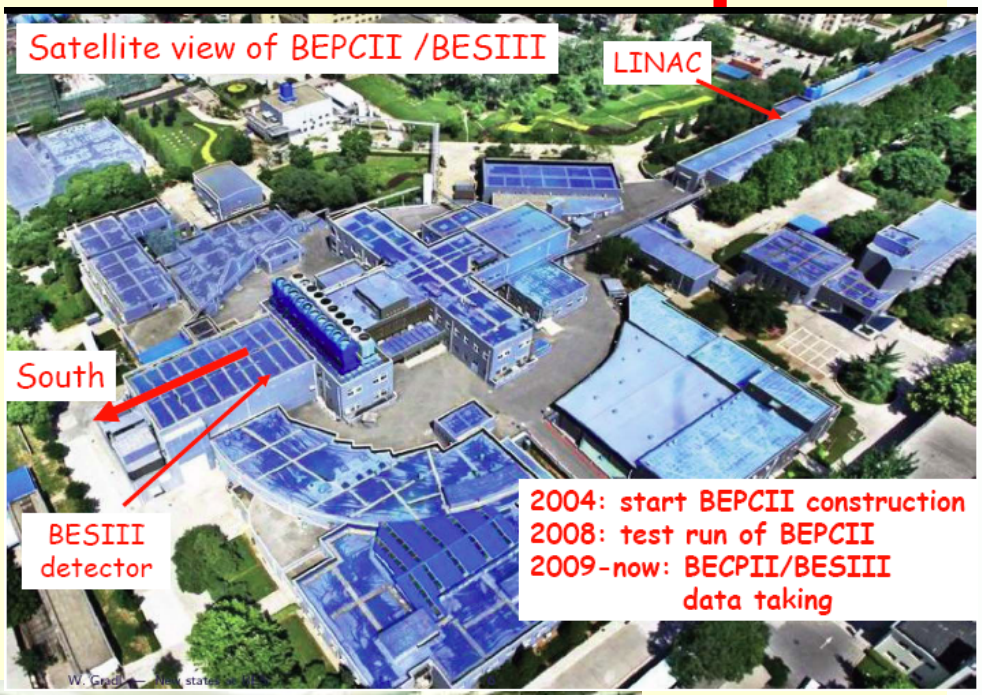
Data Processing and Analysis on the BESIII Experiment

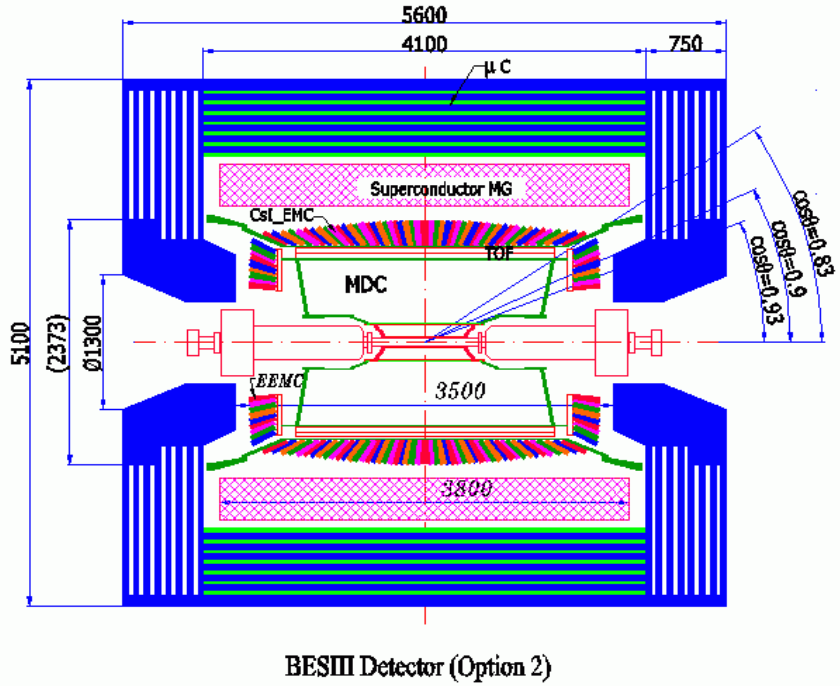
Weidong Li

Institute of High Energy Physics, CAS

ChengDu, July 4th 2011

BESIII Experiment at BEPCII





BESIII Detector (Option 2)

