



Data Acquisition at the LHC experiments

Sylvain Chapeland – CERN / ALICE

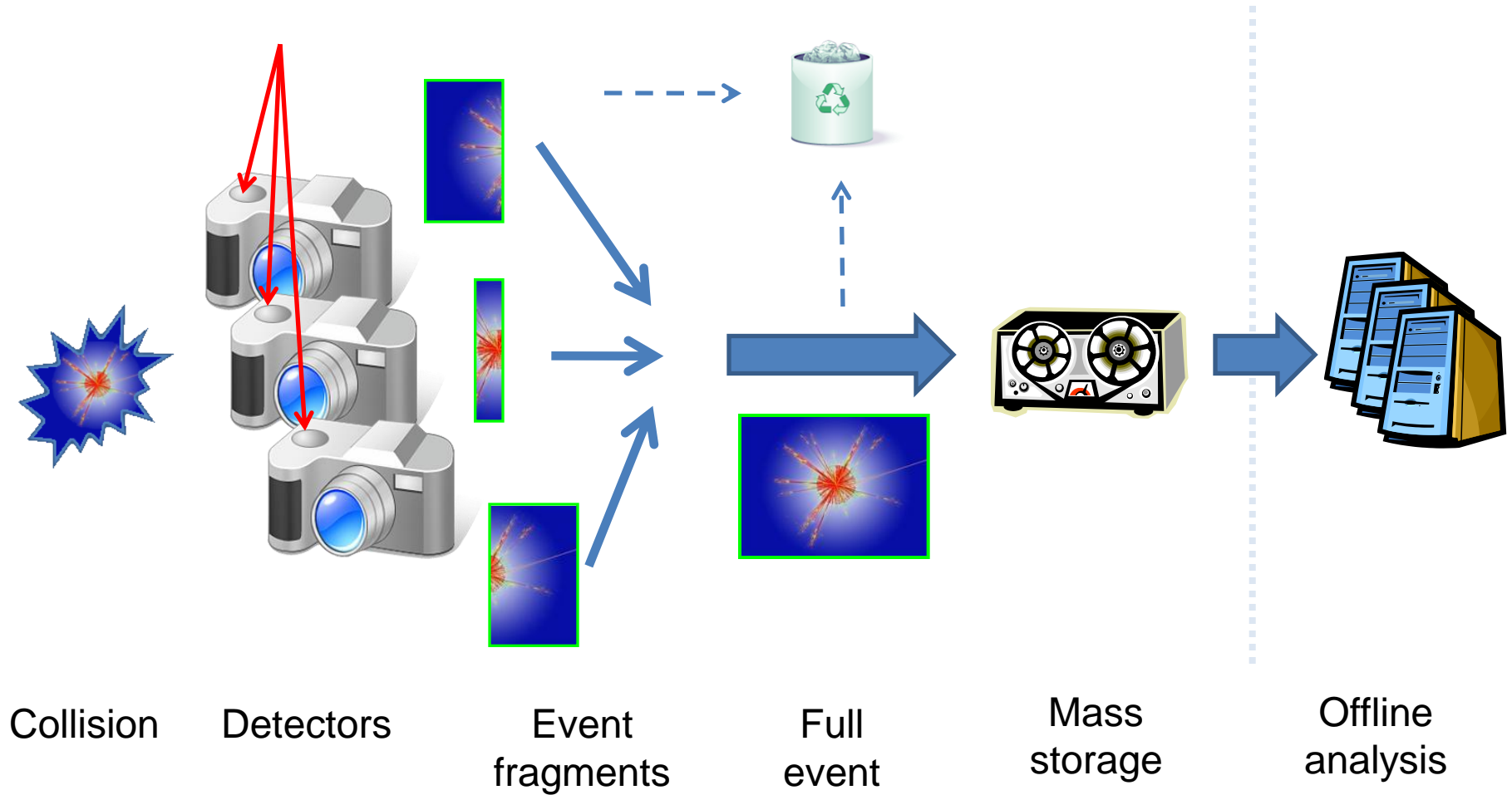
Warning

- Because time is short, some simplifications were necessary
- This presentation does not reflect the full complexity of the systems
- Effort was made to be as close as possible to reality, and to outline the specificities of each data taking environment
- Please refer to appropriate talks in the parallel sessions for further information

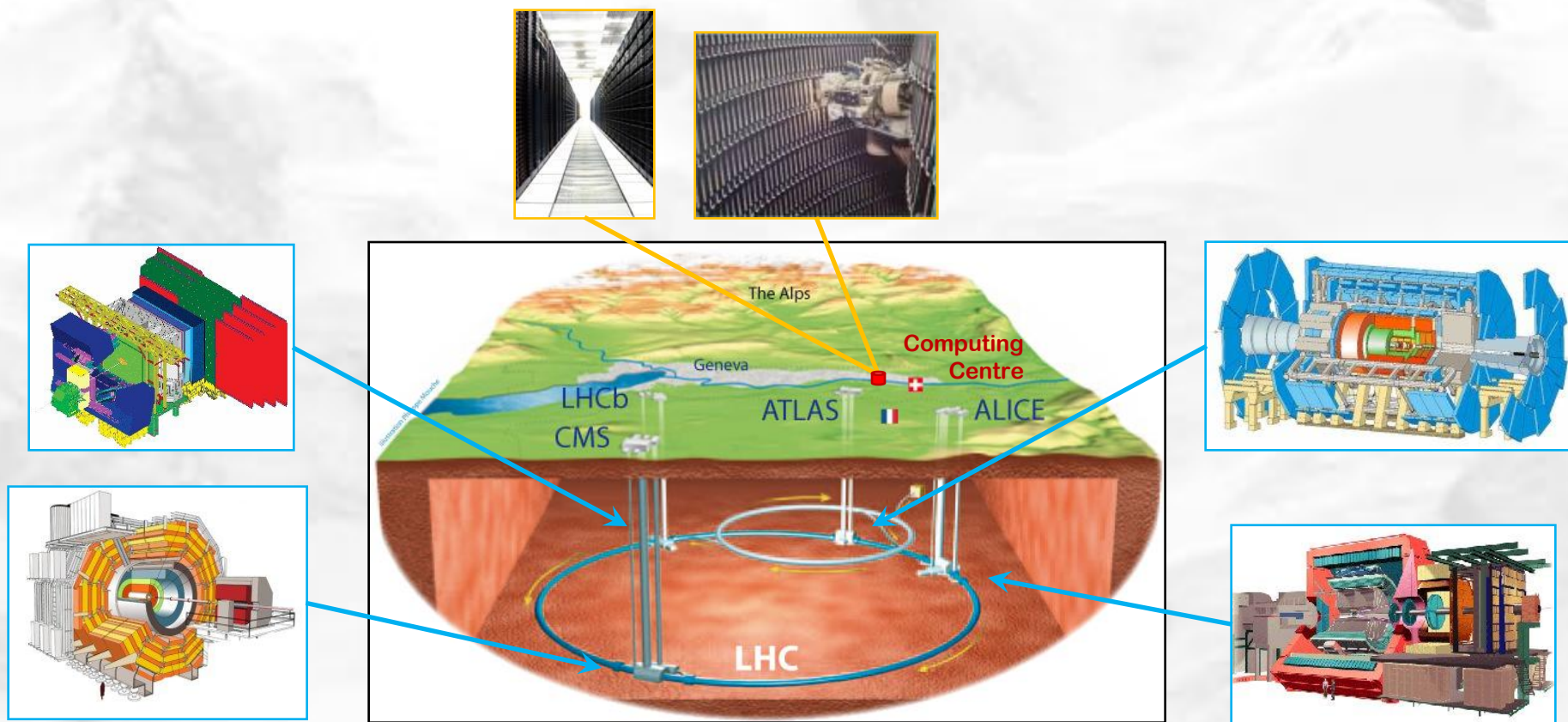
Outline

- Trigger / DAQ requirements for LHC
- Data flow designs for the experiments
- Hardware / Software implementations
 - Similarities / Specificities
- Operational aspects and commissioning
 - Conclusions

