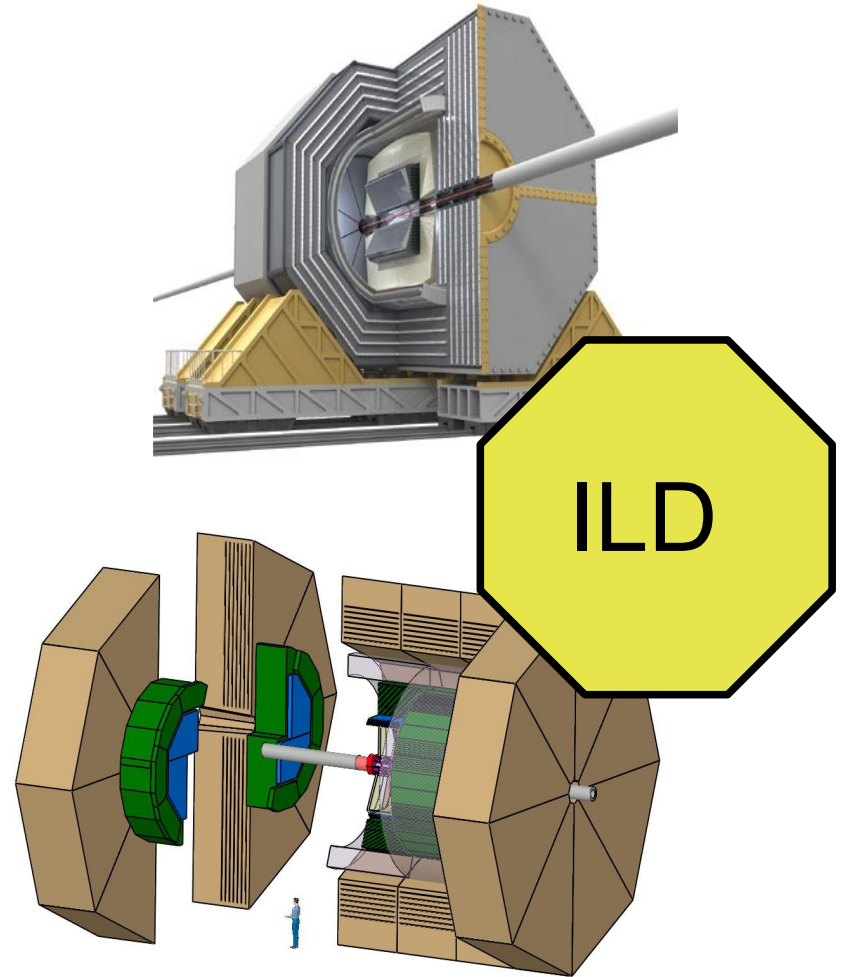# Grid tools for ILD_LDC Monte Carlo Mass Production

Frank Gaede
J.Engels, A.Gellrich, I.Marchesini
DESY
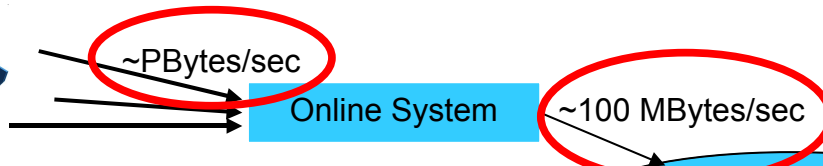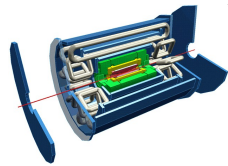ILD Meeting, Sendai, Japan
6-7 march 2008

