

CALICE Computing

(Status and Practical Tips)



Roman Pöschl
LAL Orsay



- Introductory remarks
- Infrastructure and data availability - Grid
- Software Versions
- Conditions Data Handling
- Summary and Outlook

CALICE Days - DESY Hamburg February 2007

Introductory Remarks

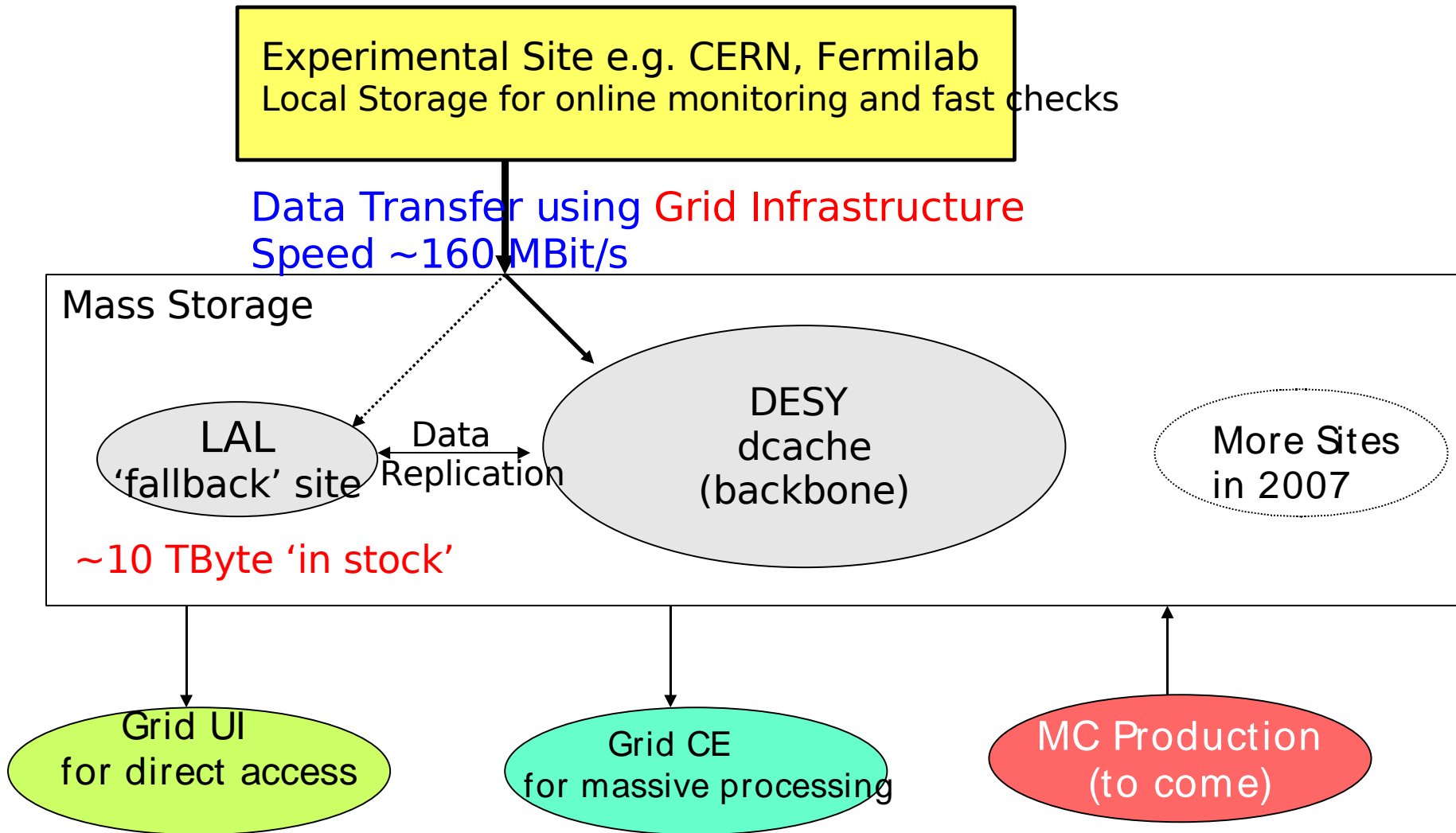
CALICE testbeam effort are an excellent environment to test 'software concepts' available for the ILC

- Confronting real data with 'Mokka based' Monte Carlo
CALICE has already lead to considerable improvements in Mokka and G4 and/or point clearly at waek points
- Testing the usability of LCIO for real data and pointing cleary to open issues of the currently available general software
 - > Input for concepts for treatment of 'low level data'
- Large data volumes and world wide user community require distributed computing
 - > Learn how to use grid tools

Employing the tools means encountering short comings, problems and sometimes slow progress (as in hardware development)

