The LHC Situation

Chris Bee

Centre de Physique des Particules de Marseille, France,

Contents

- **First collisions: July 2005!**

Event Filter Farms in the LHC Experiments

Chris Bee

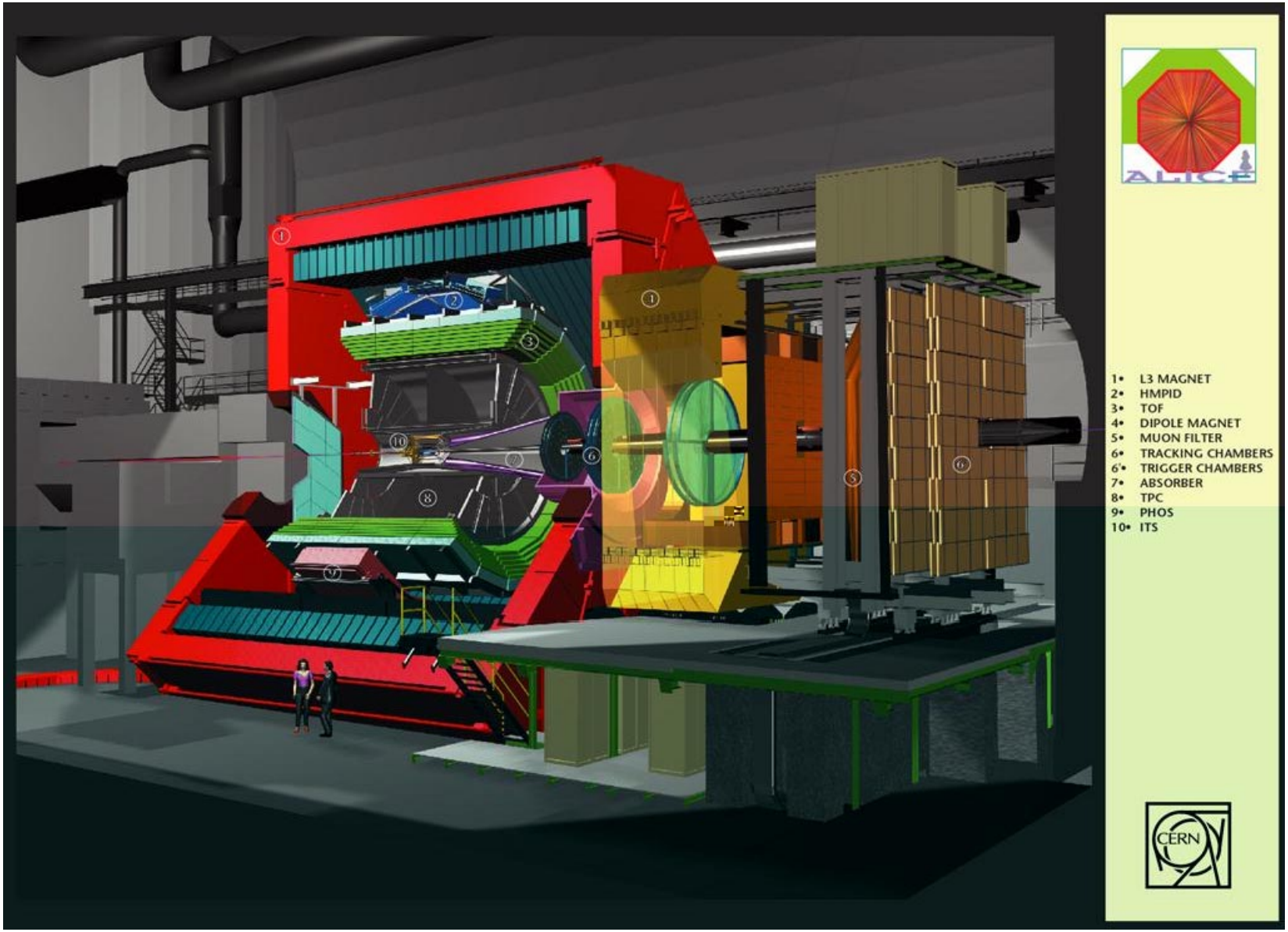
Centre de Physique des Particules de Marseille, France,