Trigger *(decisions)* and Data Acquisition *(flow)*



CERN LHC experiments

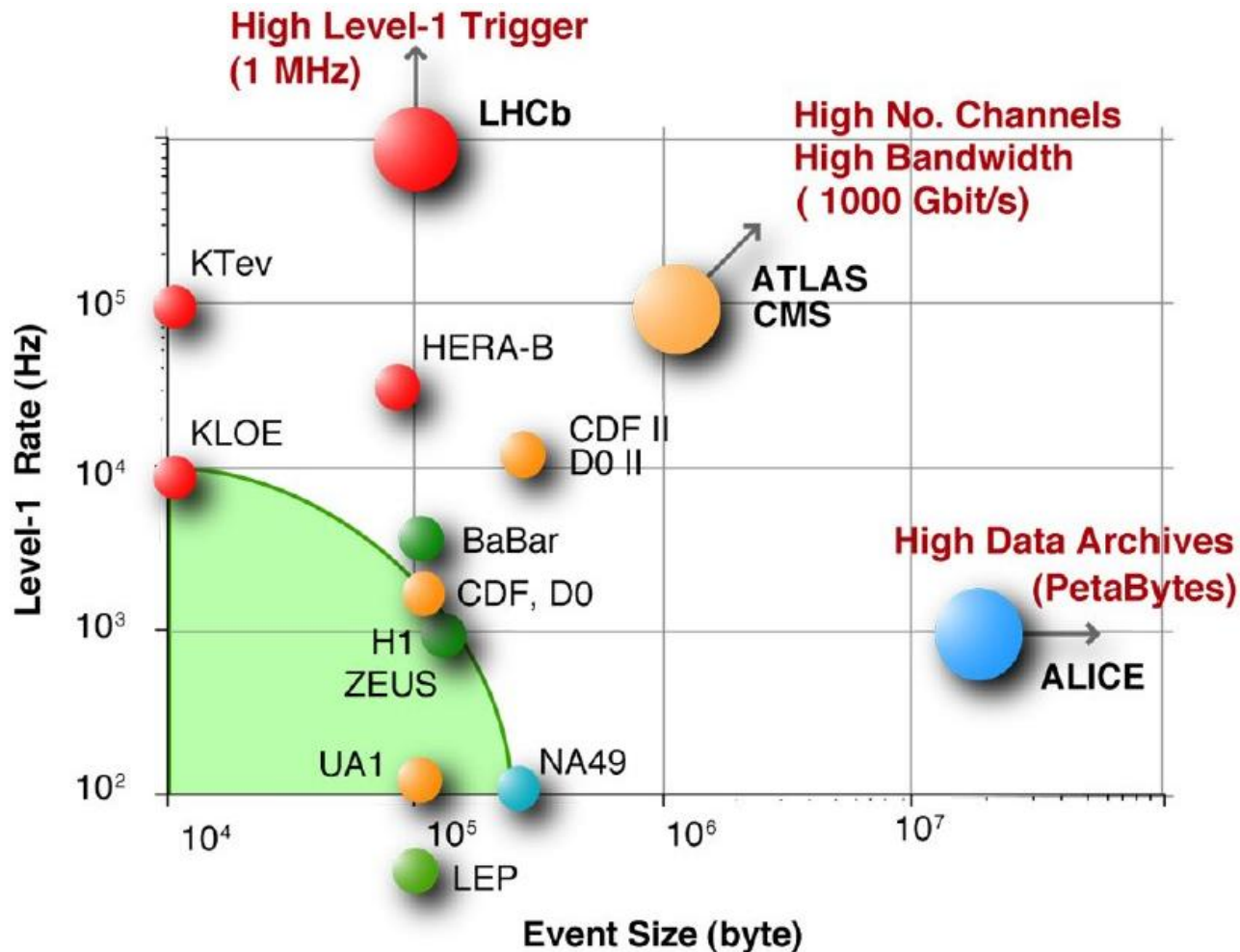


LHC experiments - DAQ needs

	ALICE		ATLAS	CMS	LHCb
Number of detectors	18		9	7	11
Number of Trigger levels (HW / SW)	3/1		1/2	1/1	1/1
Event size	86.5 MB	2.5MB	1.5 MB	1 MB	40kb
L1 Trigger rate	10 KHz	200 KHz	75 KHz	100 KHz	1 MHz
Detector readout	Trigger/Busy Partial readout		Synchronous		
Bandwidth to mass storage	1.25 GB/s	200 MB/s	300 MB/s	100 MB/s	100 MB/s
	Pb-Pb	p-p			