Calice Days DESY Hamburg Feb. 2007

Data Handling and Processing

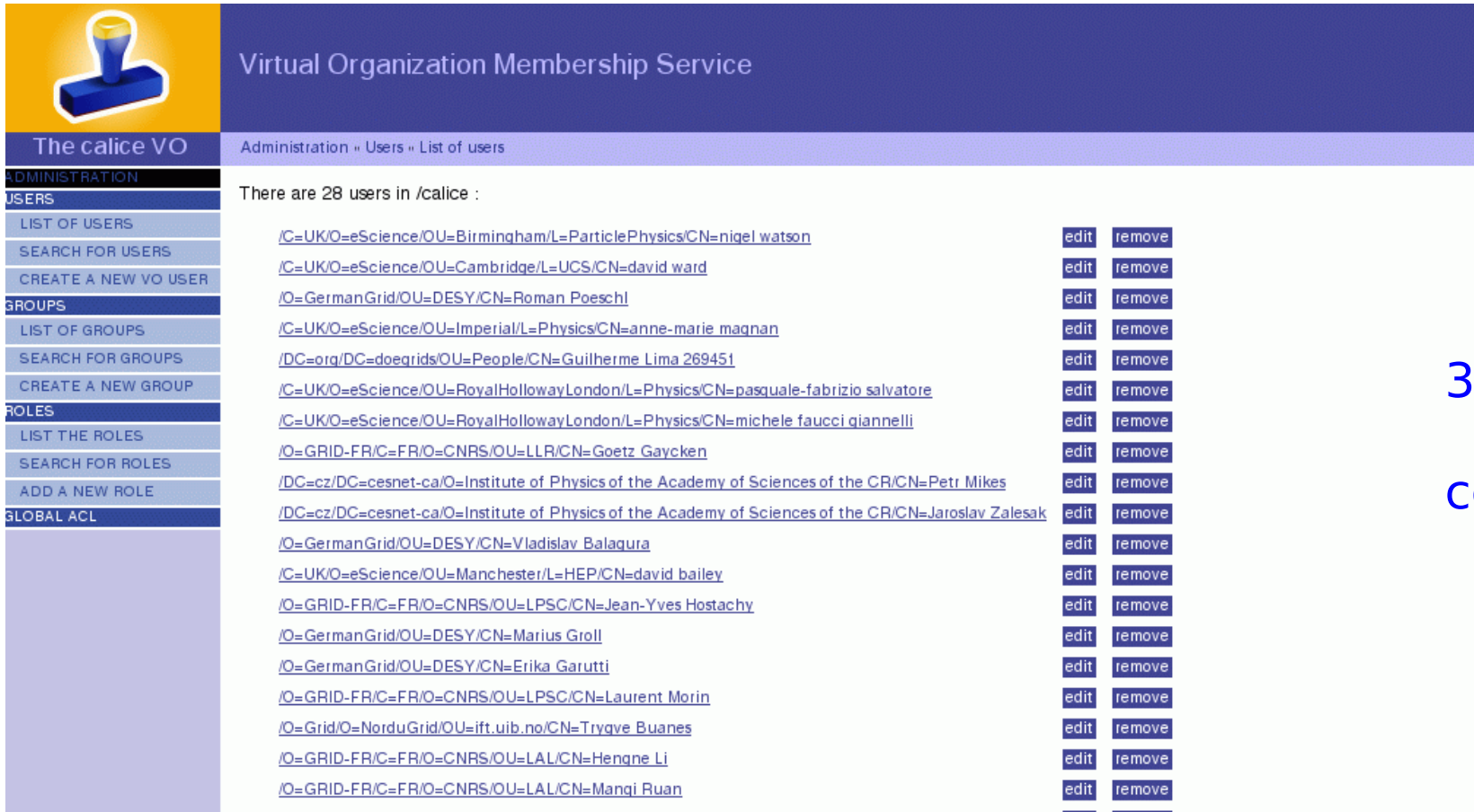


Data access independent of experimental site
Grid is the only 'environment' where all data are available

The Virtual Organisation - vo calice

Hosted by DESY:

Page for registration is <https://grid-voms.desy.de:8443/voms/calice>



Virtual Organization Membership Service

The calice VO Administration » Users » List of users

There are 28 users in /calice :

/C=UK/O=eScience/OU=Birmingham/L=ParticlePhysics/CN=nigel watson	edit	remove
/C=UK/O=eScience/OU=Cambridge/L=UCS/CN=david ward	edit	remove
/O=GermanGrid/OU=DESY/CN=Roman Poeschl	edit	remove
/C=UK/O=eScience/OU=Imperial/L=Physics/CN=anne-marie maqnan	edit	remove
/DC=org/DC=doegrids/OU=People/CN=Guilherme Lima 269451	edit	remove
/C=UK/O=eScience/OU=RoyalHollowayLondon/L=Physics/CN=pasquale-fabrizio salvatore	edit	remove
/C=UK/O=eScience/OU=RoyalHollowayLondon/L=Physics/CN=michele faucci qiannelli	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LLR/CN=Goetz Gaycken	edit	remove
/DC=cz/DC=cesnet-ca/O=Institute of Physics of the Academy of Sciences of the CR/CN=Petr Mikes	edit	remove
/DC=cz/DC=cesnet-ca/O=Institute of Physics of the Academy of Sciences of the CR/CN=Jaroslav Zalesak	edit	remove
/O=GermanGrid/OU=DESY/CN=Vladislav Balagura	edit	remove
/C=UK/O=eScience/OU=Manchester/L=HEP/CN=david bailey	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LPSC/CN=Jean-Yves Hostachy	edit	remove
/O=GermanGrid/OU=DESY/CN=Marius Groll	edit	remove
/O=GermanGrid/OU=DESY/CN=Erika Garutti	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LPSC/CN=Laurent Morin	edit	remove
/O=Grid/O=NorduGrid/OU=ift.uib.no/CN=Trygve Buanes	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LAL/CN=Hengne Li	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LAL/CN=Manqi Ruan	edit	remove

34 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 and Storage
	DESY Zeuthen	Computing and Storage
	Imperial College	Computing and Storage
	cc in2p3 Lyon	Computing and Storage
	Cambridge	Computing and Storage
	Institute of Physics	Computing and Storage
	Prague	(in preparation)
	University College	Computing and Storage
	KEK	Computing and Storage (In preparation)
	Manchester	Computing and Storage (in preparation)
	CIEMAT Madrid	Computing and Storage
	Fermilab	Offer Received
	Univ. Regina	Offer Received

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

Infrastructure and data availability – What data are available and where

DESY data taking

Binary files DESY/CERN runs (2005/2006)
Most of HCAL testruns at DESY

lfn: /grid/calice/tb-[desy](#)/native/dat
lfn: /grid/calice/tb-[cern](#)/native/dat

Converted Icio files of DESY/CERN runs 2005/2006
Most of Hcal testruns

lfn: /grid/calice/tb-[desy](#)/raw/conv_vxxxx
lfn: /grid/calice/tb-[cern](#)/raw/conv_vxxxx
currently xxxx=0402

Reconstructed files of runs 2006 (Currently Ecal only)

lfn: /grid/calice/tb-[desy](#)/rec/rec_vxxxx
lfn: /grid/calice/tb-[cern](#)/rec/rec_vxxxx
currently xxxx=0402
or xxxx-v0403-pre2

Accessing/Handling the data using grid tools?

