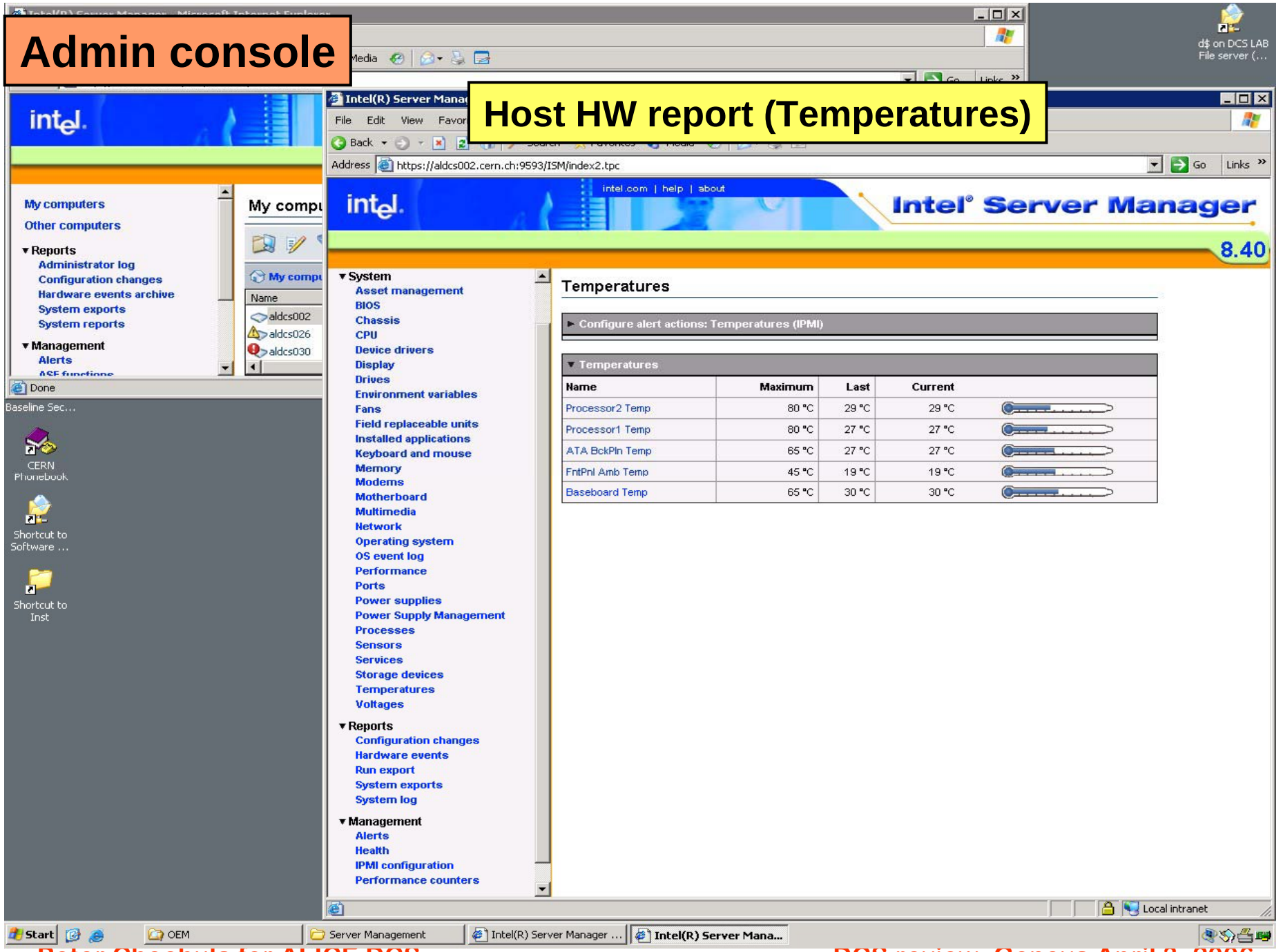
Admin console – reboot/power support




Admin console
Intel(R) Server Manager - Microsoft Internet Explorer
intel.com | help | about
Intel® Server Manager
8.40
My computers
Other computers
Reports
Administrator log
Configuration changes
Hardware events archive
System exports
System reports
Management
Alerts
ASF functions
My computers
Name
Status
Product version
Type
Address
Operating system
OS version
aldcs002
Normal
ISM 8.40
Server
192.168.39.2
Microsoft Windows Server 2003 St...
5.2.3790
aldcs026
Warning
ISM 8.40
Server
192.168.39.4
Microsoft Windows XP Professional
2002 (5.1.26
aldcs030
Critical
ISM 8.40
Server
192.168.39.30
Microsoft Windows XP Professional
2002 (5.1.26
pcaitdc14
Normal
ISM 8.40
Server
192.168.39.1;1...
Microsoft Windows Server 2003 St...
5.2.3790 - Se
Done
Host counter settings
Reports
Management
Preferences
Check your threshold values t
Update your hardware as nec
Restart your computer.
Configure alert actions: Perfor
Management
Alerts
Health
Performance counters
Preferences
Alerts
Select all
Delete
Counters
Check every
Keep history
Alert after counter is out of range
Lower threshold
Upper threshold
\Processor(_Total)\% Processor Time
2 second(s)
1 day(s)
15 time(s)
0
90
Host counter graph (CPU)
Select performance counters to monitor
Local intranet
Counters:
\Processor(_Total)\% Processor Time
View real time data
View historical data
68
65
61
54
50
Local intranet
Start
OEM
Server Management
Intel(R) Server Man...
Intel(R) Server Manage...
Intel(R) Server Manage...

Admin console

Host HW report (Temperatures)



OS Maintenance

- OS management is
  - Following CNIC architecture and
  - Based on NICEFC and LinuxFC
- Participating in evaluation of NICEFC
  - CMF
  - Remote system installation
- We appreciate the help of IT (Ivan), CMF will be used in the production cluster
- Minor concern is connected to application packaging – distribution of PVSS patches