



ALICE

Trigger and DAQ

LHC Days in Split
October-2004

P.Vande Vyvre - CERN/PH

for the ALICE DAQ project
(Birmingham, Budapest, CERN,
Mexico, Split, Zagreb)



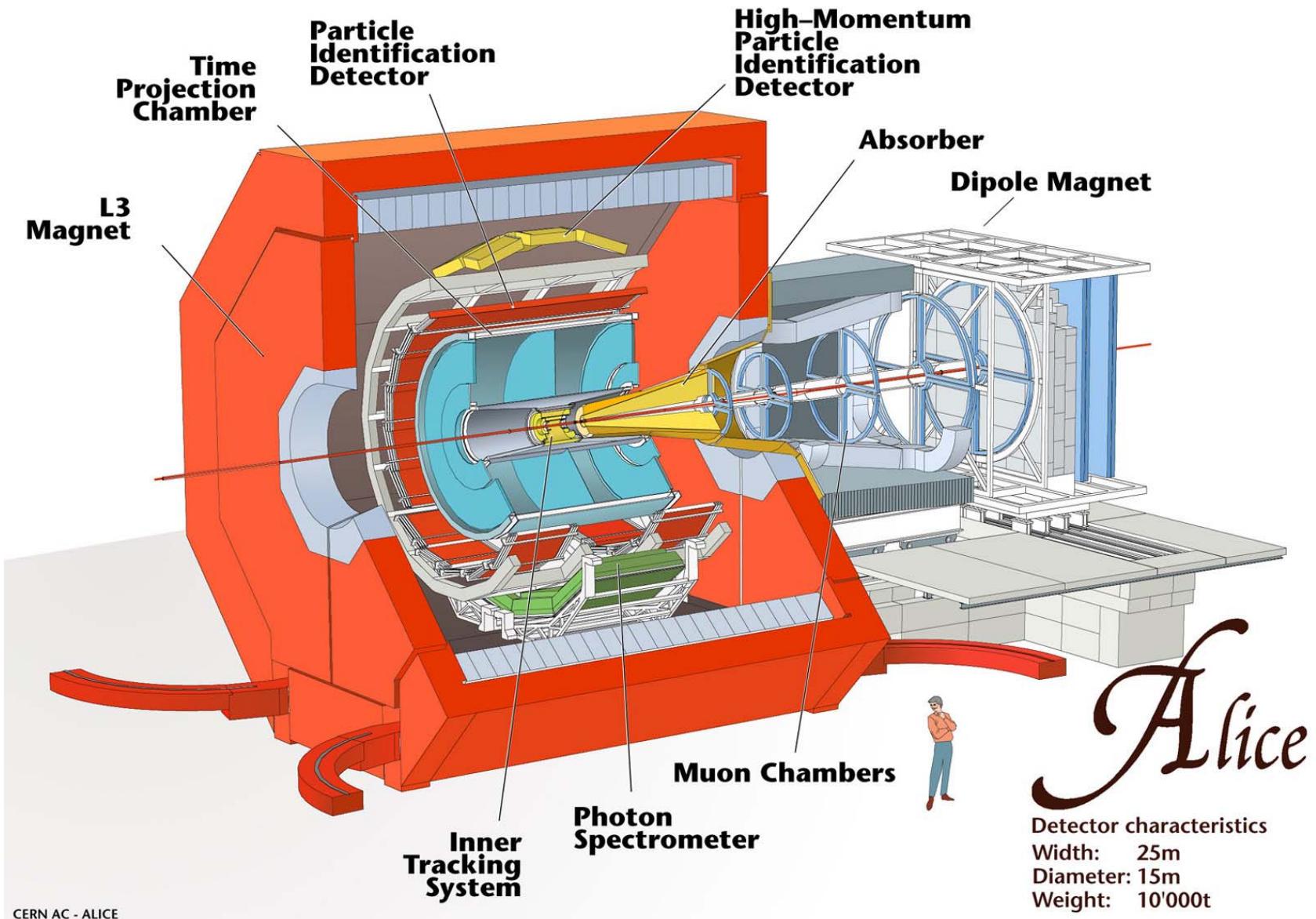
Outline

- ALICE DAQ
 - Trigger/DAQ logical model - Requirements
 - Trigger/DAQ at LHC
 - ALICE Trigger/DAQ Architecture
- Hardware components
- Software packages
- Towards the final ALICE
- Conclusions



Outline

- ALICE DAQ
 - Trigger/DAQ logical model - Requirements
 - Trigger/DAQ at LHC
 - ALICE Trigger/DAQ Architecture
- Hardware components
- Software packages
- Towards the final ALICE
- Conclusions





Trigger & DAQ logical model

**Low-Level
Trigger
System**

**Detector
Electronics**

**High-Level
Trigger
System**

**DAQ
System**

Storage (TDS)



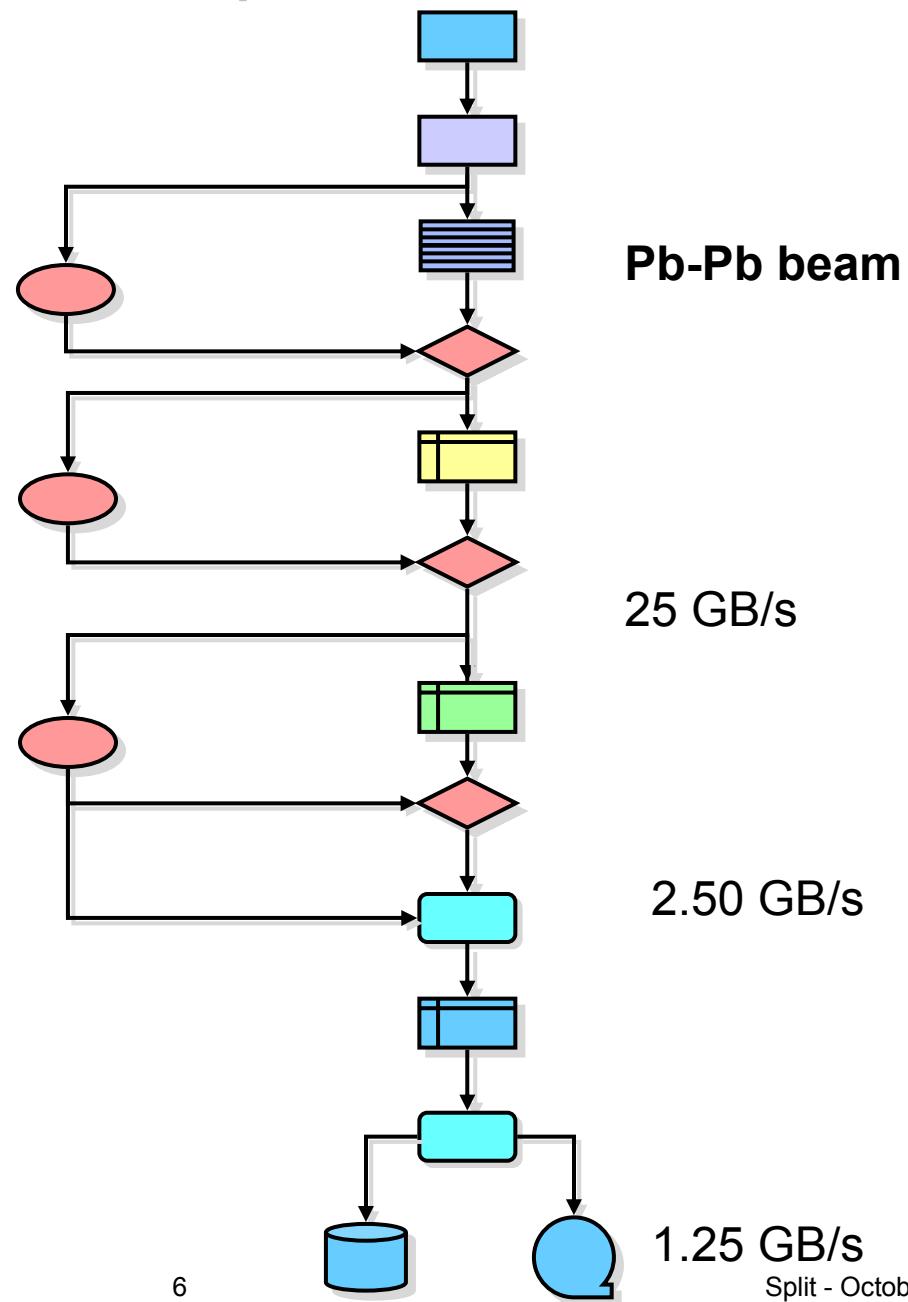
ALICE Physics requirements

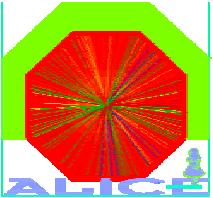
Pb-Pb beam

	Rate	Max. ev. size
- Central	20 Hz	86.0 MB
- MB	20 Hz	20.0 MB
- Dimuon	1600 Hz	0.5 MB
- Dielectron	200 Hz	9.0 MB

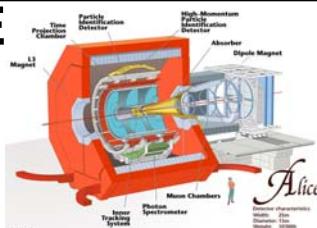
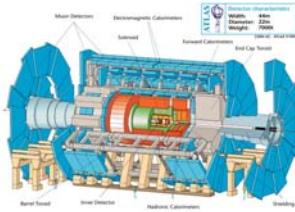
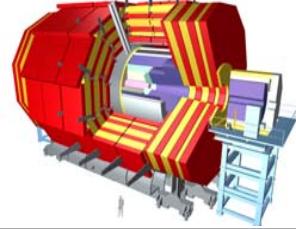
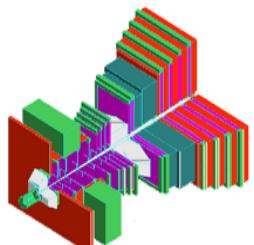
pp beam

MB	100 Hz	2.5 MB
----	--------	--------





DAQ @ LHC

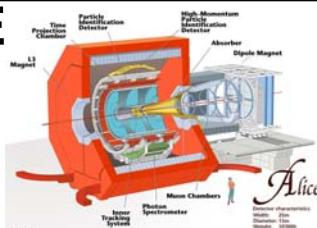
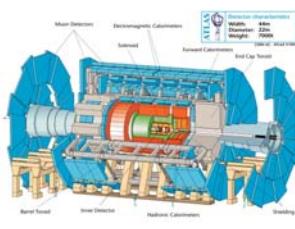
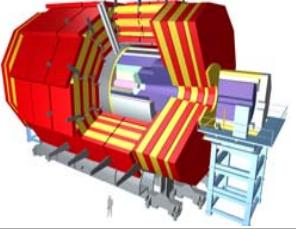
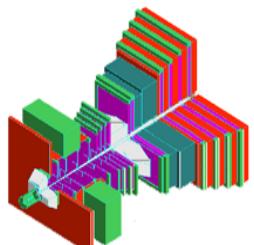
	LHC Beam	Event Size (Byte)	Readout (HLT input) (Events/s.)	Readout (GB/s)	
ALICE		Pb-Pb	4×10^7	2×10^3	25
ATLAS		pp	10^6	2×10^3	10
CMS		pp	10^6	10^5	100
LHCb		pp	2×10^5	40×10^4	4

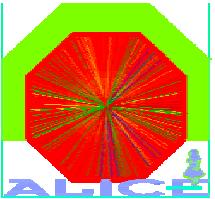


Mass Storage @ LHC

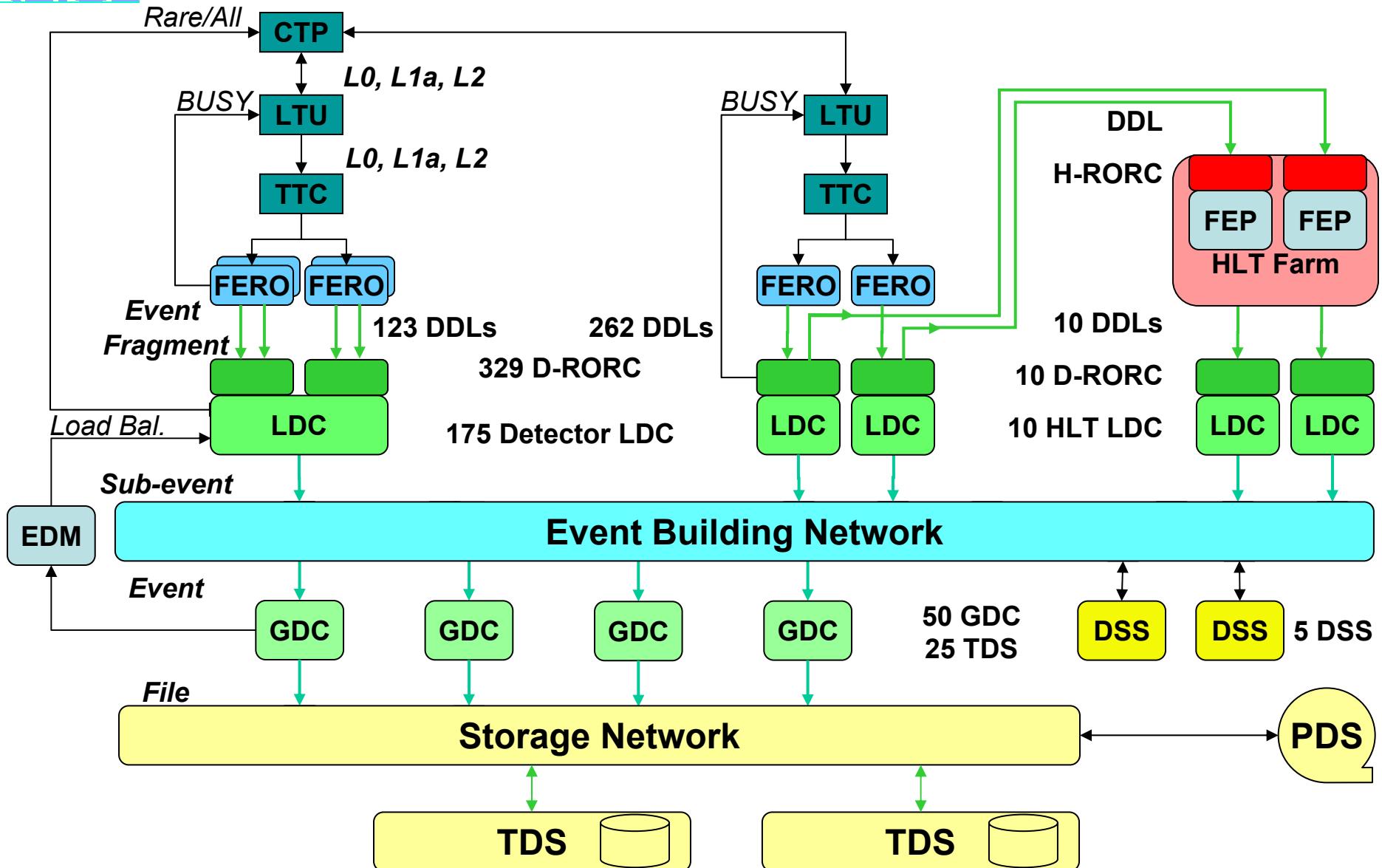
Recording
(Mass Storage)
(Events/s.) (MB/s)

Data archived
Total/year
(PBytes)