Binary data and LCIO files converted/reconstructed are registered on the Grid

Using the LCG software together with LFC file catalogue
Organized in a unix-like directory structure

e.g. `lfc-ls /grid/calice/tb-cern/native/dat`

Access to the data:

Naive copy (should always work)

e.g. `lcg-cp -v -vo calice lfn:/grid/calice/tb-desy/native/dat/ \`
`RunXXXXXX.nnn.bin file:<myfile>`

-v means verbose and is helpful to trace problems

Listing replicas of a given file (e.g.)

`lcg-lr -vo calice lfn:/grid/calice/tb-xxxx/raw/conv_v0402`

- In case of a failure try to access other replicas (while trying to report the failure)

- Try to fetch file from closest Storage Element (SE)

`lcg-cp -v -vo calice -d <your_preferred_se>`

- List storage elements with: `lcg-infosites -vo calice se`

Further Tips

- User output should be stored on physically closest SE
quick individual access but yet visibility of your exploits to collaboration
- Before submitting a large bunch of jobs
Test your scripts on your local grid-ui
During test: Run test jobs in virgin environment
- Grid jobs are still sent into a black hole
Try to put some intelligence into scripts
-> See my example scripts

Obeying these rules gives you access to virtually unlimited computing resources

Grid for CALICE – Next steps

- **Qualification of sites which have recently joined the club**
In particular integration of North American, i.e. Fermilab and Asian sites (KEK)
Compatibility of grid middleware !?
- **Full exploitation of voms features**
multiple vo memberships – no multiple certificates needed
locking of valuable files versus the outside world
(effectively done now by handmade restriction of access to DESY dcache)
- **Transition from SL3 to SL4**

SL4 will become default Grid Computing platform during 2007
- **Increase of active user community**
- **Test and qualification of more sophisticated Job submission tools**
e.g. ganga

Some useful Environment Variables

A loose compilation based on recent user 'problems'
Please check whether/how set on case of problems

BDII Host: export LCG_GFAL_INFOSYS=grid-bdii.desy.de:2170

LFC File catalogue: export LFC_HOST=grid-lfc.desy.de

Ressource broker: grid-rb2.desy.de

Check/Modify settings in

\$EDG_LOCATION/etc/edg_wl_ui_cmd_var.conf

\$EDG_LOCATION/etc/_VO_/edg_wl_ui.conf

voms server (Host certificate): \$X509_VOMS_DIR/grid-voms.desy.de

voms server (specific to vo): \$GLITE_LOCATION/etc/vomses/calice-grid-voms.desy.de

Note, that on sites which fully support calice, these variables and settings should be correctly setup by your IT division.

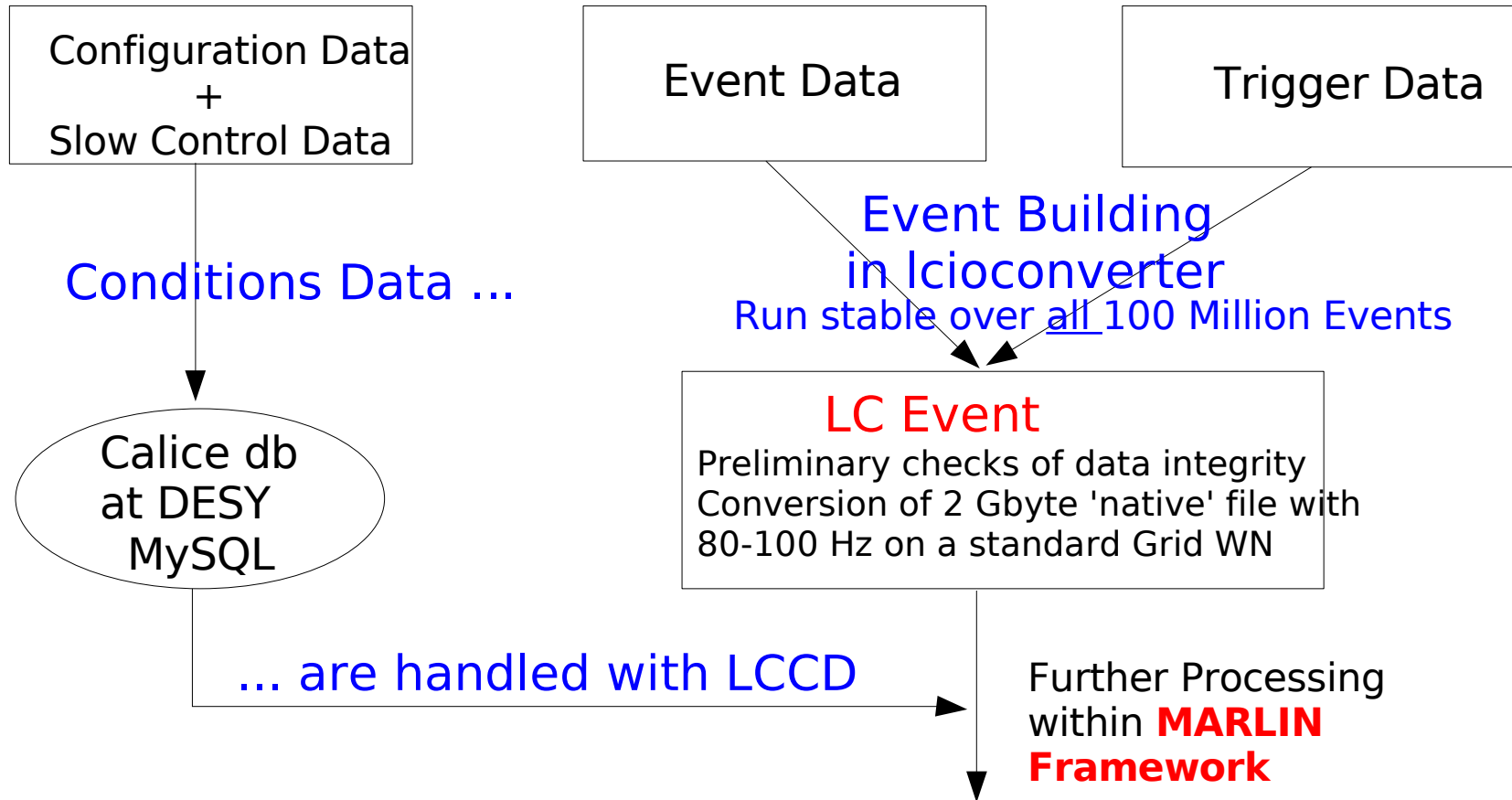
These hints are more useful for users which run a (more or less) private UI or one of the resources has a failure

