

# Controlling a large CPU farm using industrial tools

Alba Sambade Varela

*On behalf of the LHCb Online group*

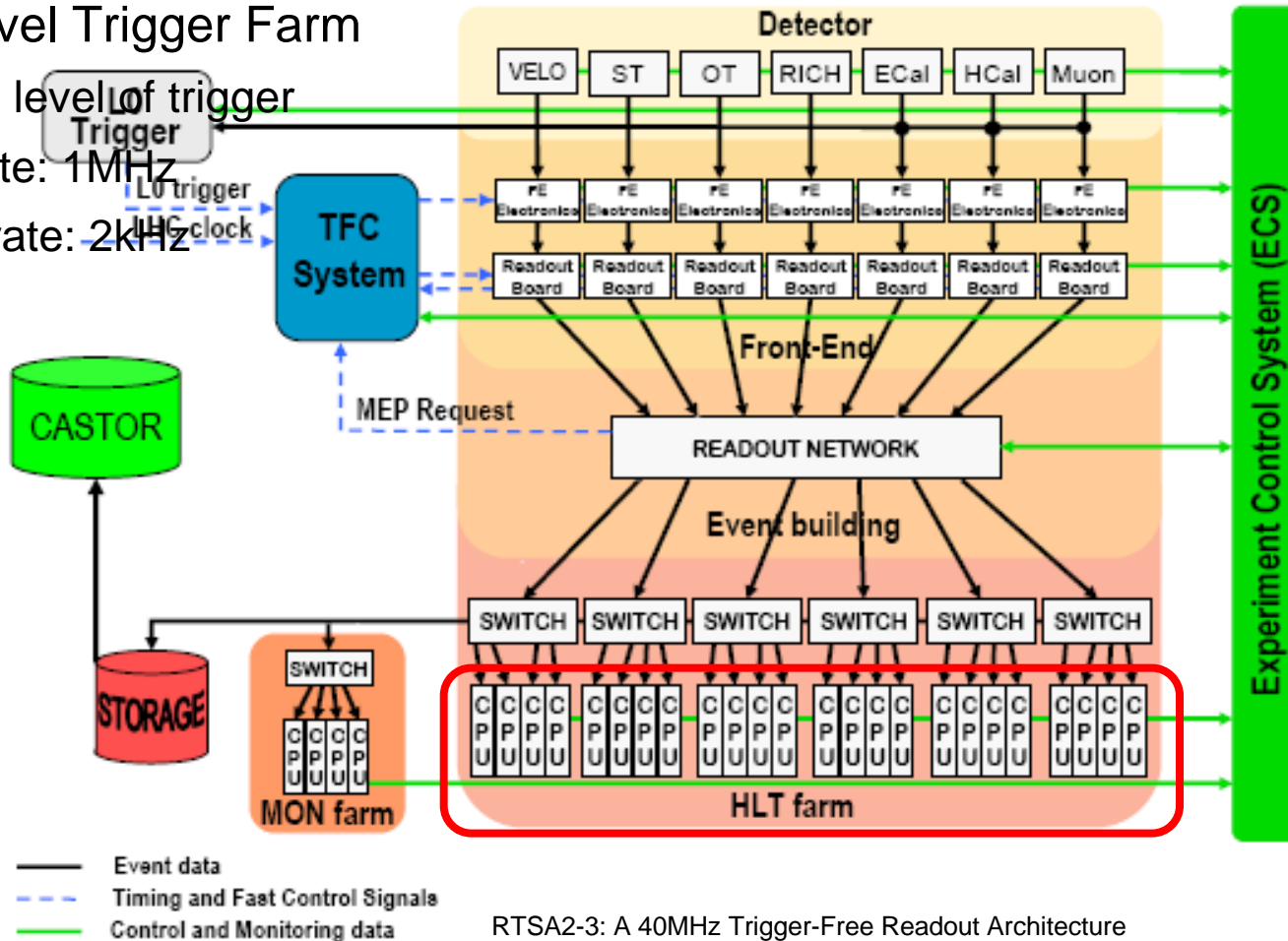
***Real Time 2009***

***May 10-15 IHEP Beijing***

# ECS and data flow in LHCb

## High Level Trigger Farm

- Second level of trigger
- input rate: 1MHz
- output rate: 2kHz

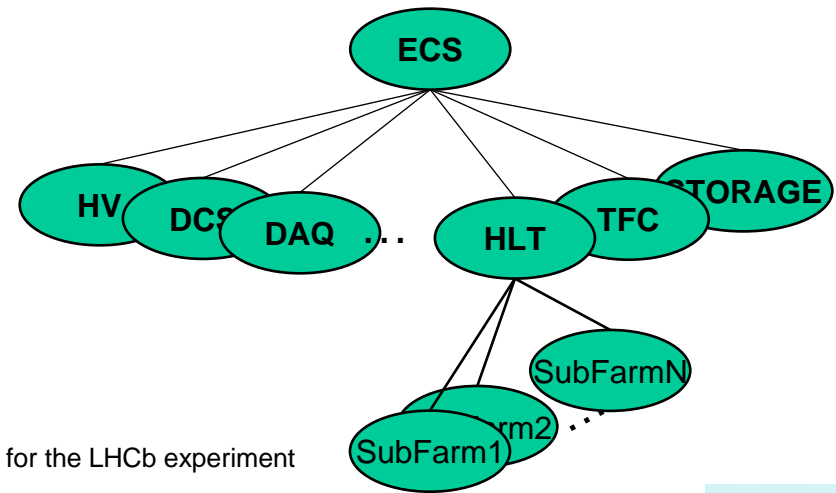


RTSA2-3: A 40MHz Trigger-Free Readout Architecture for the LHCb Experiment at CERN

# Experiment Control System



- ECS is in charge of the configuration, operation and supervision of all the online components in LHCb.
  - Industrial SCADA system: PVSS