Experiment	Collision Type	Events/s.	MB/s	PBytes
ALICE	Pb-Pb	200	1250	2.3
				
ATLAS	pp	100	100	6.0
				
CMS	pp	100	100	3.0
				
LHCb	pp	200	40	1.0
				



ALICE TRG/DAQ architecture





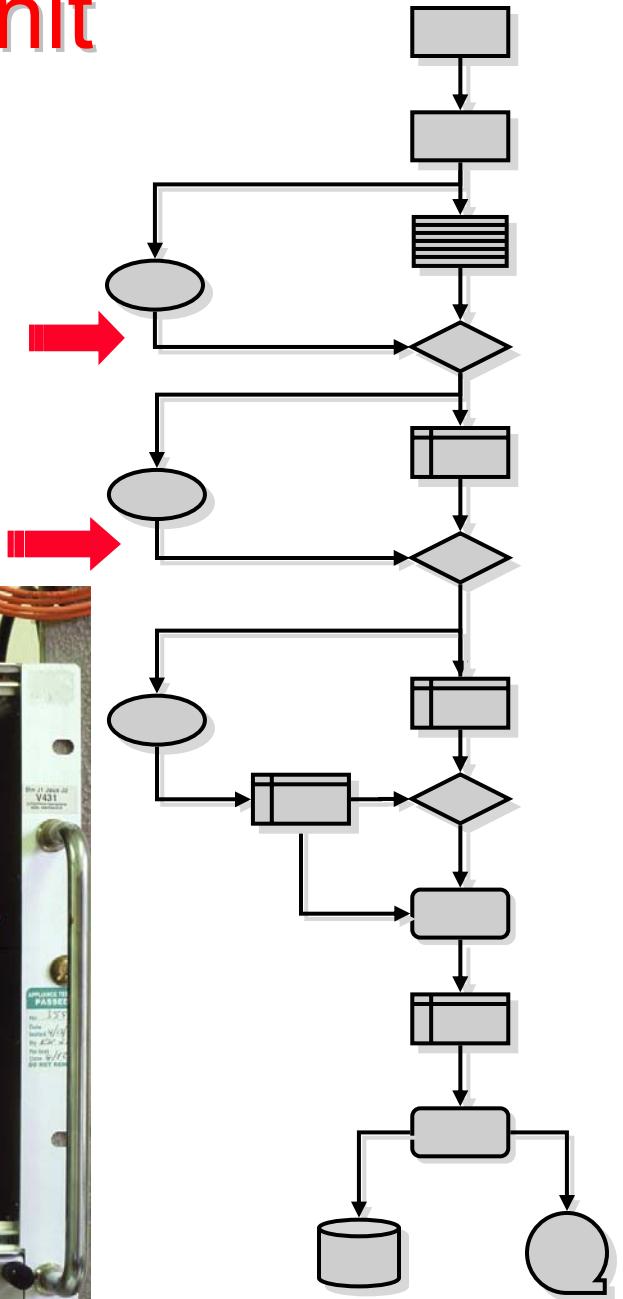
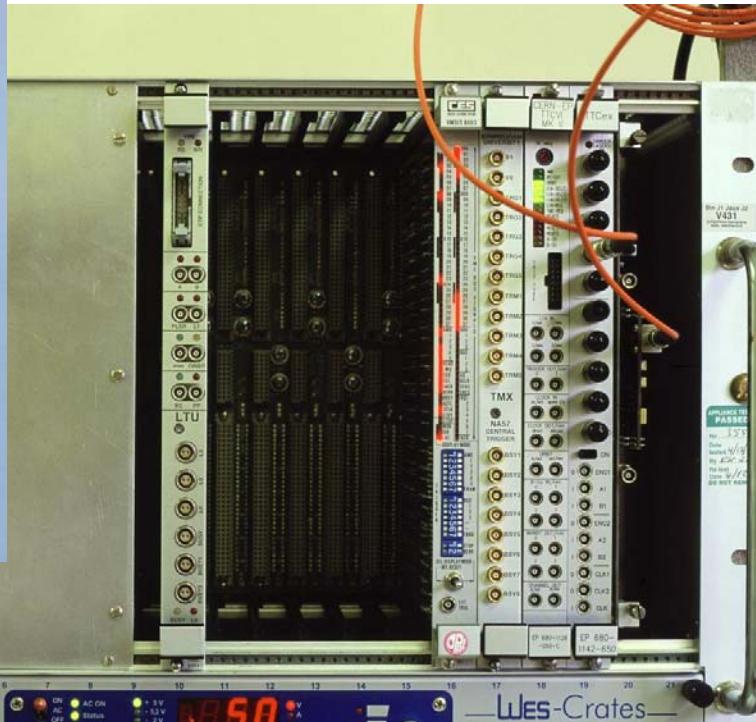
Outline

- ALICE DAQ
 - Trigger/DAQ logical model - Requirements
 - Trigger/DAQ at LHC
 - ALICE Trigger/DAQ Architecture
- **Hardware components**
- Software packages
- Towards the final ALICE
- Conclusions



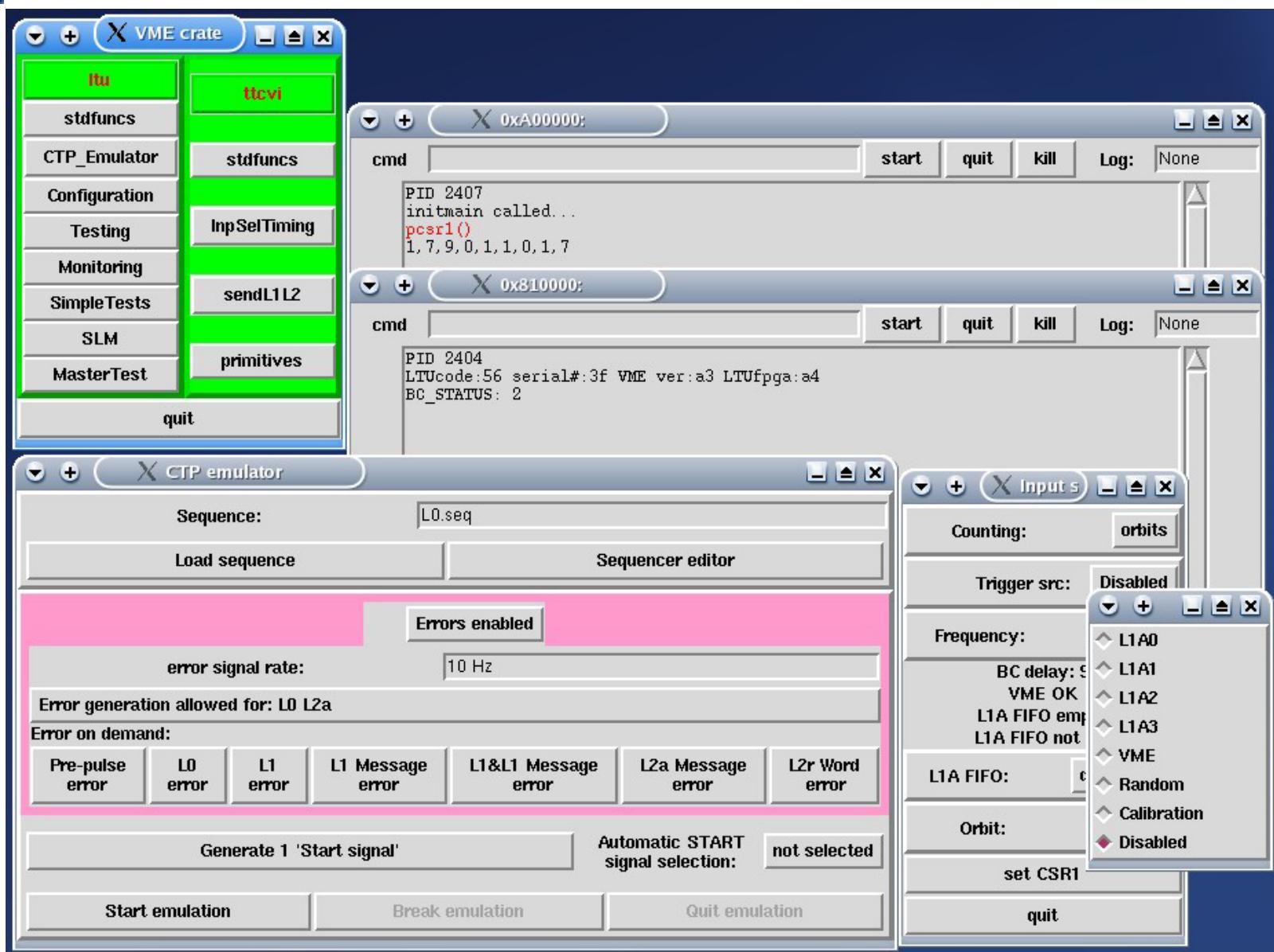
Local Trigger Unit

- ◆ Common interface between Central Trigger Processor to TTC tree of each detector





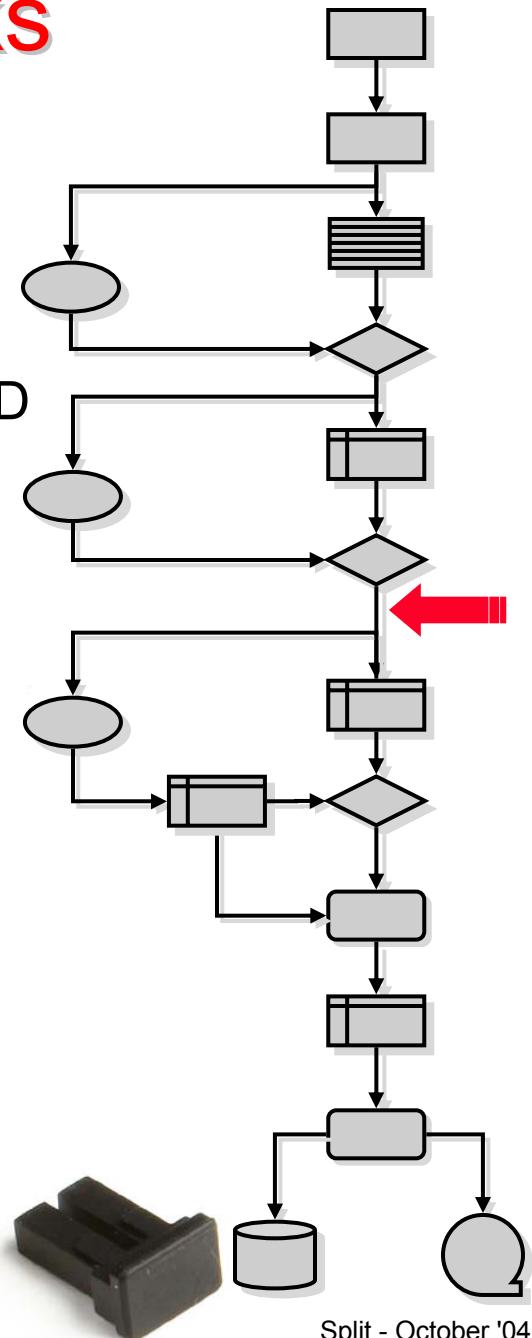
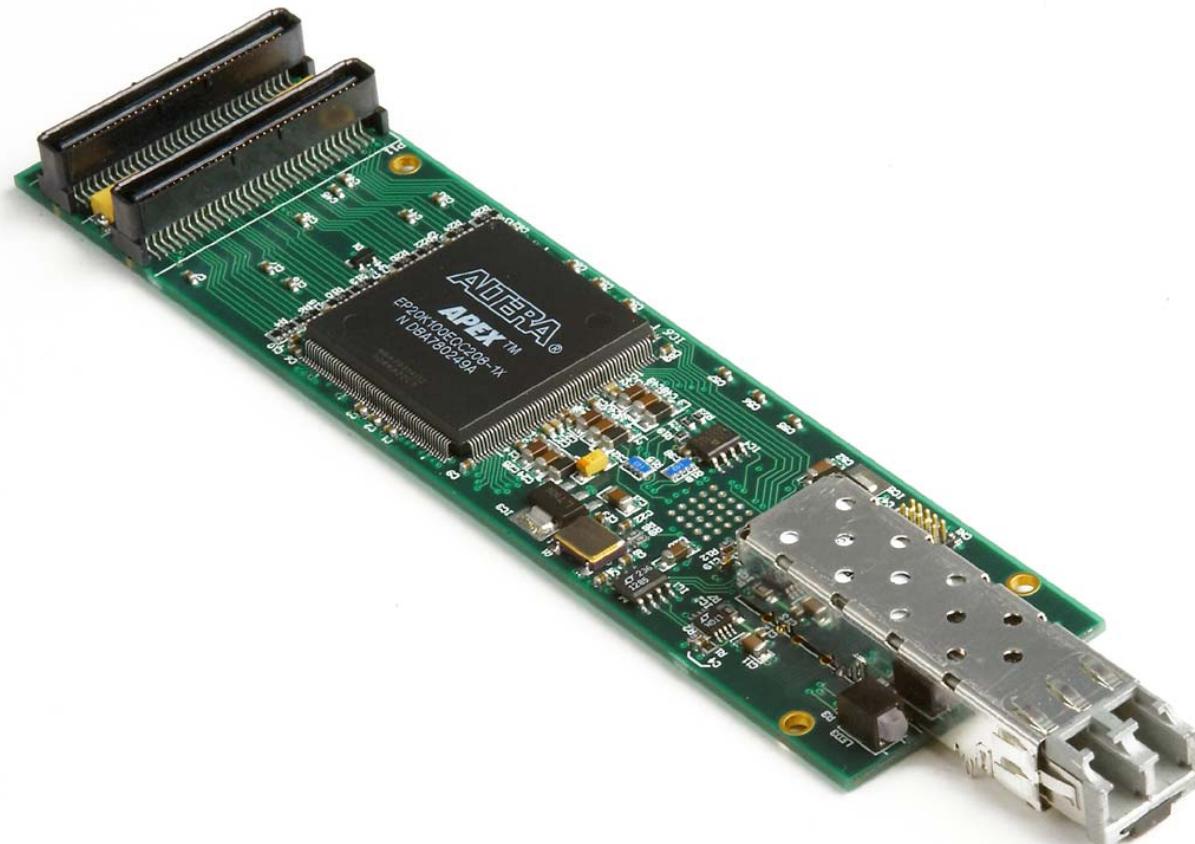
Trigger Control Software





Detector Data Links

- ◆ Standard and stable interface detector/DAQ
 - ◆ Point-to-point full-duplex digital data link
 - ◆ Massive parallelism (100's)
 - ◆ Integrated with SDD, TPC, TRD, TOF, Muon, HMPID