Contents

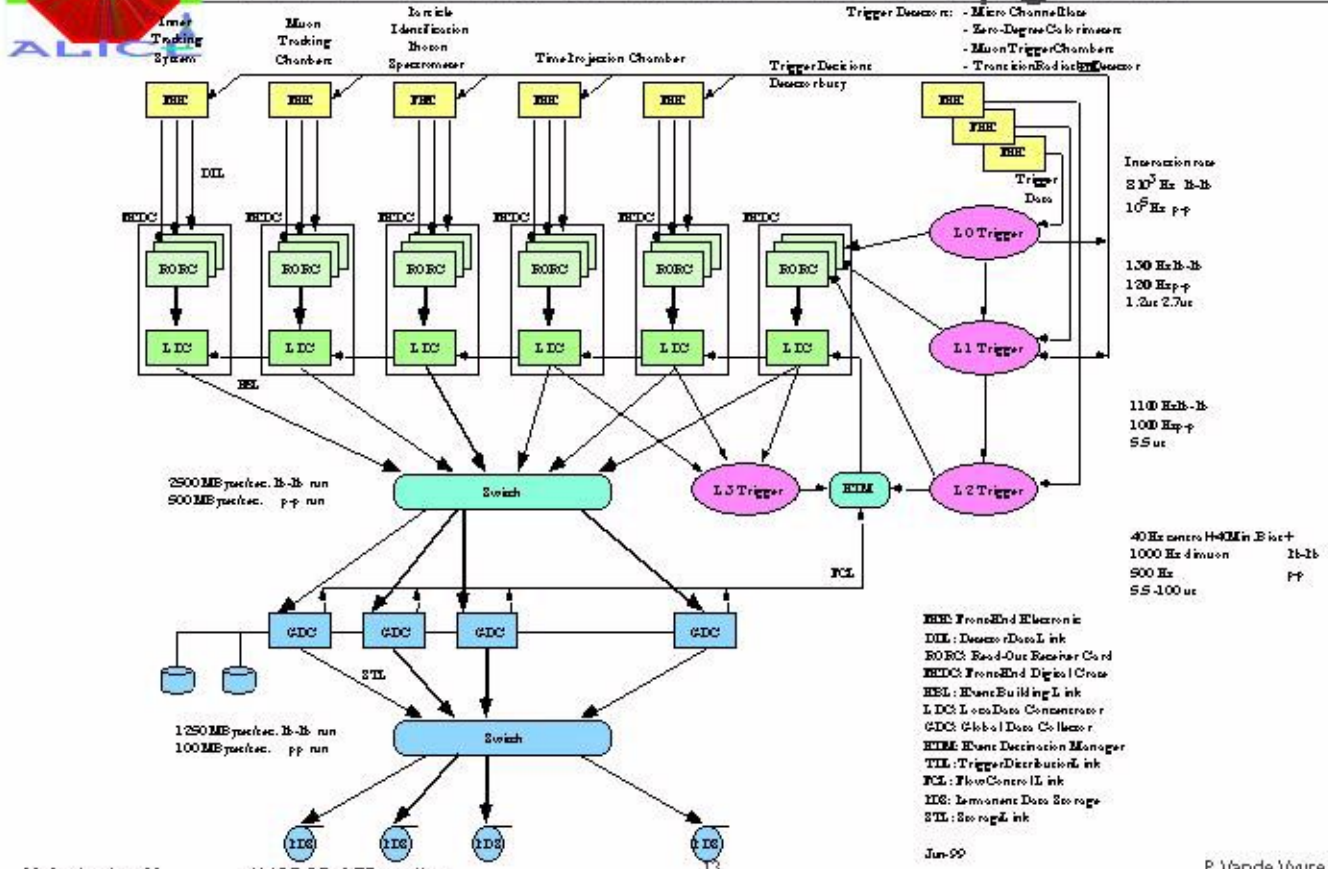
- ALICE - Pierre Vande Vyvre
- ATLAS - yt
- CMS - Johannes Gutleber
- LHCb - Themis Bowcock, Beat Jost
- LCB Event Filter Farms Common Project
- Conclusions

ALICE

- Special purpose experiment to study physics of strongly interacting matter at extreme energy densities. New phase of matter -> quark-gluon plasma
- Heavy ion (Pb-Pb, Ca-Ca) collisions
- Equivalent of 274M channels
- L0 accept rate: 1.3 kHz (Pb-Pb & Ca-Ca) @ 10^{27}, 1.2 kHz (p-p) @ 10^{30}
- L1 accept rate: 1.1 kHz (Pb-Pb, Ca-Ca)
- Event size: 500 kB (p-p), 67 MB (Pb-Pb - Central)
- Event building bandwidth: ~ 3 GB/s
- Pb-Pb L2 rate: 50 Hz - Central & min. bias, 1 kHz - Dimuon, 200 Hz - Dielectron
- L3 accept rate: ~as above: factor 2 data compression + ...
- L2 + L3 cpu power: 2.10^{6} MIPS
- Data rate to storage: 100 -> 1,250 MB/s


ALICE
1• L3 MAGNET
2• HMPID
3• TOF
4• DIPOLE MAGNET
5• MUON FILTER
6• TRACKING CHAMBERS
6'• TRIGGER CHAMBERS
7• ABSORBER
8• TPC
9• PHOS
10• ITS
CERN

TRG-DAQ architecture upgrade



Event Filter Farm - ALICE

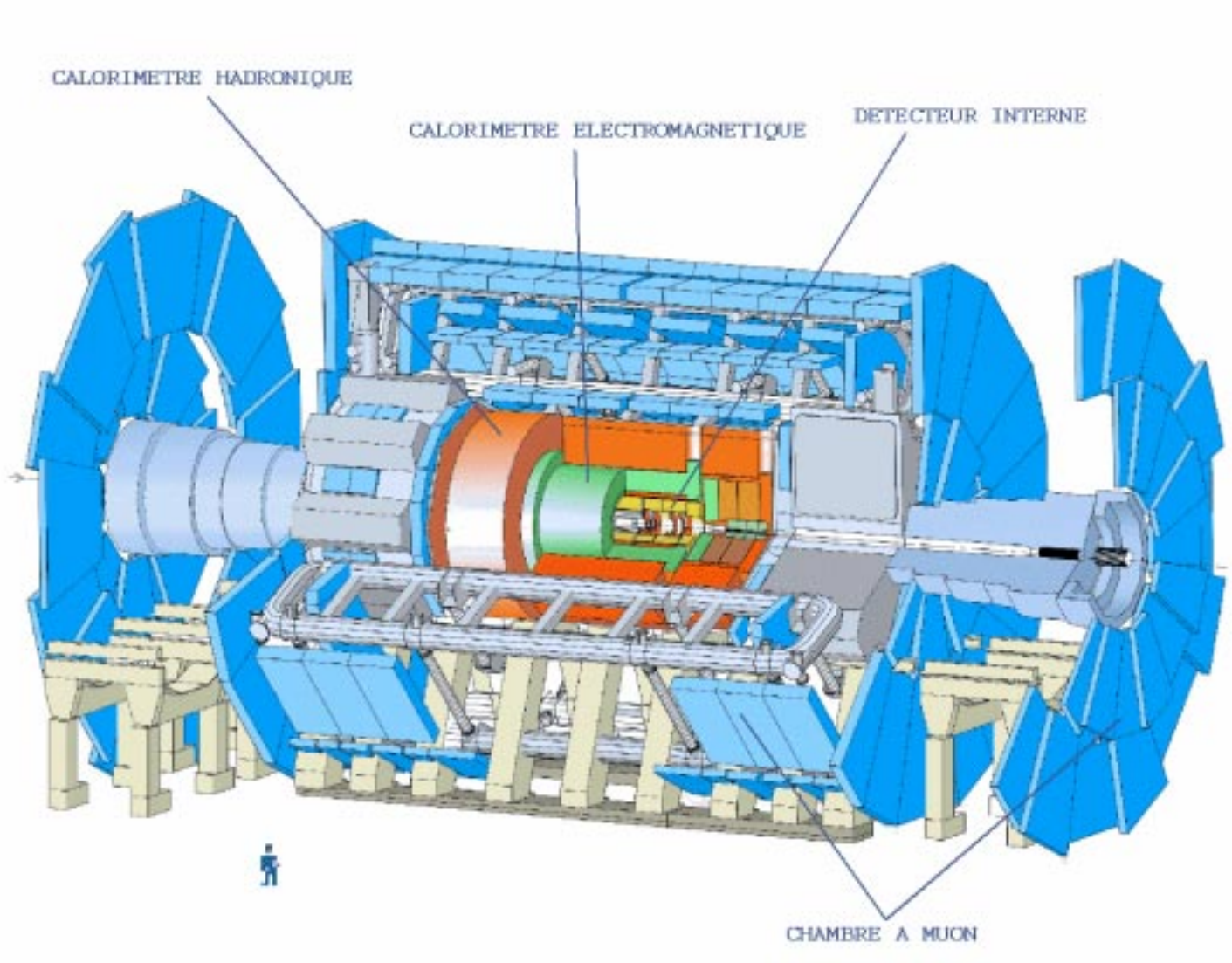
- Data compression on all data (lossless, must work on day 1) factor 2
- Originally, no trigger decision after L2 was planned - but after review of physics rates -> introduction of new TRD detector, higher average TPC occupancy + increase of TPC data volume
- Storage capacity max. limit of 1.25 GB/s but event building bandwidth ~ 3 GB/s
- -> Partial readout & filter: Identify TPC sectors containing tracks (indicated by TRD) & execute trigger algorithm ... select tracks & make decision -> factor 10 for di-electron triggers ... 'realistic'
- -> Reduce total (largely TPC) data volume (~67 MB/event in Pb-Pb central events) by online reconstruction - factor of 5 on central triggers ... 'ambitious'
- No specific EFF developments done yet

