CALICE Data Processing
(From Raw to Physics Data)

Roman Pöschl
LAL Orsay

- Calice Testbeam Data Taking
- Data Management
- Event Building and Reconstruction Software
- Pros and Cons ...
- A look into the (near) future
- Summary and Outlook

ILC Workshop Valencia/Spain November 2006

CALICE Testbeam Data Taking

CALICE collaboration is preparing/performing large scale testbeam
Data taking between 1.August and 31.October 2006

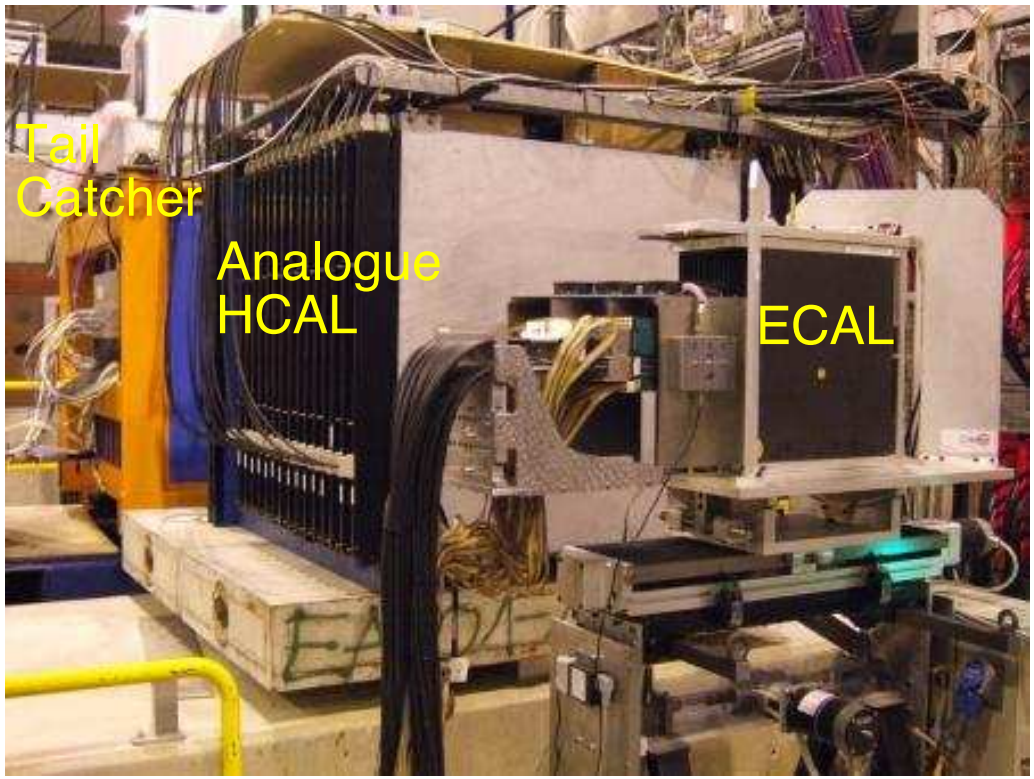
Testbeam program poses software/computing " challenges"

- Data processing from Raw Data to final Clusters in a coherent way

- Handling of Conditions Data Detector Configuration Calibration, Alignment etc.