Beam energy: 1.0-2.3 GeV

Luminosity: ~1×10^{33} $cm^{-2}s^{-1}$

Optimum energy: 1.89 GeV

No. of bunches: 93

Bunch length: 1.5 cm

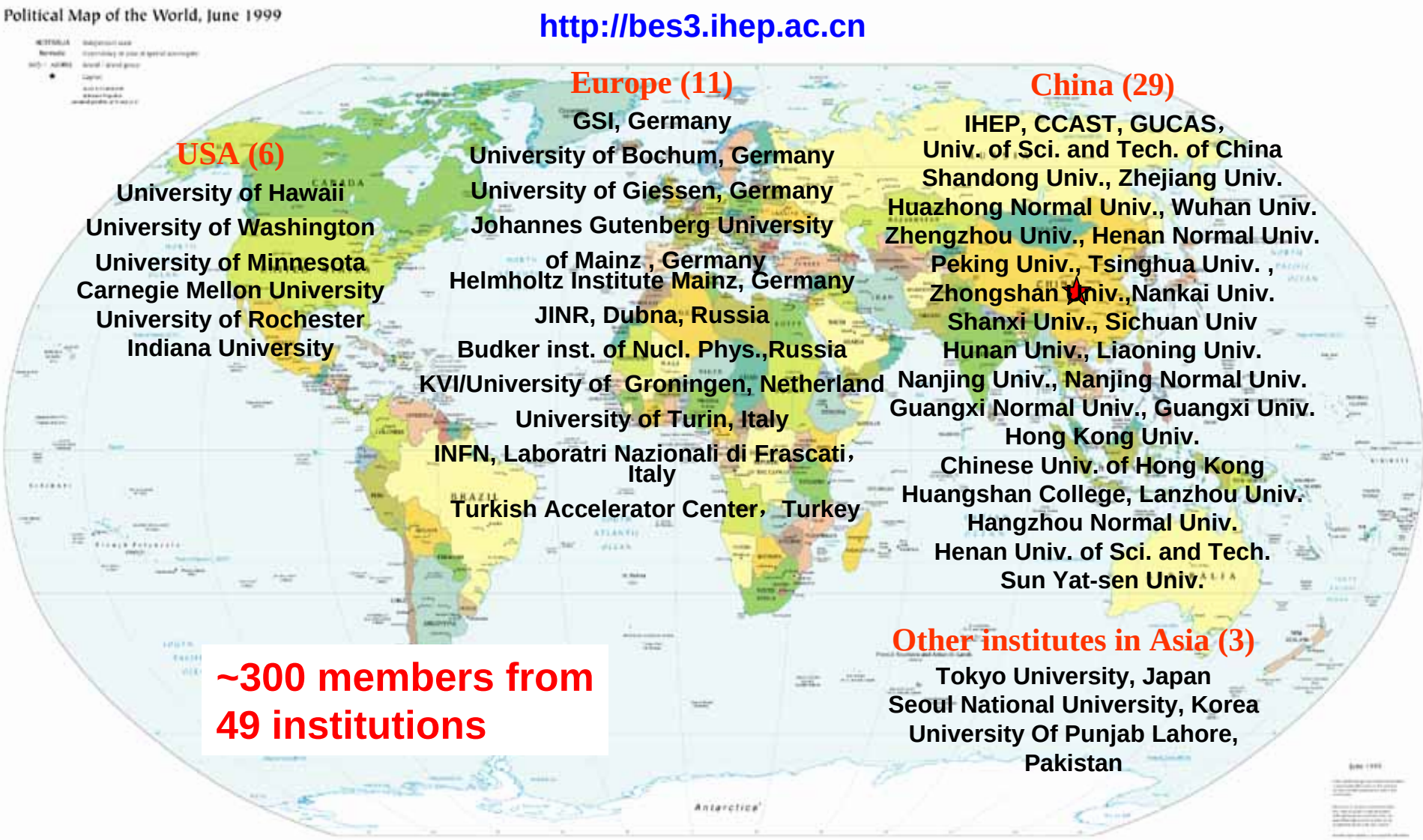
Circumference: 237 m

BESIII detector:

- Main Drift Chamber
- Time of Flight
- CsI Calorimeter
- Muon Chamber
- 1T SC Magnet

BESIII Collaboration

http://bes3.ihep.ac.cn



USA (6)

University of Hawaii
University of Washington
University of Minnesota
Carnegie Mellon University
University of Rochester
Indiana University

Europe (11)

GSI, Germany
University of Bochum, Germany
University of Giessen, Germany
Johannes Gutenberg University of Mainz , Germany
Helmholtz Institute Mainz, Germany
JINR, Dubna, Russia
Budker inst. of Nucl. Phys.,Russia
KVI/University of Groningen, Netherland
University of Turin, Italy
INFN, Laboratri Nazionali di Frascati，Italy
Turkish Accelerator Center，Turkey

China (29)

IHEP, CCAST, GUCAS，
Univ. of Sci. and Tech. of China
Shandong Univ., Zhejiang Univ.
Huazhong Normal Univ., Wuhan Univ.
Zhengzhou Univ., Henan Normal Univ.
Peking Univ., Tsinghua Univ. ,
Zhongshan Univ.,Nankai Univ.
Shanxi Univ., Sichuan Univ
Hunan Univ., Liaoning Univ.
Nanjing Univ., Nanjing Normal Univ.
Guangxi Normal Univ., Guangxi Univ.
Hong Kong Univ.
Chinese Univ. of Hong Kong
Huangshan College, Lanzhou Univ.
Hangzhou Normal Univ.
Henan Univ. of Sci. and Tech.
Sun Yat-sen Univ.

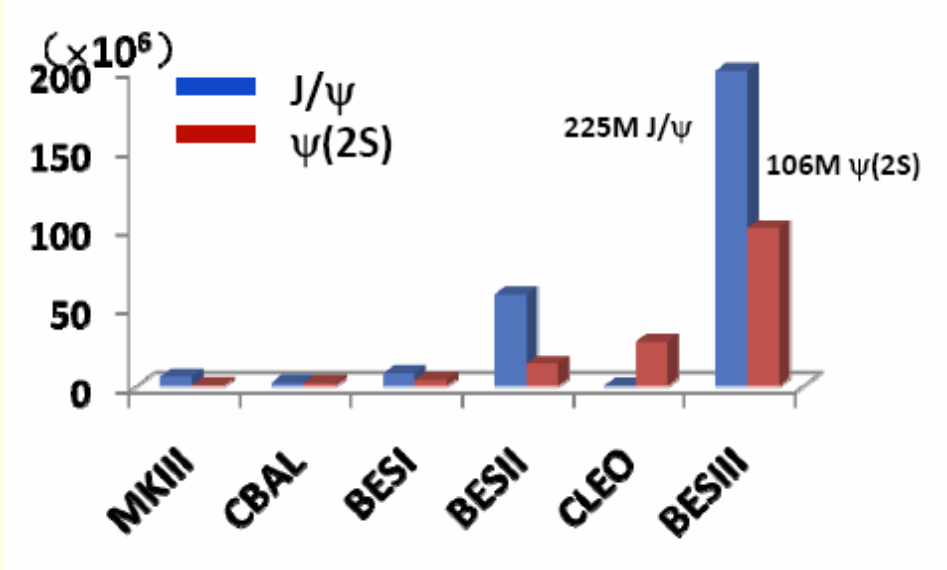
Other institutes in Asia (3)

Tokyo University, Japan
Seoul National University, Korea
University Of Punjab Lahore, Pakistan

~300 members from 49 institutions

BESIII Physics and Data

- Physics research program includes
  - Charmonium Physics
  - Light Hadrons Spectrocopy
  - Charm (D/Ds) Physics
  - QCD/R Value measurements etc.
- Experimental Data that have already been accumulated
  - ~106 M ψ(2s) events and 225 M J/ψ events
  - ~2.8 fb^{-1} at $\psi(3770)$
  - Event data at 4.01 GeV ~500 pb^{-1}
  - more J/ψ and ψ(2s) data will be taken



- Data Size
  - Monte Carlo Events: 5-7 KB/event for raw event, 18-25 KB/event for DST event
  - Real data (one reconstruction version) ~300 TB

BESIII Physics Publications

Title	
Measurement of $h_c(^1P_1)$ in ψ' decay	PRL 104, 132002 (2010)
Branching fractions of χ_{c0} and χ_{c2} to $\pi^0\pi^0$ and $\eta\eta$	PRD 81, 052005 (2010)
Observation of a p p-bar mass threshold enhancement in $\psi' \to \pi^+\pi^- J/\psi$, $J/\psi \to$ pp-bar	Chinese Phys C 34, 421 (2010)
First observation of the decays $\chi_{cJ} \to \pi^0\pi^0\pi^0\pi^0$	Acc. by PRD
Evidence for ψ' decays into $\gamma\pi^0$ and $\gamma\eta$	PRL 105, 261801 (2010)
Study of $a_0(980)$ - $f_0(980)$ mixing	Acc. by PRL
Measurement of the matrix element for $\eta' \to \eta\pi^+\pi^-$	PRD 83, 012003 (2011)
Confirmation of the X(1835) and observation of the X(2120) and X(2370) in $J/\psi \to \gamma\pi^+\pi^-\eta'$	Acc. by PRL