Always report problems to the calice-vo users mailing list!!!
calice-vo-users@desy.de

Conversion to LCIO

DAQ data types are converted/wrapped into LCIO on the basis of [LCGenericObjects](#)

DAQ Data Files/Types



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

CALICE Software Packages for Data Processing

- LCIO Conversion

All data of 2006 (DESY/CERN) have been converted using the version v04-02-xx of the converter

- Reconstruction

Many runs have been reconstructed for the Ecal using Version v04-02-01 of the reco package
see `lfn:/grid/calice/tb-xxxx/reco/rec_v0402`

'Unexperienced users' are encouraged to use these as an entry point to the data analysis

Hcal Reconstruction v00-01-17 (S. Schmidt)

Latest Reco Version: v04-03-pre2
comprises Hcal Reco - Test files available

- userlib (Common to all packages)

Current version v04-05-01

userlib comes with example processor in examples subdirectory

CALICE Software Packages – Outlook on Coming Major Release

Current release

reco v04-03-pre2
userlib v04-05-01

is pre-release !!!

Issues to be clarified

- Correct implementation of new coordinate frame
- Correct implementation of Hcal Reconstruction
- Compatibility of results with previous release

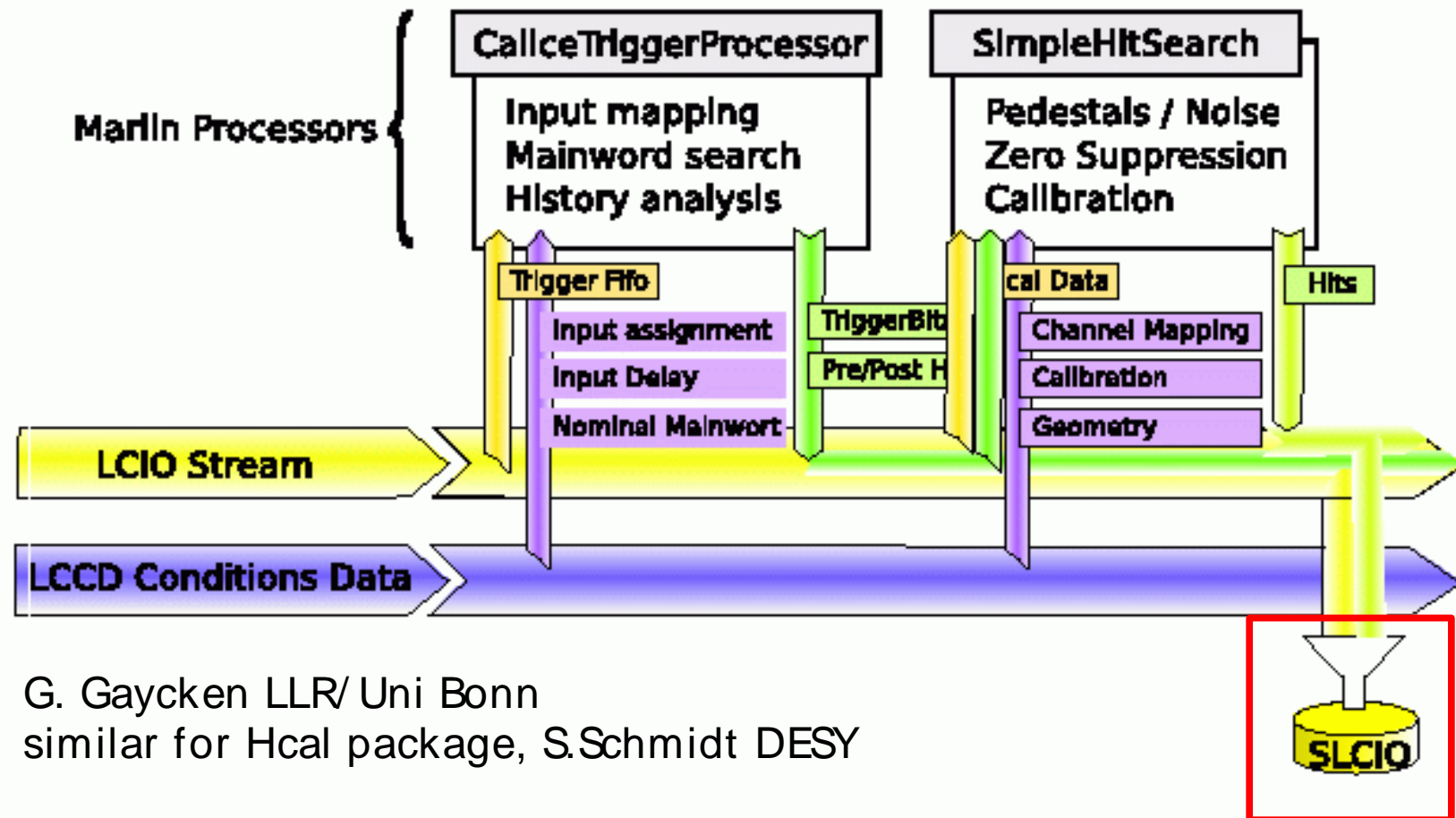
Major update in 'real' release

- Driftchamber Reconstruction DESY/CERN (Michele)

To Come:

Integration of Tail Catcher

Data Processing and Reconstruction



G. Gaycken LLR/ Uni Bonn
similar for Hcal package, S.Schmidt DESY

Reconstructed LCIO files are entry point for newcomers
... and starting point of high level analysis
Contain 'familiar' CalorimeterHits
Though not the whole story – Still have to understand fundamentals
See e.g. CaliceExampleProcessor on how to access lower level data