Future Work

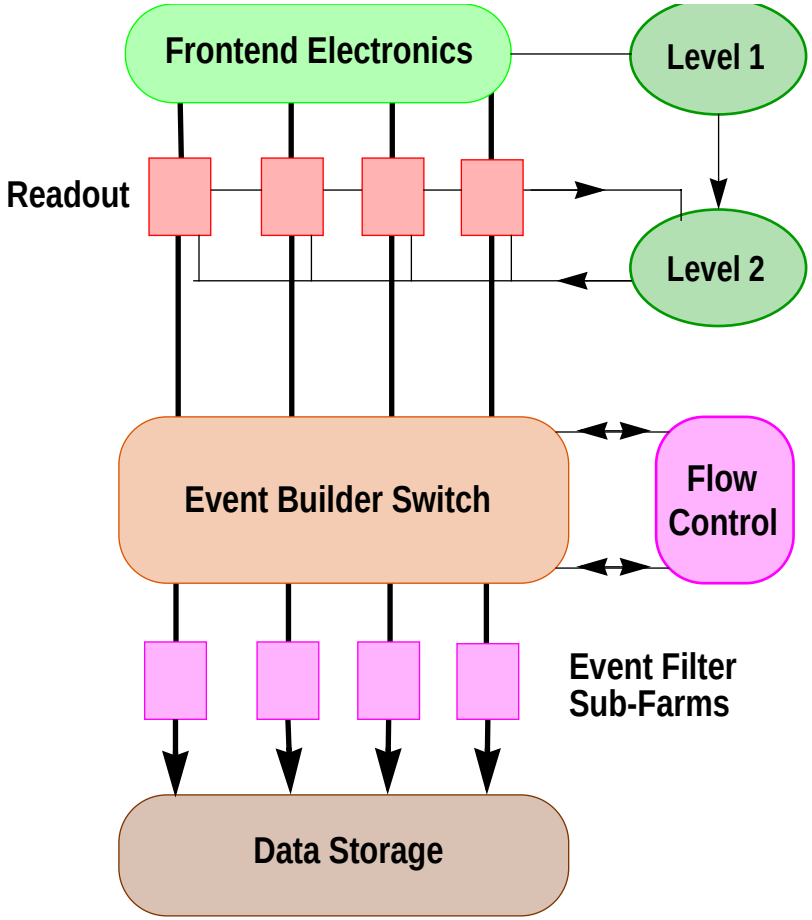
- Discuss the physics requirements and perform simulation
  - Provide a more precise estimation of the new requirements
  - Fundamental parameters: event sizes, trigger rates, TPC/TRD partial readout
  - Optimised format for TPC/TRD zero-suppressed raw data
  - Develop L3 trigger algorithm on realistic simulated data
  - Produce digitised data
- Prototyping
  - The existing test-bench is the ALICE Data Challenge
  - Includes now DAQ, compression, ROOT I/O and Mass storage in the centre
  - Use the digitised data to build a realistic prototype including L3
  - L3: partial readout, trigger and online reconstruction
- Report to the collaboration
  - Physics simulation
  - Cost and behaviour of different scenarios of TRG-DAQ architecture
  - Make a proposal to the collaboration