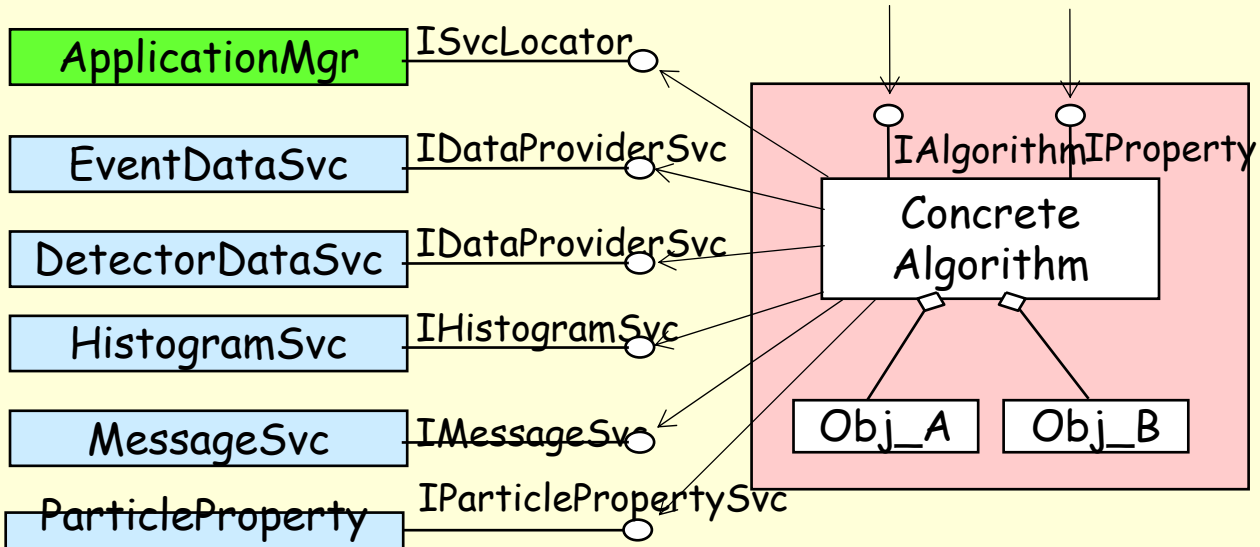
1. Framework and Infrastructure

Requirements and Software Environment

- Computing Requirements
  - Storage: ～5 PB
  - CPU : ～3000-6000 CPU cores
  - Disk and Tape I/O: ~1GB/s
- Software requirements
  - During 10-year experiment, should be stable and easy to maintain.
  - Good flexibility and extensibility. Easy to plug-in new software components.
  - Straightforward for end-users.
- Underlying framework
  - GAUDI (originally developed by LHCb)
- Simulation
  - GEANT4
- Other external LIBs:
  - CERNLIB, CLHEP, ROOT, AIDA, XercesC, GDML …
- Database: MySQL
- Computer language: C++
- Operation system:
  - Scientific Linux CERN 5/ GCC 4.3.2

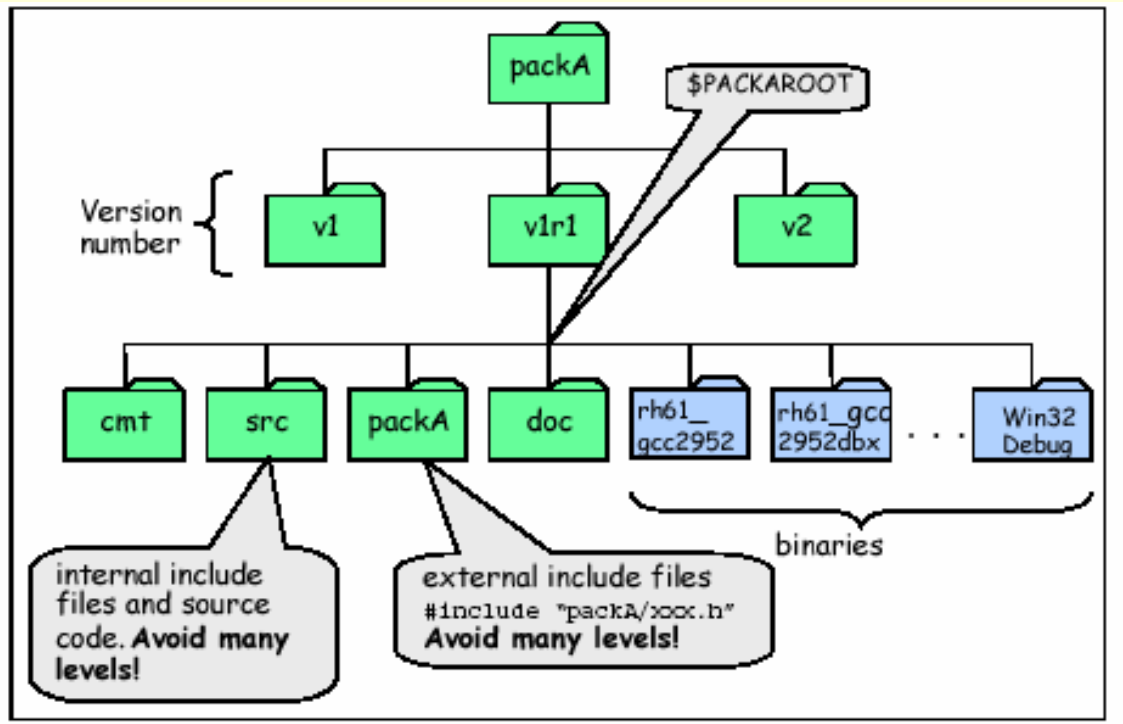
GAUDI-based Software System

- GAUDI is adopted as the software framework — a skeleton of an application into which developers plug in their code and provides most of the common functionality.
- Only need to develop experiment-dependent software packages.
- Clear separation between “algorithms” (key components) and “data” (“persistent data” and “transient data”)
- Algorithms are executed once per physics event in sequence