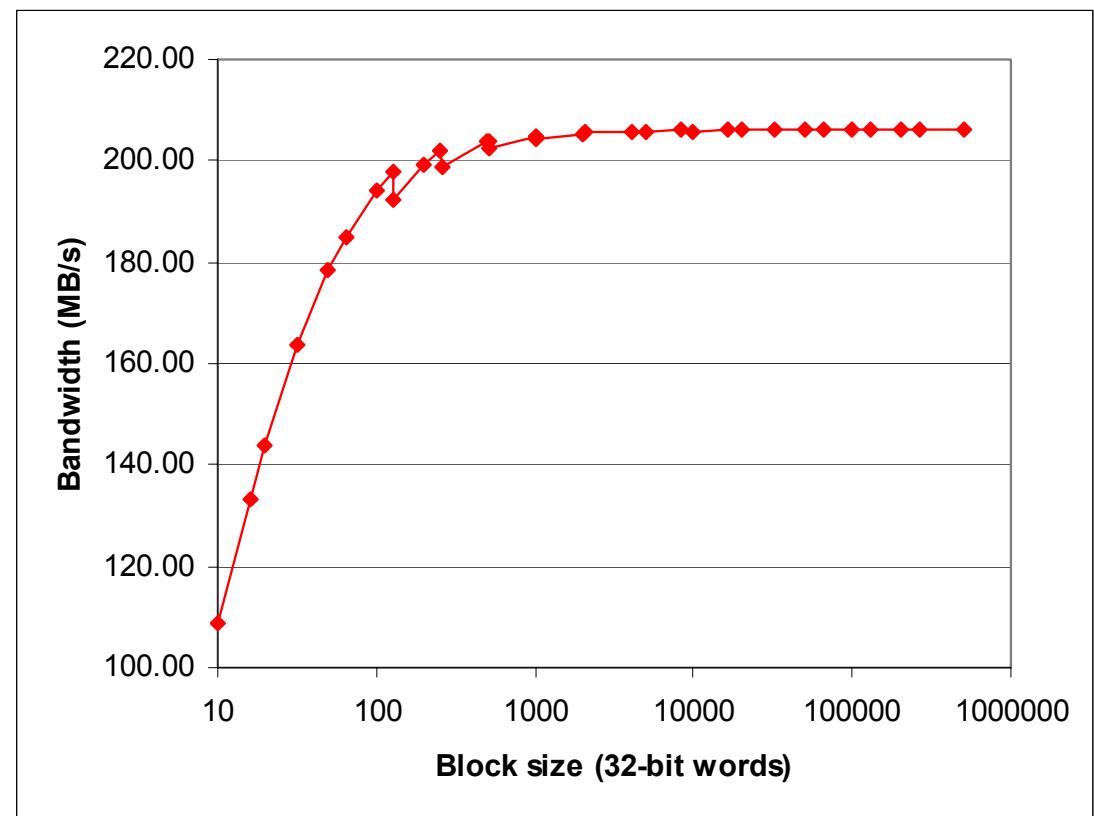




Link performance

- ◆ HEP development based on commodity components:
 - Fiber Channel or Gig. Ethernet: 2.125 Gb/s
 - Optical transceiver 850 nm VCSEL
 - Flash-based FPGA (Radiation tolerant)

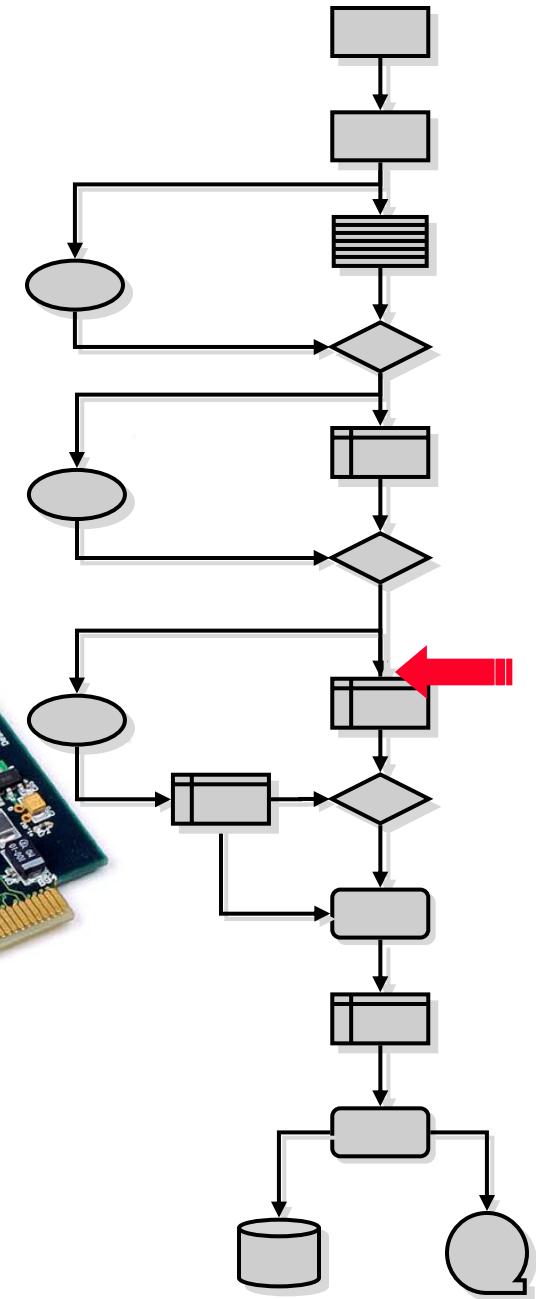
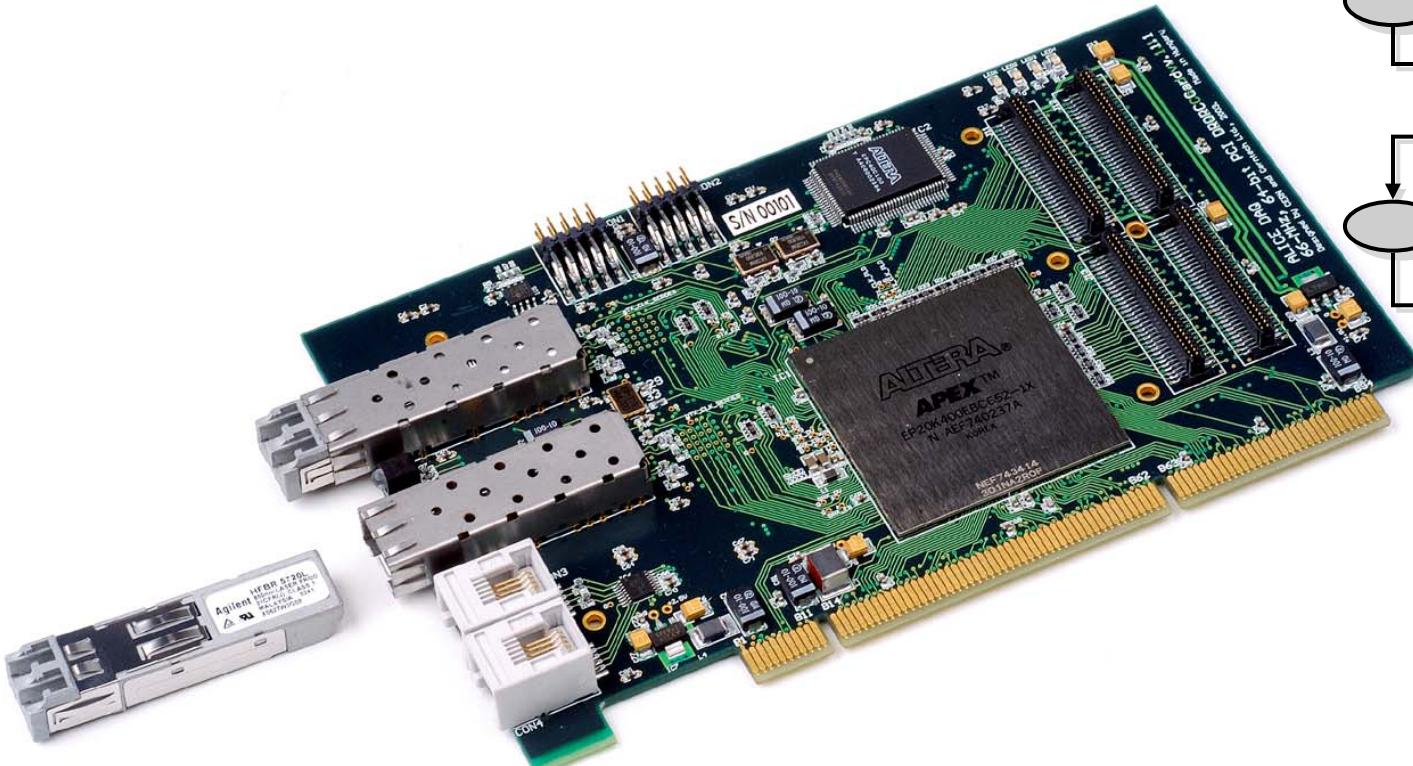
200 MB/s sustained
Lots of bw available.
Major fraction available to
end application.





Links Adapters

- ◆ Adapter for 1 or a few links to PC I/O bus
- ◆ A few-to-one multiplexer
- ◆ Massive parallelism (100's)



Split - October '04



Link and adapter performance

- ◆ Currently PCI and PCI-X busses. PCI-XP (PCI Express) in the future.
- ◆ No large local memory. Fast transfer to PC memory
- ◆ PCI interface
 - IP core (VHDL code synthesized in FPGA)
 - PCI 64 bit 66 MHz. Master enabled.

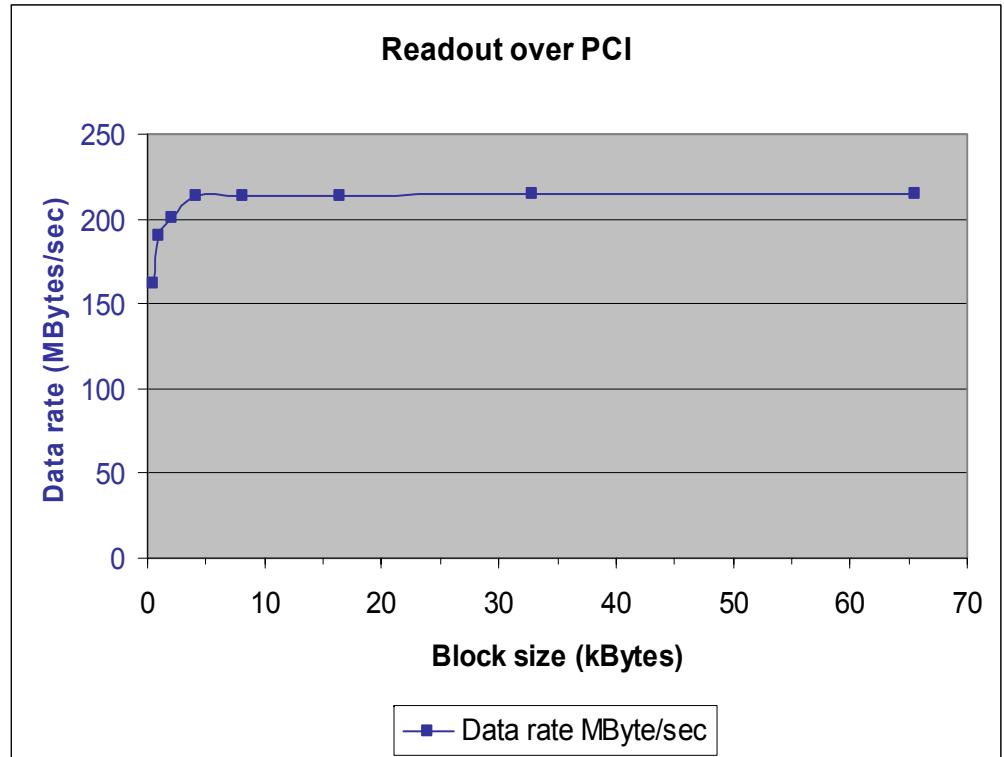
200 MB/s sustained

Total PCI load: 92 %

Data transfer PCI load: 83 %

Lots of bw available.

Major fraction available to
end application

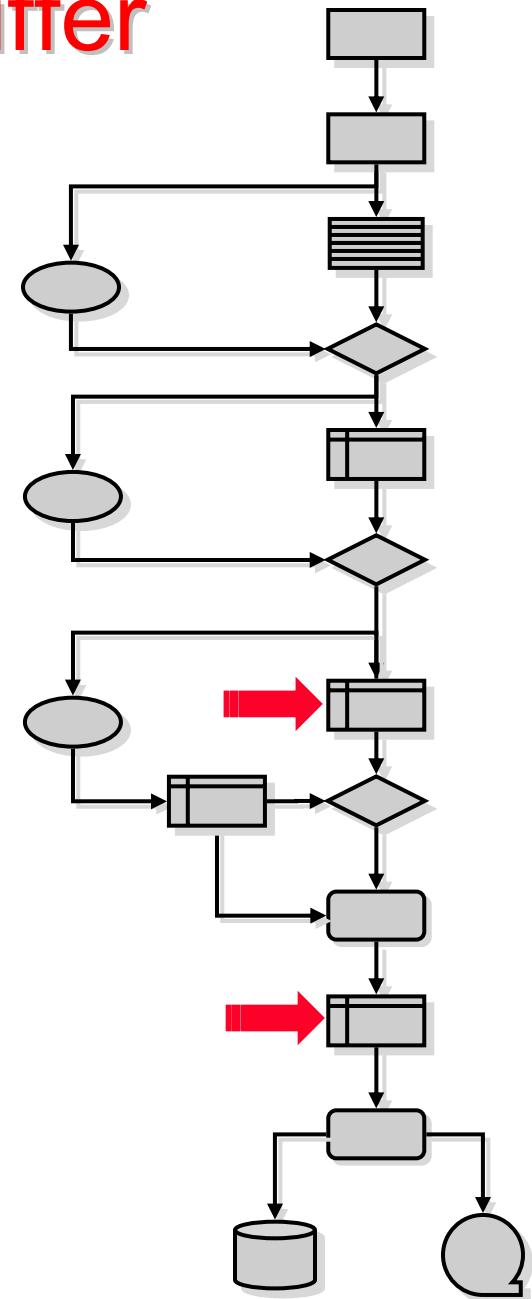




Subevent & event buffer

- ◆ Baseline:
 - Fast dual-port memories
 - Electronics racks are over
 - Extensive use of dual-CPUs PCs

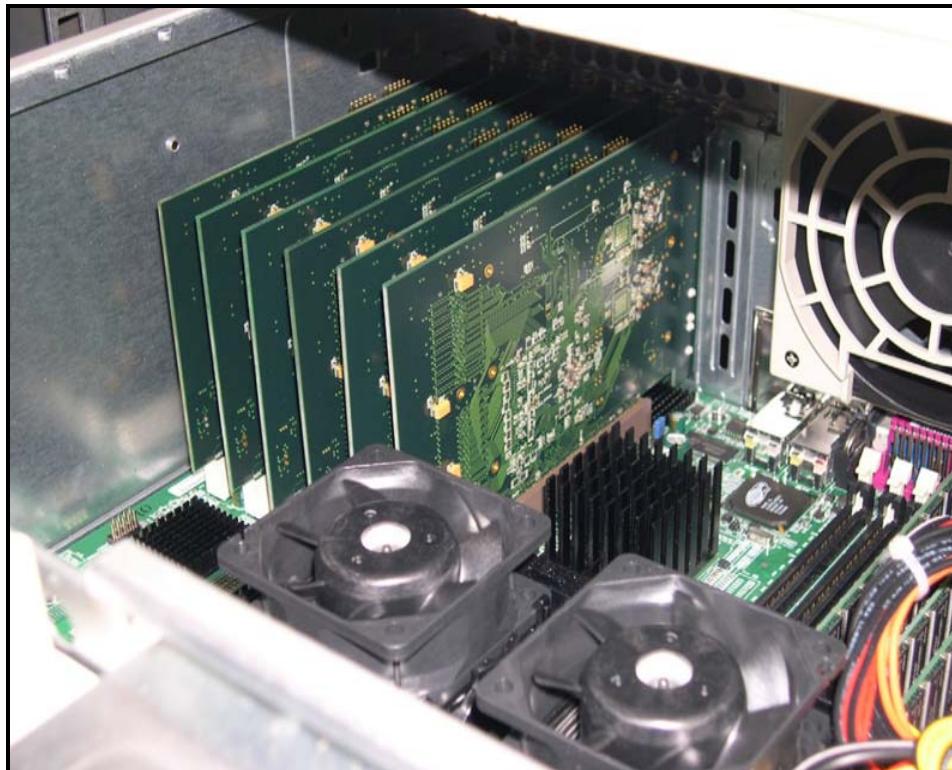
- ◆ Key parameters:
 - Cost/performance
 - Performance: I/O and memory bandwidth