Conditions Data Handling – Some Reminders

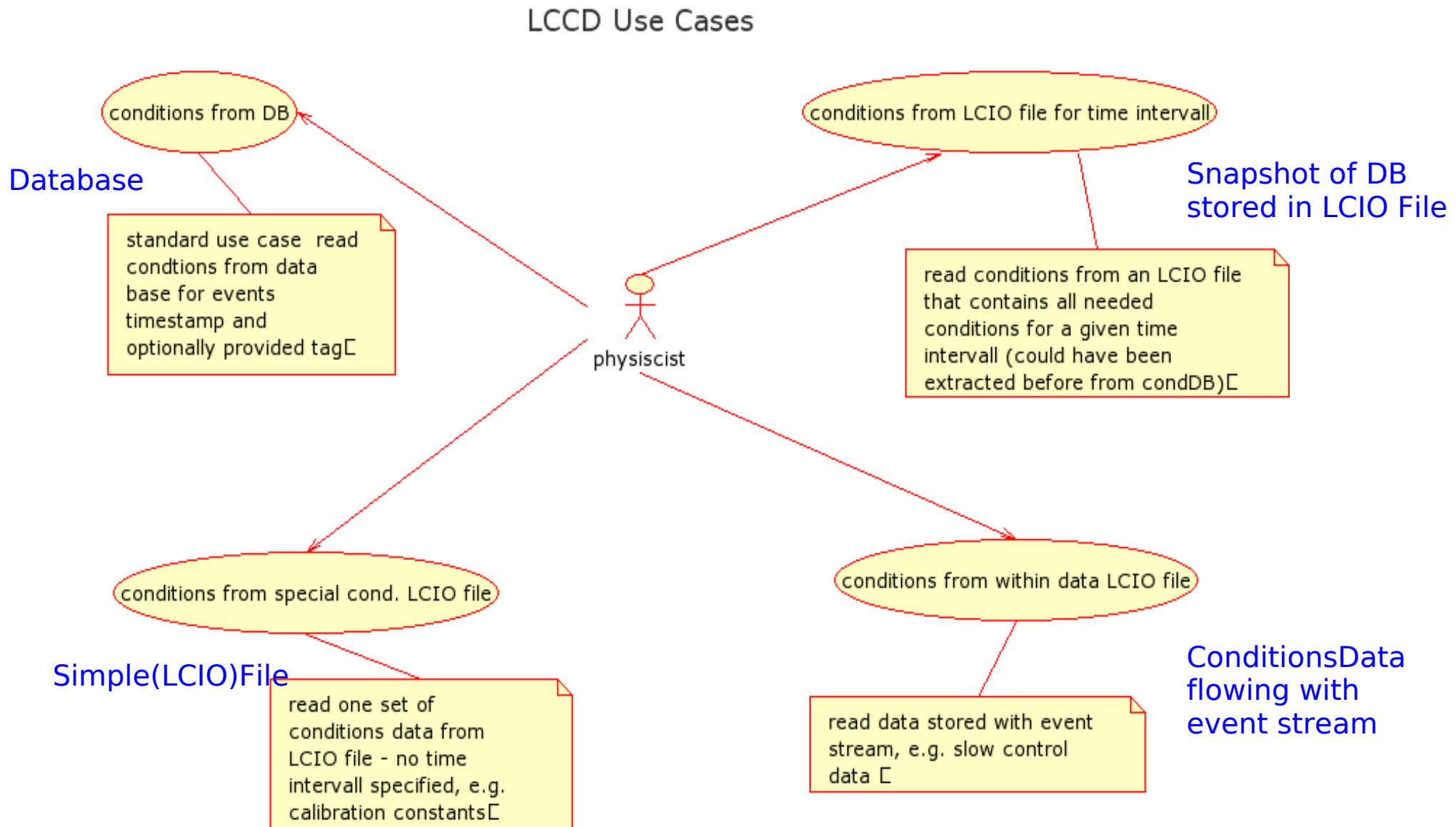
- LCCD – Linear Collider Conditions Data Framework:
 - Software package providing an Interface to conditions data
 - database
 - LCIO files

Author Frank Gaede, DESY
- Conditions Data:
 - all data that is needed for analysis/reconstruction besides the actual event data
 - typically has lifetime (validity range) longer than one event
 - can change on various timescales, e.g. seconds to years
 - need for tagging mechanism, e.g. for calibration constants