Software Configuration Management

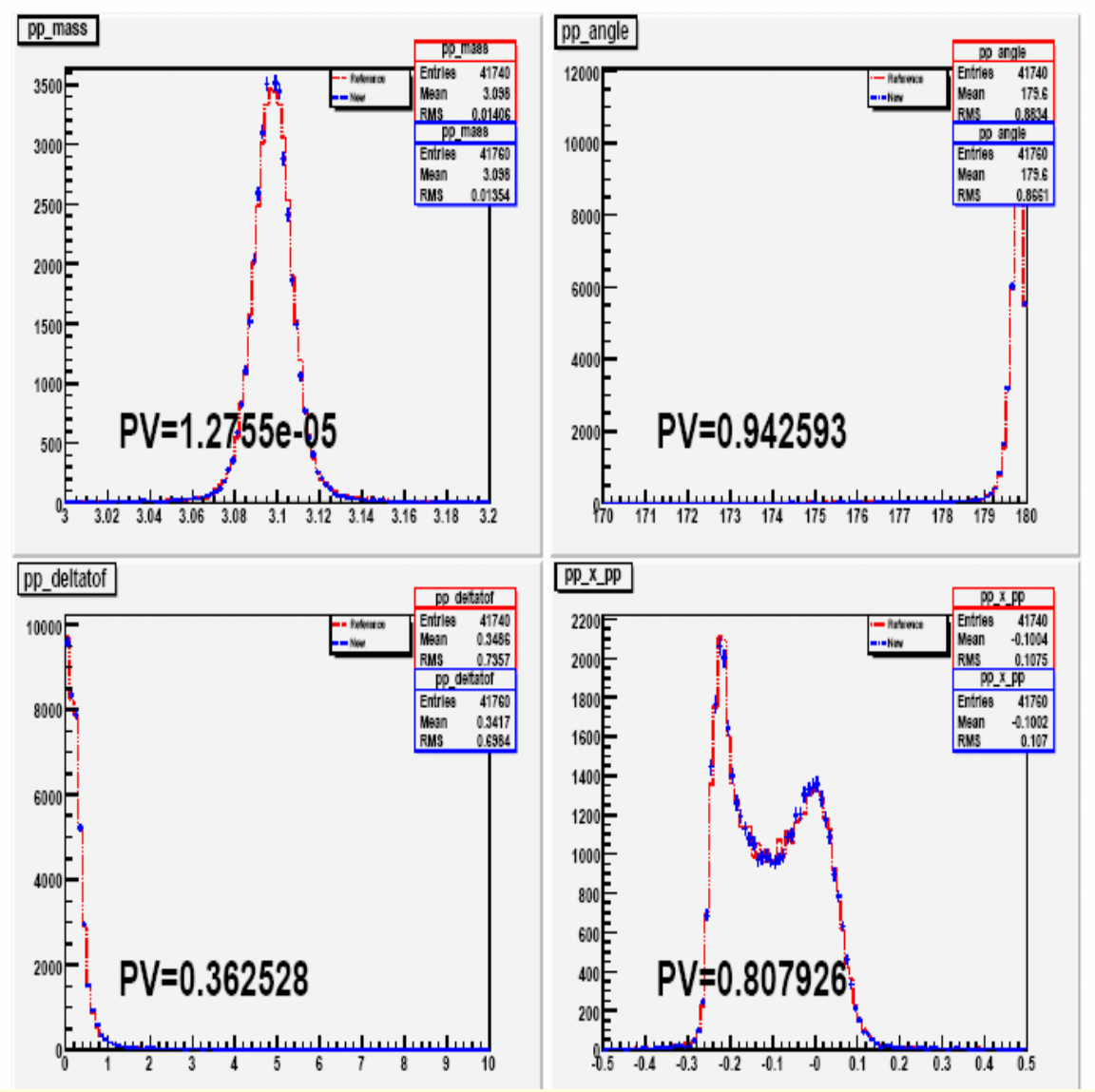
- CMT (Configuration Management Tool)
  - structures software development (concepts of areas, packages, versions, constituents)
  - organises software into packages
  - describes package properties
  - describes package constituents



- operates the software production (management, build, import/export, etc...)

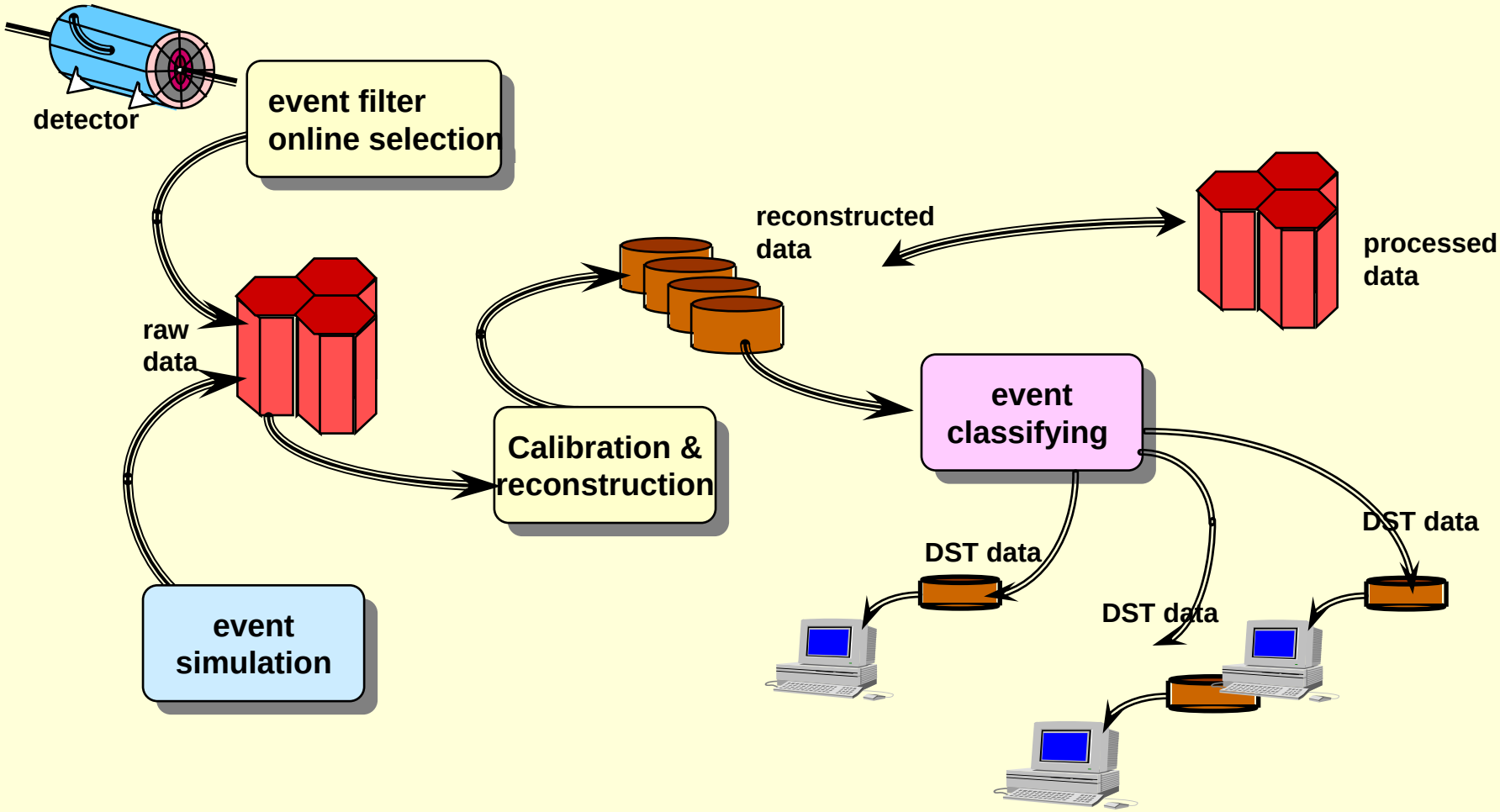
Software Validation and Distribution

- Before a software release is published, validation is done in a systematic way using both Monte Carlo and real data.
- PACMAN is used to install the software in other sites.