Interaction rate : 40 MHz
Large number of channels

LHC experiments - DAQ needs

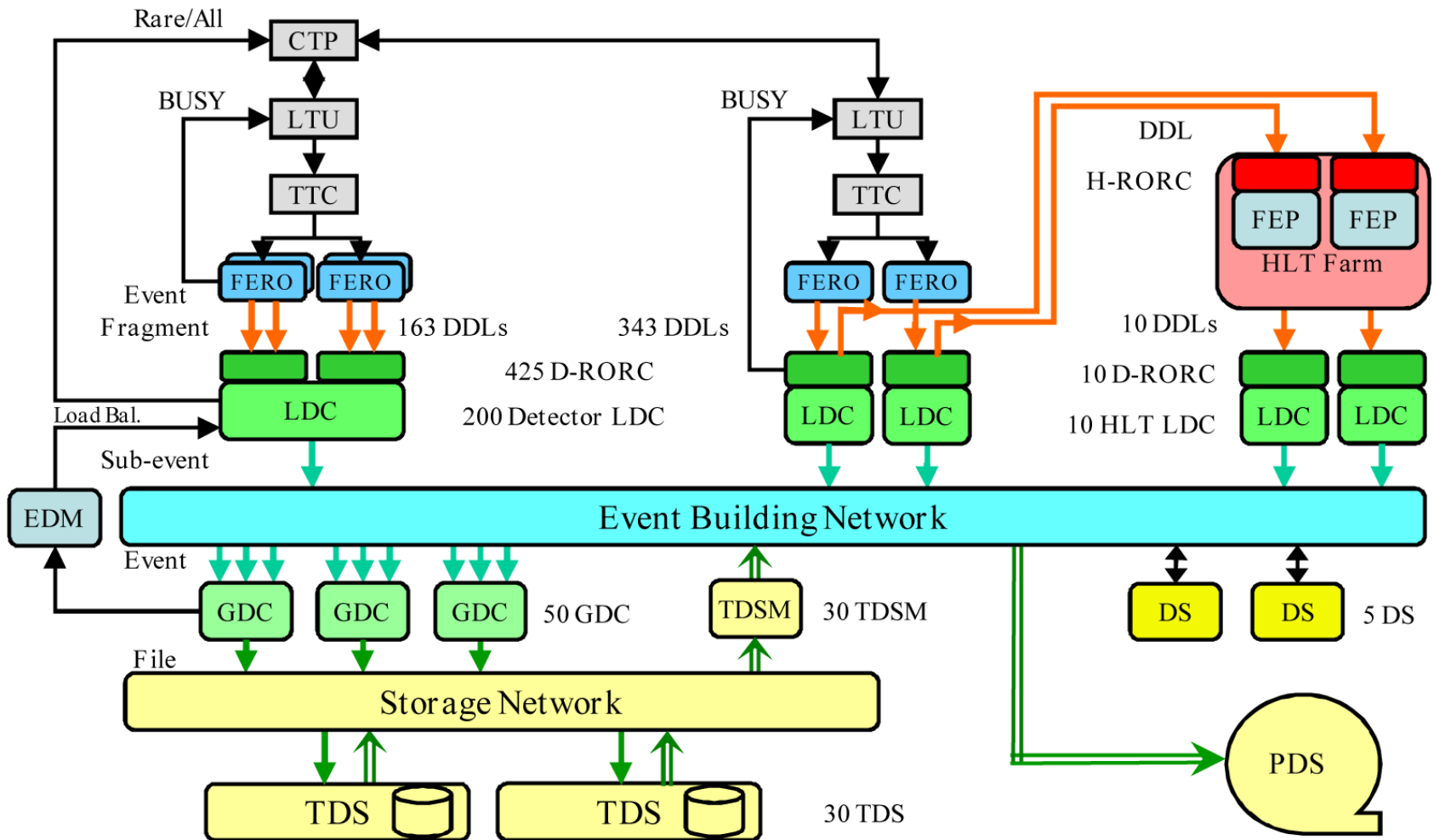


DAQ design challenges

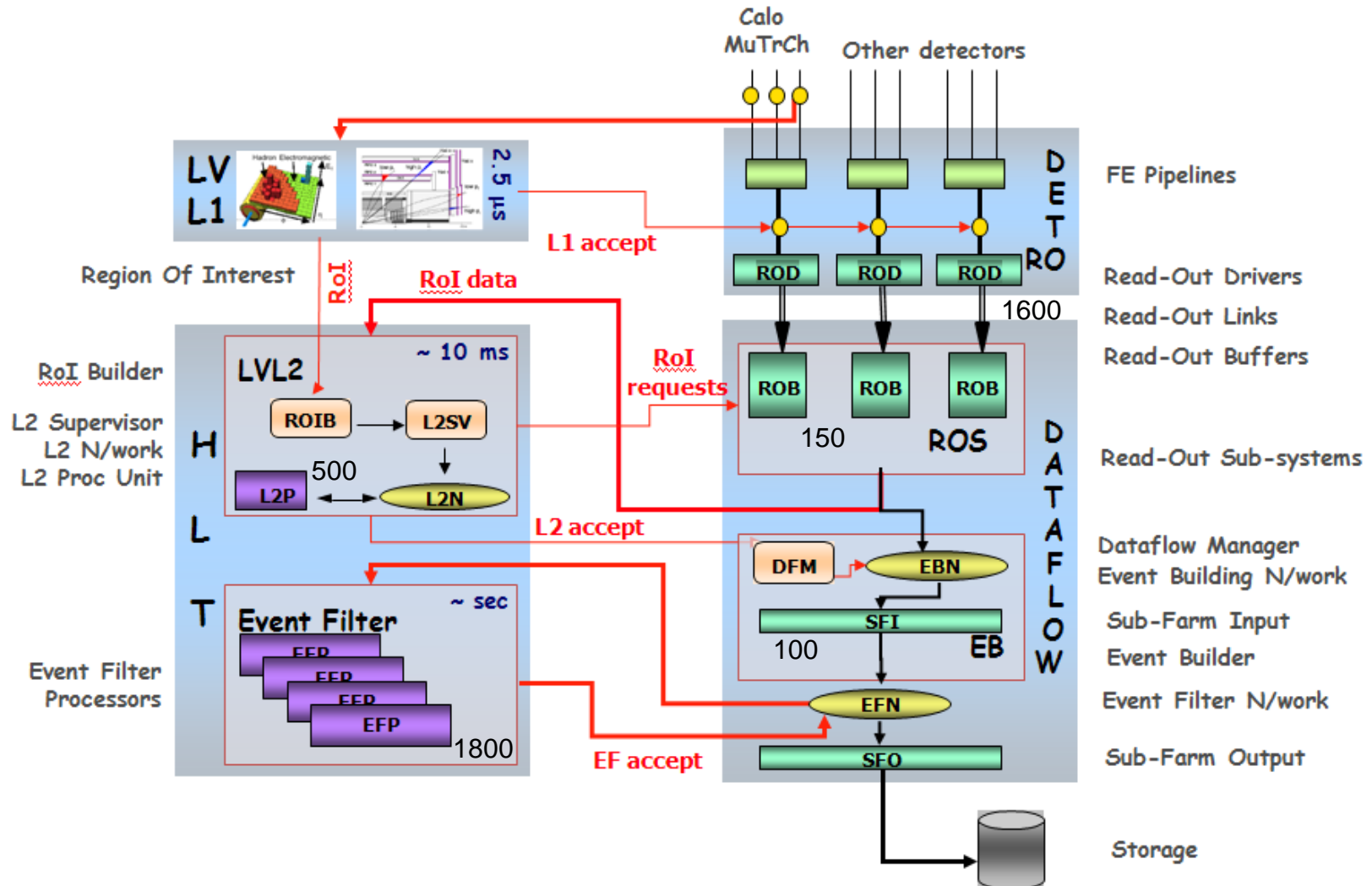
Data flowing from the detectors:

- Cope with the huge quantity
- Select the appropriate events
- Ensure measurements integrity
- Monitor to check quality
- Record for analysis and archive
- Operate such a complex system