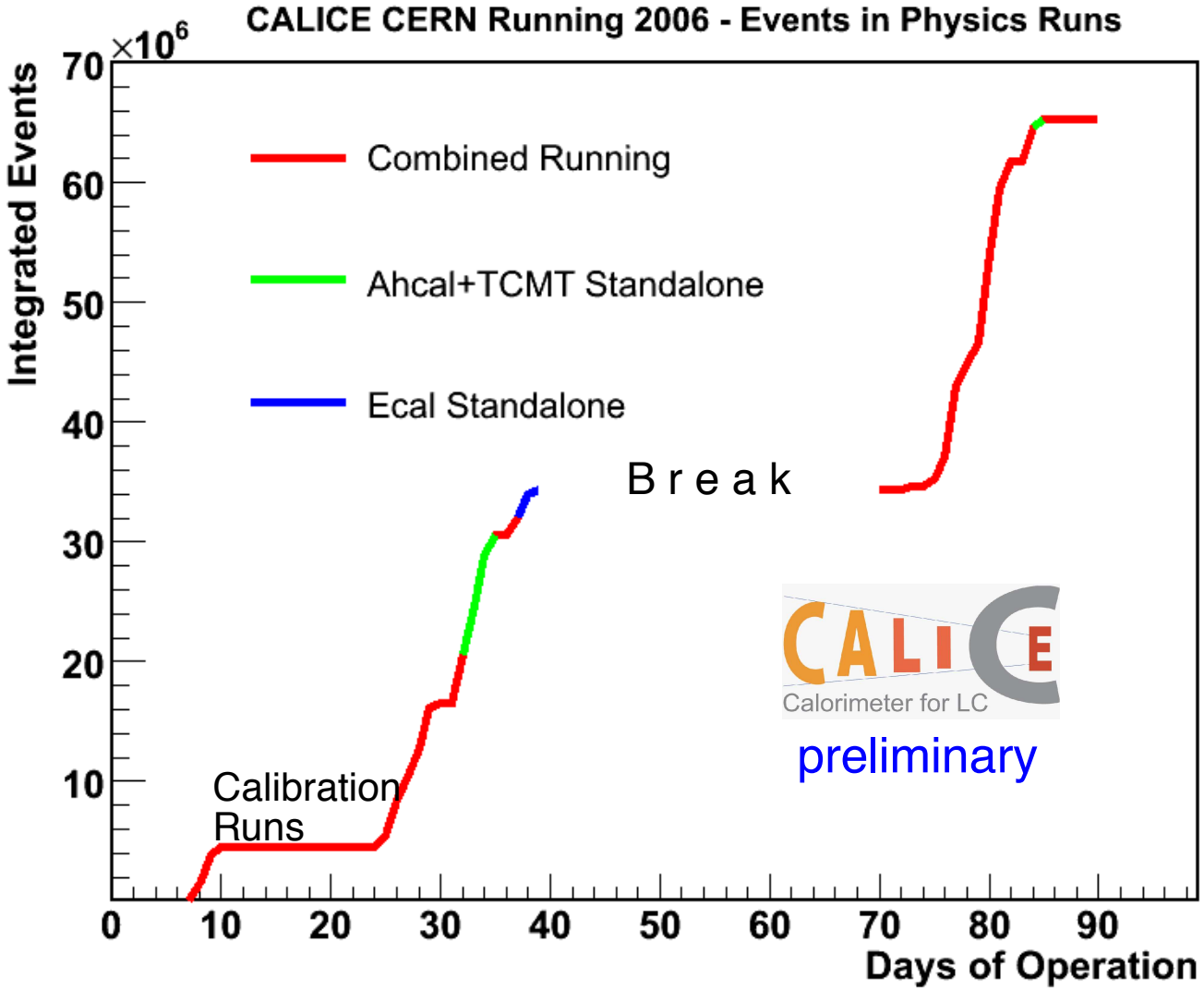
-Comparison with simulated data
'Physics' Output

