

# BES-III Computing Model

**First Workshop on Data Preservation and Long  
Term Analysis in HEP**

# Introduction to IHEP

**Comprehensive and largest research center in China for Particle physics, Accelerator, Synchrotron radiation technologies, with wide range of scientific program.**

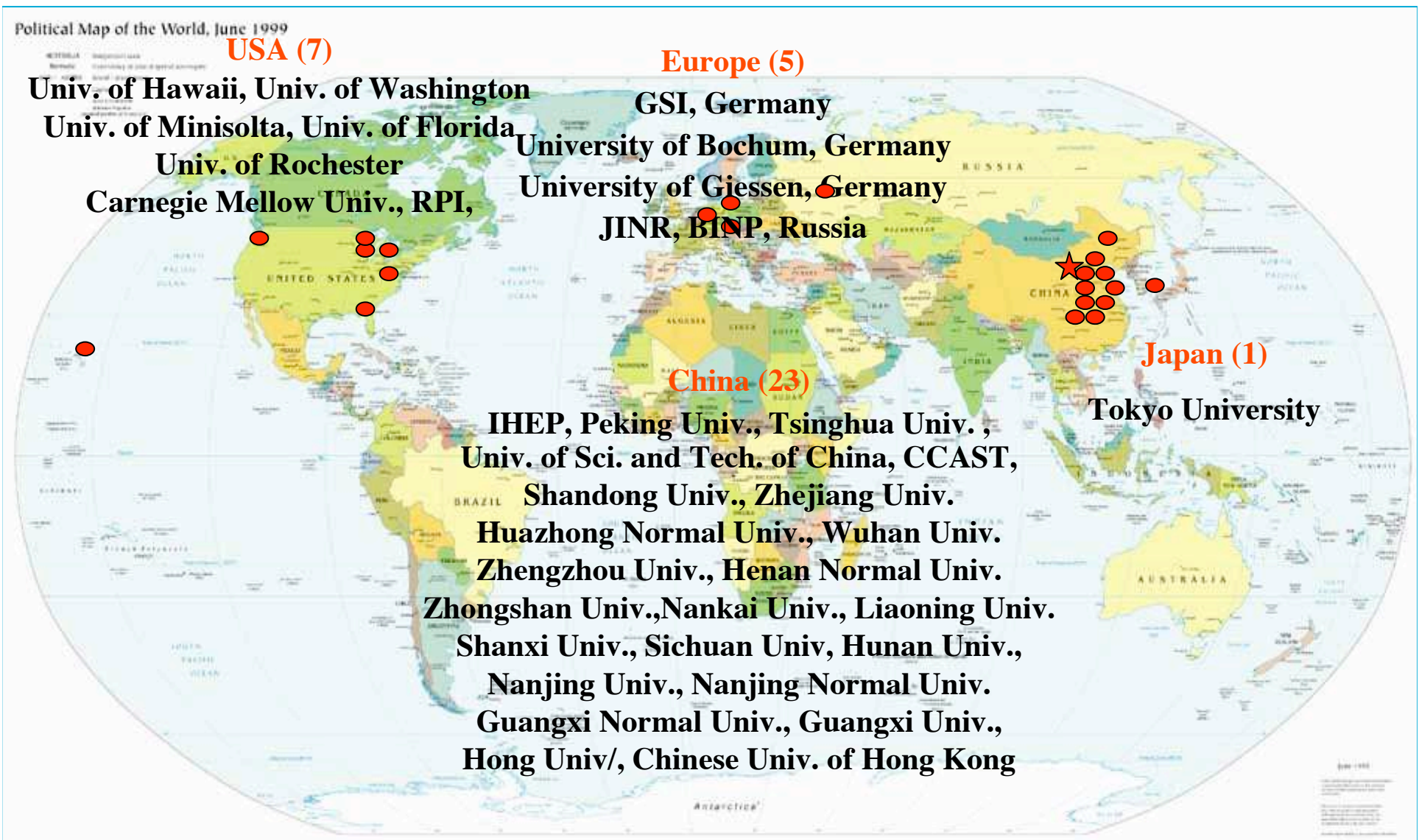
**1000 employees, ~ 650 physicists and engineers,  
300 PhD Students and post-doctors**

**Established at 1950, and became an independent institute at 1973 .**

# BEPCII/BESIII

- **BEPCII/BESIII is one of particle physics experiments.**
- **BEPCII luminosity  $10^{33} \text{ cm}^{-2}\text{s}^{-1}$  at 1.89 GeV, as well as major upgrade of BES, can insure an important roll in world HEP, especially in charm physics.**
- **Research topics of BESIII on BEPCII:**
  - **Precision measurement of CKM matrix elements**
  - **Precision test of Standard Model**
  - **QCD and hadron production**
  - **Light hadron spectroscopy**
  - **Charmonium physics**
  - **Search for new physics/new particles**

# BESIII collaboration



# Estimation of BESIII Data Volume

- Event rate is about 3000 Hz with 12KB raw event size.
- BESIII will be running on Jpsi for one year and on other particles for 4 years.
- Raw Data will be recorded to tape library at Computing Center with 36MB/s in full luminosity.