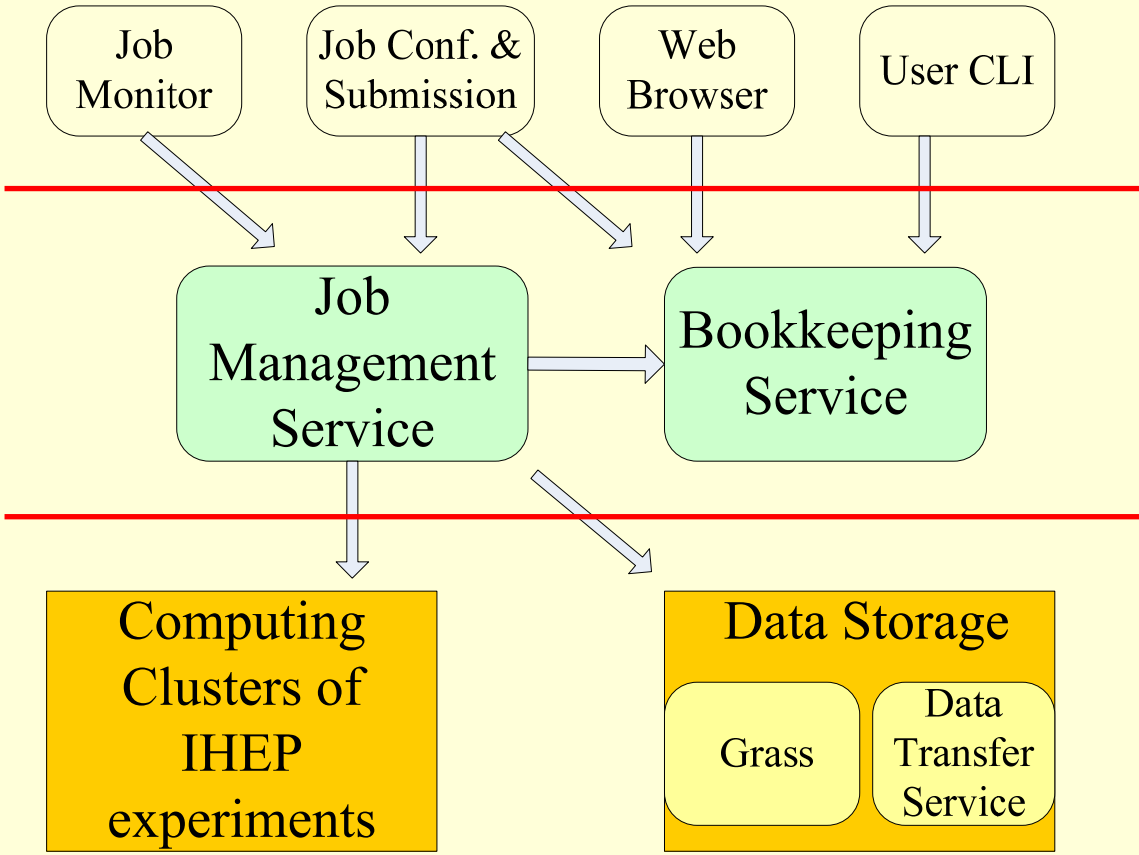
2. Data Processing and Analysis

Offline Data Processing



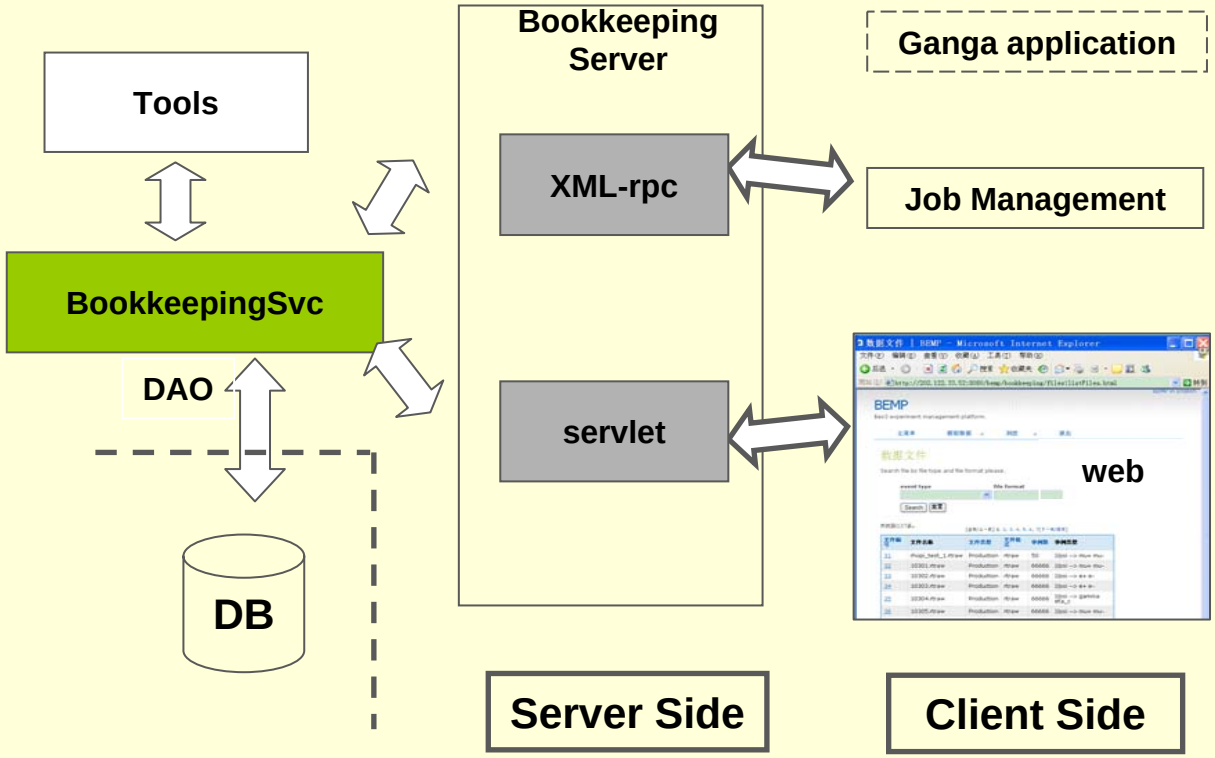
Data Production Jobs

- Job Configuration and Submission Module
  - Setting up the job running environment
  - Converting the input dataset to a list of input data files
  - Defining the policies for dividing the job into sub-jobs and for combining outputs.
  - Providing the functionality for automatic/manual re-submitting the failed jobs.
- When all the sub-jobs are successfully finished, bookkeeping database can be updated through web or command line.