Testbeam Setup at CERN



O(15000) calorimeter cells readout by Calice DAQ
No Zero Suppression

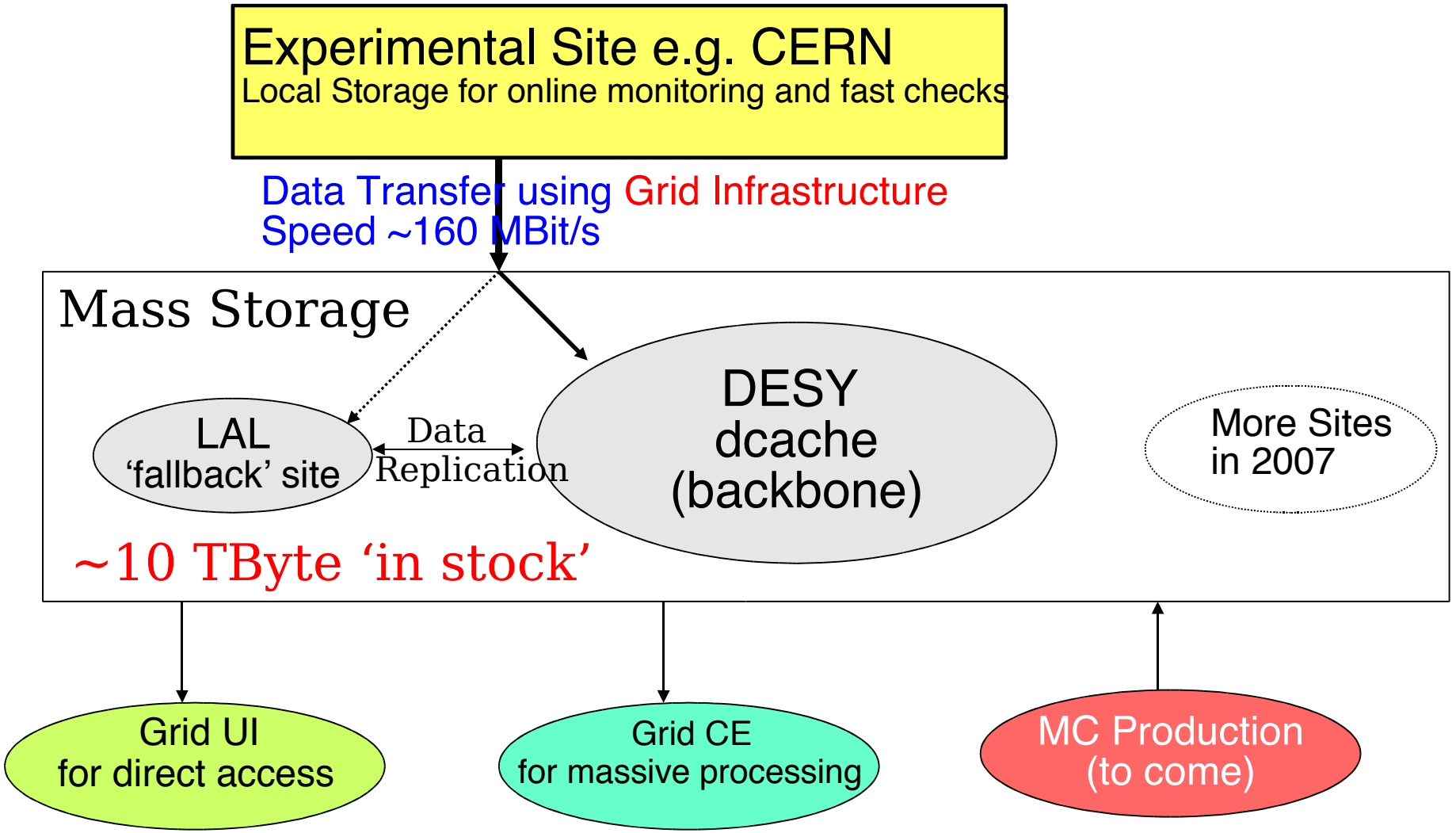
CALICE - CERN Data taking 2006



~65 Mio Events
in 'Physics' Runs
+
O(35 Mio). Muon
Calibration Events)

Efficient and fast
way of data distribution
and processing ?