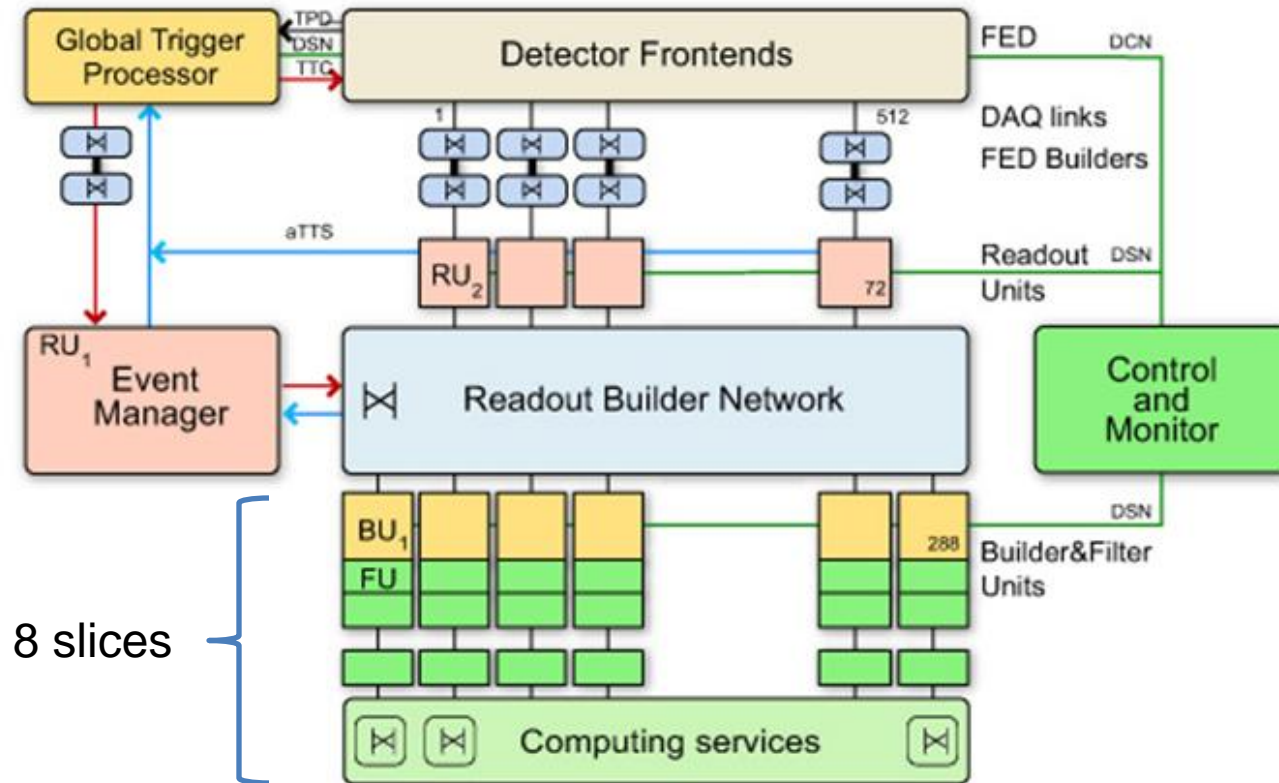
ALICE DAQ



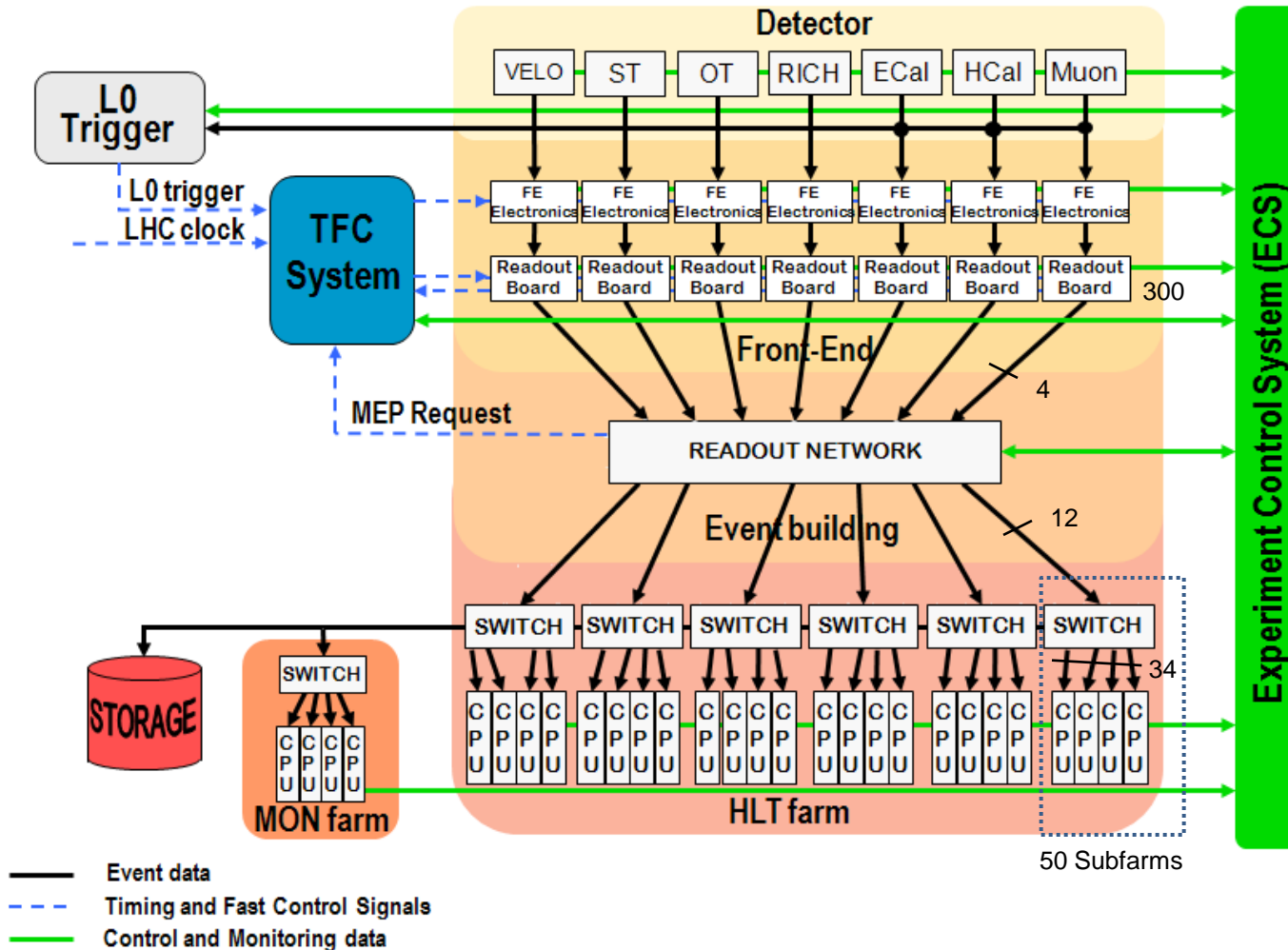
ATLAS DAQ



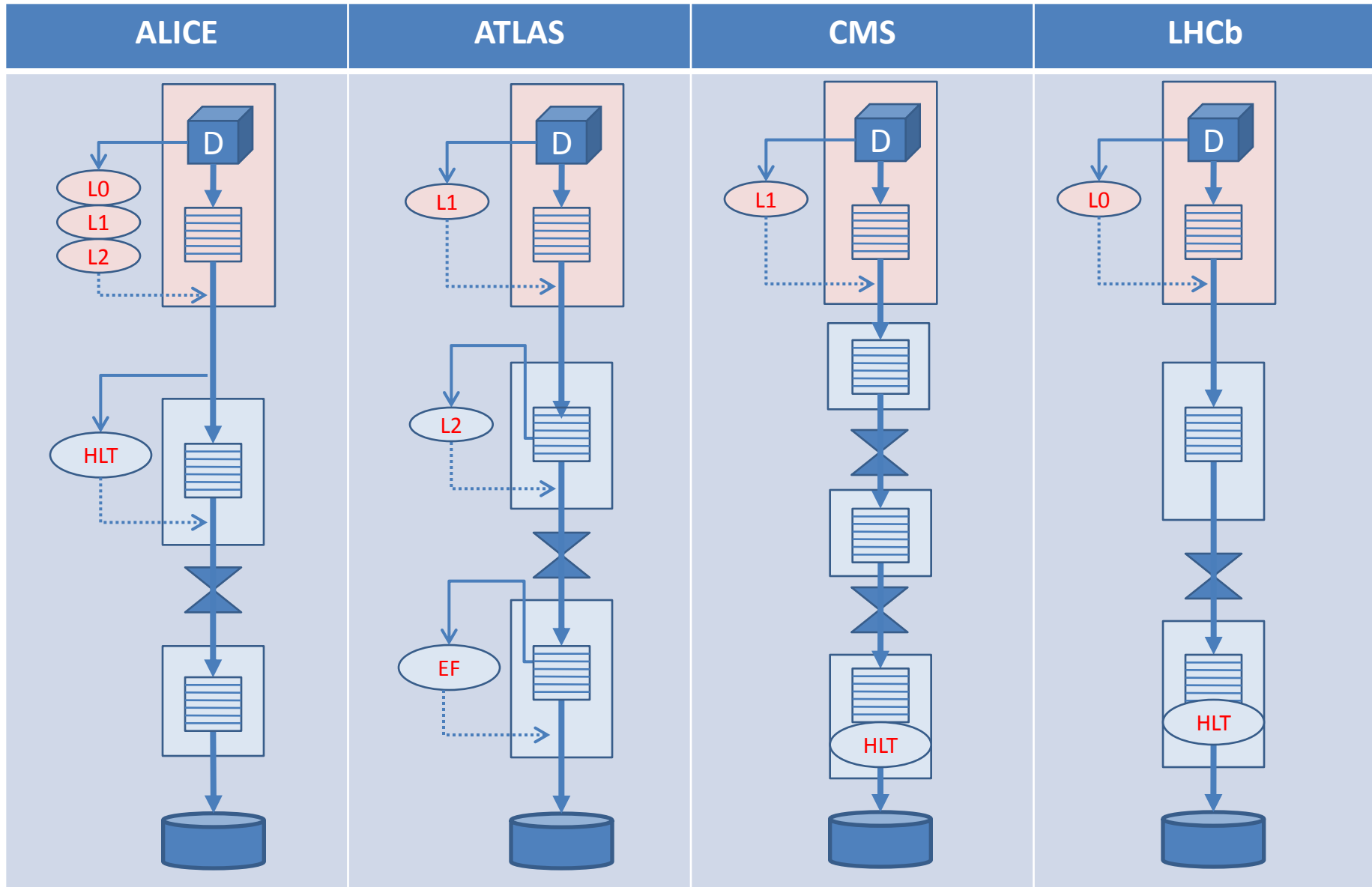
CMS DAQ



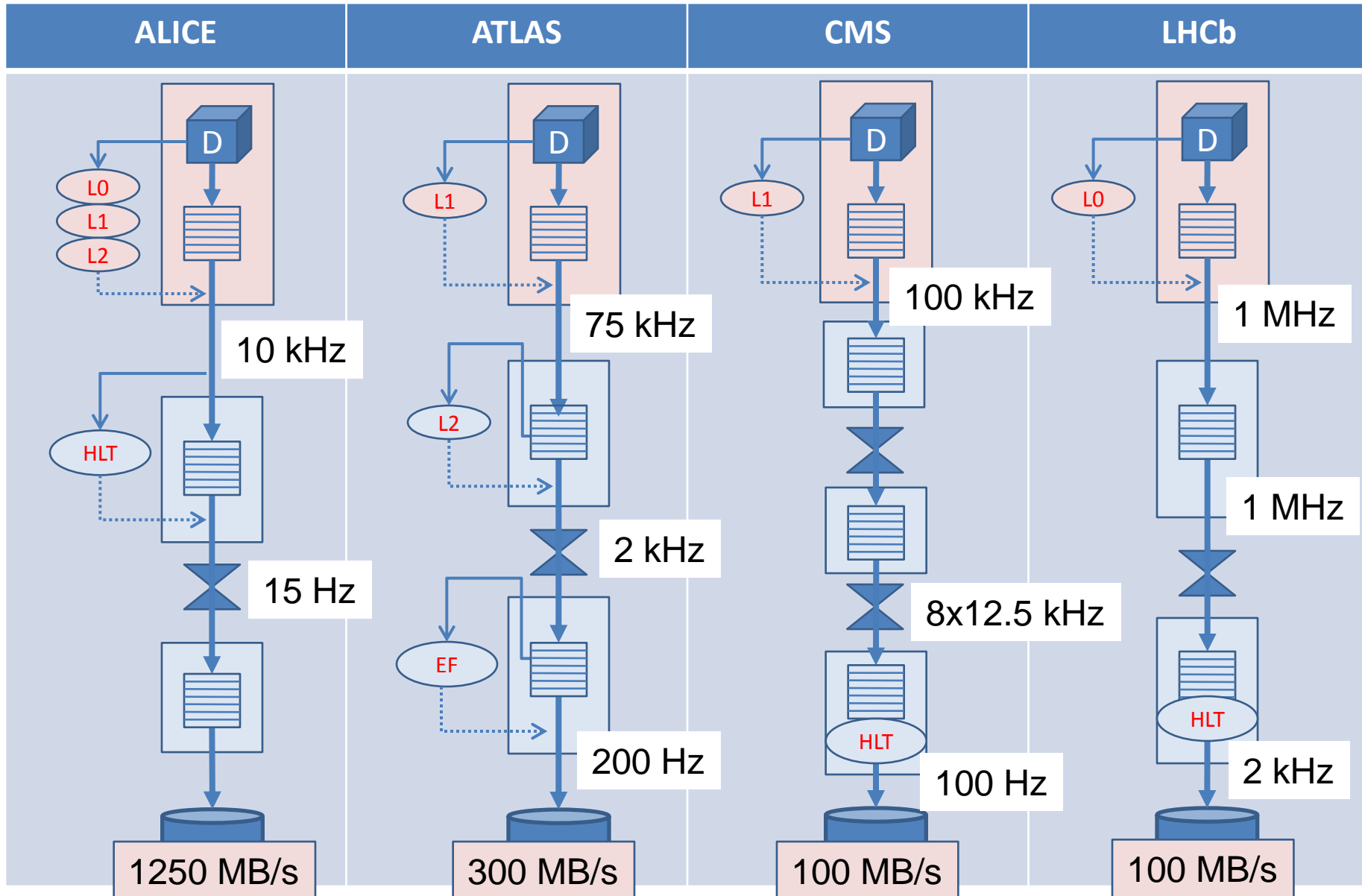
LHCb DAQ



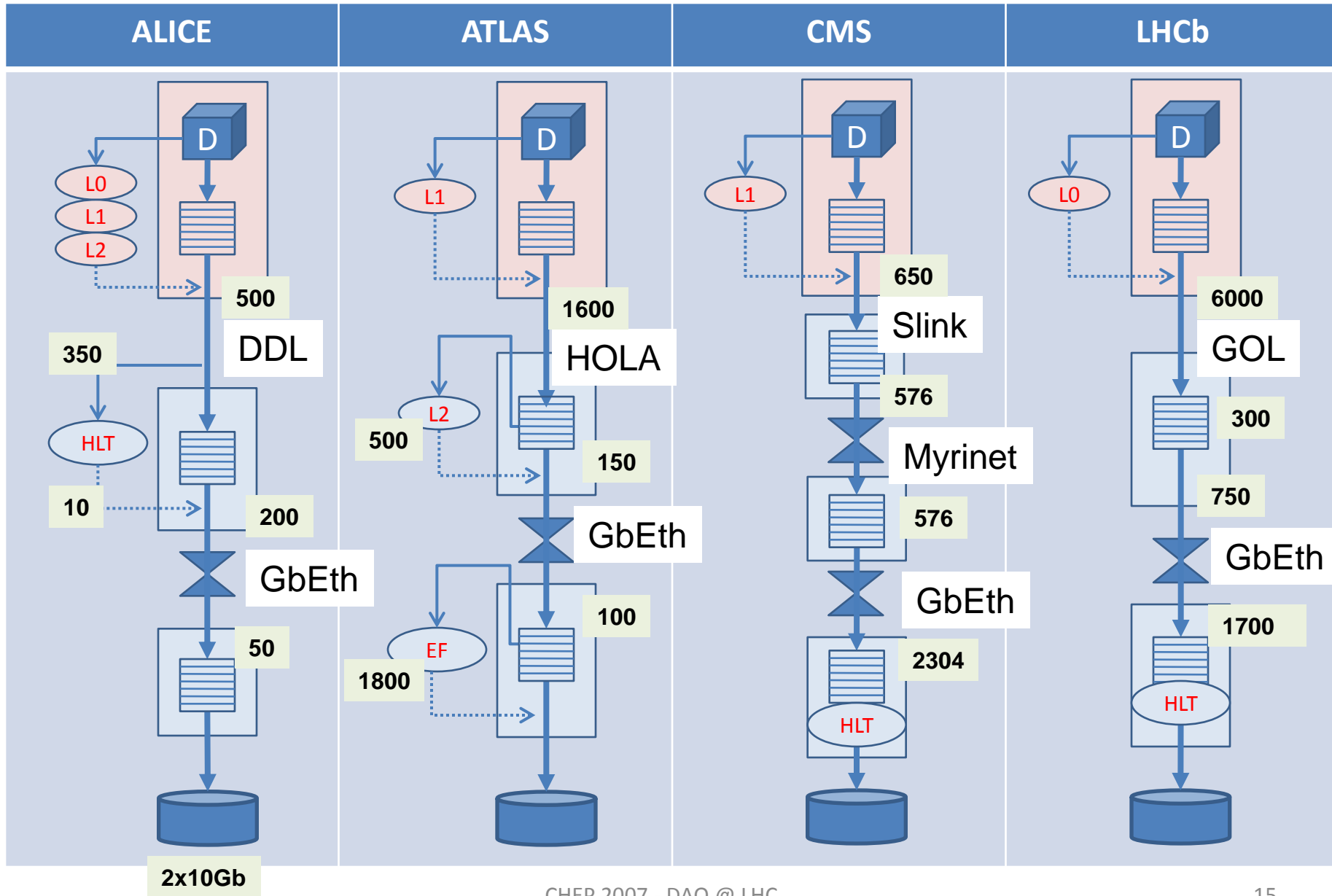
DAQ data flow



DAQ data flow - rates



DAQ data flow - links

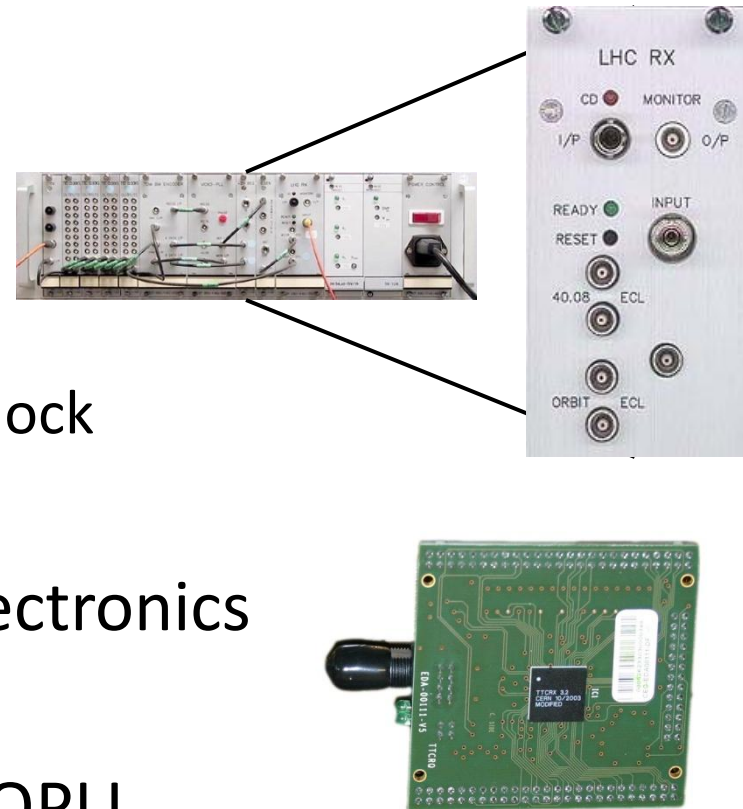


Trigger

- Base components common to all LHC experiments:
Timing, Trigger and Control (TTC) system
 - Synchronization with LHC clock
 - Distribution to detector electronics of:
 - Level-1 trigger
 - Broadcasted and individually-addressed control signals
- Completed by experiment custom electronics:
 - Work out low level trigger decisions
 - Pipeline buffers on data stream to handle latencies