1

Outline

ILD

LHC Grid Computing

~PBytes/sec

Online System

~100 MBytes/sec
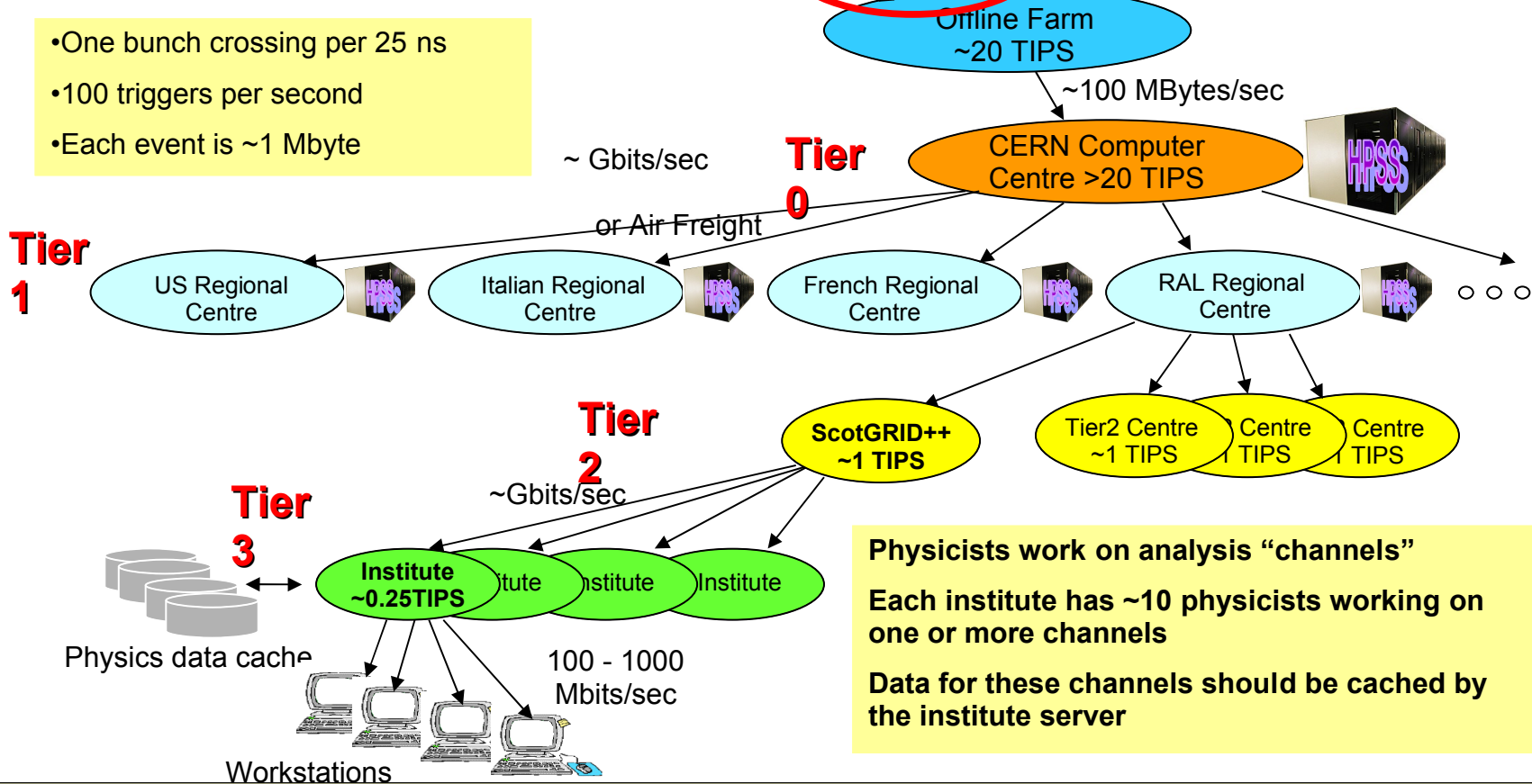
1 TIPS = 25,000 SpecInt95