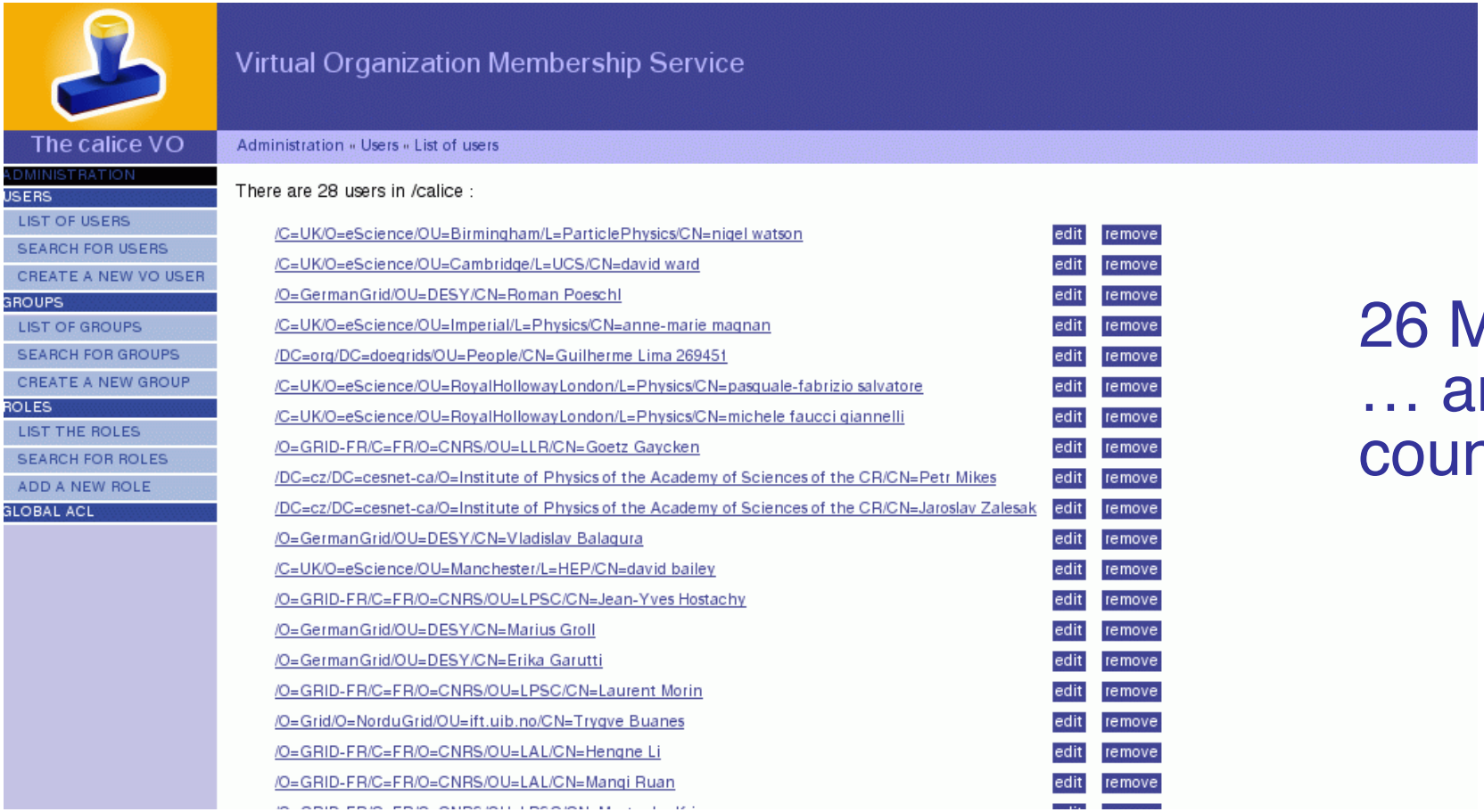
Data Handling and Processing



Data available to whole collaboration ~20 Min. - 1h after run end
Data access independant of experimental site
Grid is the only 'environment' where all data are available

The Virtual Organisation - vo calice

Hosted by DESY:
Page for registration is https://grid-voms.desy.de:8443/voms/calice