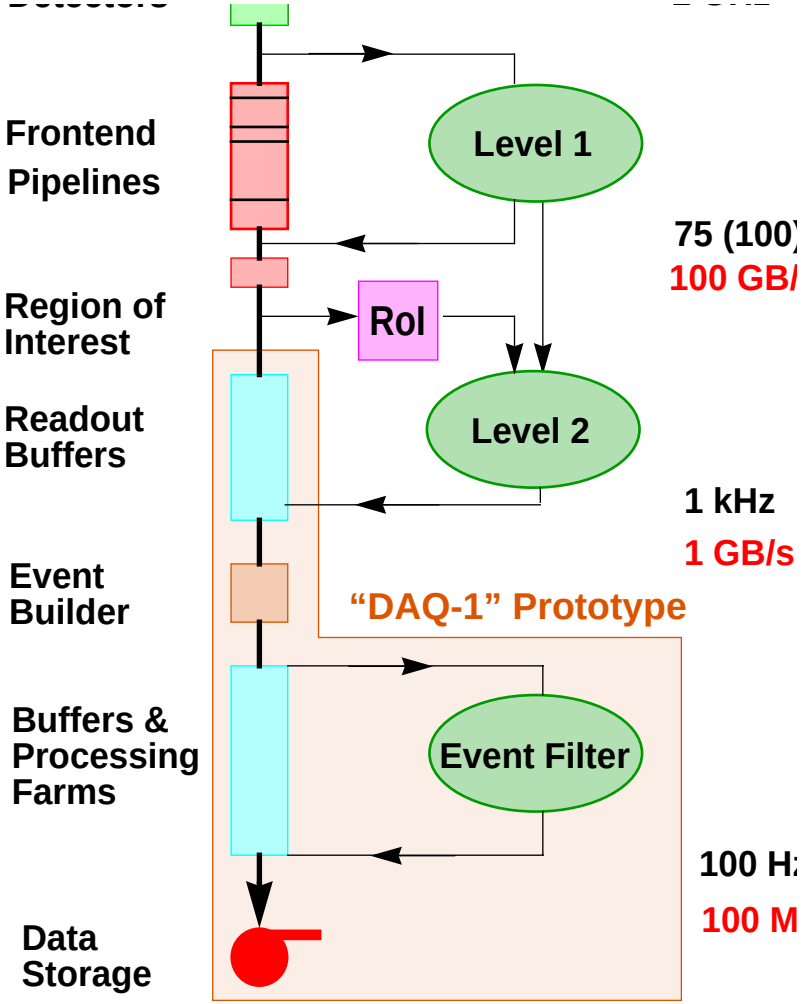
ATLAS

- General purpose p-p detector
- 150M channels
- L1 accept rate: 100 kHz -> readout
- L2 accept rate: 1 kHz
- Event size: 1 MB
- Event building bandwidth: 1 GB/s
- L3 accept rate: ~100 Hz
- L2 + L3 cpu power: > 2.10^6 MIPS
- Data rate to storage: 100 MB/s


CALORIMETRE HADRONIQUE
CALORIMETRE ELECTROMAGNETIQUE
DETECTEUR INTERNE
CHAMBRE A MUON

Trigger DAQ architecture





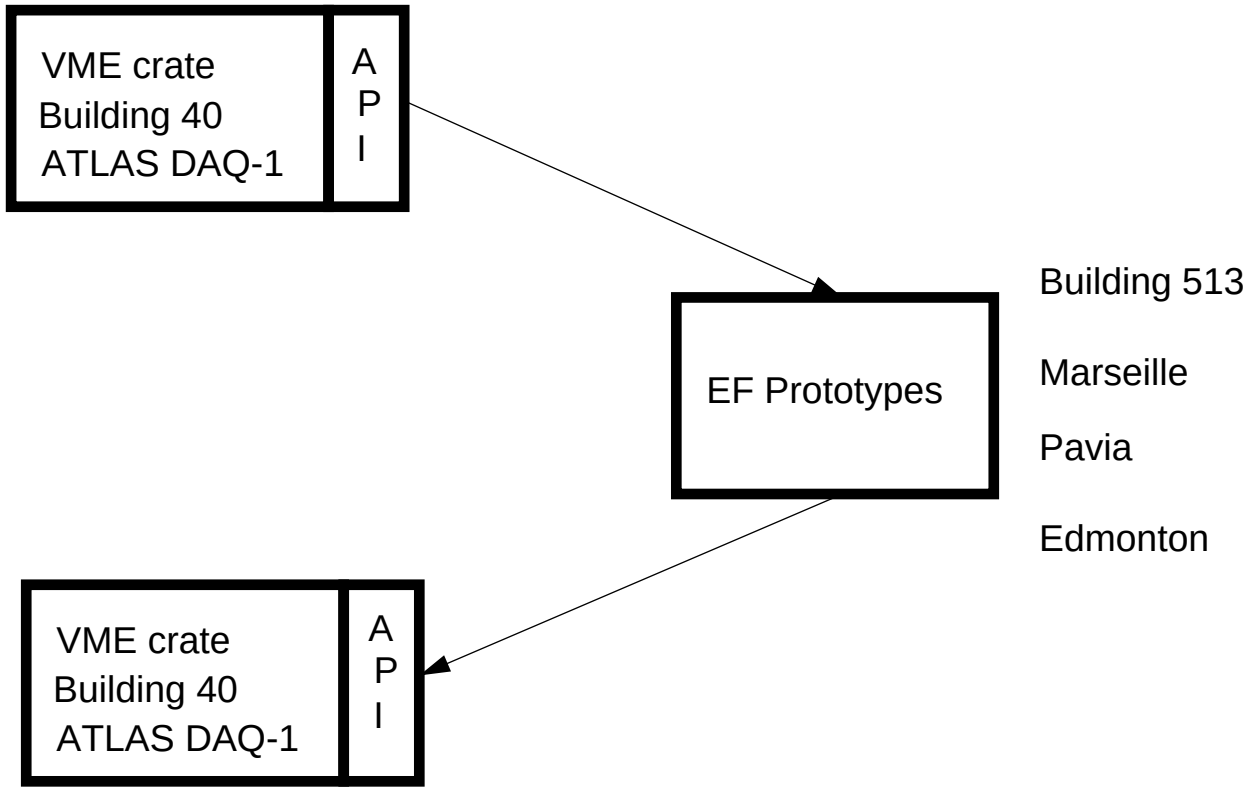
Event Filter Farm - ATLAS

- **Event filtering**
- **Reduction of rate from LVL2 output of ~1 kHz to ~ 100 Hz**
- **Online 'full' detector monitoring**
- **Online physics monitoring**
- **Calibration procedures**

Developments - ATLAS

- **3 small prototypes built**
  - Commodity PCs (Dual Pentium II), fast ethernet switch (NT)
  - HP 4 cpu SMP (HP-UX)
  - Pentium quads (Linux)
- **Single software design for dataflow - different implementation at inter-proc comms level**
- **Monitoring & Control prototype using Java agents**
- **Running realistic ATLAS offline code**

- **Tests done in IT on PCSF (~15 bi-nodes)**
- **All 3 prototypes integrated into ATLAS “DAQ-1” prototype**