PC (1999) = ~15 SpecInt95

• One bunch crossing per 25 ns

• 100 triggers per second

• Each event is ~1 Mbyte

Offline Farm
~20 TIPS

~100 MBytes/sec

~ Gbits/sec

or Air Freight

Tier 0

CERN Computer
Centre >20 TIPS

HPSS

Tier 1

US Regional
Centre

Italian Regional
Centre

French Regional
Centre

RAL Regional
Centre

○ ○ ○

Tier 2

~Gbits/sec

ScotGRID++
~1 TIPS

Tier2 Centre
~1 TIPS

Centre
~1 TIPS

Centre
~1 TIPS

Tier 3

Physics data cache

Institute
~0.25TIPS

Institute

Institute

Institute

100 - 1000
Mbits/sec

Workstations

Physicists work on analysis "channels"

Each institute has ~10 physicists working on one or more channels

Data for these channels should be cached by the institute server

• Grid invented to cope with huge amount of LHC data
• do we need the Grid for the ILC at this stage ?

3

Monte Carlo production for LDC-LOI

- LOI-benchmark channels defined by WWS-SW panel

- also need SM backgound

- -> use large Whizard SM data set produced at SLAC

- for example use ~50/fb (lower limit):

| Process	fb^{-1}	#events (*10^3)	Process	#events
ee->2f	20	~2500	light quark 2f at 91.2 GeV	20000
ee->4f	50	~6000		
ee->6f	50	~100	tt (6f) at 350 GeV	20000
ee->hX	50	~20		
ee->γγ(n*γ)	10	~600		
νν(n*γ)	20	~1700		
ee->ee	0.1	~200		
eγ->eγ	0.1	~650		
γγ->X	0.1	~1000		
rest	1	~600		

- 14 M events
- assume 2-4min./event :
- -> **50-100 CPU years**
- per detector configuration !

using the Grid is the only way forward in order to to get the resources that we need for the LOI mass production !

Grid Definition

- I. Foster: What is the Grid? A Three Point Checklist (2002)

- "A Grid is a system that:

- coordinates resources which are not subject to centralized controls …

    - integration and coordination of resources and users of different domains vs. local management systems (batch systems)

- … using standard, open, general-purpose protocols and interfaces …

    - standard and open multi-purpose protocols vs. application specific system

- … to deliver nontrivial qualities of services."

    - coordinated use of resources vs. uncoordinated approach (world wide web)