Readout System Performance

- ◆ Supermicro server motherboard with dual Xeon CPUs @ 2.4 GHz
 - In the future: multicore CPUs from Intel and AMD
- ◆ Six PCI-X slots, 4 bus segments (3+1+1+1)
- ◆ Linux OS
- ◆ ALICE Data-Acquisition software (DATE)





Performance: 6 D-RORCs

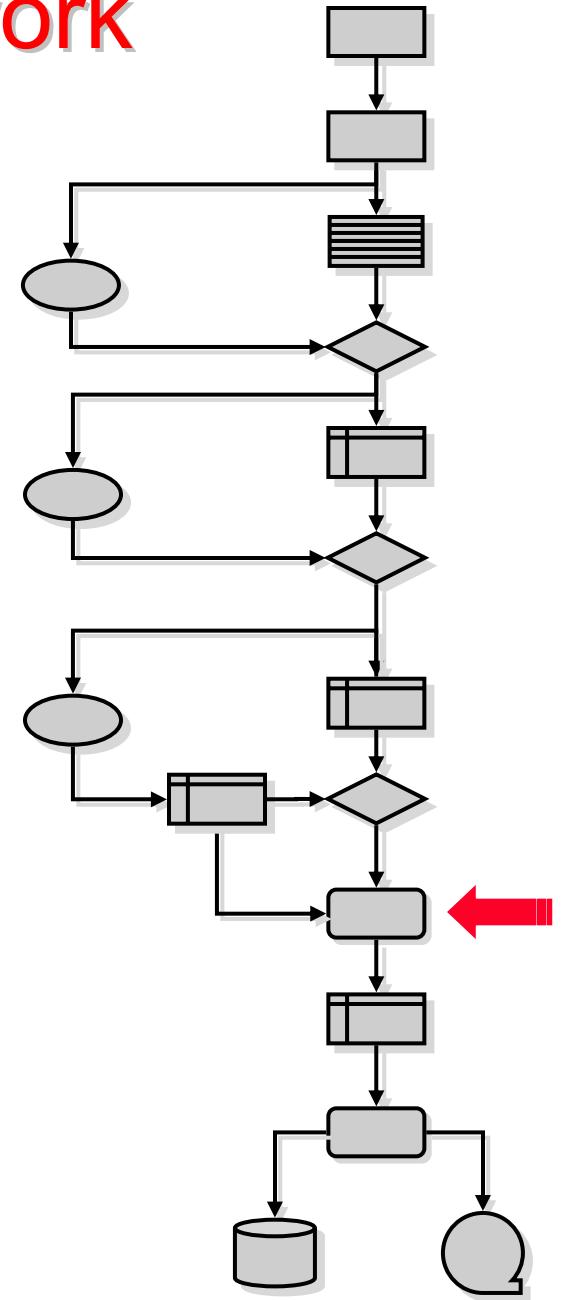
- ◆ Testing the fully populated PC using data source internal to PCI interface
 - Interoperability test
 - Measure the maximal input bandwidth

		#4	PCI #6						1 Ch	1 Ch
		#3	PCI #5						1 Ch	1 Ch
Controller #1	#2	PCI #4							1 Ch	1 Ch
	Segment #1	PCI #3							1 Ch	1 Ch
		PCI #2							1 Ch	1 Ch
		PCI #1							1 Ch	1 Ch
									1 Ch	1 Ch
		Bandwidth [MB/s]								
		Normalized Bandwidth [MB/s/Ch]								
		264	464	424	528	792	1045	840		
		264	232	141.3	264	264	261.3	140		



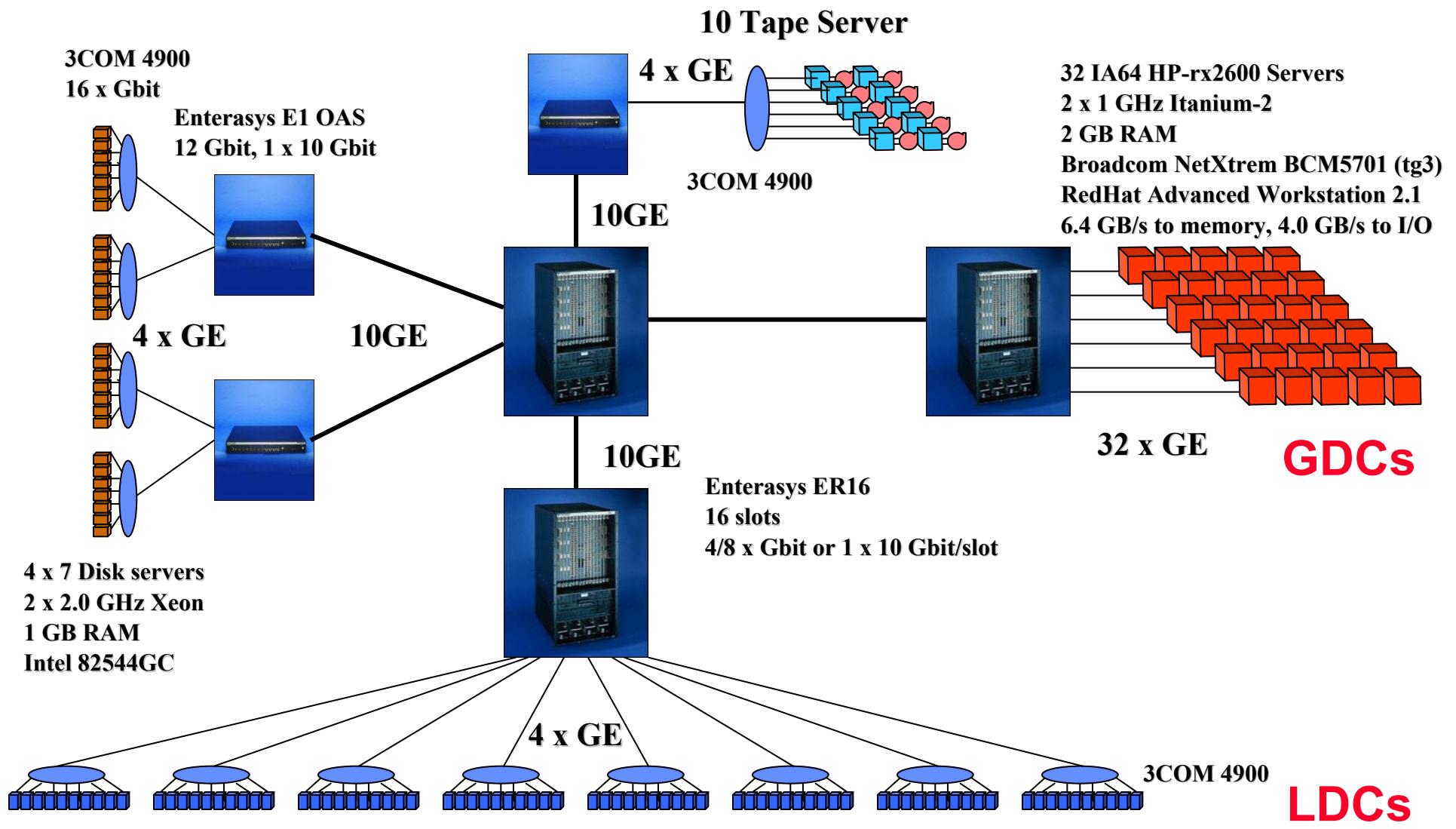
Event Building Network

- ◆ Baseline:
 - Adopt broadly exploited standards
Switched Ethernet and TCP/IP
- ◆ Motivations for switched Ethernet:
 - Performance of Gigabit Ethernet switches already adequate: 2 Tbit/s of aggregate bandwidth
 - Use of commodity items: network switches and interfaces
 - Easy (re)configuration and reallocation of resources
- ◆ Network Interface Card (NIC)
 - ◆ TCP/IP Offload Engine (TOE)
Dedicated processor to execute IP stack
 - ◆ 10 Gbit Ethernet in the PCs



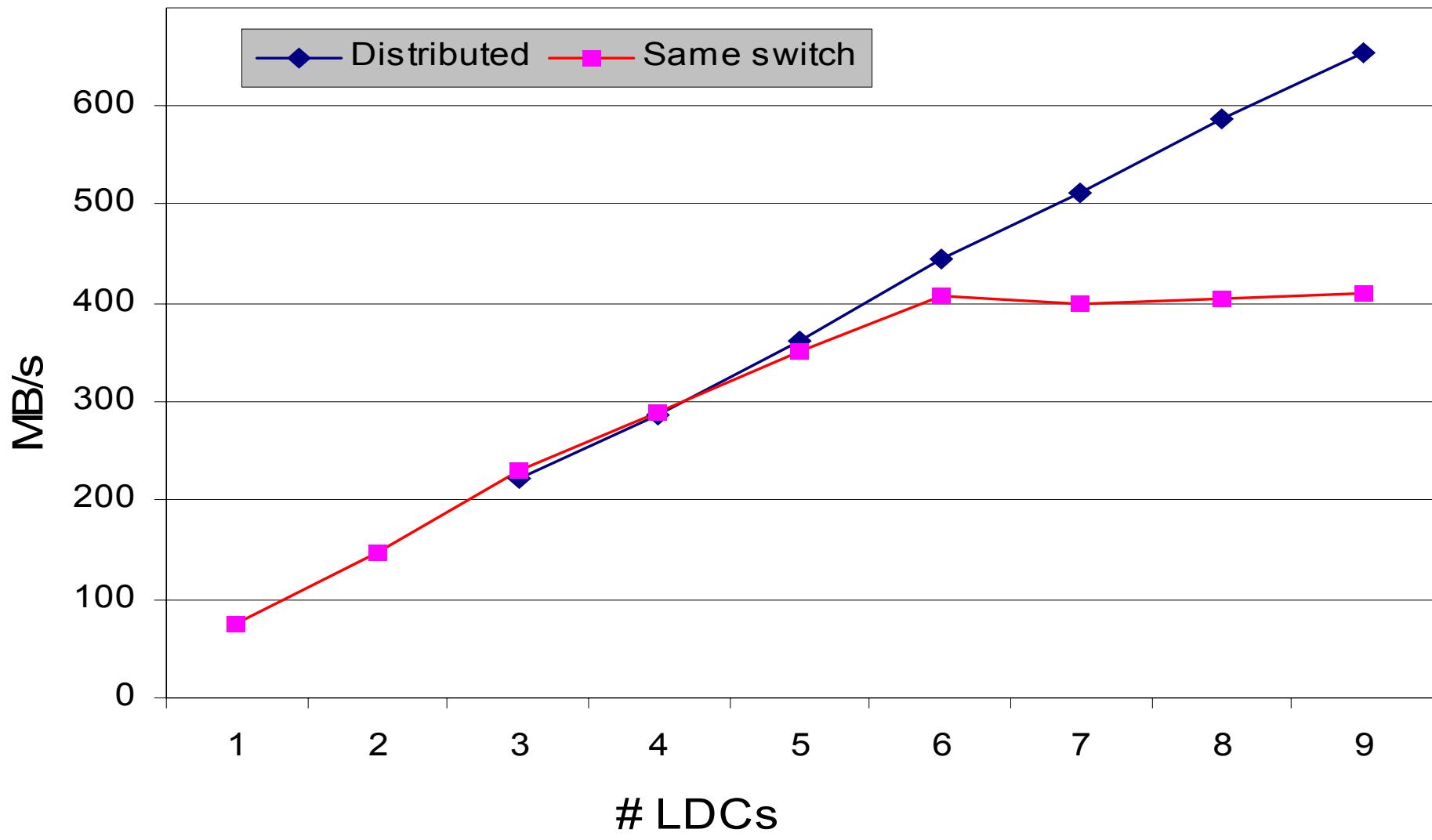


2003 ALICE Data Challenge (ADC V)





ADC V Trunking

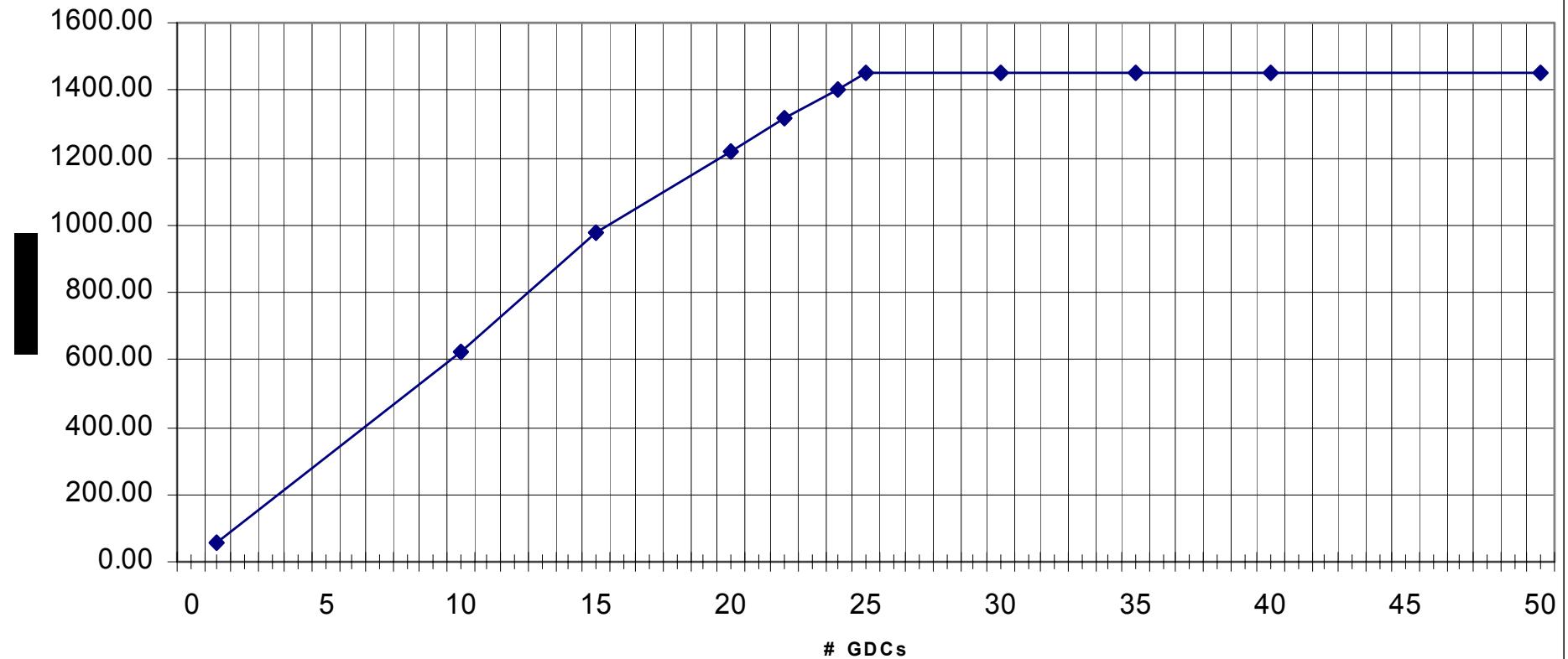


Trunk of 4 x Gb Eth



Scalability of network-based event building

DATE COLE EQUIPMENT FLAT, 1 MB events, 21 LDCs



- Reliability and scalability of the whole system
- Throughput limited by data sources