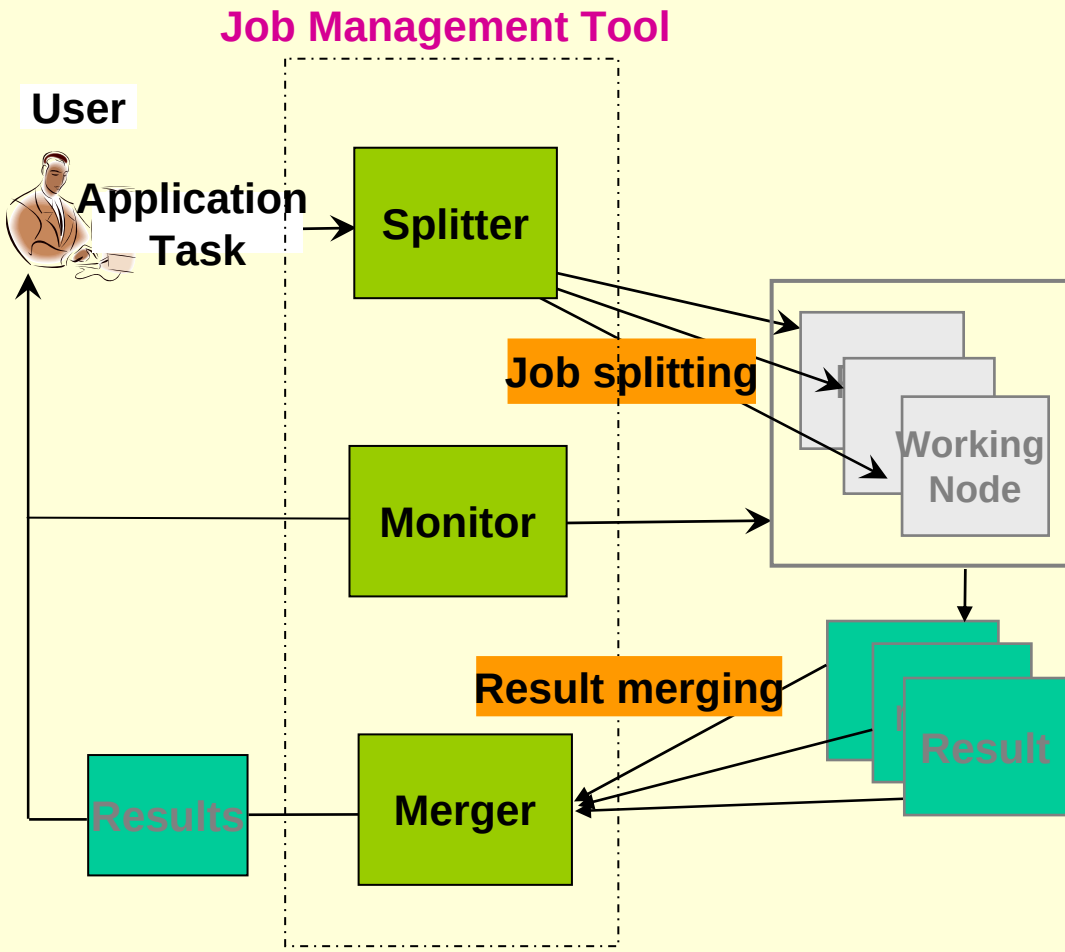
Data Management

- The Bookkeeping system keeps track of data processing history and provides facilities to query the information needed by users.
  - File, job, relation between them, various types of conditions for data processing
- Public interfaces
  - Web browser - ***servlet***
  - Communication with other components (Ganga/Job Management Service)
    - ***XML-rpc***



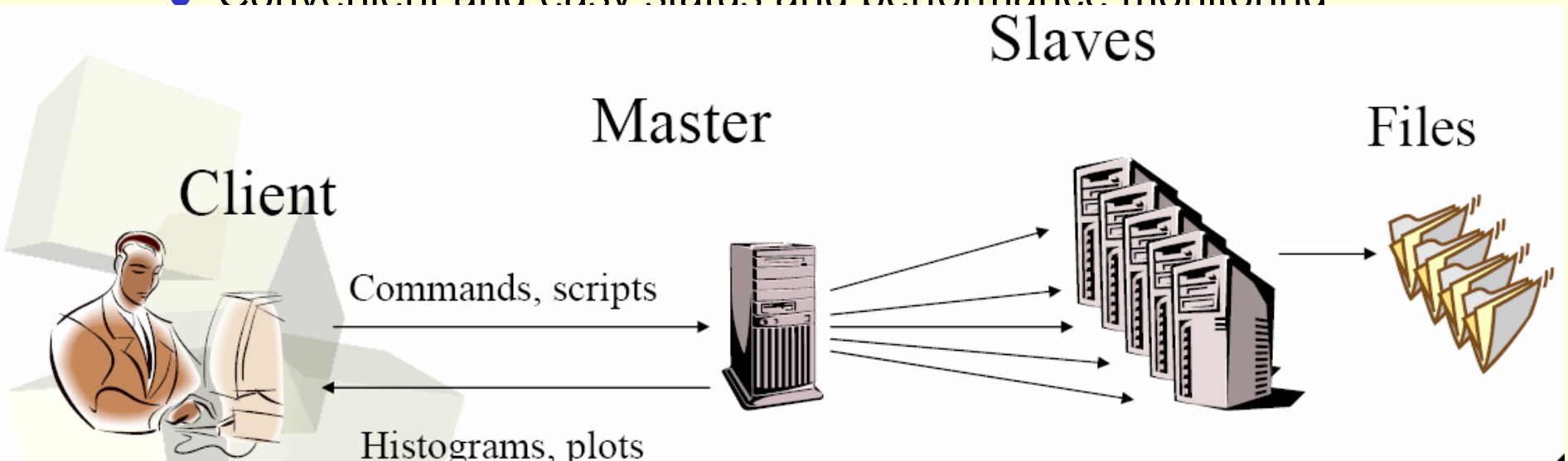
Analysis Jobs

- Ganga-based Job Management Tool
  - Setting up the job running environment
  - Contacting Bookkeeping Service to get input data files
  - Depending on the need, the job can be split into sub-jobs by default or according to user's own policy.
  - Monitor helps users to keep track of their job status.
  - Providing the functionality for automatically/manually re-submitting the failed jobs.
- The tool has already been used in MDC calibration.
- Should be ready for analysis jobs after completing some work.