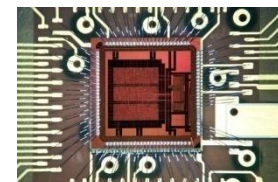
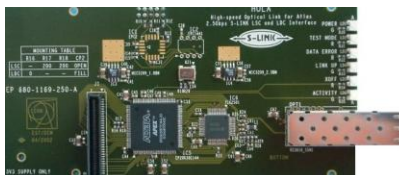
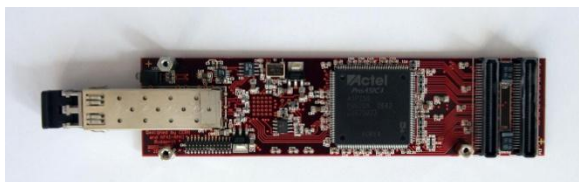
TTC system

- Dedicated electronics developed in particular:
 - TTCmi : machine interface
 - 40.08 MHz LHC bunch-crossing clock
 - 11.246 kHz orbit signals
 - TTCrx : receiver ASIC for the electronics
 - Radiation tolerant
 - VME interface, laser modules, QPLL ...




Detector links

	ALICE	ATLAS	CMS	LHCb
Link	DDL	HOLA	SLINK	GOL
Support	Optical	Optical	Copper	Optical
Maximum transfert rate	265 MB/s	160 MB/s	640 MB/s	200 MB/s
Notes	<ul style="list-style-type: none"> • Full duplex • Radiation tolerant sender unit • PCI-X interface • 1.6 GB/sec/PC 	<ul style="list-style-type: none"> • Duplex Slink • Single optical pair 	<ul style="list-style-type: none"> • SLink 64bit @100MHz • LVDS signals 	<ul style="list-style-type: none"> • ASIC • Radiation hard • 1.6-Gbit/s serializer • separate link for some analog data

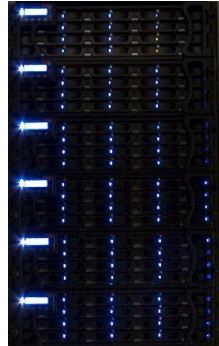


Event building

	ALICE	ATLAS	CMS	LHCb
Fragment input	6 - 12 links in PC (DDL -> PCI)	12 links in PC (HOLA -> PCI)	1 link in PC (Myrinet -> PCI)	Up to 36 links (GOL) per readout board (12 fibers ribbon)
Sub-events output	1 x Gigabit	1 x Gigabit	1 x Gigabit	4 x Gigabit
Full event building	Gigabit Ethernet switch			
	Not same list of detectors for each event		First layer of fragment multiplexing with Myrinet: data balance and data to surface	Multiplexing before / after router with switches

Storage

- Local transient data storage (~100 Terabytes)
 - Accommodate data flow peaks
 - 1-7days buffer to cope with uplink unavailability
 - Concurrent read/write by multiple hosts
- Mass storage (Petabytes)
 - CASTOR software
 - Appropriate grid registration tools
 - Tape robots



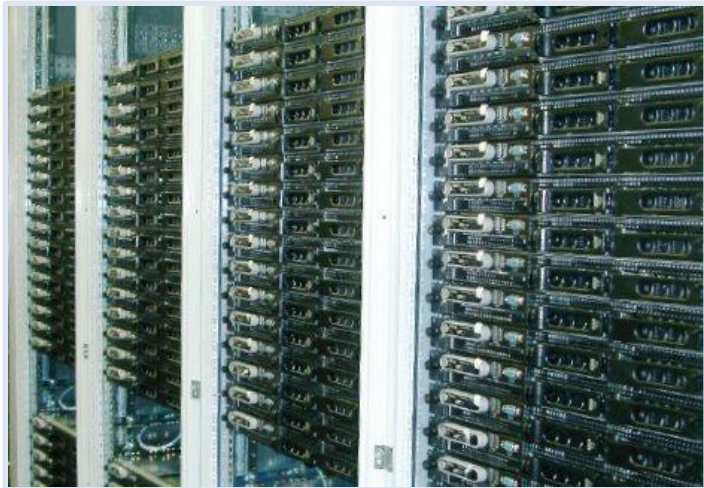
DAQ computing facilities

- Number of nodes: 200 – 2500 per experiment (ramping up)
 - Handled independently by each experiment
- Rack mounted PCs : limited space especially in underground locations
- Cooling doors with horizontal airflow
 - Includes temperature /smoke detection
- Large power requirements – currently 200-500 kW per experiment
 - Partial UPS coverage (<20% , for 10 min)
- Remote operation (control rooms and off-site)
 - Power control : dedicated hardware (PDU) or software (IPMI)
 - Remote console access: KVM switches (local station and IP reach), SSH on dedicated Ethernet-100 links
- Hardware maintenance
 - contracts with machine providers
 - in-house CERN support for some critical equipment (e.g. network routers)

DAQ computing facilities



DAQ computing facilities

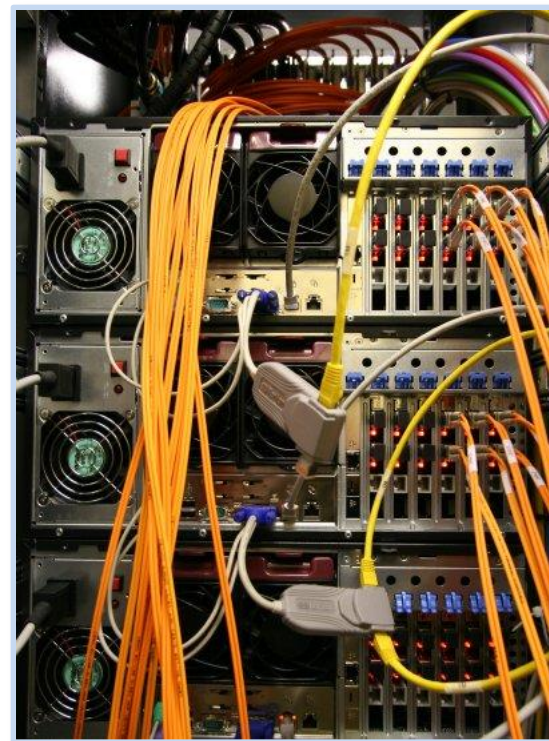
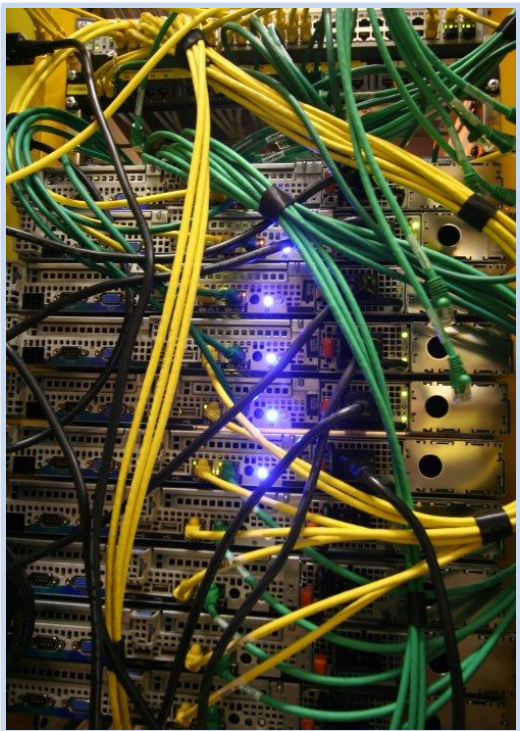


DAQ computing facilities



DAQ computing facilities

côté cuisine



Fabric management

- Operating system: CERN Scientific Linux – SLC4
- Machine installation – 2 flavors:
 - PXE boot, kickstart, YUM / RPM
 - Diskless installation
- Use of some dedicated control PCs in each rack (not all)
- Configuration:
 - Custom or existing software (e.g. Quattor)
 - Database: Oracle, MySQL
 - Users management: LDAP
- Monitoring
 - Custom or existing software (e.g. Nagios, Lemon, IPMI)
 - Appropriate hooks in DAQ software