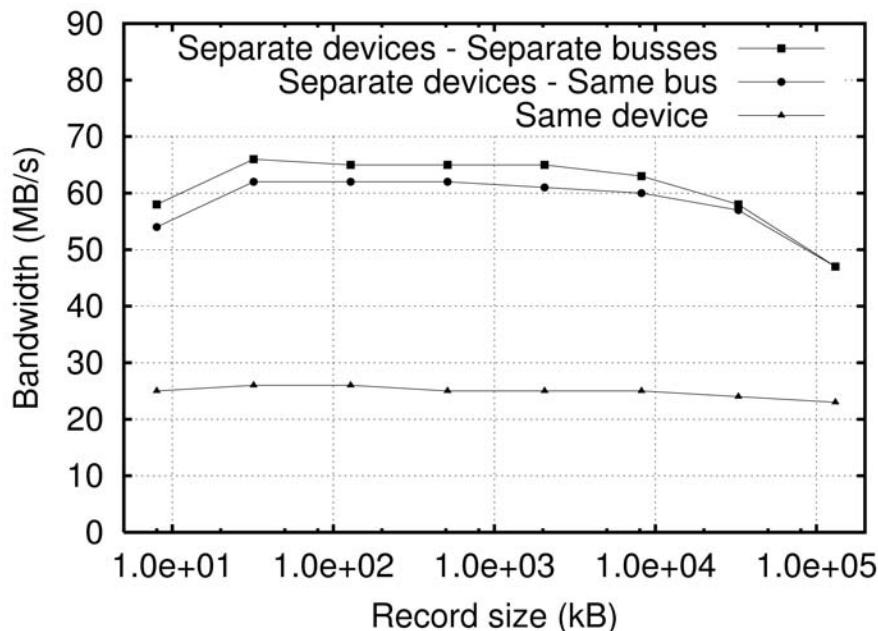
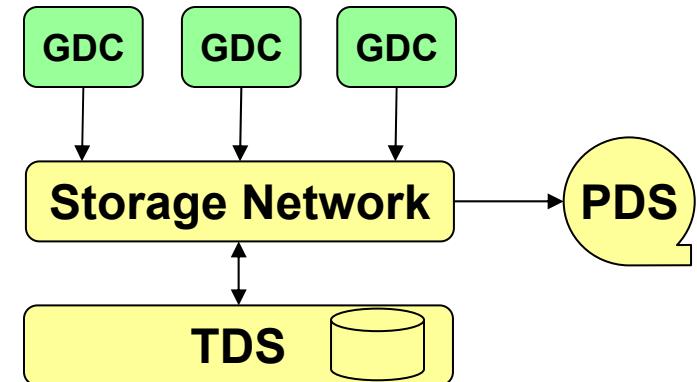
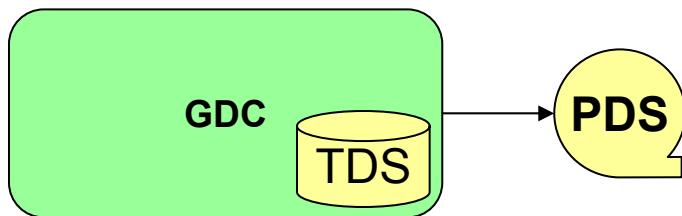


Mass Storage System

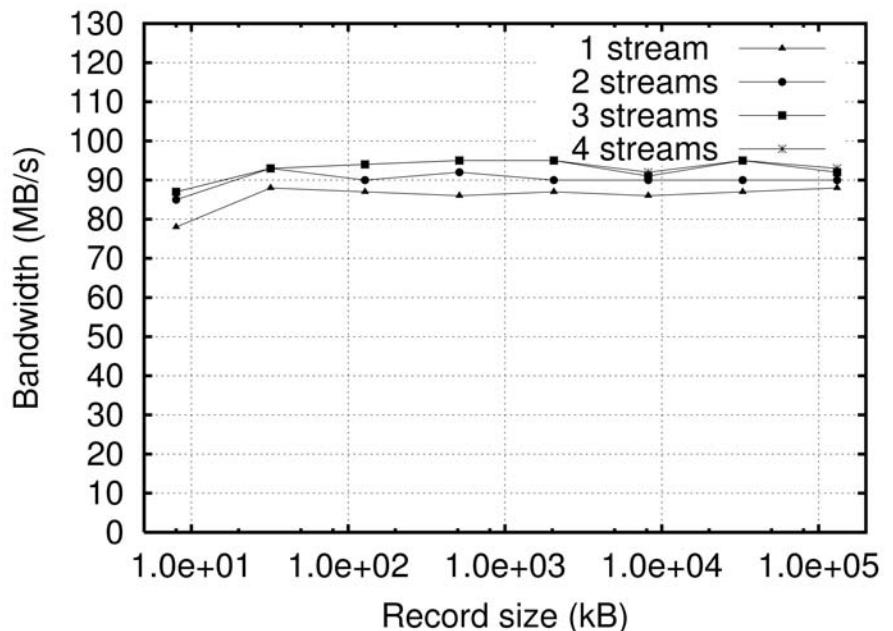
- ◆ Logical model: distributed file-system
- ◆ Software implementation
 - CASTOR system developed by CERN/IT
 - Accessibility via the GRID
- ◆ Baseline hardware implementation
 - Transient Data Storage
 - Located at the experimental area
 - Capacity: a few hours of autonomous data taking
 - Before archiving to tertiary storage, if any
 - Permanent Data Storage
 - Located in the computing centre
 - Infinite capacity, very low cost
 - Single write and a few read
 - Sufficient performance to achieve performances with reasonable number of parallel streams and media operations
 - ◆ 1 GByte/s: 40 active streams at ~30 MB/s
 - ◆ 1 media operation every 2'30" with 200 GB/volume
 - Open to new implementations thanks to software model



Possible Transient Storage



- ◆ TDS: local GDC storage
- ◆ Storage attachment: internal IDE

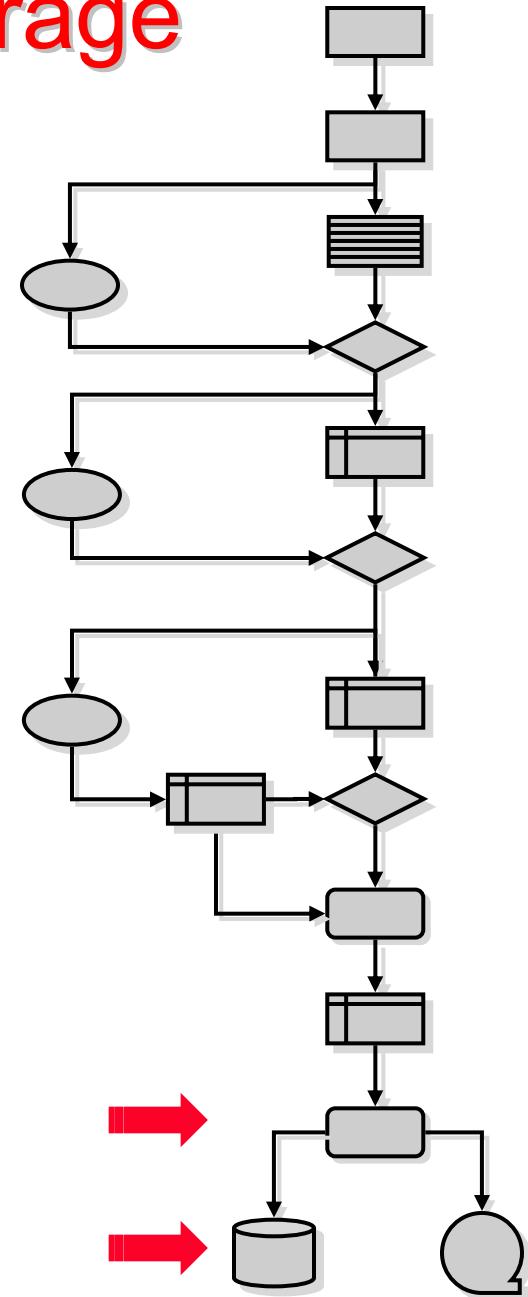


- ◆ TDS: storage array
- ◆ Storage attachment: Fibre Channel



Transient Data Storage

- ◆ Transient data storage at the experimental area
 - ◆ Baseline
 - Storage arrays of commodity disks
 - Box attachment: Fibre Channel
 - Disk attachment: IDE or serial-ATA
 - RAID-level
 - ◆ Partnering with industry for test of equipment
 - ◆ Key selection criteria:
 - Cost/performance
 - Bandwidth/box
 - Robustness

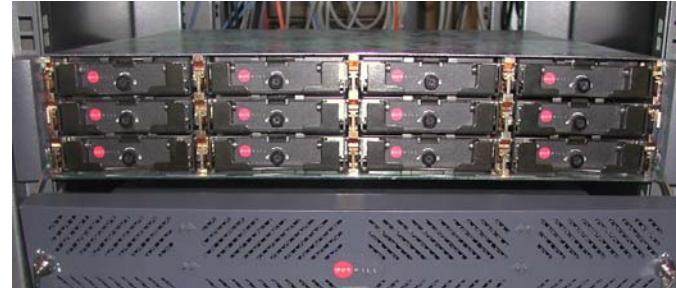




Storage Arrays

◆ dotHILL SANnet II 200 FC

- 12 fiber channel disk slots
- 1 GB cache
- 1 x 2Gbit fiber host channel



◆ Infortrend IFT-6330

- 12 IDE drive slots
- 128 MB cache
- 2 x 2Gbit fiber host channels



◆ Infortrend EonStor A16F-G1A2 (INFN CASPUR Storage Lab)

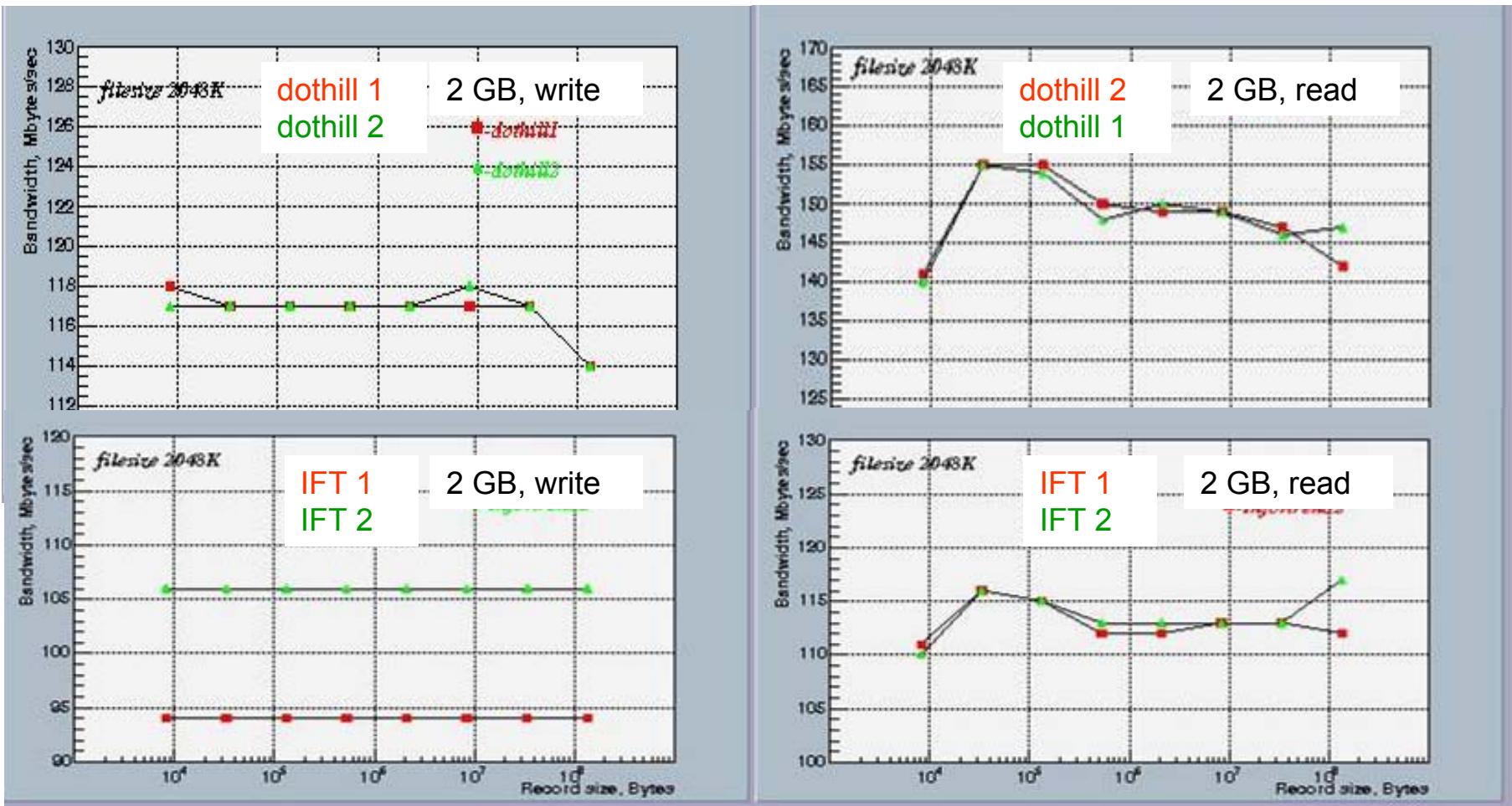
- 16 SATA drive slots
- 1GB cache
- 2 x 2Gbit fiber host channel





Storage Arrays Performance

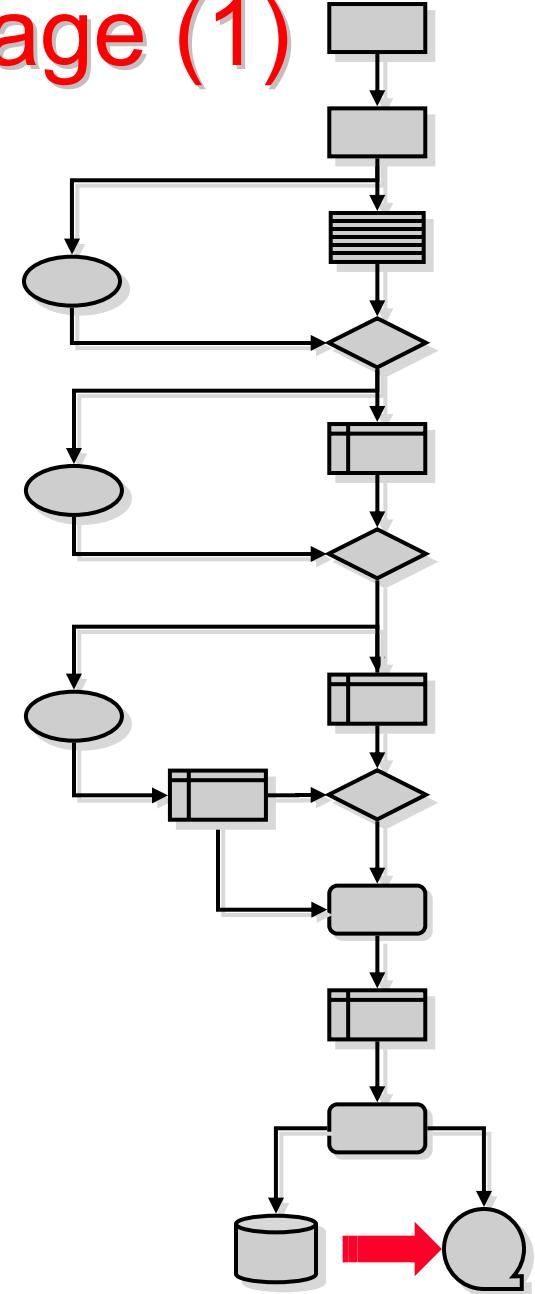
- Aggregate throughput measured for
 - Set of 5 disks configured as RAID 5
 - Filesizes of 100, 300, 1024 and **2048 MB**
 - Recl=8 kB, 32 kB, 128 kB, 512 kB, 2 MB, 8 MB, 32 MB and 128 MB