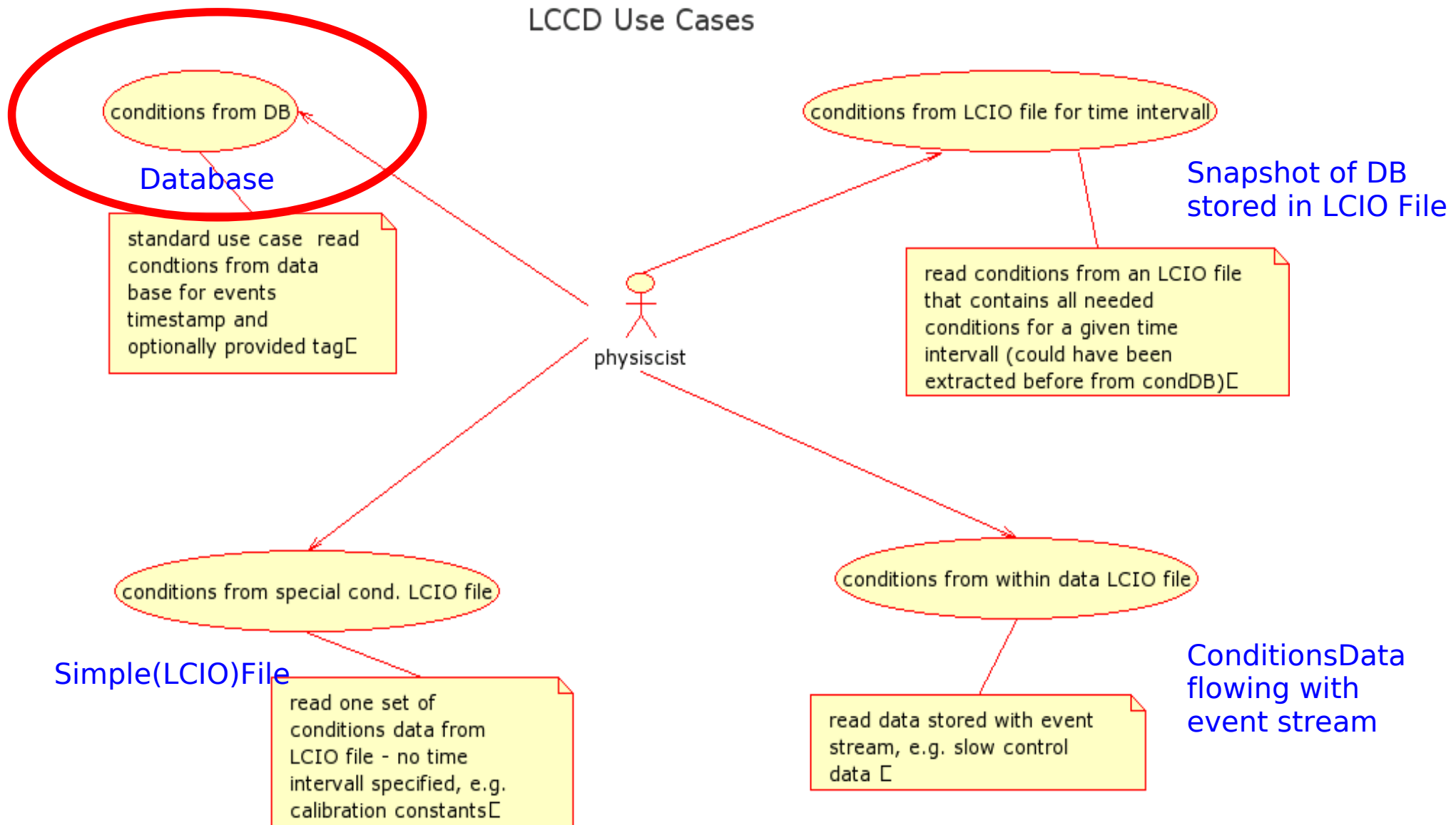
Material 'stolen' from Frank Gaede

Calice Days DESY Hamburg Feb. 2007

Sources of Conditions Data – Use Cases



Sources of Conditions Data – Use Cases

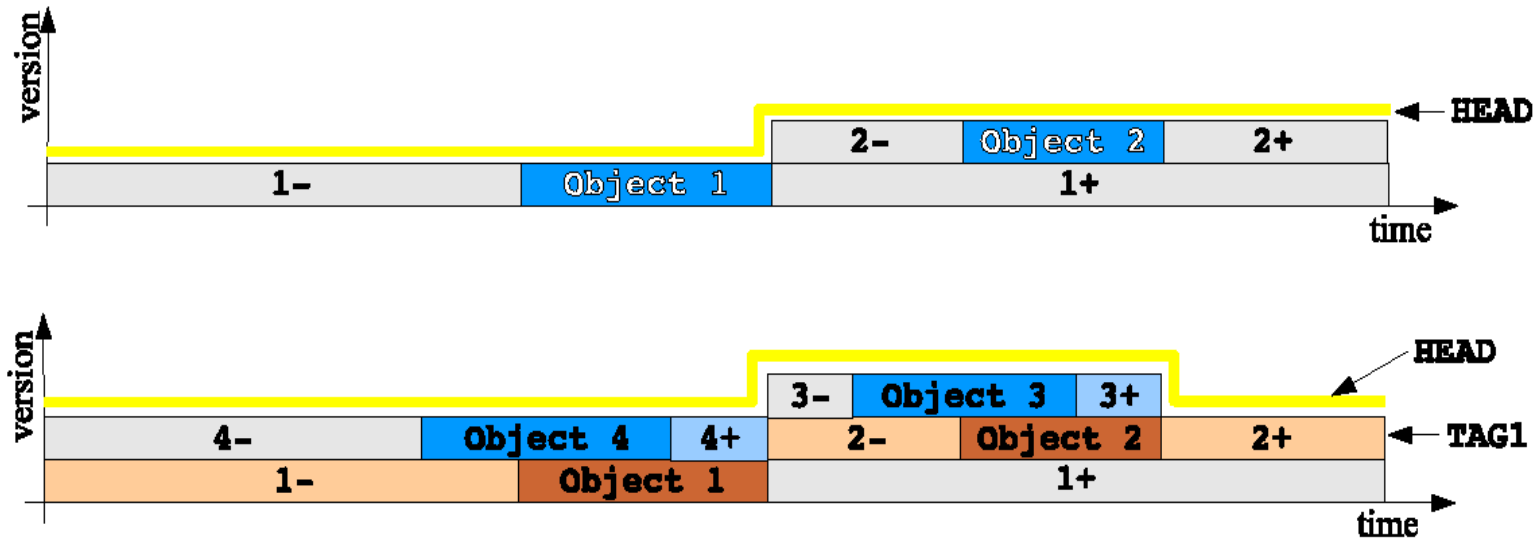


ConditionsDBMySQL – Overview

Digged out and explored out by Frank Gaede for us
Interfaced to LCCD by Frank

- Open source implementation of CondDB API
 - Conditions data interface for ATLAS (Cern IT)
- developed by Lisbon Atlas group
- features
 - C++ interface to conditions database in MySQL
 - data organized in folder/foldersets
 - objects stored as BLOBs (binary large objects)
e.g. LCIO objects or std::vector
 - tagging mechanism similar to CVS
 - scalability through partitioning options
 - outperforms implementation based on Oracle

ConditionsDBMySQL – Versioning of Conditions Data



Browse objects in HEAD



Browse objects in tag TAG1



Figure 3: tagging and browsing example in the ConditionsDB mySQL's implementation.