The Grid dream

Mobile Access

Desktop

Visualizing

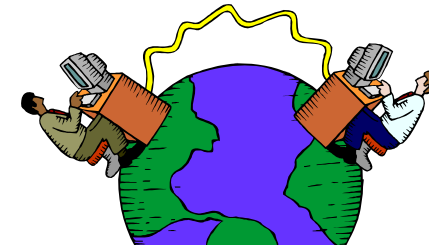**GRID MIDDLEWARE**
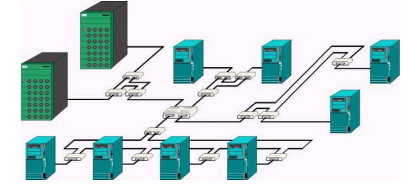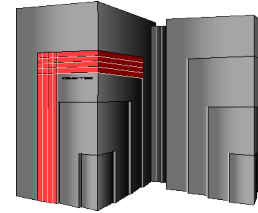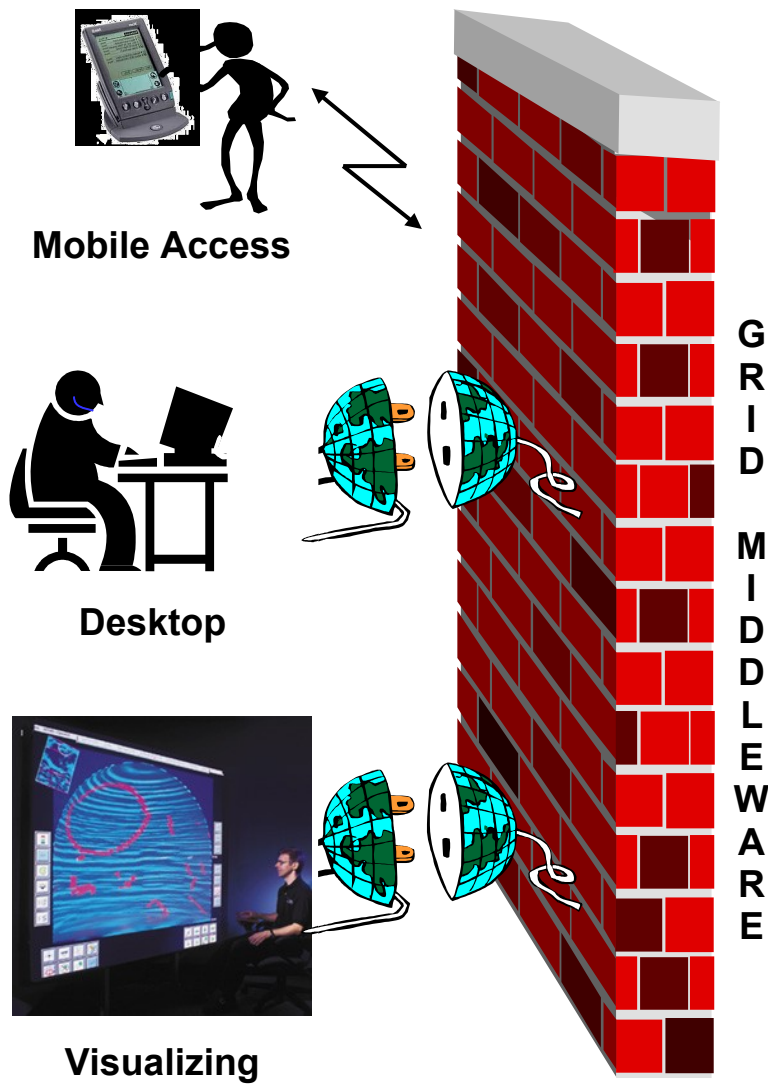
Supercomputer, PC-Cluster
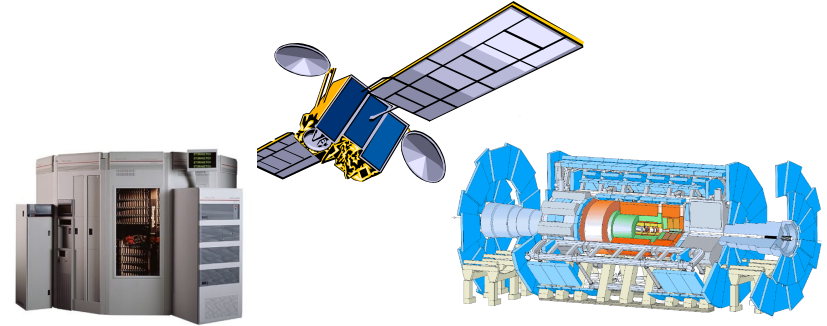
Data Storage, Sensors, Experiments

Internet, Networks

Grid: Authentification & Authorization

- a user is uniquely identified through a ***certificate***

  - an encrypted electronic document, digitally signed by a Certification Authority (CA)

    - a certificate is your passport to enter the grid world
    - example: /O=GermanGrid/OU=DESY/CN=Frank Gaede

- access to resources is provided (controlled) via membership in a **Virtual Organization**

  - a dynamic collection of individuals, institutions, and resources which is defined by certain sharing rules

    - the VO a user belongs to is not part of the certificate.
    - ILC related VOs: ilc, calice