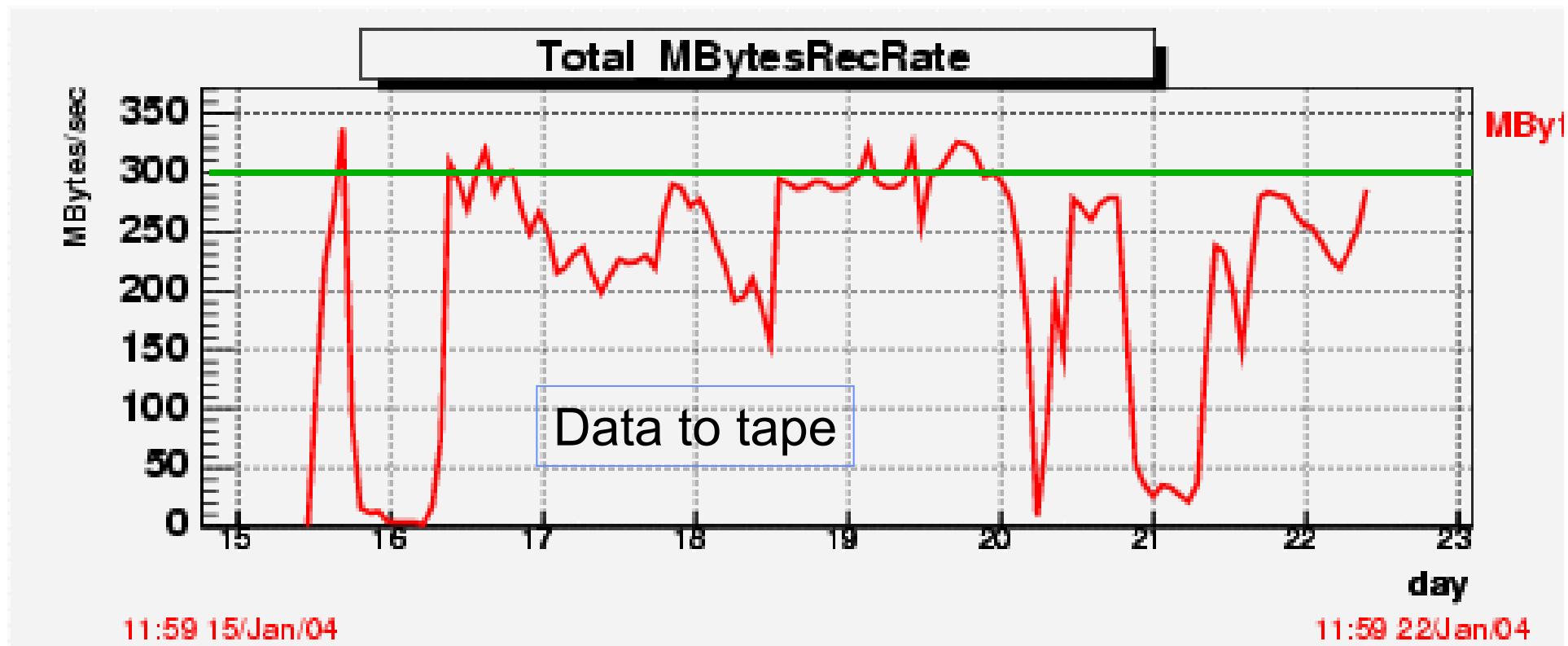
Permanent Data Storage (1)

- ◆ Permanent data storage in the computing centre
- ◆ Baseline solution
 - Magnetic tape
- ◆ Critical areas
 - High Energy Physics peculiar use of tapes
 - Infrastructure hidden by a hierarchical storage management sw
 - Limited market, different application
 - Limited competition
 - No demonstrated alternative yet
- ◆ Demonstrated solution for LHC
 - 15 parallel streams





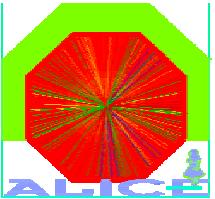
Performance of Permanent Data Storage



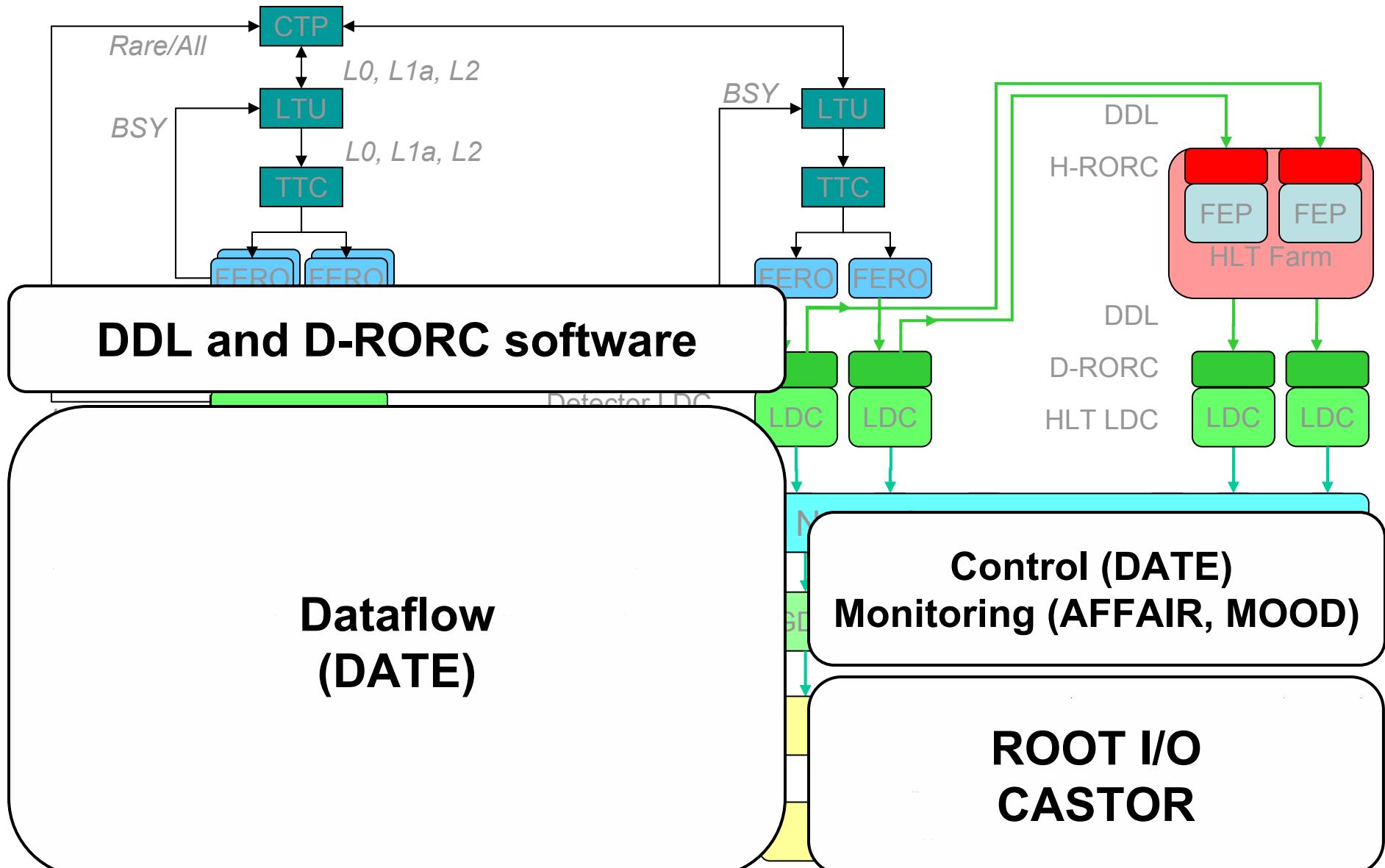


Outline

- ALICE DAQ
 - Trigger/DAQ logical model - Requirements
 - Trigger/DAQ at LHC
 - ALICE Trigger/DAQ Architecture
- Hardware components
- Software packages
- Towards the final ALICE
- Conclusions



ALICE DAQ Software





DAQ Software Framework

- ◆ DAQ Software Framework
 - Common interfaces for detector-dependant applications
 - Address all configurations and all development phases
- ◆ DAQ Software
 - Complete ALICE DAQ software framework in 3 packages:
 - DATE:
 - ◆ Data-flow: detector readout, event building
 - ◆ System configuration, control (1000's of programs to start, stop, synchronize)
 - AFFAIR: Performance monitoring
 - MOOD: Data quality monitoring
 - Production-quality releases
 - Evolving with requirements and technology ⇒ home-development
- ◆ Key issues
 - Scalability (1 to 1000, demonstrate it)
 - Support and documentation



Data Flow - DATE

RUN Control

DAQ - Run Control

DOMAIN: divia23073

Configuration **Run Parameters** **Ready to start** **Data Taking**

Start run **Stop run**
 Run AutoStart
 Autoset GDC
 Recording Enabled
 AFFAIR EDM
 ALIMDC HLT

Show **Show**

RUN NUMBER : 1785 DAQ Logic Engine Status : RUNNING

Info: Run 1785 running

Trace Fri 13 11:07 Run 1785 running
Clear Fri 13 11:07 Run number saved on /dateSiteAdc/configurationFiles/runNumber.config
Fri 13 11:07 Starting run 1785
Debug Fri 13 11:07 * Message from tbcd0029gdc: TRACE STOP_PROCESS: EVB 3223 has been killed as r
Fri 13 11:07 * Message from tbcd0029gdc: ACTION End of run requested with error
Pause Fri 13 09:10 * Message from tbcd0049ldc: ERROR file /date/runControl/Linux/checkProc.sh problem
Bigger Fri 13 08:05 Run 1784 running
Smaller Fri 13 08:05 Run number saved on /dateSiteAdc/configurationFiles/runNumber.config

root@tbcd000001:~

```
11:21am up 78 days, 22:29, 1 user, load average: 1.73, 1.69, 1.62
90 processes: 87 sleeping, 3 running, 0 zombie, 0 stopped
CPU0 states: 2.0% user, 50.5% system, 1.2% nice, 46.1% idle
CPU1 states: 3.0% user, 75.3% system, 2.0% nice, 21.0% idle
Mem: 384356K av, 374564K used, 9792K free, 3020K shrd, 147540K buff
Swap: 1044184K av, 26364K used, 1017820K free          152456K cached
```

PID	USER	PRI	NI	SIZE	RSS	SHARE	STAT	%CPU	%MEM	TIME	COMMAND
15208	nobody	14	5	4080	4080	3644	R N	99.9	1.0	13:09	recorder
1334	root	9	0	2332	2284	1592	S	0.5	0.5	21:39	sshd
1574	root	9	0	1060	1060	820	R	0.3	0.2	30:05	top
3	root	19	19	0	0	0	SWN	0.1	0.0	13:47	ksoftirqd_CPU0
1337	root	9	0	2368	2364	1856	R	0.1	0.6	3:17	xterm
5070	nobody	8	0	4004	3976	1468	S	0.1	1.0	10:00	rcServer
1	root	9	0	496	448	448	S	0.0	0.1	0:12	init
2	root	8	0	0	0	0	SW	0.0	0.0	0:00	keventd
4	root	19	19	0	0	0	SWN	0.0	0.0	12:06	ksoftirqd_CPU1
5	root	9	0	0	0	0	SW	0.0	0.0	1:58	kswapd
6	root	9	0	0	0	0	SW	0.0	0.0	0:00	kreclaimd
7	root	9	0	0	0	0	SW	0.0	0.0	0:00	bdflush
8	root	9	0	0	0	0	SW	0.0	0.0	0:01	kupdated
9	root	-1	-20	0	0	0	SW<	0.0	0.0	0:00	mdrecoveryd
15	root	9	0	0	0	0	SW	0.0	0.0	0:00	scsi_eh_0
16	root	9	0	0	0	0	SW	0.0	0.0	0:00	scsi_eh_1

LDC status display

LDC name	tbed0001ldc	tbed0013ldc	tbed0030ldc	tbed0037ldc
Event rate	13	13	14	13
Bytes recorded rate	40.182 M	41.203 M	41.938 M	40.163 M
Bytes in buffer	C 1192% M 1195%	C 1188% M 1193%	C 1192% M 1194%	C 1187% M 1193%
Number of events	10453	10462	10457	10450
Events recorded	9816	9825	9820	9813
Bytes injected	31'031'205'136	31'057'922'896	31'043'079'696	31'022'299'216
Bytes recorded	29'141'863'284	29'175'752'364	29'154'396'136	29'140'480'912
Readout SOR/EOR phases	0	0	0	0
Recorder SOR/EOR phases	0	0	0	0

GDC status display

GDC name	tbed0003gdc	tbed0004gdc	tbed0014gdc	tbed0015gdc
Events received	4924	5170	5438	3505
Events recorded	622	639	673	432
Bytes received	14'588'026'944	15'347'910'144	16'167'256'896	10'428'860
Bytes recorded	14'392'096'256	15'175'728'576	15'983'200'832	10'259'648
Event builder SOR/EOR phases	0	0	0	0
Status	FULL	FULL	FULL	FULL

EDM status display

EDM name	tbed0015edm
wakeUpId received	(nblnRun:10442)
maxWakeUpId	(nblnRun:10442)
lastThresholdSent	(nblnRun:10454)
lastUpperBoundSent	(nblnRun:10464)
edmMask	[0]:00040000 [1]:00000100
Excluded	3 4 14 26 29 41 50 51 64 65 74 75 96 97



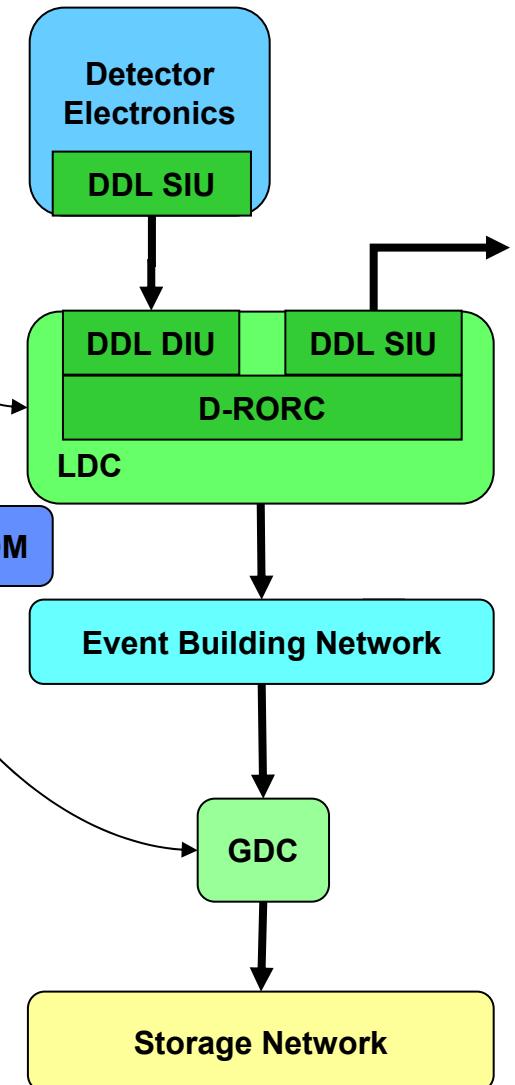
Control - DATE



- ◆ DATE software
 - Operator console
 - State machines
 - Control of distributed system
- ◆ Home-made development based on free software

Experiment
Control
System

Run Control
State Machines





DATE Scalability in ADC V

ADCV_CONTROLS::ALLDETECTORS_CONTROL

File View Options Windows

DAQ - Run Control

HI running on tbed0001.cern.ch with PID 29758
RC running on tbed0001.cern.ch

ALICE

Disconnected Configuration Connected Run Parameters Ready to start Data Taken

Start processes Start Stop Abort

Recording Enabled ALIMDC GDC ON AFIAIR HLT

RUN NUMBER : 5443 Run Control Status : RUNNING

Trace Tue 26 15:42:46 (RC) Run number saved on /dateSiteAdc/configurationFiles/ALLDETECTORS/runNumber.tbed
Clear Tue 26 15:42:45 (HI) Run Options loaded from : /dateSiteAdc/configurationFiles/ALLDETECTORS/runOptions.tbed
Debug Tue 26 15:42:45 (HI) Start processes time : 32 seconds
Pause Tue 26 15:42:13 (RC) Starting run 5443
Bigger Tue 26 15:42:13 (RC) Run Options loaded from : /dateSiteAdc/configurationFiles/ALLDETECTORS/runOptions.tbed
Smaller Tue 26 15:42:13 (HI) RC options saved as : /dateSiteAdc/configurationFiles/ALLDETECTORS/AD
Tue 26 15:41:58 (RC) Connected to remote hosts
Tue 26 15:41:58 (HI) Connection time 34 seconds
Tue 26 15:41:29 (RC) Starting Logon Engines at 26 Aug 2003 15:41:26 (Wait...)