CVS-like management system

'Horizontal' and vertical browsing in time possible
Time Stamp (by LCCD) in units of nanoseconds

Accessing ConditionsData Using LCCD – Users Point of View

Update of Conditions data handled within MARLIN
(Note: LCCD does not depend on MARLIN and vice versa)

- Source of ConditionsData defined in MARLIN steering File
e.g. ConditionsData for Cell Mapping from DB

```
DBCondHandler CALDAQ_TriggerAssignment  
/cd_calice_cernbeam/CALDAQ_TriggerAssignment HEAD
```

- Handling of Conditions Data (updating etc.) within a
ConditionsProcessor (provided by **MARLIN**)

- Steps to access ConditionsData

- a) Register Pointer to a CellMap and its name

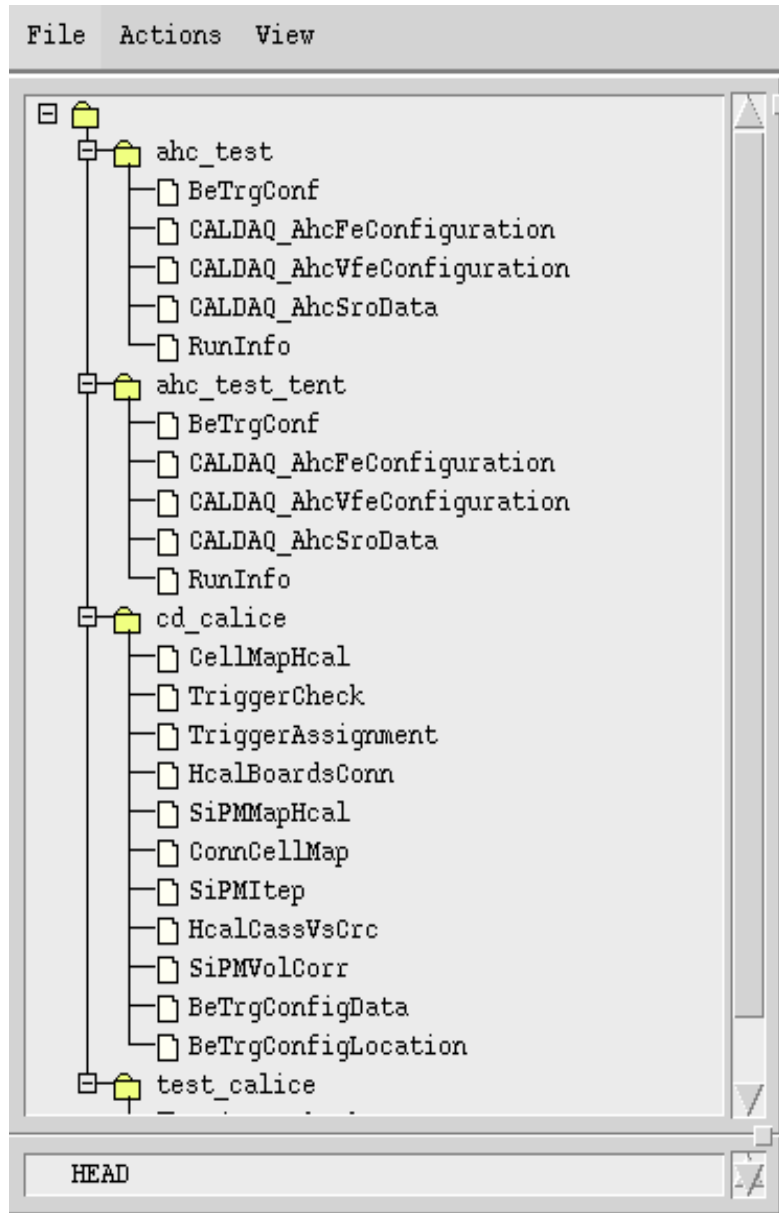
```
marlin::ConditionsDataProcessor->registerChangeListener \  
( &_yourListenerMethod , "CALDAQ_TriggerAssignment" );
```

- b) Obtain CellMap within event

```
_yourListenerMethod( col ){do sthg.;} 
```

- c) See example Processor and/or my recent posting
to the calice sw mailing list for examples

CALICE Database Hosted by DESY



Trigger Info: Assignment of triggerbits
Trigger Configuration
Info to validate Trigger
information

Calibration Data

Cell Mappings: Relation electronic channel
and
geometrical channel
i.e. Cabling of devices

Hardware configuration during data taking.

Database server: flccaldb02.desy.de
flccaldb01.desy.de: access by experts only !!!