DAQ software

- Lots of glue to interface the hardware components
- Software provides high flexibility
- Data flow handling:
 - “Simple” operations: pack and ship
 - No payload processing, excepted HLT / monitoring
 - High-speed throughput, low latency
- Distributed control and communication
- Process synchronization
- Dynamic configuration

DAQ software

- Huge software packages
- Usually low-level languages for processes on the data path
- Higher complexity tools for “slow control”
 - State machines
 - Databases for configuration
- Complex libraries to process and filter events
- Non negligible part of graphics interfaces
- Extensive use of code management and building tools (e.g. CVS, RPM, ...), 100ks LOC

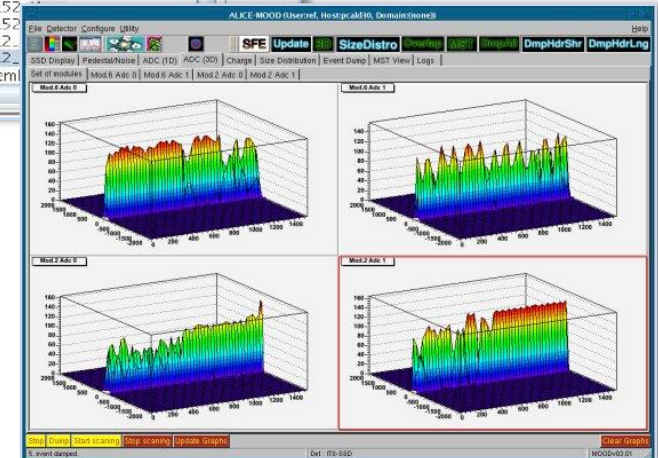
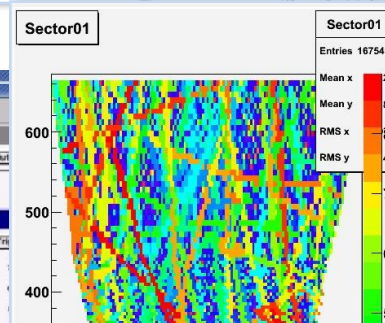
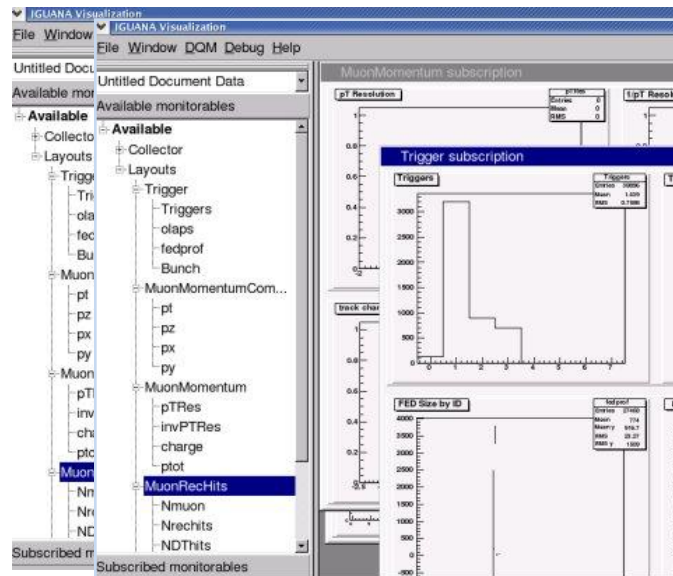
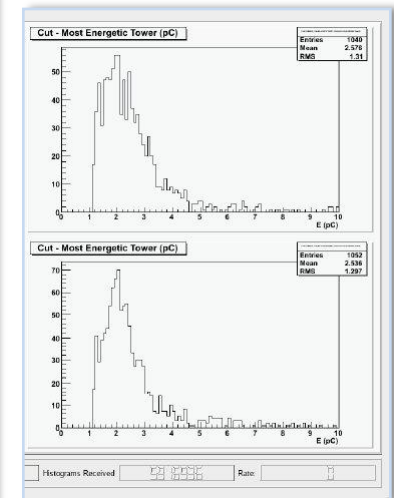
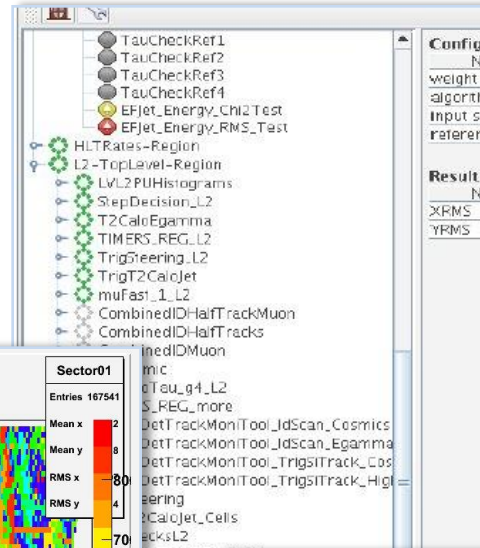
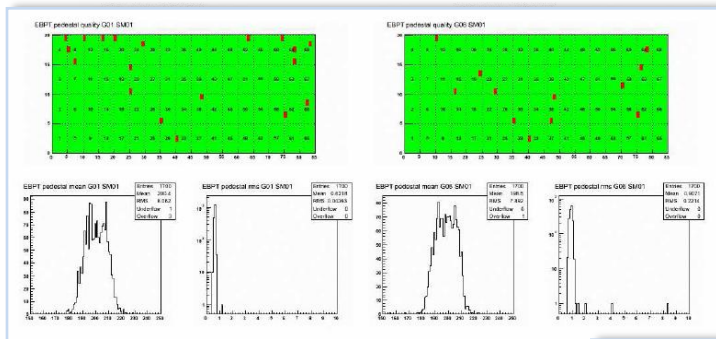
Online event / data quality monitoring

	ALICE	ATLAS	CMS	LHCb
Data source	Before/after Event building	All levels	Full events	Full events Readout boards
Graphics	ROOT	ROOT - GTK	ROOT -IGUANA	ROOT
Storage	MySQL	RDB	Oracle	Oracle
Access	Display process	Display process	Web	Display process
Technologies	C++ DIM / SMI	C++, IPC Java	Web services	DIM

- Quite some activity going on
 - Several successive frameworks
 - Actual needs showed up late
- Complete information systems: collect, process, publish
- Wide range of display, sharing, and archiving mechanisms

Online event / data quality monitoring

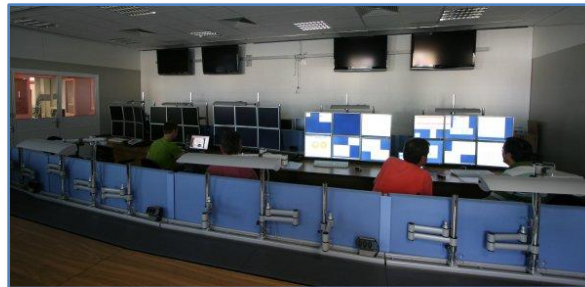
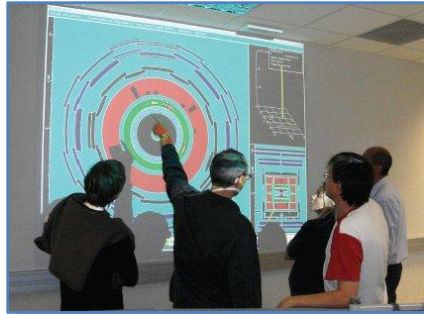
Example screenshots



DAQ operation

- Control rooms
 - Dedicated detector positions
 - Main operation with limited crew once stable
 - Summary to detailed status interfaces
 - Remote tools
- Bookkeeping
 - Existing (e.g. Elog) or custom developed tools

DAQ operation



DAQ operation

TPC

NOT_RUNNING
SYNCHRONOUS
STARTING
STARTING_PDSREC
STARTING_EVB
RUNNING
RUNNING_ERR
STOPPING_EVB
STOPPING_PDSREC
WAIT_STOPPED
STOPPED

LDC (19)

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

DATEALLDETECTORS.DAQ-ALLODETECTORS-CONTROL

File View Options Windows Status updated

ALLDETECTORS
DAQ - Run Control
HI running on pcald37 with PID 5906
RC running on pcald37 with PID 5855