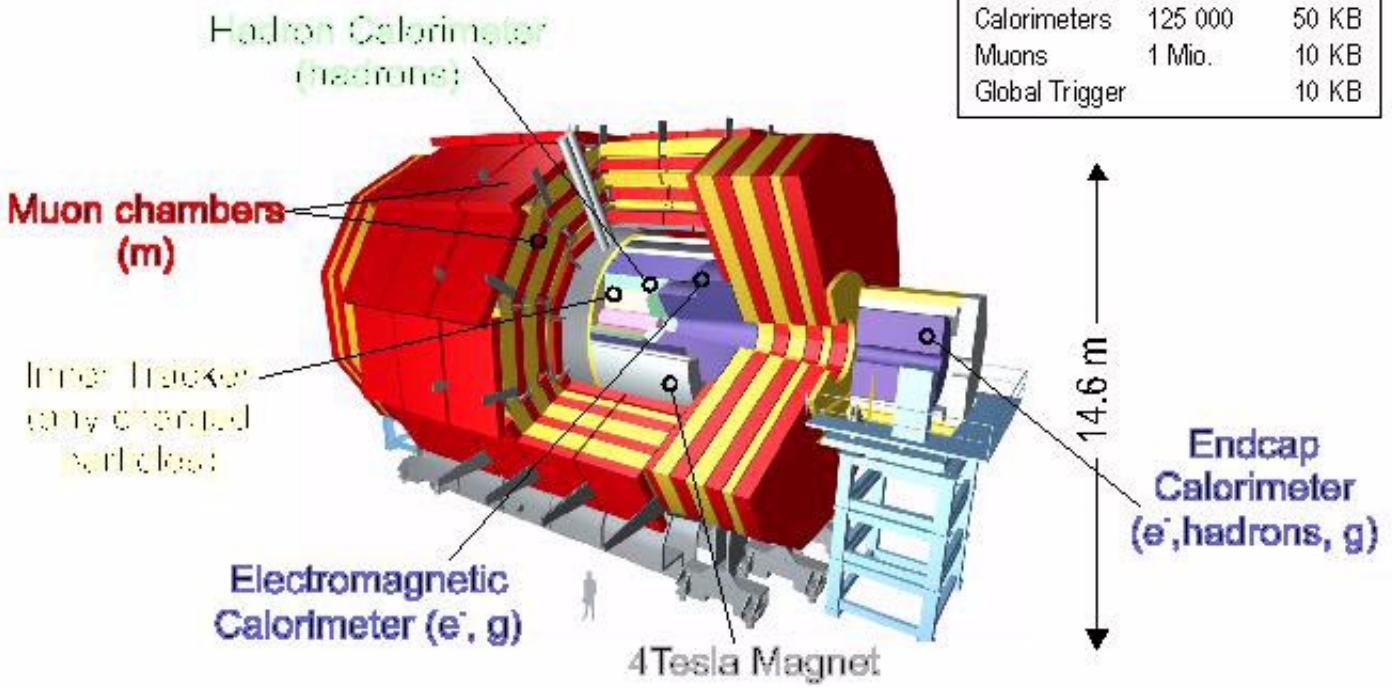


CMS

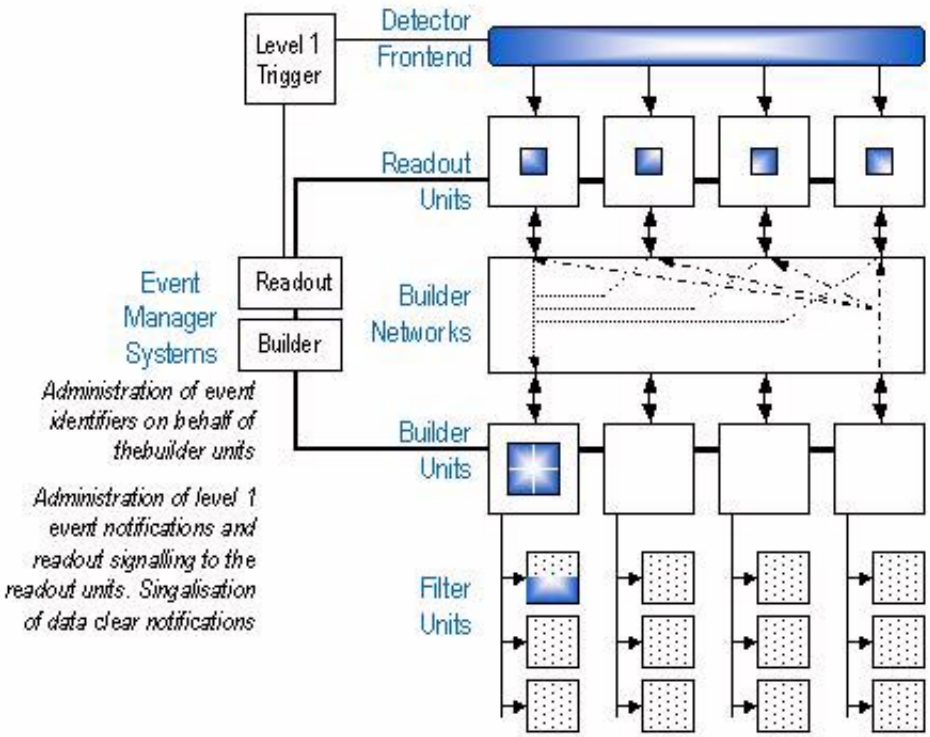
- General purpose p-p detector
- 98M channels
- L1 accept rate: 100 kHz -> readout
- Full event size: 1 MB
- Readout network bandwidth: 60 GB/s
- EF accept rate: ~100Hz
- EF cpu power: 5.10^6 MIPS
- Data rate to storage: 100 MB/s

Compact Muon Solenoid

Pixel	80 Mio.	100 KB
Inner Tracker	16 Mio.	700 KB
Preshower	512 000	50 KB
Calorimeters	125 000	50 KB
Muons	1 Mio.	10 KB
Global Trigger		10 KB



DAQ Overview



Event Filter Farm - CMS

- ‘LVL2’ triggering + Event filtering - multi-level process
- Reduction of rate from LVL1 output of ~100 kHz to ~ 100 Hz
- Full events only built at ~ 100 Hz
- Online ‘full’ detector monitoring
- Online physics monitoring
- Calibration procedures

Filter Farm Requirements

- The farm must have a sustained computing capacity of 5 Tera OPS.
- The farm must be built from general purpose computing equipment in order not to enforce a specific operational environment.
- The inter farm connection must be capable of handling a peak load of messages at 100 kHz up to 2 KB
  - software patterns may be used to achieve this rate (e.h. scatter gather) – timeliness is not an issue.
- The connection to the computing services must be capable of performing data transfer at 100 MBytes/sec.