  - FSM package
    - Definition in terms of hierarchies of Finite State Machines.
  - Distributed Information Management System (DIM)



See talk CMS1-1  
An Integrated Control System for the LHCb experiment

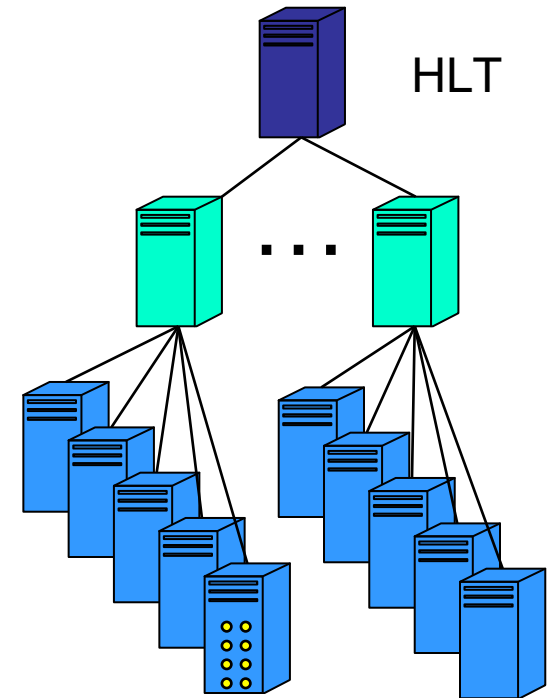


# What do we have to control?

# High Level Trigger CPU Farm



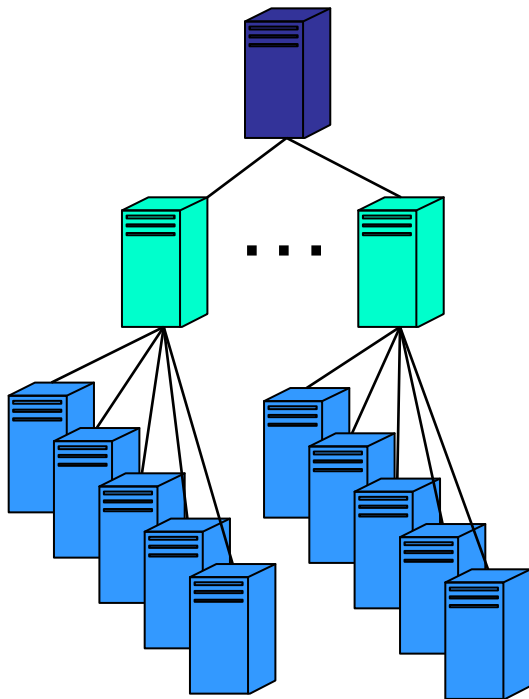
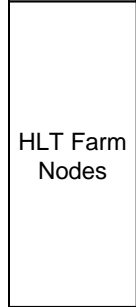
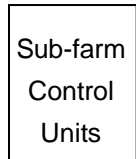
- Composed out of up to 2000 boxes (= nodes)
- grouped into 50 subfarms (50 racks)
  - up to 40 nodes/subfarm
    - 8 cores/node \*
    - 1 HLT algorithm running/core