7

Grid schematic view
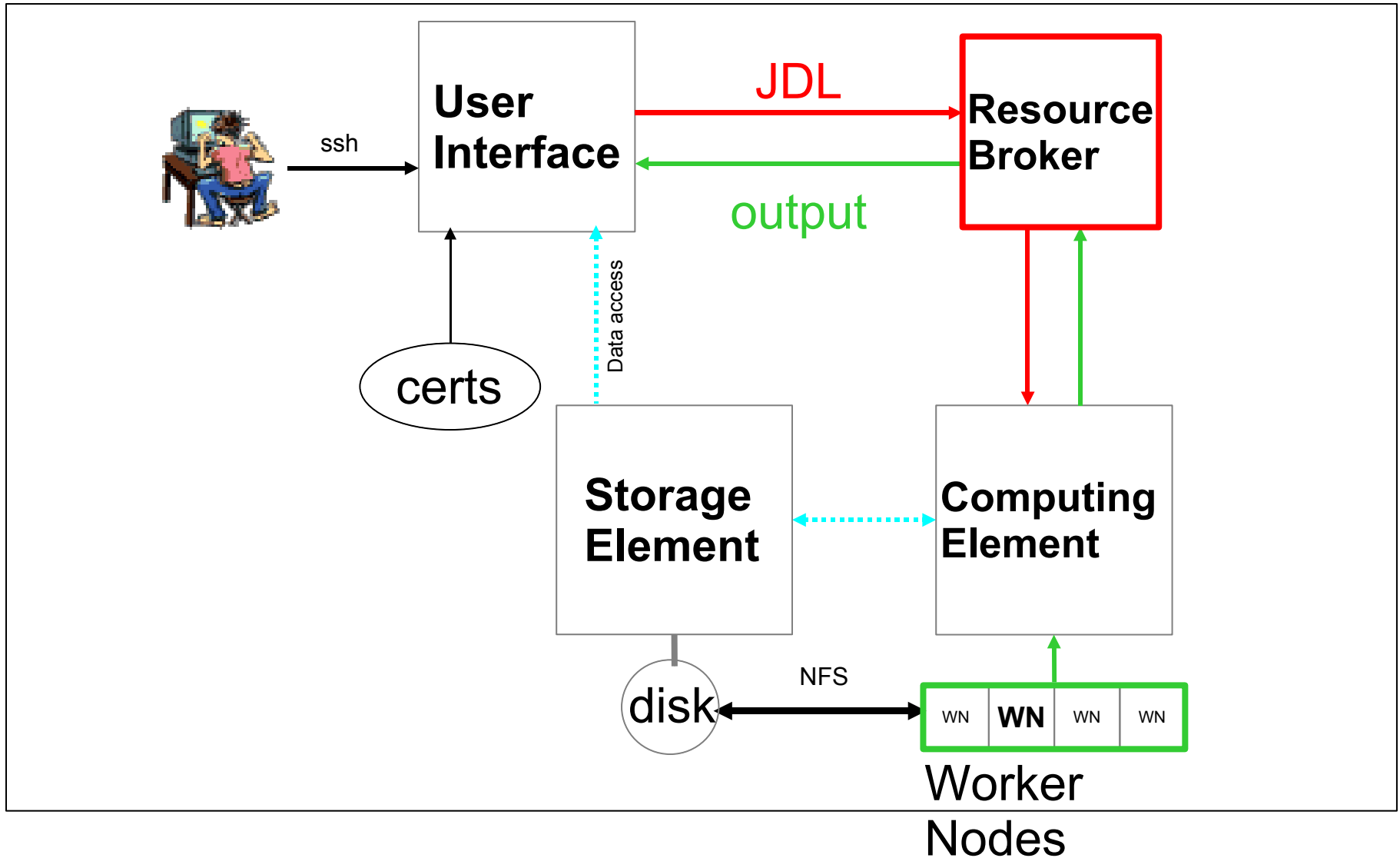
ilcsoft job submission scripts

- automated job submission scripts for **Mokka & Marlin** with direct access to MySQL database   I.Marchesine

  - keep track of job status

                                                              (see next slide)
  - register produced output files

- job input:

  - steering files

  - input files: stdhep, slcio

  - binaries (either on SE or installed locally)

- job output

  - slcio files – stored on Grid SE

  - tar ball: log files, steering files

Monte Carlo production DB

I.Marchesine

Mysql database.

MC database

Input files table

All about generator files:
- location
- process
- #events
- ...