CMS Specific Farm Requirements

- The event filter farm shall be partitionable to subfarms. Each subfarm shall be connected to one builder unit.
  - increase per builder network port throughput
  - benefit from fast local communication between FU and BU
- The on-line processing framework for the event filter farm shall use the CMS off-line data structures for event access and calibration.
- The on-line framework shall be platform independent.
- The on-line framework shall be decoupled from data transfer and data storage management systems.

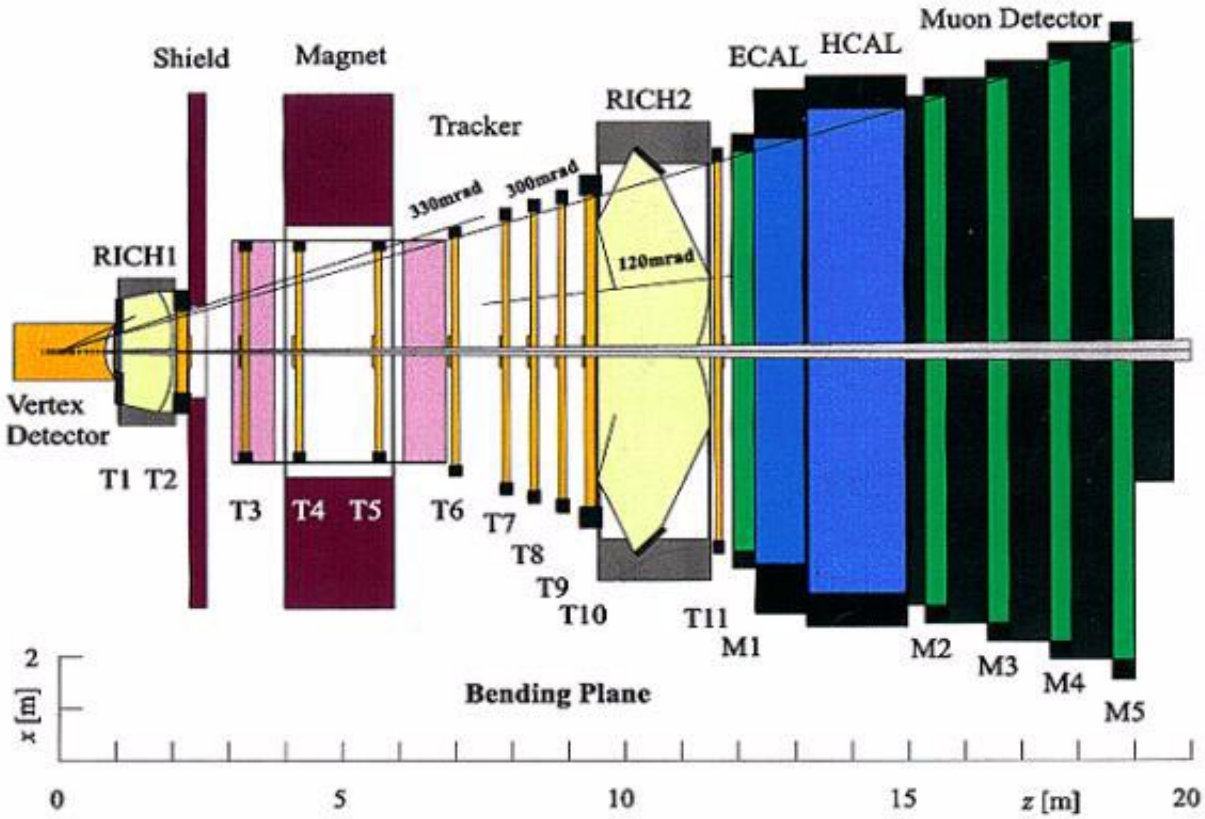
Farm Control

- The filter farm control system shall allow to include any task in its scope
- The filter farm control system shall allow to define policies freely
  - no hardcoding of how commands are sent
  - extension of the system with new commands
- The farm control system shall be able to communicate in platform independent information format
  - XML, GIOP
- The farm control system shall be able include control of the readout units, builder units and event managers.