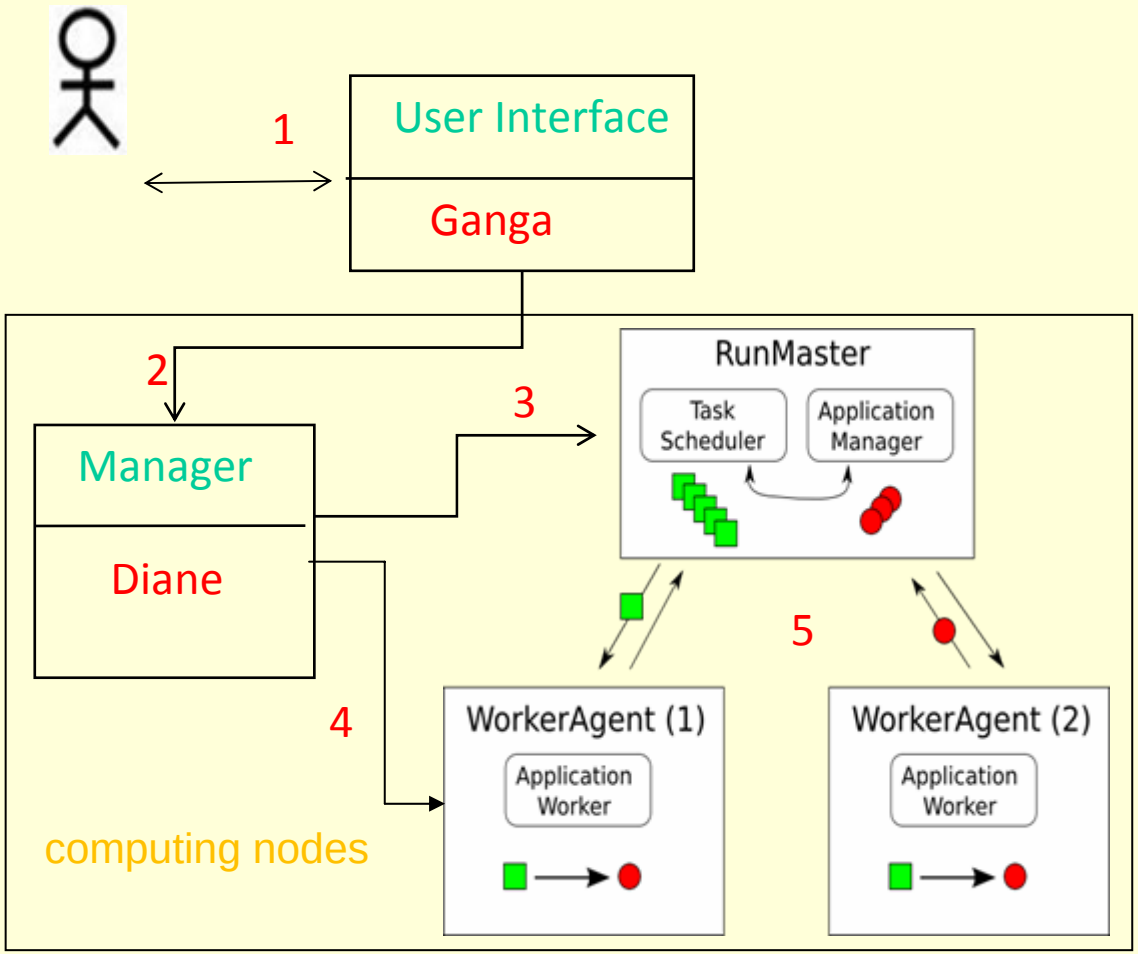
PROOF for Physics Analysis

- Parallel ROOT Facility
- Interactive parallel analysis on a cluster
- Advantages:
  - Interactive
  - Auto splitting and merging
  - Efficient, good performance with dynamic load balancing via master and slave architecture
  - Convenient and easy status and performance monitoring



3. Other Computing Issues

Locally-distributed Computing: DistBoss



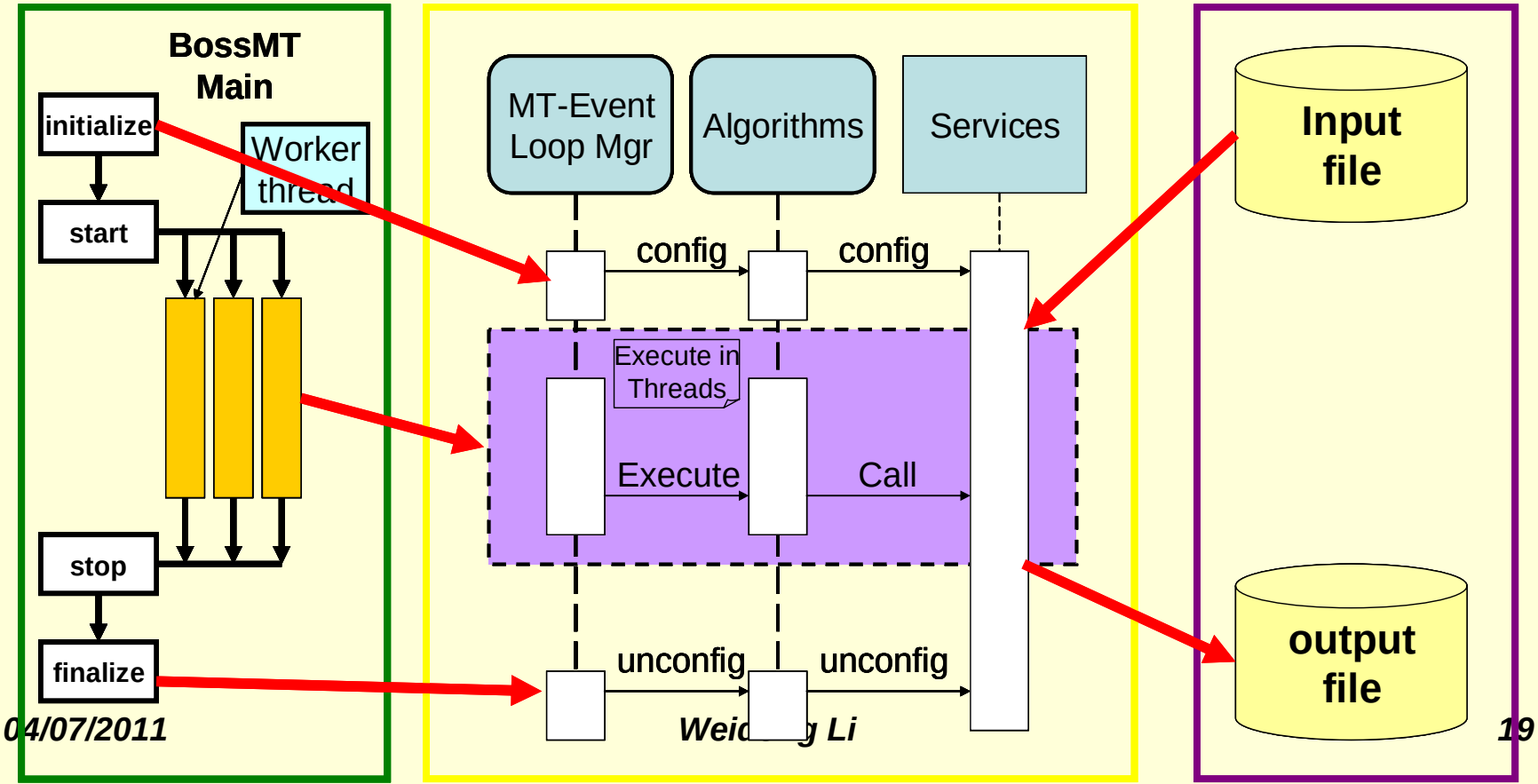
- The DistBoss is a distributed system developed for fast calibration, reconstruction, software validation etc.
- Ganga is used as User Interface.
- Diane is used to control and manage the running of master and workers.
  - The master schedules tasks of the workers.
  - The workers consume jobs and return the results.
- The data processing is paralleled at event level.