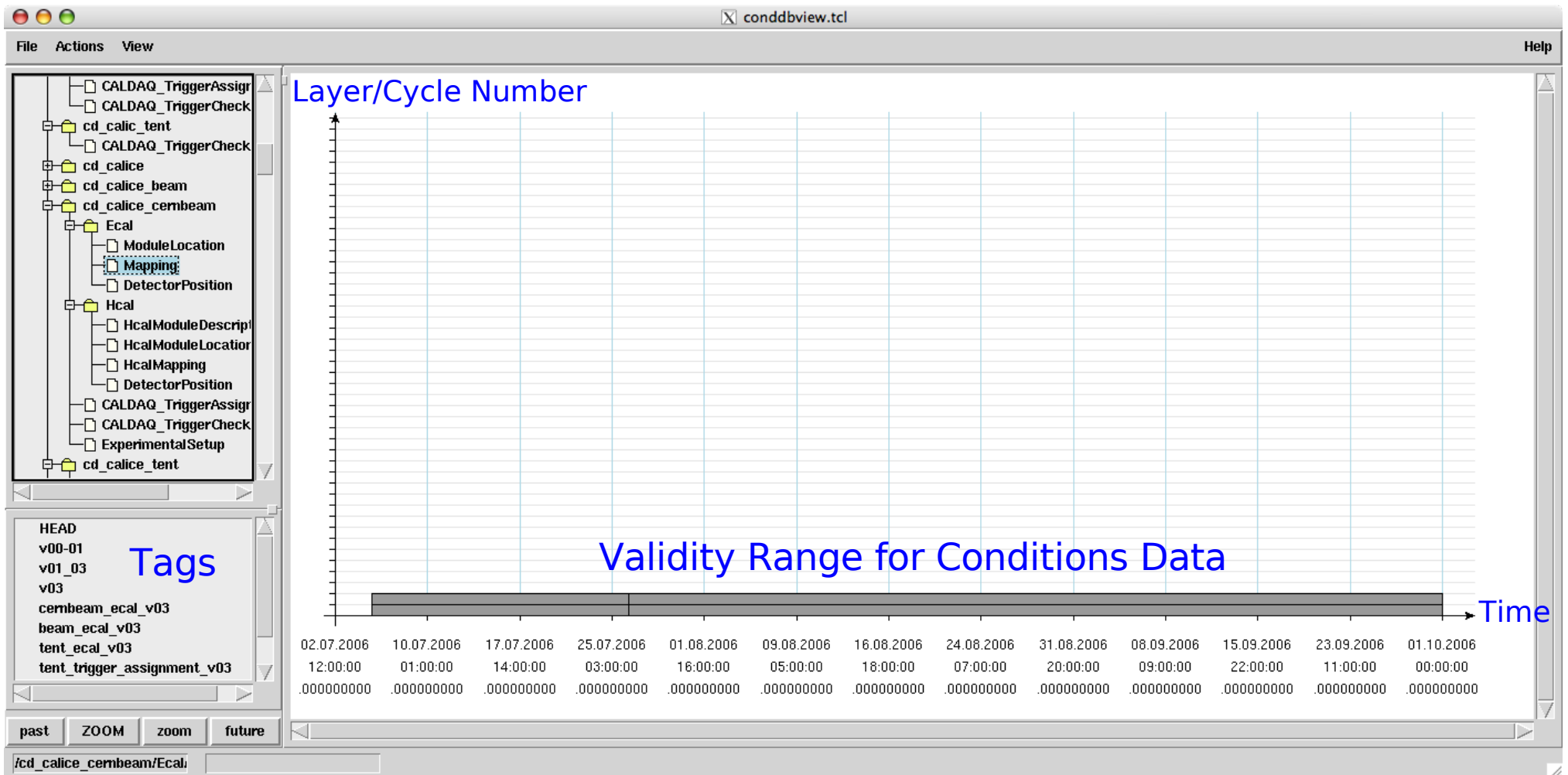
Behind DESY firewall

Accessible from 'calice' institutes

First attempt to visualize
Conditions Data
(S.Schmidt, M.Schenk, R.P.)

Calice Days DESY Hamburg Feb. 2007

Conditions Data in CALICE Database



Conditions Data Handling – General Issues

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
LCCD is first attempt into that direction

Issues to be addressed:

- Type safety (Data are stored as LCGenericObjects)
- 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!

Testbeams are ideal environment to develop a working
Conditions Data Handling before ILC starts

Conditions Data Handling – Practical Hints for Users

- Use lccd version v00-03-04

Convenient print outs if conditions data are missing
Improved management on database connections

- CondDBMySQL

calice runs own version of this package, available via Zeuthen cvs server
CondDBMySQL_ILC-0-5-10

Example for installation (after unpacking of tarball):

```
./configure --with-mysql-inc=/usr/include/mysql --with-mysql-lib=/usr/lib/mysql --with-conddbprofile=localhost:condb_1:condb:condb --prefix=/data/poeschl/extern/CondDBMySQL/ilc-0.5.10
```

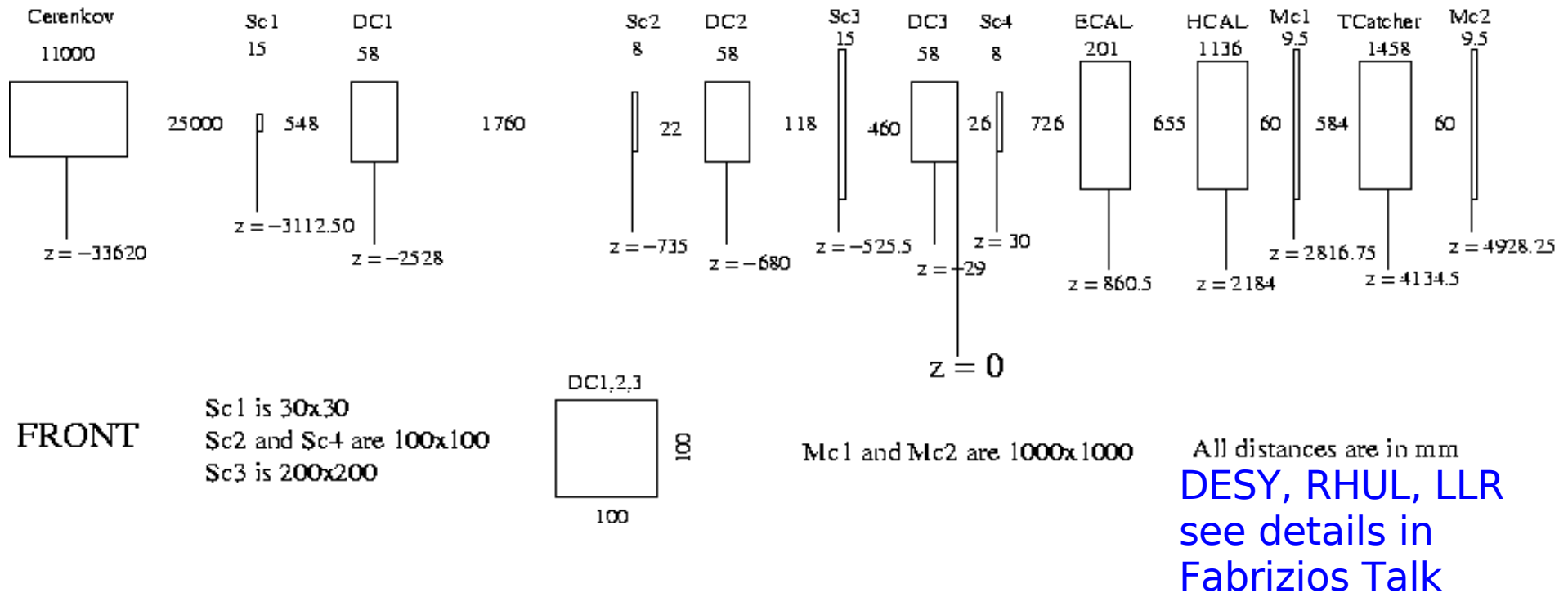
where conddbprofile can be arbitrary expression, only needed to satisfy CondDBMySQL

```
make  
make install
```


A view to the Monte Carlo Branch

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

TOP – CERN August 2006



Will use grid for MC production

Estimation ~ 5000 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 European ILC Software for processing of Testbeam Data

Full dataprocessing chain in operation including conditions data handling

Very important input for current and future developments of ILC Software

Allows for stringent tests 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

CPU consumption still tiny but will grow fast when starting e.g. MC production

- Need to give answers to questions and rumours on performance of ILC Software

Is it too slow, if yes why – Test Scenarios???

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 developers

Overhead generated by usage of ILC Software
- Slower program execution?
- Profiling of ILC Software needed

Source of (potential) errors unclear