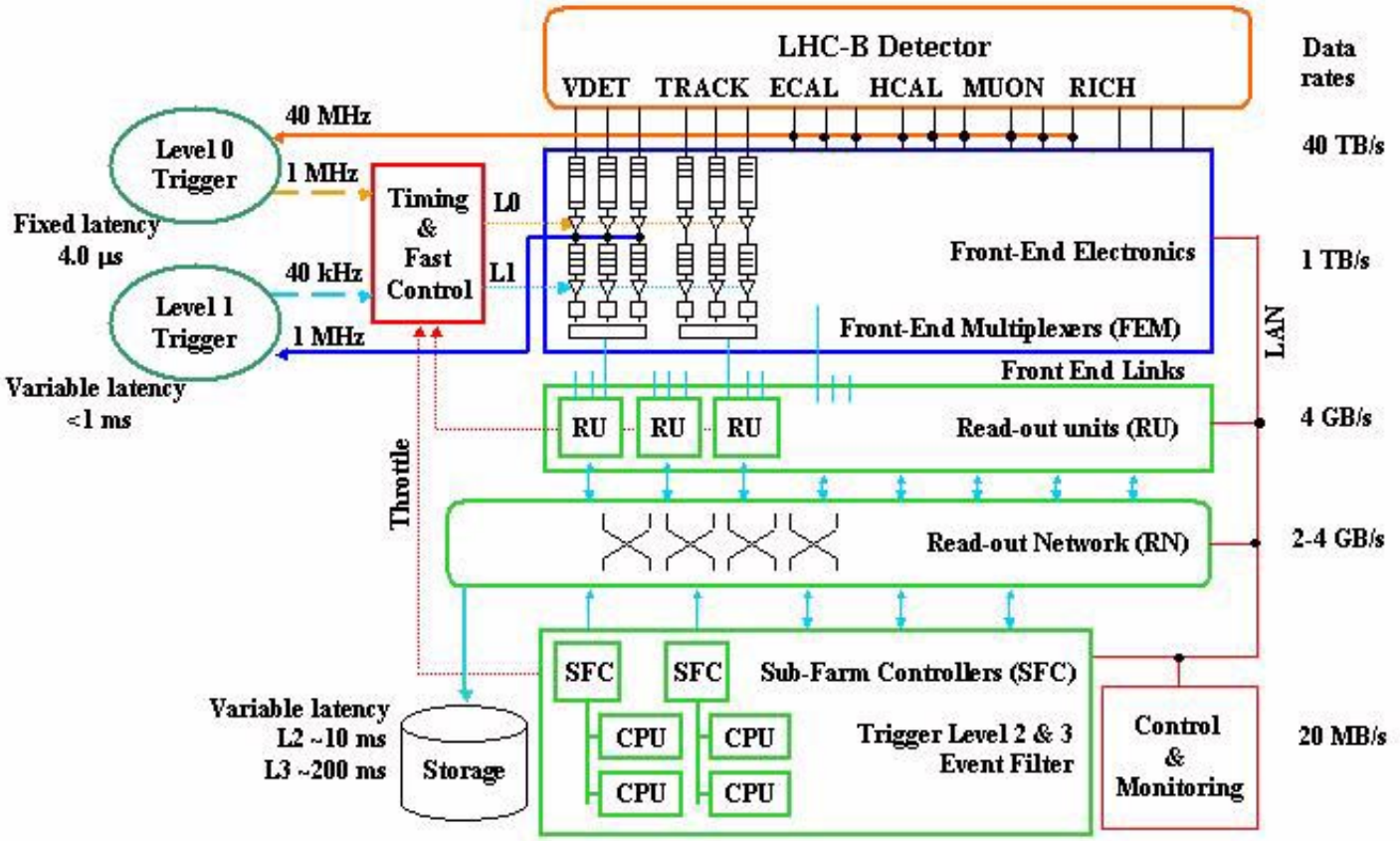
LHCb

- Special purpose experiment to measure precisely CP violation parameters in the $b\bar{b}$ system
- Single arm spectrometer with one dipole
- 0.9M channels
- L0 accept rate: 1 MHz
- L1 accept rate: 40 kHz -> readout
- Event size: 100 kB
- Event building bandwidth: 4 GB/s
- L2 accept rate: ~5 kHz
- L3 accept rate: ~200 Hz
- L2 + L3 cpu power: 2.10^6 MIPS
- Data rate to storage: 20 MB/s

LHCb Detector



Trigger/DAQ Architecture



Event Filter Farm - LHCb

- LVL2 + LVL3 triggering
- Reduction of rate from LVL1 output of ~40 kHz to ~ 5 kHz (LVL2) & then ~ 200 Hz (LVL3)
- Special triggering challenges: inelastic p-p interactions & events with b-mesons very similar -> very complex trigger algorithms
- Online ‘full’ detector monitoring
- Online physics monitoring
- Calibration procedures

No specific EFF developments done yet

MAP cont'd

MAP hardware

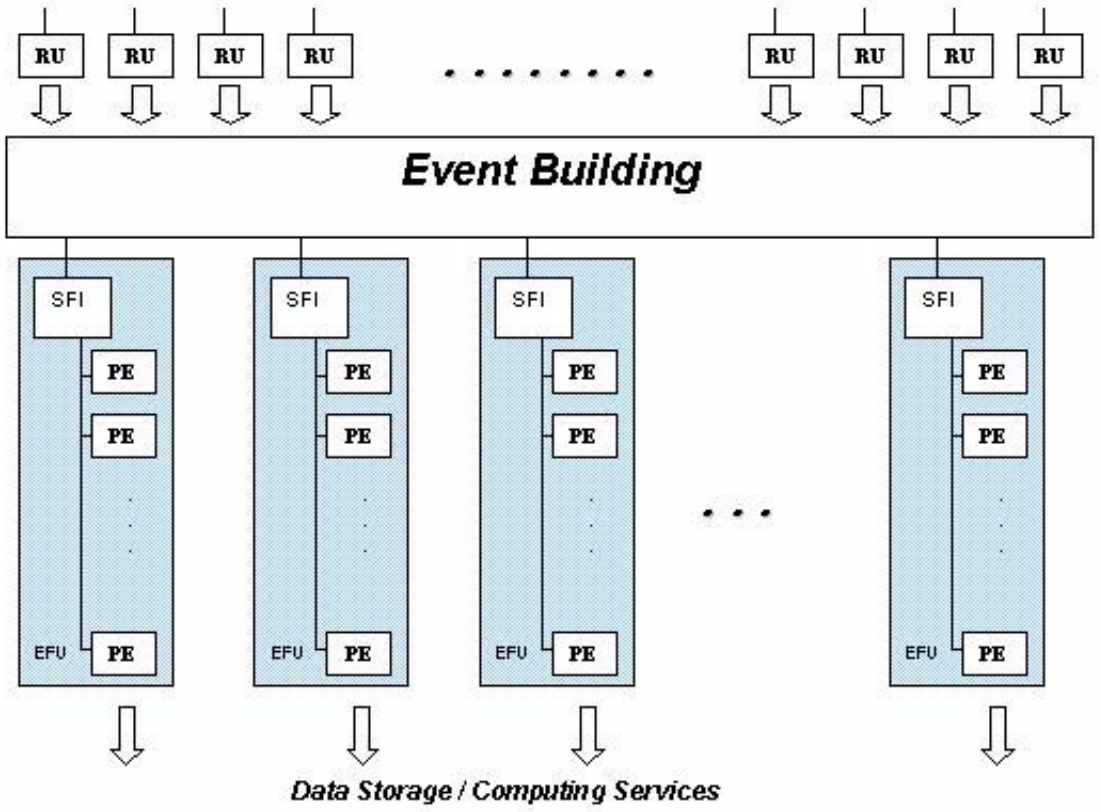
- 300 processors
  - 400MHz PII
  - 128 Mbytes memory
  - 3 Gbytes disk
  - 100BaseT ethernet +hubs
  - commercial units BUT
    - custom boxes for packing and cooling