Mokka jobs table

All about the running Mokka jobs:
- grid id
- status
- input file
- process
- ...

Simulations table

Successful simulations information:
- location (grid and SE)
- process
- #events
- ...

Marlin jobs table

All about the running Marlin jobs.

Reconstr. table

Successful reconstructions information.

Php based v

MySQLCC – [MCAdmin@flcweb01.desy.de:3306] Query Window

File Edit View Query Options HotKeys

	ob	Output_File_Name	Status	Mokka_Version	Geant4_V	Co	M	Date	Resubnum	Co
235	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
236	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
237	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
238	rid-	MCVS_2007_12_07_c	Submitted	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2008-0	2	
239	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
240	rid-	MCVS_2007_12_07_c	new	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
241	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
242	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
243	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
244	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
245	rid-	MCVS_2007_12_07_c	Submitted	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2008-0	3	
246	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
247	rid-	MCVS_2007_12_07_c	checked	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2007-1	0	
248	rid-	MCVS_2007_12_07_c	Submitted	cvsHEAD_2007_12_07	9.0version	i.ma	ivar	2008-0	3	

Result 1

Columns
- ☑ B_Factor
- ☑ Barrel_...
- ☑ Compu...
- ☑ Contac...
- ☑ Date_o...
- ☑ Date_o...
- ☑ Detect...
- ☑ Geant4...
- ☑ Grid_J...
- ☑ ID
- ☑ Input_Dir
- ☑ Input_F...
- ☑ Mail_of...
- ☑ Mokka

350 rows in set (0.02) sec
Query OK, 1 row affected (0.00) sec

Messages History Explain

Executing Query Read Only

10

Monte carlo file catalogue - DB

International Linear Collider Database – Mozilla Firefox

File   Edit   View   History   Bookmarks   Tools   Help

http://www-flc.desy.de/simulation/database/   Google

Getting Started   Latest Headlines

International Linear Collider Database

Generator Files Simulated Files Reconstructed Files

More links:

- XML Files

  Gear geometry files for the latest detector models and the steering files adopted in the recent reconstructions. These .xml files can, anyway, **always be found in the tar archives** associated to each simulated .slcio file and each reconstructed .slcio file (see below "Contents").

- Make a request

  To request some simulation or reconstruction still missing in the database.

Contents: (more details in the web interfaces for the three different tables)

- **Generator Files.**

  Generator files used as input for the simulations.

- **Simulated Files.**

  This table of the database contains:

  - The simulated **.slcio files** from Mokka.
  - Each .slcio file is paired to a **tar.gz archive** containing:
    - The **GearOutput.xml** gear geometry file from Mokka.
    - The **FILE_ID.g4** and the **FILE_ID.steer** files: this files, named as the .slcio simulated file itself, are respectively the macro and the steering files used in the job.
    - The **FILE_ID.log** file: this log file, named as the .slcio simulated file itself, is the log file from Mokka.
    - The standard output files (**err** and **out**) from the job.

- **Reconstructed Files.**

Done

http://www-flc.desy.de/simulation/database/

Monte carlo file catalogue – DB II

International Linear Collider Simulations Database – Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www-flc.desy.de/simulation/databasesimulation/

Getting Started   Latest Headlines

International Linear Collider Simulations Database

Search Database   **Browse Database**   **XML Files**   **Make a request**

search DB for files wrt.: tag,run,process,E,date,B-field,...

PARAMETER	INPUT	EXAMPLE
Tag:		For the new productions select the tags: • **Test_500**: first tests with the Whizard samples from SLAC • **Test_SinglePar**: test single particle production • **Test_ZPole**: test Z pole production • **Test_350_Zh120**: test Z higgs (120), CMS 350, pythia • **Test_500_Zh120**: test Z higgs (120), CMS 500, pandora pythia TAGS SUMMARY
Run ID:		m-5-4_cb_1000_noisr_ldc00sc_3.00t_r1690._l2730._qgsp_bert
Process:		cb,n1n1h,...
Center of Mass Energy [GeV]:		1000,500,...
Date of Production:		2006-02-19,2007,12,2006-05,...
Event Generator:		pythia,...
Detector Simulation:		mokka,mokka 5.4,...
Detector Model:		ldc00sc,ldc01_02sc,...
Physics List:		qgsp_bert, lcphys,...
B Field [T]:		3.0,4.0,...