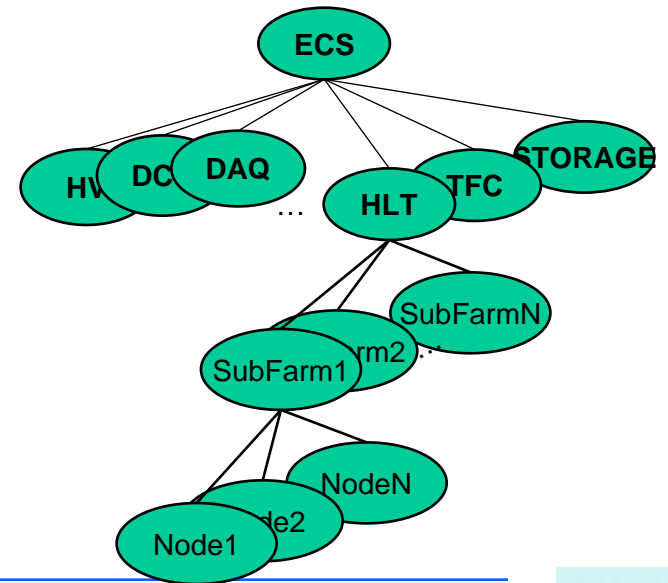
\* Nowadays

# HLT Top level control

Control levels



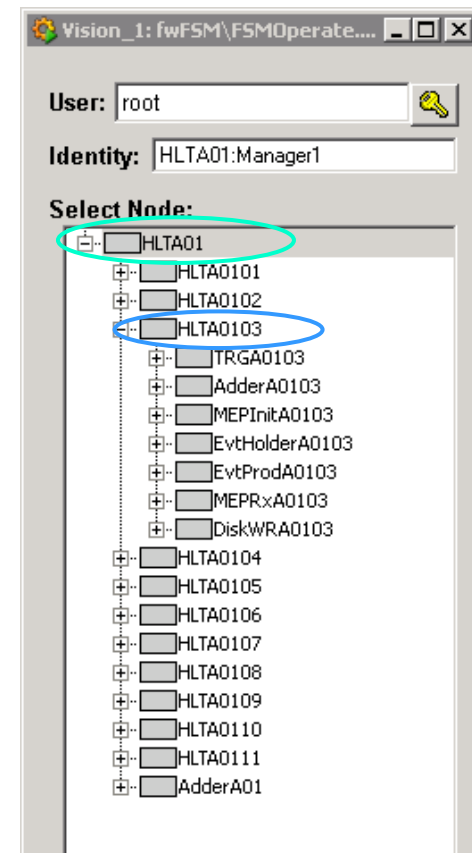
- HLT control PC
  - Running main HLT project
- Control PC per sub-farm
  - Running its own control project
  - Corresponds to a control unit in ECS