Disconnected Configuration Connected Run Parameters Ready to start Data Taking

Start processes
Start
Stop
Pause
Continue

HLT mode R: DAQ only
Recording disabled

RUN NUMBER : 3 Run Control Status : RUNNING

Trace
Clear
Debug
Pause
Bigger
Smaller

Ps1 10 09:40:57 (RC) Starting Data Taking for run 3
Ps1 10 09:40:57 (HI) Current RC options loaded from : DATE_CONFIG
Ps1 10 09:40:57 (HI) Start processes time : 6 seconds
Ps1 10 09:40:51 (RC) Starting run 3
Ps1 10 09:40:51 (RC) Get and update run number from database
Ps1 10 09:40:51 (RC) New Run options loaded from : Database DATE_CONFIG
Ps1 10 09:35:13 (RC) Connected to remote hosts
Ps1 10 09:35:13 (HI) Connection time 2 seconds

ATLAS DAQ Software Graphical User Interface - Expert Control

File Commands Access Control Tools Settings

Run control

RUN CONTROL STATE: RUNNING

Shutdown Start

Stop Start

Pause Continue

Run Parameters

Run type: calibration
Run number: 210
Event number: 4826400
Event rate: 6.692 kHz
Recording: Disable
Run Start Time: 14/02/07 10:12:39
Run Stop Time:
Integrated active run time: 00:12:22

Monitor Segment & Resource Data Set Tags Infrastructure PMG DataFlow

SFI-Segment-All
SFI-1
SFI-2
SFI-3
SFI-4
SFI-5
SFI-6
SFI-7
SFI-8

Rate stability

2042	07.07.07 12:10	valid	Oleg Solovyanov	Information	Online	open	TDAQ	Re: Start of Run 14848	/det/lar area is not v
2041	07.07.07 12:01	valid	Oleg Solovyanov	Information	Online	open	TDAQ	Start of Run 14848	Disabled EP-C
2040	07.07.07 11:42	valid	Oleg Solovyanov	Information	Online	open	TDAQ	End of Run 14848	PT is always c
2039	07.07.07 11:29	valid	Oleg Solovyanov	Information	Online	open	TDAQ	Re: Start of Run 14843	Run stopped b

InfoViewer - DATE_SITE = localdetectors/site

Level Date Time Host Facility Message

11:00:39 pcald37.com.ch pcnControl New configuration loaded from : Database DATE_CONFIG
11:00:39 pcald37.com.ch pcnControl Connecting to gpc2:
11:00:39 pcald37.com.ch pcnControl Current configuration loaded from : DATE_CONFIG
11:00:40 pcald37.com.ch pcnControl Connecting to gpc2:
11:00:40 pcald37.com.ch pcnControl Starting Log: Expires at 13 Dec 2005 11:00:40 (Mat...)
11:01:11 pcald37.com.ch pcnControl Connection problem with LDC2
11:01:11 pcald37.com.ch pcnControl shutdown (DATEALLDETECTORS-CONTROL)
11:01:23 pcald37.com.ch pcnControl shutdown (DATEALLDETECTORS-CONTROL)
11:01:23 pcald37.com.ch pcnControl pcnControl is unatched

Archive Errors

Select Clear Time Level Database Username System Facility Stream Run Message Query

Create Save max min exclude

Delete Load

Status: Idle
Query: SELECT * from messages ORDER BY timestamp
150 messages, 1 errors, 1 fatal

DATE Online Electronic Logbook v1.07

Views Actions Help Links Logout

Run Statistics

1-20 of 196 (Page 1 of 10)

	Run	Start Time	Duration	LDCs	GDCs	Detectors	Total Events	Total Data (MB)	Data Rate (MB/s)	Events/s	Run Type
(1)	912	21/08/2007 11:26:57	na	1	1	1	na	na	na	na	DAQ
(2)	911	21/08/2007 10:52:41	24 s	1	1	1	2 955	123.14	334.29	123.14	DAQ
(2)	910	21/08/2007 10:47:27	36 m	1	1	1	443 407	240 801	112.16	206.52	DAQ
(2)	909	21/08/2007 10:45:01	22 s	1	1	1	4 673	2 538	115.35	212.41	DAQ
(2)	908	20/08/2007 23:37:47	15 m	1	1	1	93 707	102 566	112.22	102.52	DAQ
(2)	907	20/08/2007 23:09:15	27 m	1	1	1	167 120	182 918	112.08	102.40	DAQ

Detector: TPC

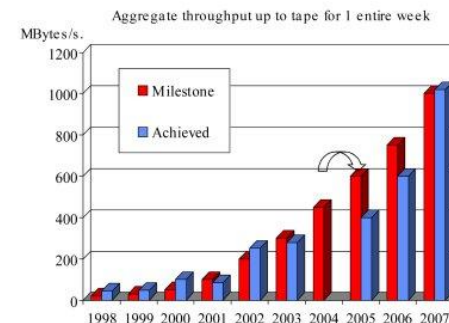
Data source: DEL Internal

Mc-TPC-A00-0	768	763	848	841	842	843
Mc-TPC-A01-0	770	771	844	845	846	847
Mc-TPC-A02-0	772	773	848	849	850	851
Mc-TPC-A03-0	774	775	852	853	854	855
Mc-TPC-A04-0	776	777	856	857	858	859
Mc-TPC-A05-0	778	779	860	861	862	863
Mc-TPC-A06-0	780	781	864	865	866	867
Mc-TPC-A07-0	782	783	868	869	870	871
Mc-TPC-A08-0	784	785	872	873	874	875
Mc-TPC-A09-0	786	787	876	877	878	879
Mc-TPC-A10-0	788	789	880	881	882	883
Mc-TPC-A11-0	790	791	884	885	886	887
Mc-TPC-A12-0	792	793	888	889	890	891
Mc-TPC-A13-0	794	795	892	893	894	895
Mc-TPC-A14-0	796	797	896	897	898	899
Mc-TPC-A15-0	798	799	900	901	902	903
Mc-TPC-A16-0	800	801	904	905	906	907
Mc-TPC-A17-0	802	803	908	909	910	911

Refresh Disconnect all Select All Quit

Commissioning

- Previous years
 - First tests on small independent systems
 - Integration with electronics
 - Data challenges using central facilities
- Now real and large-scale hardware installed and running
 - Standalone DAQ runs
 - Detector commissioning runs
 - Daily operation
- Ramping up of processing farms
 - Waiting gives you better equipment for a given budget
 - Gradually equip to keep up with data flow / money
- Cosmics data taking in progress or starting this fall



Summary ^(1/2)

- Heterogeneous trigger/bandwidth requirements
 - Ad-hoc architectures
- Common base trigger components
- Custom electronics for :
 - Low level trigger
 - First layers of pipeline buffers in data flow
 - Input multiplexing
 - Detector data links

Some parts reused in non-LHC experiments

Summary (2/2)

- Commodity hardware (according to experiments needs) for:

- Networking / Event Building
- High level trigger
- Storage
- Operation and control

Accessibility, performance, long-term maintenance

- Common operating system
- Mostly custom software
excepted for few standard tasks (e.g. database)
- Common Mass storage archiving system

DAQ design challenges solutions

- Cope with the huge quantity : *appropriate hardware, flexible/scalable architecture*
- Select the appropriate events : *different filter layers*
- Ensure measurements integrity : *radiation tolerance, data headers, software checks and fault recovery*
- Monitor to check quality : *automatic histograms and displays, in parallel to data flow*
- Record for analysis and archive : *disks and tape pools*
- Operate such a complex system : *dedicated control tools*

Conclusions _(1/2)

- Main trend: use standard equipment whenever possible
 - Well done, especially where big numbers involved
 - Follow industry standards allows to benefit from best rates (bandwidth and price!)
 - The less cables, the better
- Custom hardware limited where necessary
 - specific constraints at the beginning of data flow
 - critical synchronization / latency, radiation, number of channels

Conclusions _(2/2)

- Flexibility will show benefits very soon
 - good designs proved when facing the unexpected
 - all look to have such a quality
- For the future: hope for more standard software ?
 - Although no “consumer-like” tasks in DAQ
 - But PC-based cluster computing technologies have significantly progressed: some tools are getting mature
 - Integration of distributed computing tools at the level of OS?

Thanks to the DAQ project leaders and their teams
for providing valuable information and feedback to
prepare this talk !