Done

12

Monte carlo file catalogue – DB III

International Linear Collider Simulations Database

Search Database   Browse Database   XML Files   Make a request

lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0014.tar.gz

List of all .slcio files on the Grid

lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0001.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0002.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0003.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0004.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0005.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0006.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0007.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0008.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0009.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0010.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0011.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0012.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0013.slcio
lfn:/grid/ilc/mc-2008/simulated/LDC01_05Sc/singleParticle/M-6-5p2_eta_gammagamma_Theta4-176_1-50GeV_LDC01_05Sc_LCP_Test_SinglePar_0014.slcio

get list of **logical grid file names** for
• downloading data from the grid
• further processing ion the grid

13

ilc sw-installation

- ilc software requirements and complexity has grown
  - ~30 packages with sometimes optional dependencies
- **ilcinstall** tool used to facilitate installations
  - script to install all of the LDC software in one go
  - "**start script – go to lunch – run application**"
  - fully configurable:
  - versions, dependencies/build options, links to existing packages/tools, e.g. root, CLHEP,...
- used for reference installations in afs (SL3/SL4)
  - user can link their packages against these
  - even w/o installing any software on their computer
- /afs/desy.de/group/it/ilcsoft/v01-03-02   New

try to use this also on the grid for user convenience

ilcsoft installations on the grid

J.Engels

Grid

CE

WN1 WN2 ... WNn

ESA
$VO_ILC_SW_DIR

SE
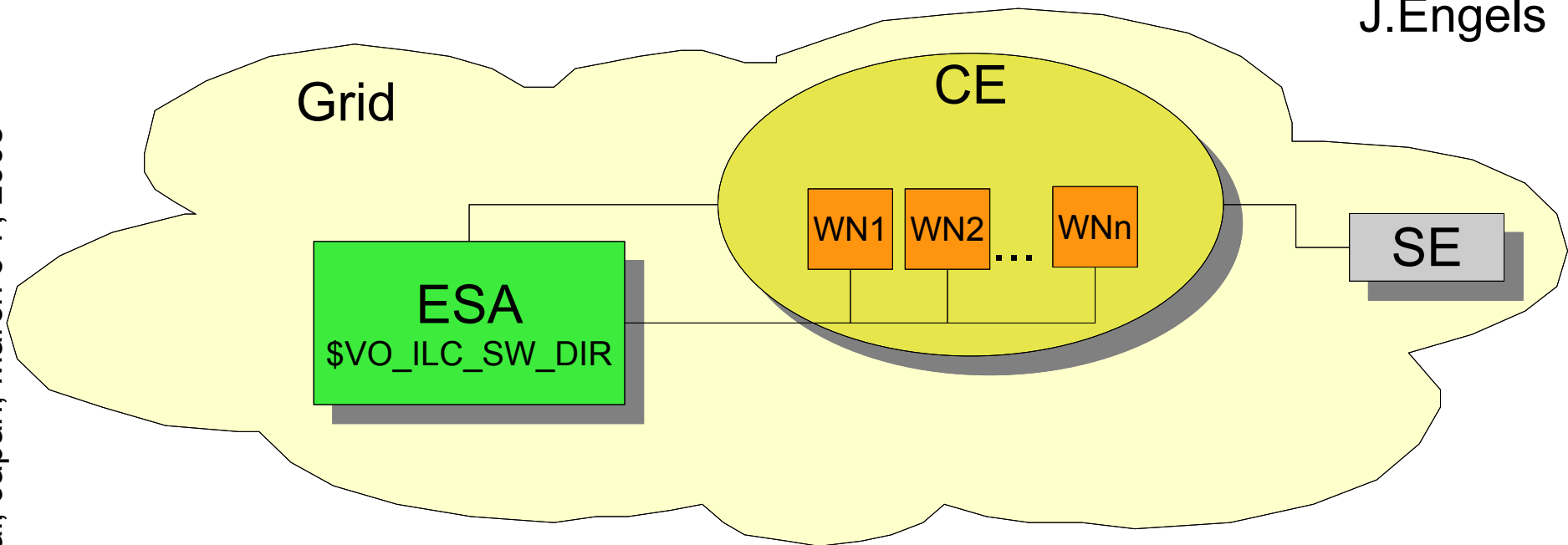
- ilcsoft v01-03-02 reference installation is already available on the grid !