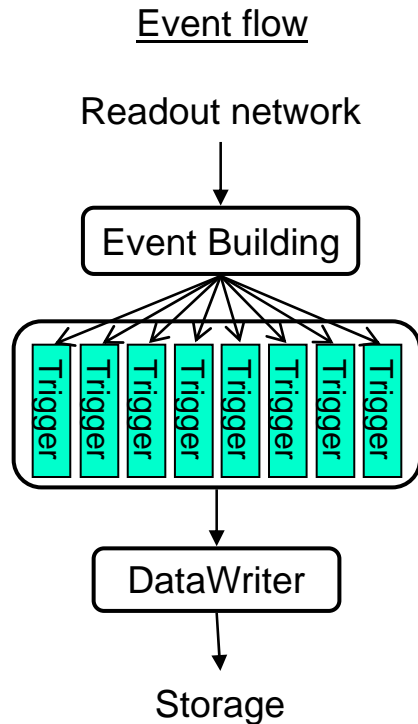
# HLT low level control



- Logical aggregation
  - Functional algorithms
- Division of HLT node by tasks



# Task architecture on HLT node

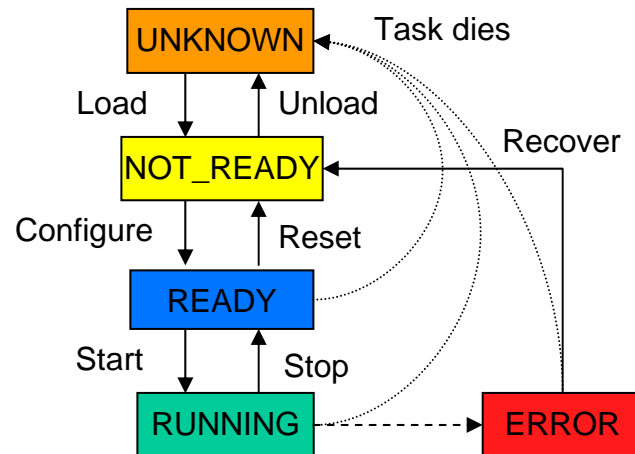


- Each node in the EFF runs one event-builder process.
  - It distributes the assembled events to trigger processes.
- As many trigger-processes as there are CPU cores.
  - Compute trigger decision and declare accepted events.
- Each node also runs one instance of the data-writer.
  - Sends accepted events to Storage system.
- Algorithms implemented with GAUDI.
  - GAUDI: data processing experiment independent framework.
  - Same software used as for offline analysis.



# HLT Tasks control

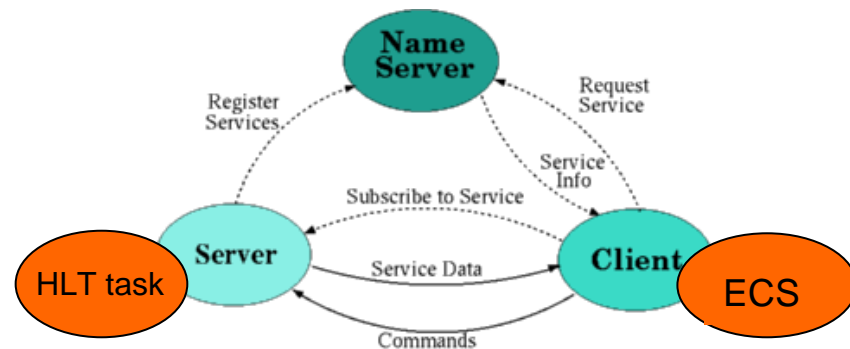
- Algorithms are treated by the ECS as hardware devices.
  - Integrated as Device Unit (DU)
- DU behavior modeled with Finite State Machines
  - Common state diagram for all algorithms
  - Transitions mapped to Gaudi transitions
    - Command parameters



# Communication layer



- Communication via Distributed Information Management System (DIM)
  - Communication mechanism based on client-server paradigm
- Task behaves as a DIM server
  - Publishes services
    - Algorithm status
    - Counters
  - Receives commands
- Controls based on HLT naming convention
  - Sub-farm (row and index)
  - UTGID (task name)