26 Members
… and
counting

VO Manager: R.P./LAL, Deputy: A. Gellrich/DESY

The Grid in/for Calice

Large Data Volume => Significant Computing Ressources required
Decentralized Organization <=> Decentralized Computing

Virtual Organization calice

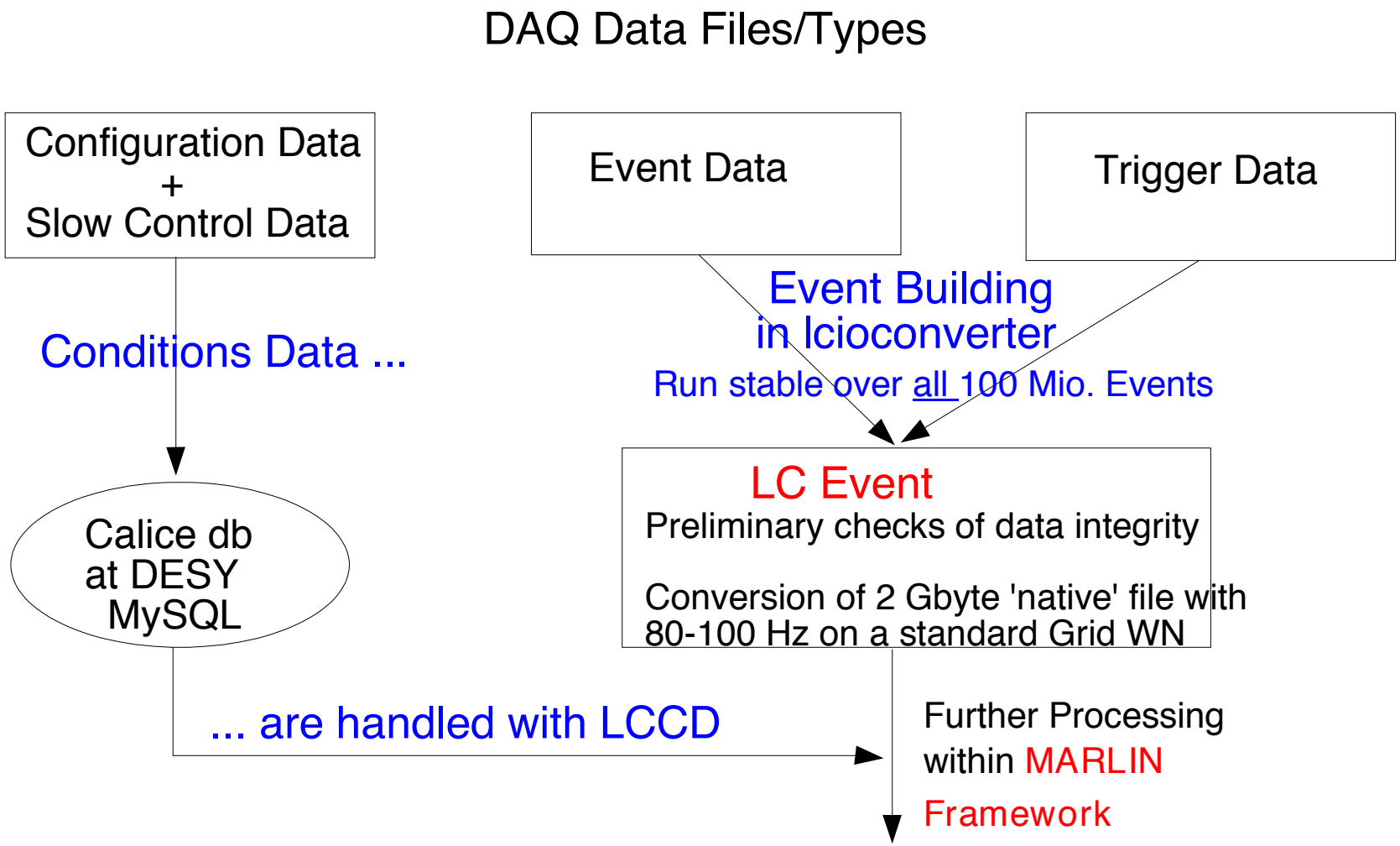
Supported by:	DESY Hamburg	Hosting, Computing and Storage
	LAL	Computing and Storage
	LLR	Computing (Storage to come)
	DESY Zeuthen	Computing and Storage
	Imperial College	Computing and Storage
	cc in2p3 Lyon	Computing (Storage to come)
	Institute of Physics Prague	Computing and Storage (in preparation)
	University College	In preparation
	CIEMAT Madrid	Offer Received
	Univ. Regina	Offer Received