- installed in ESA (Experiment Software Area) for ILC

  - directory visible to all worker-nodes in a CE

  - referred to through ${VO_ILC_SW_DIR}

possible on all LCG grids supporting VO 'ilc'

15

example: using v01-03-01 on the grid

```
#mytest.jdl

VirtualOrganisation = "ilc";

Executable = "myscript.sh";

StdOutput = "out";

StdError = "err";

InputSandbox = {"myscript.sh"};

OutputSandbox = {"out","err"};

Requirements = Member( "VO-ilc-ilcsoft-v01-03-01-sl4",
        other.GlueHostApplicationSoftwareRunTimeEnvironment);
```

```
#!/bin/sh

# myscript.sh


export LCSOFT=$VO_ILC_SW_DIR/ilcsoft/sl4/v01-03-01

export PATH=$LCSOFT/bin:$PATH

export LD_LIBRARY_PATH=$LCSOFT/lib:$LD_LIBRARY_PATH

export MARLIN_DLL=libMarlinReco.so:libLCFIVertex.so:libSiliconDigi.so:libOverlay.so:libPandoraPFA.so:libEutelescope.so


Marlin -x
```

glite-wms-job-submit -a ./mytest.jdl

16

grid sites with ilcsoft v01-03-02

CE	SW-VER	SW-OS	DATE	TIME	SAM	JOB	TAGGED	HIST-LOGS
cclcgceli02.in2p3.fr	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
cclcgceli03.in2p3.fr	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
dg10.cc.kek.jp	v01-03-02	sl3	2008-03-03	12-38-40	OK	OK	VO-ilc-ilcsoft-v01-03-02-sl3	History
grid-ce3.desy.de	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
heplnx206.pp.rl.ac.uk	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
lcg-ce0.ifh.de	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
lcg-ce1.ifh.de	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
lcgce02.gridpp.rl.ac.uk	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
node07.datagrid.cea.fr	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History
t2ce03.physics.ox.ac.uk	v01-03-02	sl4	2008-02-29	18-54-12	OK	OK	VO-ilc-ilcsoft-v01-03-01-sl4 VO-ilc-ilcsoft-v01-03-02-sl4	History

http://ilcsoft.desy.de/grid/results/User.html

17

MC production status

- tools for grid production are available

- Mokka development to be frozen next week

- -> need to test full chain:
  - start with some simple events:
    - singles O(10k) of g, e, mu+-, pi+-, K_l, K_s
    - O(10k) Z->uds @ 90,250,500 GeV
    - O(10k) ZH @ 250, 500 GeV
    - both for LDC and LDCPrime ?
    - (use also to put together and test a standard reconstruction)

- will provide important feedback on actual throughput that can be achieved and allow better estimate of what is possible

Summary

- the Grid is the only way to get a reasonably large set of Monte Carlo for LOI/detector optimization

- grid tools for Mokka/Marlin available:
  - software installation
  - job submission
  - data catalogue/web interface to database

- will be used for mass production soon

- tools can of course be adapted also to be used with other simulation tools such as Jupiter
- let us know if you like to contribute to the Monte Carlo mass production