# Sub-Farm Control GUI



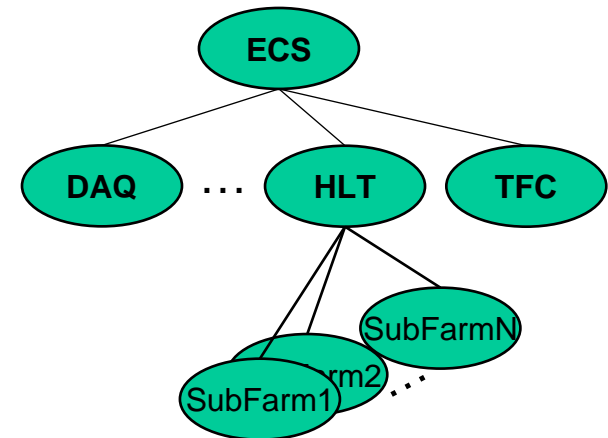
The screenshot displays the 'HLTA01: TOP' control interface. At the top, the system is identified as 'HLTA01' and is in a 'RUNNING' state. The date and time are 'Tue 05-May-2009 19:42:10'. The user is 'root'. A tree view on the left allows selecting nodes, with 'HLTA01' expanded to show sub-nodes like 'HLTA0101' through 'HLTA0111' and 'AdderA01'. A central table lists these sub-systems and their states, all of which are 'RUNNING'. On the right, a 'GauchoJob' window shows performance metrics for various components.

Sub-System	State
HLTA0101	RUNNING
HLTA0102	RUNNING
HLTA0103	RUNNING
HLTA0104	RUNNING
HLTA0105	RUNNING
HLTA0106	RUNNING
HLTA0107	RUNNING
HLTA0108	RUNNING
HLTA0109	RUNNING
HLTA0110	RUNNING
HLTA0111	RUNNING
AdderA01	RUNNING

Counter/Rate	Value
EventSelector/EventsReq	705820
Output/SpaceCalls	705740
Output/SpaceErrors	0
Output/BytesOut	-800323904
EventSelector/EventsIn	705740
Output/ErrorsOut	0
Runnable/EvtCount	705820
Output/EventsOut	1411480

# Sub-farm control

- Hierarchical ECS main control + dynamic allocation
  - Easy to include new farms into global system
- Single sub-farm controls as framework package + base on naming convention
  - Easy to duplicate farm control system.



# Partitioning



- Several sub-detectors (many teams)
- Different possible configurations (commissioning stage).
  - Possibility to modify dynamically the readout components included to control.
  - Different running modes (Physics, Calibration, Cosmics, etc).
    - Run\_type parameter sent with “Configure” command
- Pool of sub-farms → Dynamic allocation
  - Different instances of run control running in parallel (readout partitions)

# Farm pool controls

Vision\_1: fwFSM:FSMOperate...  
 FARM: TOP  
 User: root  
 Identity: HLT.Manager2  
 Tue: 05-May-2009 18:12:02  
 root

**System**      **State**  
 FARM      READY

**Sub-System**      **State**  
 SUBFARMS      READY  
 PARTITIONS      READY  
 farmAlloc      READY

Select Node:  
 CALIBFARM  
 CALD07  
 FARM  
 SUBFARMS  
 PARTITIONS  
 farmAlloc

Status of subfarms:

Farm	Owner	Allocated	Remove	Insert	Comment
HLTA01	MUON				
HLTA02	MUONC				
HLTA03	IT				
HLTA04	Removed from pool			<input type="checkbox"/>	Bad state for control s
HLTA06	none		<input type="checkbox"/>		
HLTA07	none		<input type="checkbox"/>		
HLTA08	none		<input type="checkbox"/>		
HLTA09	HCAL				
HLTA10	none		<input type="checkbox"/>		
HLTA11	none		<input type="checkbox"/>		
HLTB01	none		<input type="checkbox"/>		
HLTB02	none		<input type="checkbox"/>		
HLTB03	none		<input type="checkbox"/>		
HLTB04	none		<input type="checkbox"/>		
HLTB06	none		<input type="checkbox"/>		
HLTB07	none		<input type="checkbox"/>		

colour	host up	dns up	tsrsv up
red	down	down	down
orange	up	down	down
yellow	up	up	down
green	up	up	up

Number of free subfarms: 36

View All Owners      Close

REFRESH/DO IT      Comment:

Messages

Close

# Conclusions

---



- HLT control completely defined and integrated into global LHCb control system.
- Implemented with same toolkit (PVSS & FSM) used through the ECS.
  - Keeps Homogeneity
- Fully configurable at real time.
- Automatic control of processes for the shift operator.