Run cmd:

distboss joboption [queue] [workernumber]

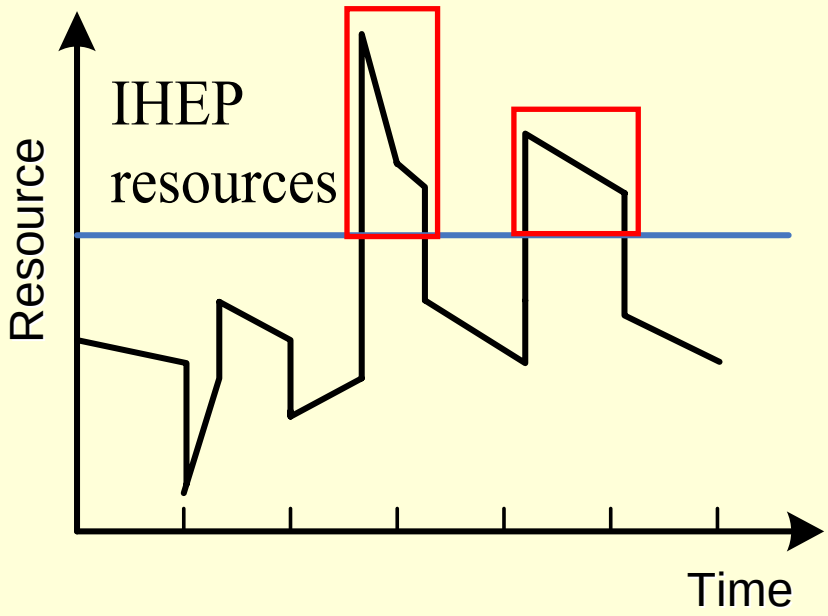
Parallel Computing

- **Performance**
  - EMC reconstruction about 2.5 times faster in the quad-core machine
  - memory usage is reasonable: 10% increase per thread

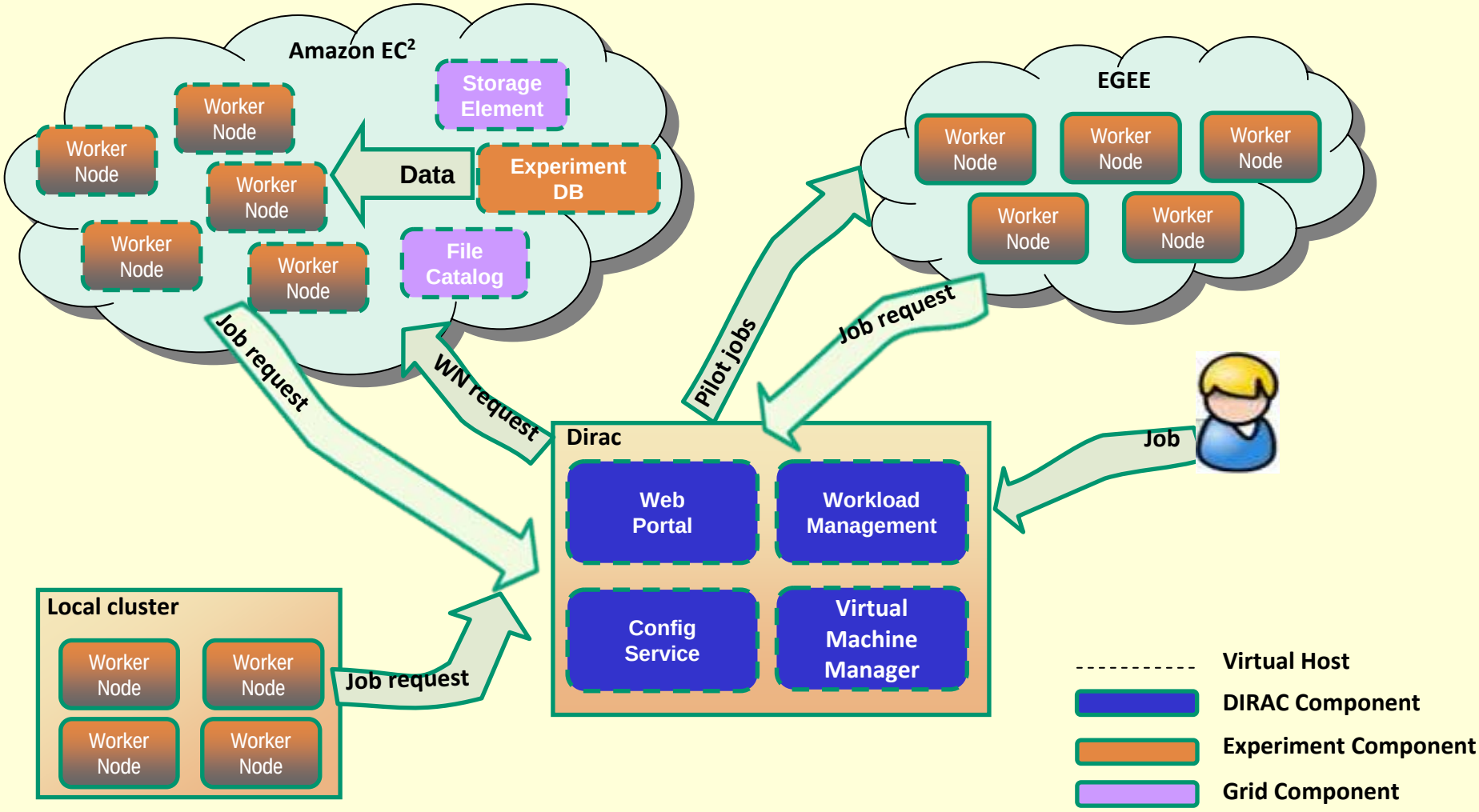


Distributed Computing（Grid/Cloud）

- Motivation
  - The more data is accumulated, more difficult for IHEP to provide all the computing resources for both raw data processing and MC production.
- A possible computing model
  - IHEP is responsible for processing and storage of all the real raw data.
  - MC production and analysis jobs will be distributed among a number of sites with enough computing resources.



A Possible Solution (from Belle II)



Summary

- The BESIII experiment completed successfully:
  - ~106 M $\psi(2s)$ events and 225 M J/ψ events
  - ~2.8 fb^{-1} at $\psi(3770)$
  - Event data at 4.01 GeV ~500 pb^{-1}
- The large-scale offline software system was developed to meet the requirements from both data processing and analysis.
- In the coming years, distributed computing (GRID/Cloud) will become a necessity to handle the ~PB level data.

Thank you !