The last four responded quickly to my request from last week, thanks

Acknowledged EGEE project: https://cic.in2p3.fr

Conversion to LCIO

DAQ data types are converted/wrapped into LCIO on the basis of LCGenericObjects



Remark: LCIO and ILC software framework is not needed to analyze calice data but using it delivers important input for future ILC s/w development
-> General ILC Concept for low level data handling

Intermezzo – Conditions Data Handling

- LCCD – Linear Collider Conditions Data Framework:
  - Software package providing an Interface to conditions data
    - database
    - LCIO files

    Author Frank Gaede, DESY

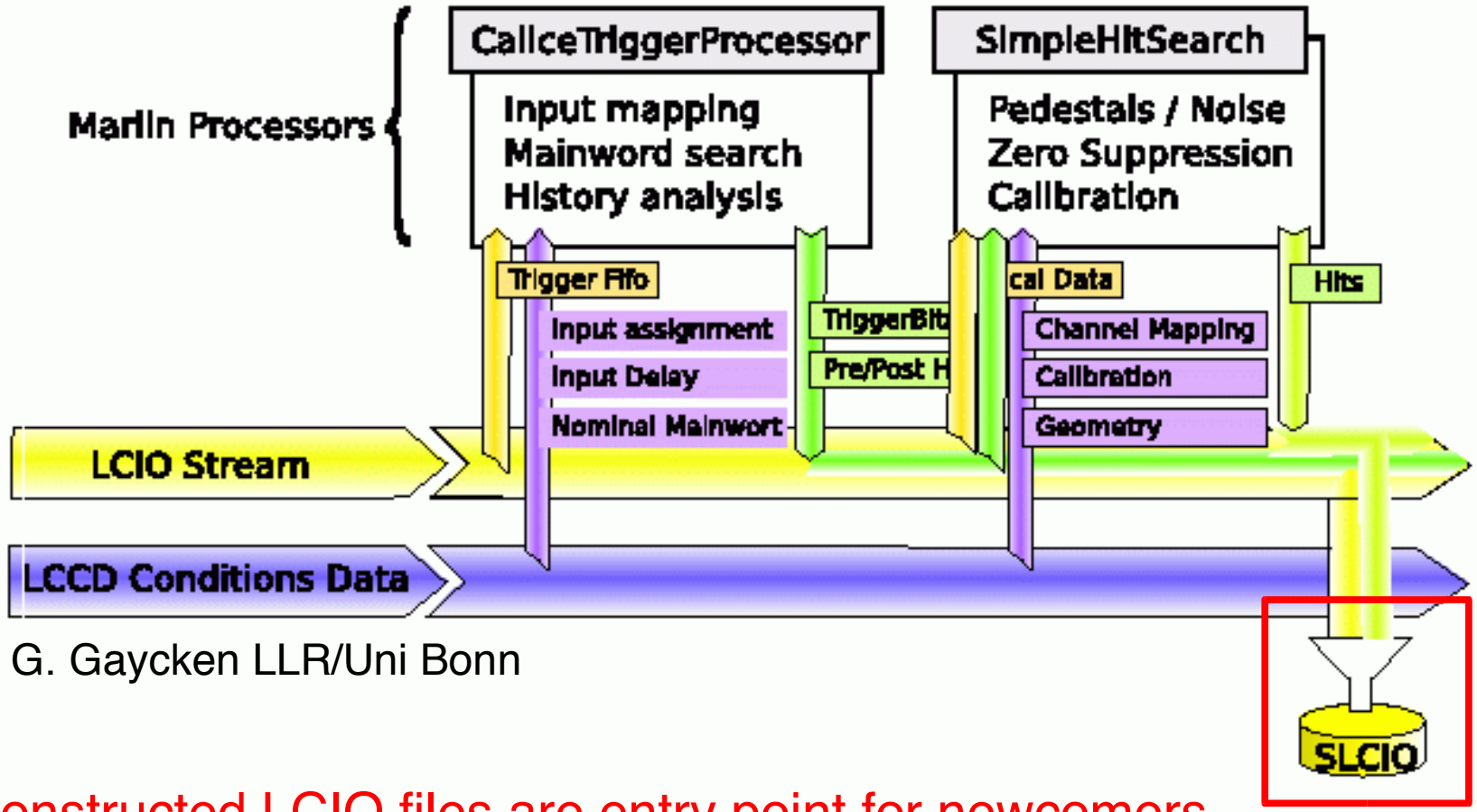
LCCD works and is heavily used within calice !!!