LCB Event Filter Farms Project

Background

- All 4 LHC experiments planning to use Event Filter Farms
- Different trigger rates, connections, data flows but output is usually a full event (with its reconstruction)
- IT/PDP already using commodity farms
  - Traditional batch & CDR (PC, RISC)
  - RISC,PC performance evaluation studies (NA45, NA48, Compass)

Simplified EF Schematic



Main challenges

- Management/Control/Monitoring of the filter applications
  - reconstruction, compression, partial tracking, etc.
  - 1000's of tasks
- Management/Control/Monitoring of the "computer system" itself
  - could be 1000's computer systems
  - or one very large SMP
  - or ...

Status of the project

- February 98 - PAP submitted
  - 3 Phases
    - Identify common issues & prepare PEP
    - Study and address the issues
    - Prototypes - demonstrate scalability
    - ATLAS, CMS (ALICE/LHCb initially observers) and IT
- December 1998 - PEP submitted

Goal & Scope

- Demonstrate that conventional (commodity) computing can be used for on-line software filtering in the context of LHC data acquisition chains
- Investigate & address issues created by L3 stages of the DAQ
  - Cluster Architectures & Management
  - Application Management/Control/Monitoring
- Exchange of information/solutions with LHC and others.
  - "Farm Computing"

Main deliverable

"Set of tools and documented recipes on how to build, operate and run a powerful, stable and friendly controlled filter farm, suitable for running experiment analysis code and usable at test beams as DAQ backend"

Cluster Architectures & Management

- This is usually work in IT
  - Systems administration
  - Hardware/software monitoring
  - Benchmarking
  - Communications
  - Architectures
  - Resilience
  - Modeling and Simulation

Systems administration

- Current IT/PDP services "only" on < 650 processors in < 400 machines
- Outsourcing the System Management would cost 1 MCHF/year for 1000 processors
- Although a lot of tools exist in IT, no coherent set, no global monitoring, etc.
- Commercial tools exists (but expensive) but not clear they offer so much advantages

ASCI Blue Pacific

- 108 Frames
- 1464 Nodes
- 5856 CPU's
- 85 M$ contract !
- Proprietary management software

Benchmarking

- A lot of benchmarking is done everywhere
  - Duplicate work
  - Measure different things

⇨Agree on metrics

⇨Develop a common set of benchmark tools

⇨Use the tools when technology becomes available (communication/computers); publish results

Communications & Architectures

- Measure, assess, test new (or existing) technologies
  - Using developed benchmarking tools
  - Allow for experiments accessing the equipment
    - example: ATLAS/Marseille scalability tests
  - Deliverable : coherent reports, access to equipment for experiments

Application Management, Control and Monitoring (1)

- This is usually the work of experiments
  - **Application monitoring**
    - e.g. rejection rate
    - challenge is here to present data collected from 1000's of tasks/threads
    - deliver prototype

Application Management, Control and Monitoring (2)

- Application management
  - Configuration and Control
    - How to control/manage/configure 1000's of tasks
    - Recovery/Decision strategy in case of failures
    - Deliver prototype with live demo
- Man Machine Interface
    - Definition of "levels"
    - Prototype
    - Automation of "Web" forms generation

"Farm Computing"

- This is experiment specific
  - Define requirements in terms of Farm management ("Node" for CMS, "subfarm" for ATLAS, GDC for ALICE) and Application monitoring/control
  - Regular workshops with LHC experiments, HEP experiments (CDF, HERA B, NA48, etc.), and IT

Potential benefits

- Common solutions when appropriate
- Possibly a system management model exportable at other institutes
- Reuse what's done elsewhere
  - IT and other laboratories
- Technology evolution tracking using production or pre-production facilities
- Common forum for discussions

But ... activities already started

- IT participation in ATLAS EF
  - Meetings
  - DAQ Note 61, DAQ Note 97; CHEP poster & paper
  - Model of Marseille farm available
  - ATLAS EF prototype scalability tests (May 98; Oct 98; Dec 98)
- IT participation in CMS Tridas
- CDF Farm presentation in May (mini workshop)
- Farm management issues being addressed in IT/PDP
- Farm networking issues being addressed in IT/CS,IT/PDP
- Modelisation/Simulation tool available form IT/PDP
- ALICE has shown recently interest

Proposed Milestones - 1999

Task	Milestones	Dates
0.0.2	Project set-up : assignments and workshops dates	10 March 1999
1.1.4	System management and administration requirements and market survey report	24 November 1999
1.2.3	Hardware software monitoring requirements and possible solutions report	13 October 1999
1.3.3	Benchmarking tools	13 October 1999
2.1.4	Application monitoring requirements and parameters report	24 November 1999
2.2.3	Application management, control and configuration requirements and market survey report	24 November 1999
2.3.3	Man machine interface definition and prototype report and demo	24 November 1999
3.2.2	Event Filter Workshop 1999	June 1999 ? - will be determined after task 002 milestone completed

Conclusions

- **All 4 LHC experiments have need of an Event Filter Farm**
- **Although the T/DAQ architectural details of the 4 experiments are different, the requirements on the Filter Farm itself are very similar**
- **All four experiments envisage full output of selected events**
- **All 4 experiments are 'watching technology' and relying on current trends in technology to continue**
- **The mechanics of the Filter Farm in all four cases will be very similar if not identical**
- **=> The LCB EFF seems the right way to go ... should be actively supported & encouraged at the highest levels**
- ***All experiments are relying on the LCB EFF project to provide them with major input for their EFF***