GDC (91)

NOT_RUNNING
STARTING
STARTING_ALIMDC
STARTING_EVB
RUNNING
RUNNING_ERR
STOPPING_EVB
STOPPING_ALIMDC
WAIT_STOPPED
STOPPED

LDC (71)

NOT_RUNNING
SYNCHRONOUS
STARTING
STARTING_EDMC
STARTING_RECORDER
STARTING_HLTAGENT
STARTING_EDMAGENT
STARTING_READOUT
RUNNING
STOPPING_READOUT
STOPPING_EDMAGENT
STOPPING_HLTAGENT
STOPPING_RECORDER
STOPPING_EDMC
STOPPED

SD

LDC status display

LDC name	tbed0001ldc	tbed0002ldc	tbed0003ldc	tbed0004ldc
Event rate	377	311	297	330
Bytes recorded rate	38.453 M	31.652 M	30.443 M	33.814 M
Bytes in buffer	C 40% M 40% 638	C 39% M 39% 623	C 39% M 39% 620	C 40% M 40% 6
Number of events	8645	7761	7580	8533
Events recorded	8008	7126	6944	7896
Bytes injected	886056720	795453792	776902740	874577616
Bytes recorded	822185216	733107796	714806848	811135200
Readout SOR/EOR phases	0	0	0	0
Recorder SOR/EOR phases	0	0	0	0

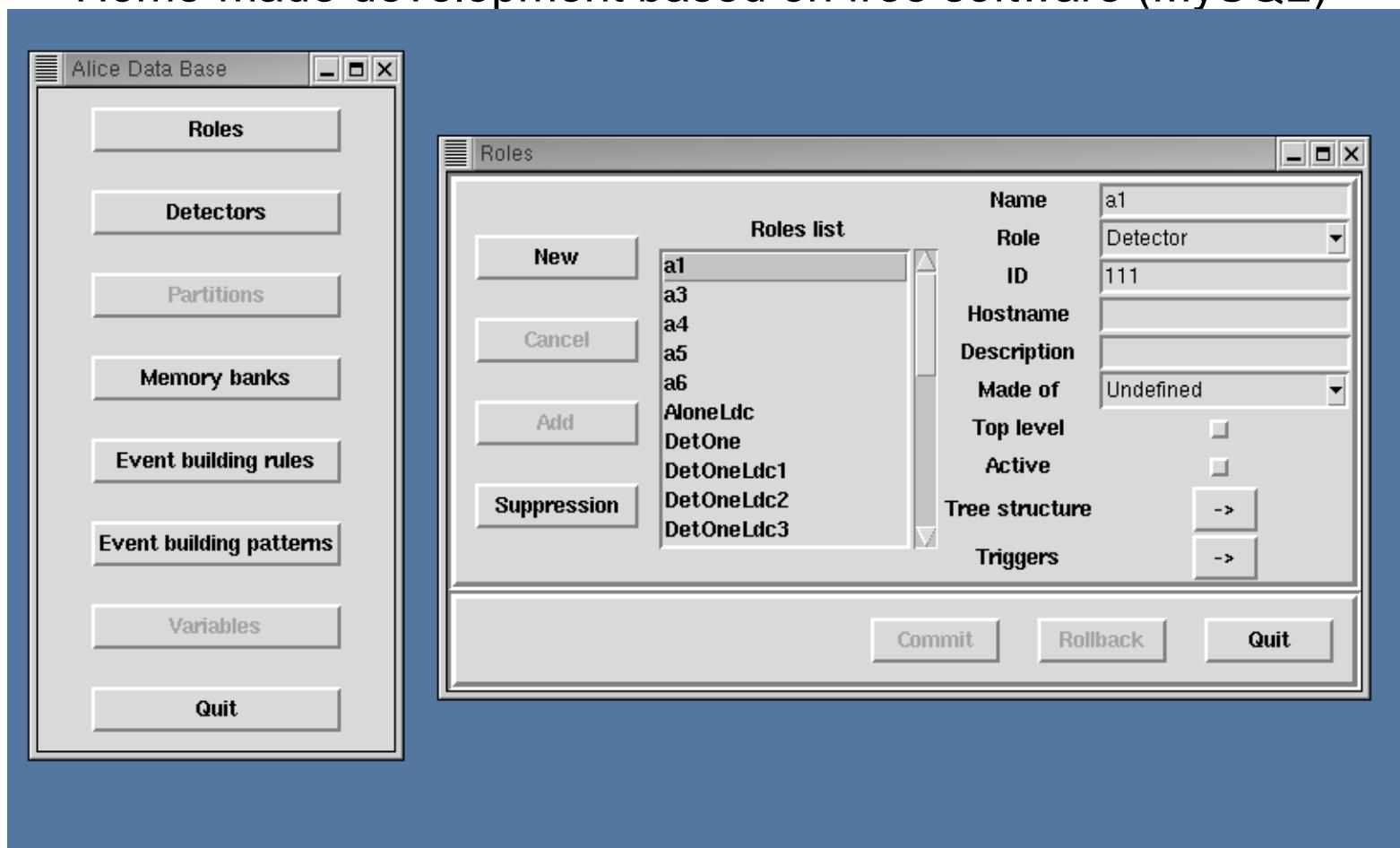
GDC status display

GDC name	tbed0001gdc	tbed0002gdc	tbed0003gdc	tbed0004gdc
Events received	5007	4872	4726	4658
Events recorded	145	3	2	2
Bytes received	563359168	560661364	542728828	533713160
Bytes recorded	65782276	64823208	43215472	43215472
Event builder SOR/EOR phases	0	0	0	0
Status				



DAQ Configuration Database

- ◆ DATE Configuration Database
 - Operator console
 - State machines
- ◆ Home-made development based on free software (MySQL)





Web access to Configuration Database

Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://localhost/date/dumpconfig.php

Home Bookmarks

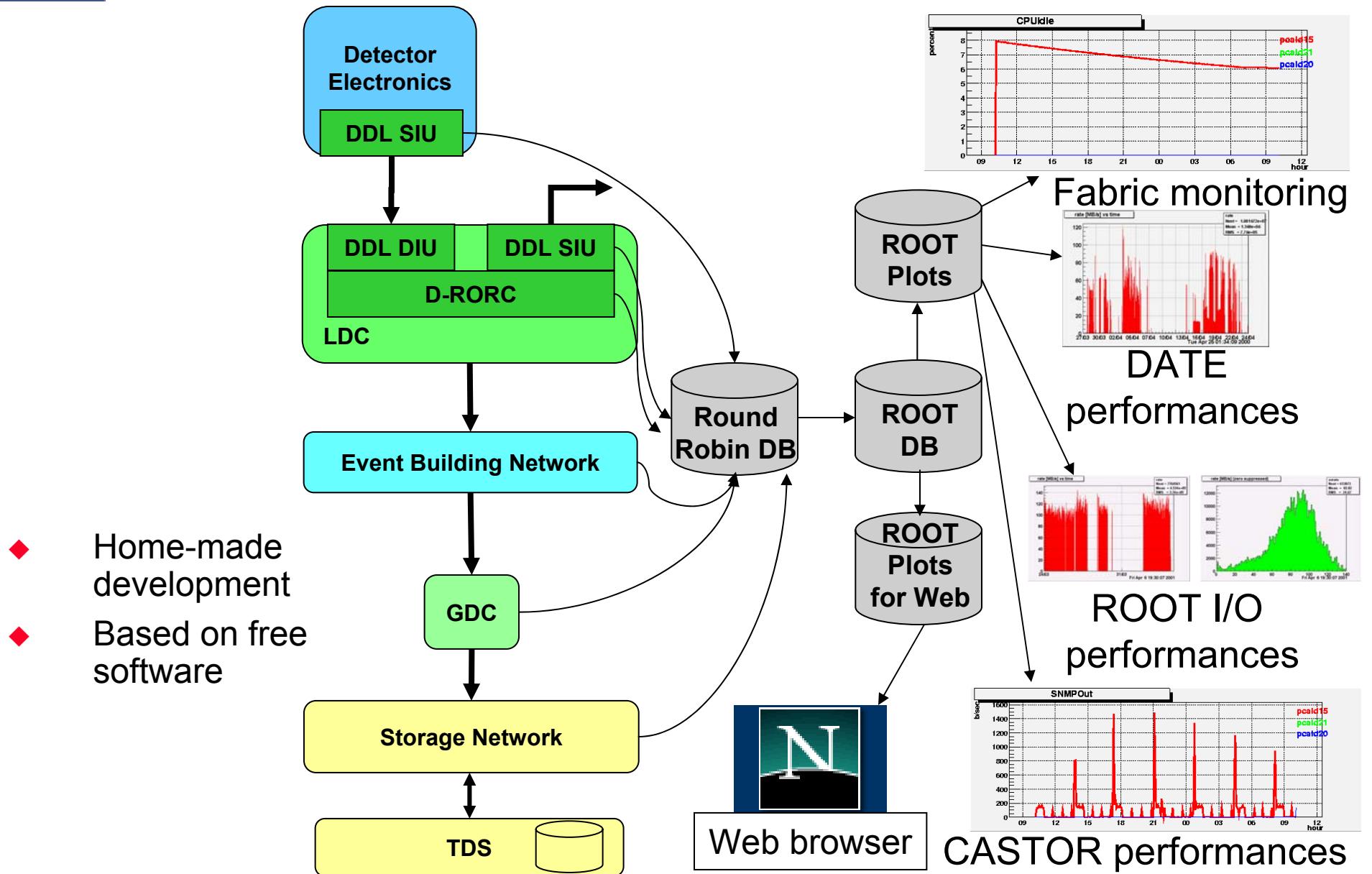
Connected successfully to DATE database

Roles

Name	Hostname	Description	Role	Id	Toplevel	Madeof	
gdc1	host8	GDC #1	GDC	1	1	1	Undefined
gdc2	host9	GDC #2	GDC	2	1	1	Undefined
DetOneLdc1	host1	DetOne LDC #1	LDC	1	0	1	Undefined
DetOneLdc2	host2	DetOne LDC #2	LDC	2	0	1	Undefined
DetOneLdc3	host3	DetOne LDC #3	LDC	3	0	1	Undefined
DetTwoLdc1	pcald37	DetTwo LDC #1	LDC	10	0	1	Undefined
DetTwoLdc2	host5	DetTwo LDC #2	LDC	11	0	1	Undefined
AloneLdc	host7	Single LDC	LDC	30	1	1	Undefined
DetThreeLdc	host6	DetThree LDC	LDC	160	0	1	Undefined
DetOne		Detector 1	Detector	1	1	1	Subdetector
DetTwo		Detector 2	Detector	2	1	1	LDC