The importance of conditions data (not only) for 'real' data renders the development of a fully functional cd data toolkit to be a fundamental !!! piece of the ILC Software

- Efficient storage and access to conditions data
  Browsing, convenient interfaces
- How to 'distribute' conditions data (e.g w.r.t to grid) ?
  BTW.: LHC does have some headache with that!

Data Processing and Reconstruction

... as it exists for Calice Ecal (LLR) ...



G. Gaycken LLR/Uni Bonn

Reconstructed LCIO files are entry point for newcomers

... and starting point of high level analysis

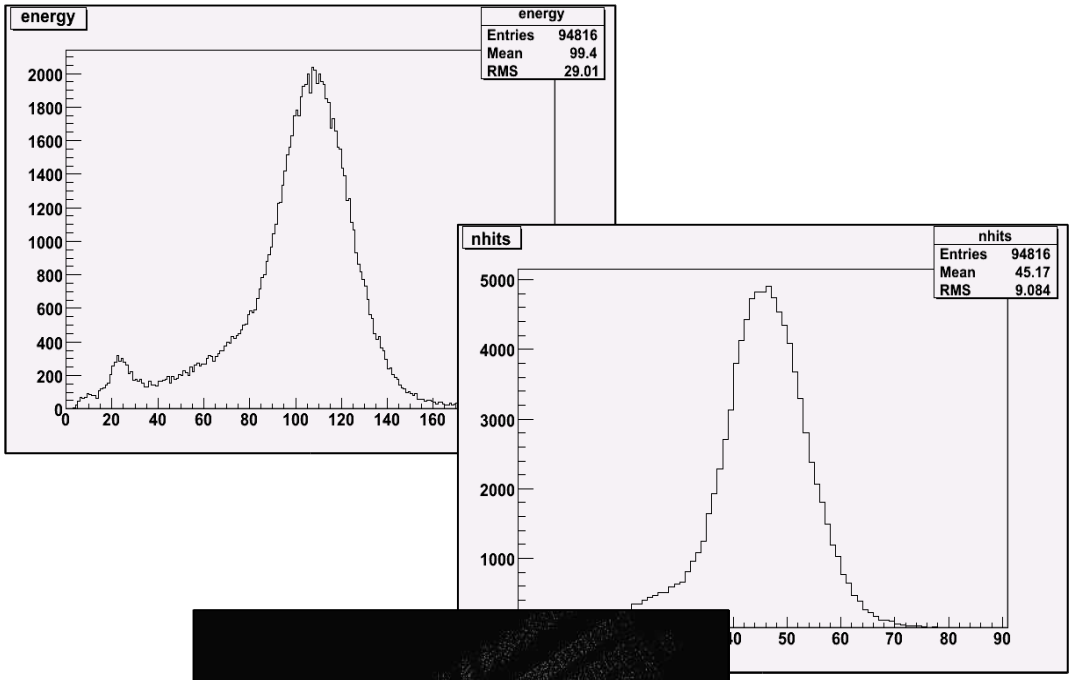
Contain 'familiar' LCCalorimeterHits

Though not the whole story – Still have to understand fundamentals

see talks in Calo Session

AHCAL Reconstruction Framework – S.Schmidt/DESY

- MappingI
  - ADCBlocks → CaliceHits1
- PedestalCalibration
  - CaliceHits1 → CaliceHits2
- GainCalibration
  - CaliceHits2 → CaliceHits3
- InterCalibration
  - CaliceHits3 → CaliceHits4
- SaturationCorrection
  - CaliceHits2, CaliceHits4 → CaliceHits5
- MIPCalibration
  - CaliceHits5 → CaliceHits6
- MappingII
  - CaliceHits6 → CalorimeterHits



Calibration steps modularized in MARLIN processors

(more details → Calo Session)

Pros and Cons using ILC Software for Calice Testbeam Data

Pros

Benefit from existing
tools/features for/of ILC Software
e.g. LCEvent allows to
gather information on event

Newcomers can work
in one software framework
for testbeam and physics studies

Define at an early stage
of the ILC R&D the needs
for a complete data processing
Coherent s/w concept at time
of ILC Detector TDR
Not just guesswork!!!

It's in the spirit of the LDC CDR!!!!

BTW: The converted LCIO files can be analyzed
on any OS (endianess) and on future
64bit architectures!!!

Cons

Need to wait for
converted files
No quick turnaround in
particular during development
of DAQ and tests
Needs tight communication between
DAQ and s/w developpers

Some overhead generated
by usage of ILC Software

Soruce of (potential)
errors unclear

(Basic) Issues for a Common Data Processing Chain