# Volumes of Various types of Data

<b>Data type</b>	<b>Size(TB)</b>	<b>Media</b>
<b>Raw</b>	<b>480 + 480 (backup)</b>	<b>Tape lib.</b>
<b>Rec.</b>	<b>2880 (twics/year)</b>	<b>Tape lib.</b>
<b>DST</b>	<b>80</b>	<b>Disk</b>
<b>MC-raw</b>	<b>480</b>	<b>Tape lib</b>
<b>MC-rec</b>	<b>1440 (once/year)</b>	<b>Tape lib</b>
<b>MC-DST</b>	<b>80</b>	<b>disk</b>
<b>Disk cache</b>	<b>400</b>	<b>disk</b>
<b>Total</b>	<b>5760TB (tape) + 560TB (disk)</b>	<b>Tape+disk</b>

# Requirement of Computing power

<b>Type of Computing</b>	<b>Number of Events</b>	<b>KSpecInt2K</b>
<b>Rec.</b>	<b><math>2 \times 10^{10}</math></b>	<b>964</b>
<b>MC-rec.</b>	<b><math>2 \times 10^{10}</math></b>	<b>350</b>
<b>Analysis</b>	<b><math>2 \times 10^{10}</math></b>	<b>500</b>
<b>Total</b>	<b><math>6 \times 10^{10}</math></b>	<b>1810</b>

It approximately is equal to 2000 Xeon 3.00 GHz CPUs, we estimate that CPU Utilization is 90%.

# Offline Software Environment

- Underlying framework
  - GAUDI (originally developed by LHCb)
- **Simulation**
  - GEANT4
- **Other external LIBs:**
  - CERNLIB, CLHEP, ROOT, AIDA, XercesC, GDML ...
- **Database: MySQL**
- **Software configuration management**
  - CMT and CVS
- **Computer language:** C++ for simulation/reconstruction/analysis and Java for web application
- **Operation system and compiler:**
  - SLC4/ gcc3.4.6



# Analysis within the Collaboration (1)

- Working groups
  - Physics groups are organized by the major physics topics
    - Charm Group, Charmonium Group, Light Hadron Spectroscopy Group, Tau and QCD Group
  - Offline Software Group
    - Responsible for the core software framework, simulation and reconstruction algorithms and various tools needed by physics analysis
  - Calibration Group
    - To provide calibration algorithms and maintain calibration constants in the database
  - Data Production Group
    - Responsible for real data processing and MC data production
  - Data Quality and Monitoring Group
    - Responsible for software validation and monitoring detector performance at online/offline level
  - Computing Group
    - Maintain the computing facilities such as computer farms, network, tape library, disks etc.

# Analysis within the Collaboration (2)

- Real data processing and MC data production
  - Real data and shared MC data are processed using the standard software release and configuration at IHEP.
- Analysis activities outside IHEP
  - Collaborators outside IHEP form different small physics groups using their own computing facilities.
  - Subsets of DST data are copied to sites outside IHEP.
  - MC production can also be done outside IHEP by setting up the identical computing environment.
  - Produced data at other sites can be sent back to IHEP for public use.

# Data Analysis Model (1)

- Event Data
  - RAW data: delivered by DAQ for reconstruction in byte stream format
  - Simulated Event Data: contain digits, hits and other MC truth information in ROOT format.
  - REC data : reconstructed data is event data written as output of reconstruction procedure in ROOT format.
  - DST Data: a reduced event representation suitable for analysis in ROOT format.
- Software releases
  - 2-4 major releases each year
  - When a new release is published, lot of checking will be done for validation using both MC and real data.
  - Release information like package tag version are stored in the database.

# Data Analysis Model (2)

- Monte Carlo simulation strategy
  - Software: detector simulation is based on Geant4 and piling-up of random trigger event as the background
  - MC mass production: Data Production Group is responsible for producing MC data for common use.
- Storage and access
  - Size of current real event: 17 KB for raw, 35 KB for REC and 5 KB for DST (noise and background level is still high now)
  - Size of MC event: 8 KB for raw, 40 KB for REC and 13 KB for DST
  - All the data are copied to disk for processing or physics analysis
  - During the mass production, the information about job and processing history are recorded in a book-keeping system.

# Data Conservation Prospects

- Current extent of data conservation at your experiment
- Which data should be conserved
  - Data about the experimental conditions and various parameters like calibration constants, detector geometry data etc
  - Event data (MC and real)
- At what level should data be conserved (raw, analysis level)
  - Raw data and DST data should be conserved when the experiment system becomes stable.
- Anticipated software headaches
  - The maintenance and improvement of software is a headache for being lack of man power.
- Timescale for readiness of preserved data
- Planned lifespan of preserved data (10, 20 years..)
  - About 15 years
- Where will the data / software be physically stored
  - Data in disk/tape and software source code in CVS at IHEP
  - Data and software releases in disk at other sites

# Collaboration prospects

- Planned end-date of your experiment
  - In about 10 years
- Expected scientific output (papers)
  - 20-30 papers each year
- Person power evolution in the next years
  - more than 100 persons, not exactly.
- Long term collaboration life/organisation
  - 10~15years
- Plans for open access
  - no plan in its life time.
- ...

**Thanks!**