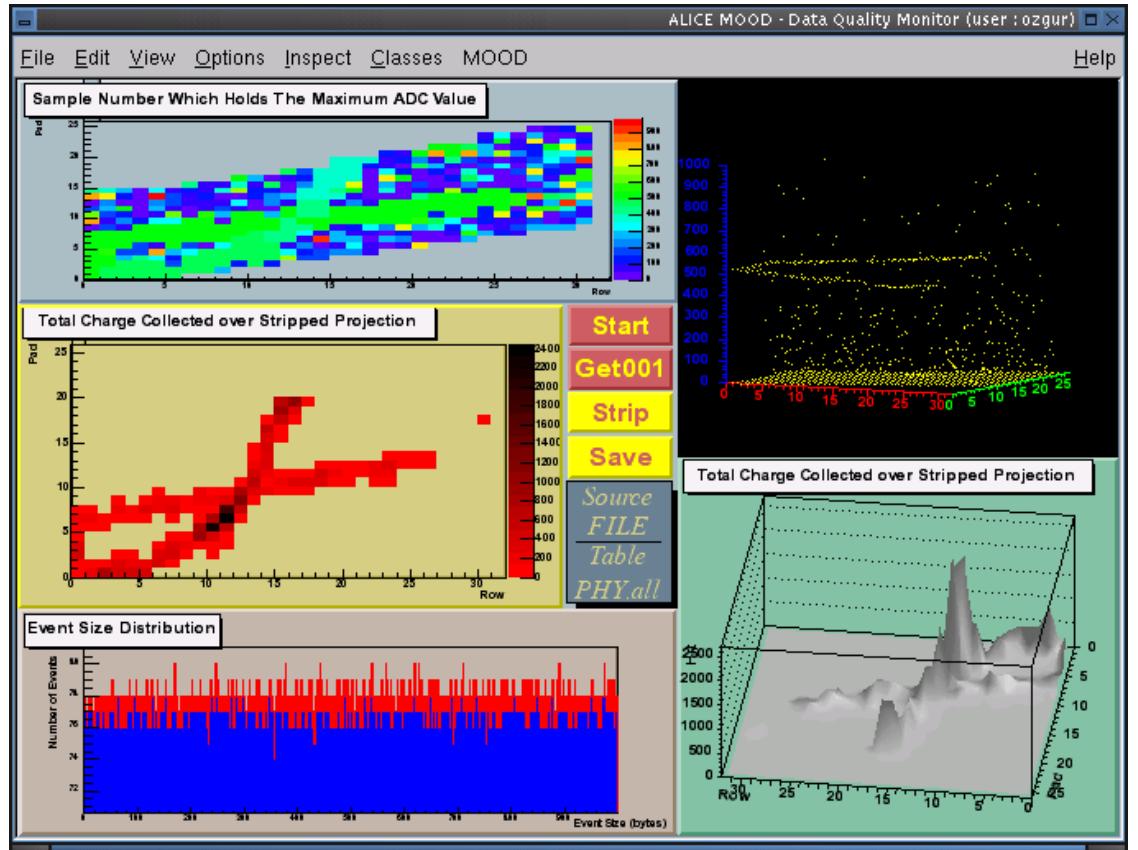
Performance Monitoring - AFFAIR





Data quality monitoring - MOOD

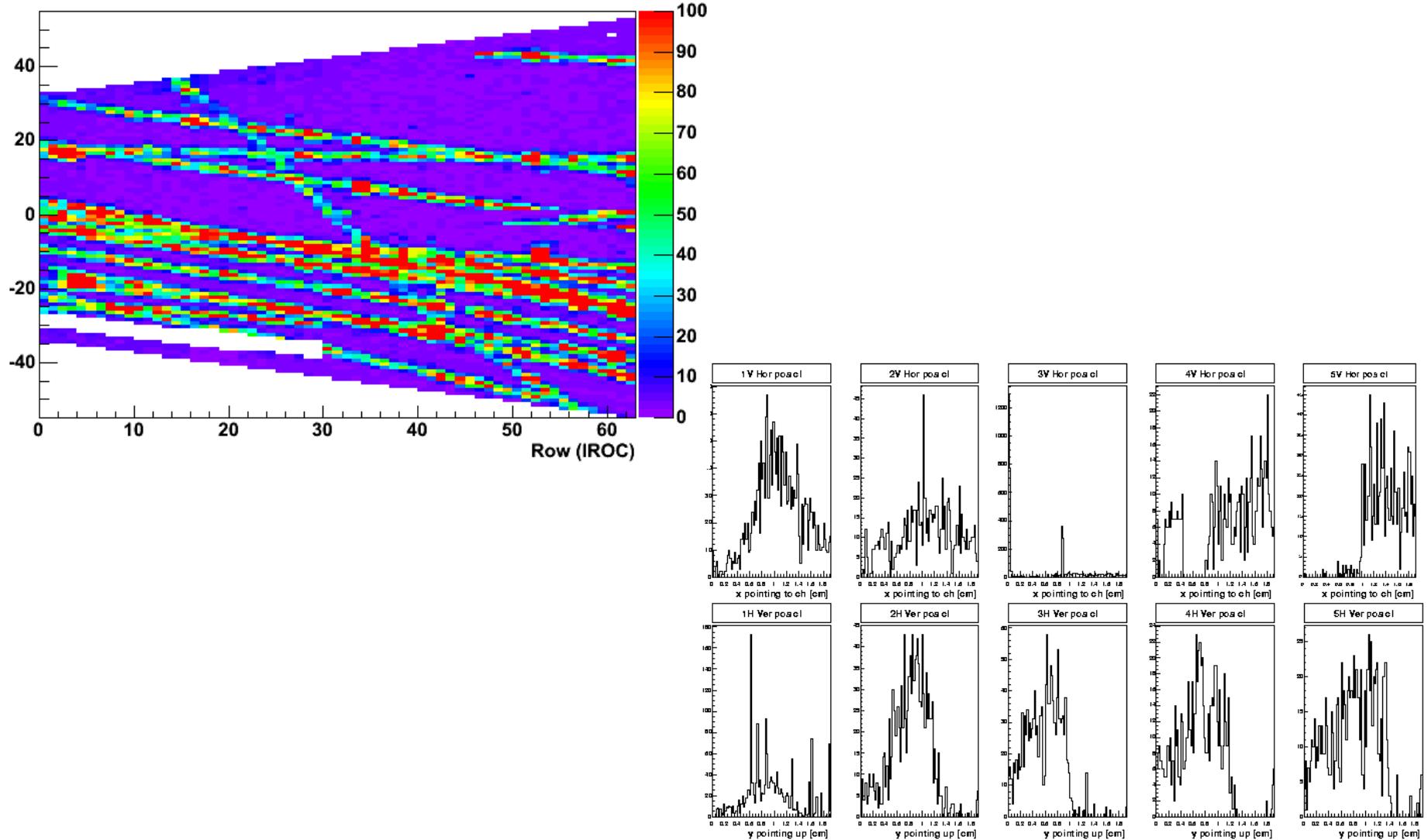
- ◆ MOOD framework
 - Interfaces to detector code
 - Software development in all institutes
- ◆ Applications:
 - Raw data integrity
 - Detector performance





Data quality monitoring - MOOD

Event: "377" Timebin: 0-1000





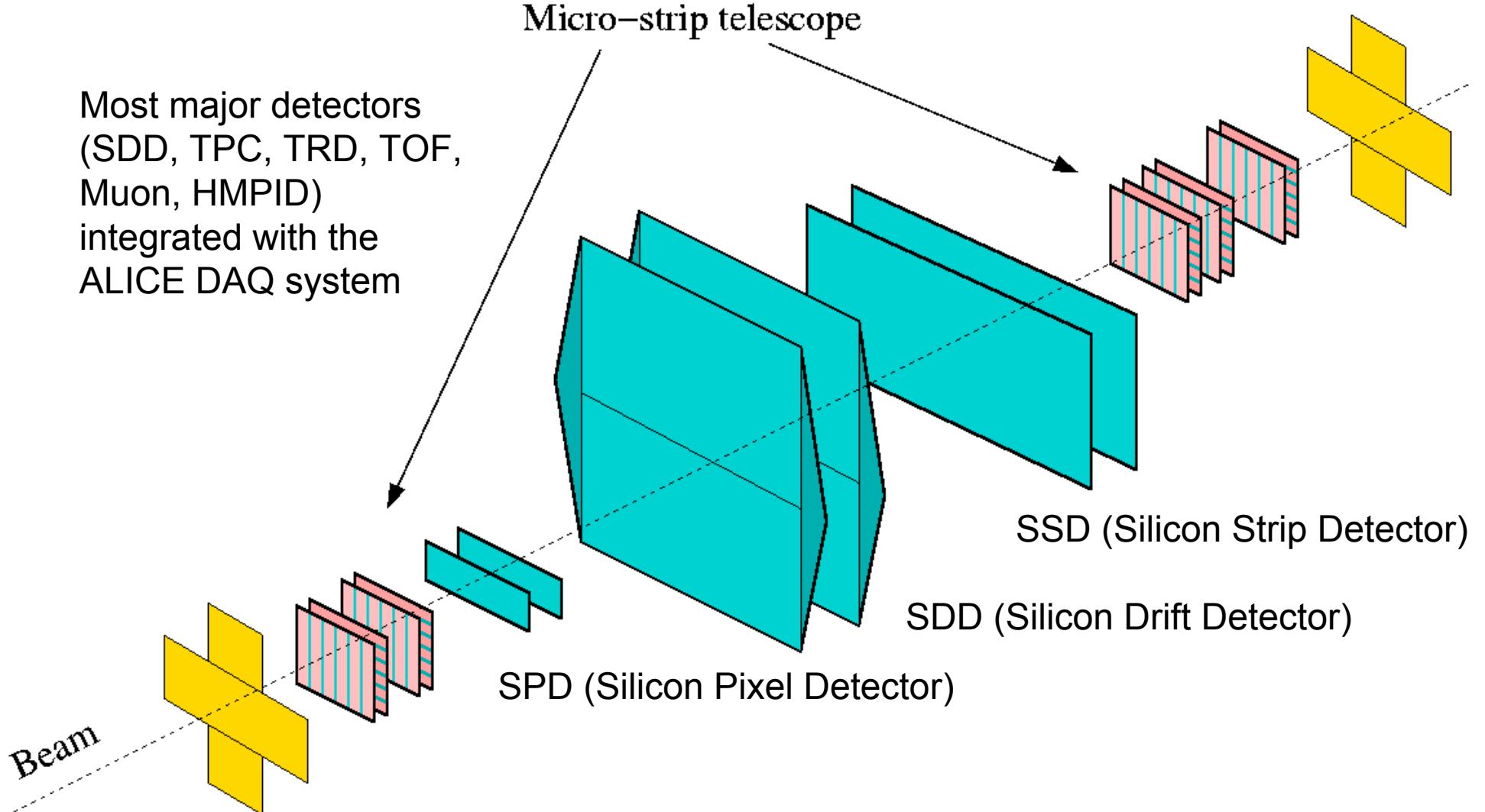
Outline

- ALICE DAQ
 - Trigger/DAQ logical model - Requirements
 - Trigger/DAQ at LHC
 - ALICE Trigger/DAQ Architecture
- Hardware components
- Software packages
- Towards the final ALICE
- Conclusions



Combined ITS test beam

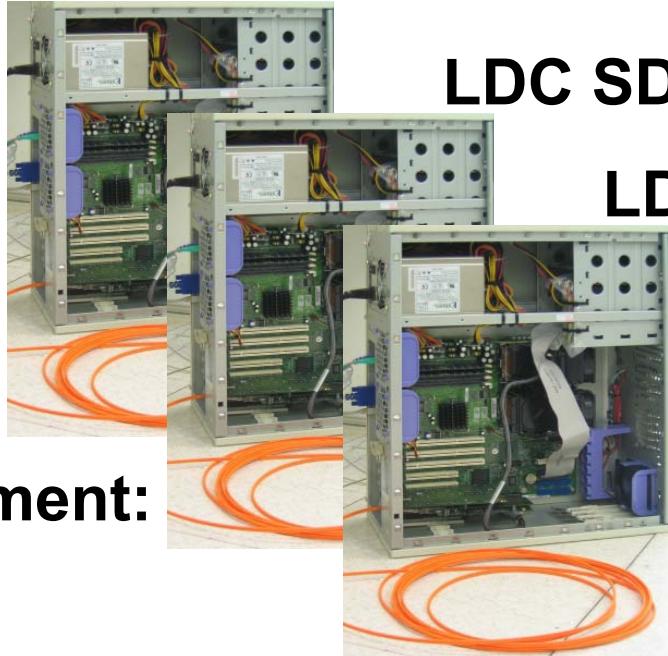
Most major detectors
(SDD, TPC, TRD, TOF,
Muon, HMPID)
integrated with the
ALICE DAQ system



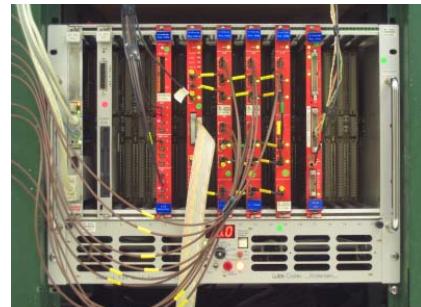


DAQ for ITS Test Beam

1 equipment:
• RORC

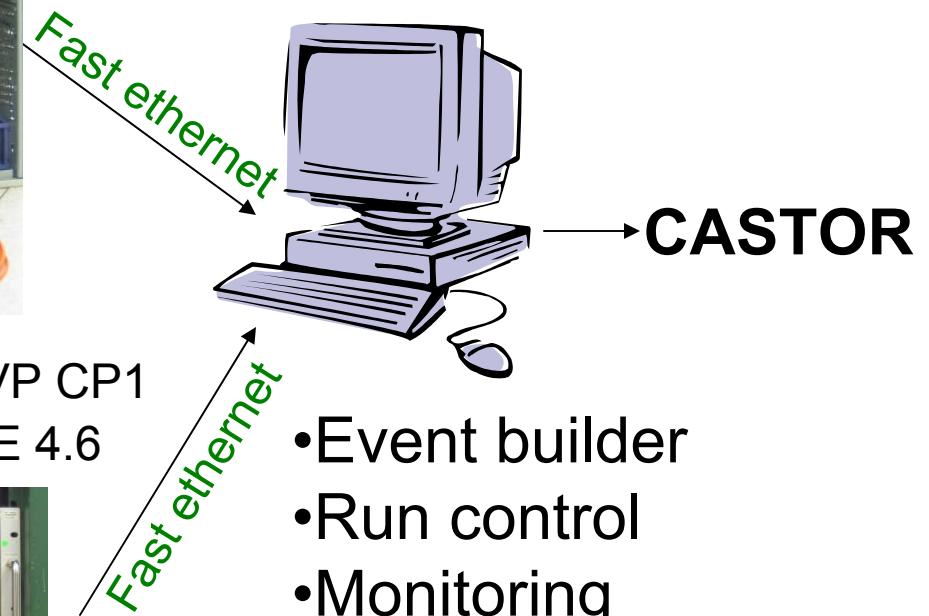


3 equipments:
• Trigger (HW/SW)
• Microstrip readout
• TDC



LDC VME: CCT VP CP1
RH linux 7.3.2 – DATE 4.6

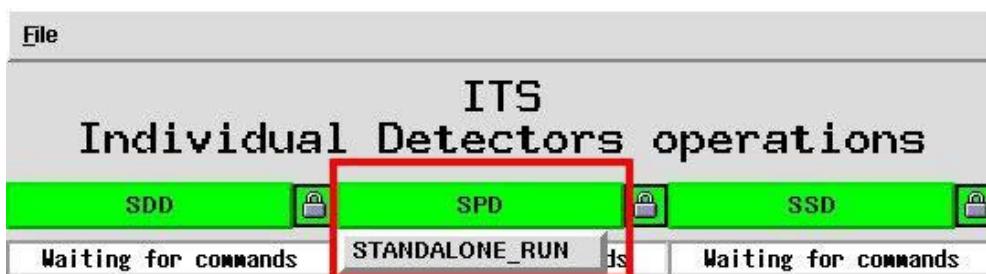
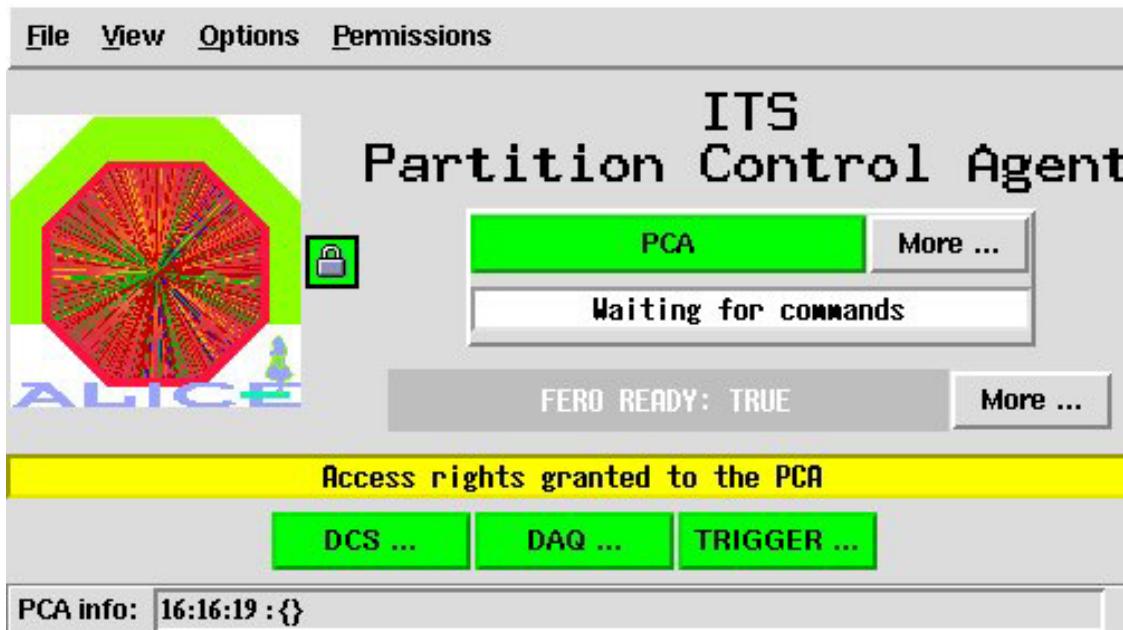
GDC:
Seil 2 Xeon 2.4 GHz 1GB
RH linux 7.3.2 – DATE 4.6



- Event builder
- Run control
- Monitoring
- NFS server



Control for ITS Test Beam





Simulation

- ◆ Complete model of the ALICE TRG/DAQ
- ◆ See next 2 talks:
 - Tome Anticic
ALICE Trigger and DAQ – Simulation
 - Linda Vickovic
ALICE Mass Storage System Simulation



Conclusions

- ◆ ALICE TRG DAQ system
 - All major hw and sw components released
 - Production phase
- ◆ Towards the final system
 - Integrated with most major detectors
 - First combined test beam
 - Simulation
- ◆ System evolution
 - New technology: PCI-XP, multicore CPU, 10 Gbit Eth
 - Isolate from technology obsolescence by logical interfaces (DDL or Mass Storage System)
- ◆ Start installation 1Q 2005. Startup 1Q 2007