'High' level data are well addressed CaloHits, Clusters, Tracks etc.

Need to develop a software concept for treatment of low level data

- Avoid several (incompatible) chains on analysis
  One source of information

- 'Model for 'close to detector' data
  Hardware information needed to judge on data quality
  Extensions to LCIO ?!

- Software management tools flexible enough to react to changes
  in DAQ system

- DAQ has to be developped true to an agreed scheme
  and embedded into overall software development

Initiative started at EUDET Annual Meeting for Calo DAQ
Next Generation R&D programs should develop concept
for entire ILC data processing!!!

A view to the Monte Carlo Branch

- Model for the simulation of the CERN test beam is available (in release 06-02 of Mokka)

Will use grid for MC production
Estimation ~ 10000 kSI2kd for simulation of CERN data

Simulation will be followed by a digitisation step
Realized as Marlin Processors within Digisim Package
A.M Magnan, G. Lima

Summary and Outlook

- Calice uses ILC Software for processing of Testbeam Data

  ILC Datataking in a (big) nutshell
  Very important input for current and future developments of ILC Software
  Allows for a revision of the ILC Software concepts on a 'living' beast

- Calice uses systematically Grid tools

  First (and only?) R&D project within ILC effort
  24h/24h 7h/7h during CERN testbeam
  So far mostly for data management
  Might be also used regularly used for conversion in future
  Example: Conversion of ~100 CERN runs within ~10 hours
  CPU consumption still tiny but will grow fast when starting
  e.g. MC production

A big thank you to all experts of IT divions, particularly DESY, CERN and LAL, who support our effort!!!!

- Experience with testbeam data clearly reveals the needs for a coherent concept to handle 'low level' data within ILC Software
  - (Latest) Next generation R&D projects should be used to develop a complete data processing/handling strategy for the ILC.
    Avoid 'island' solutions and work on an integrated effort

CALICE does not only hardware-prototyping but also 'computing prototyping'