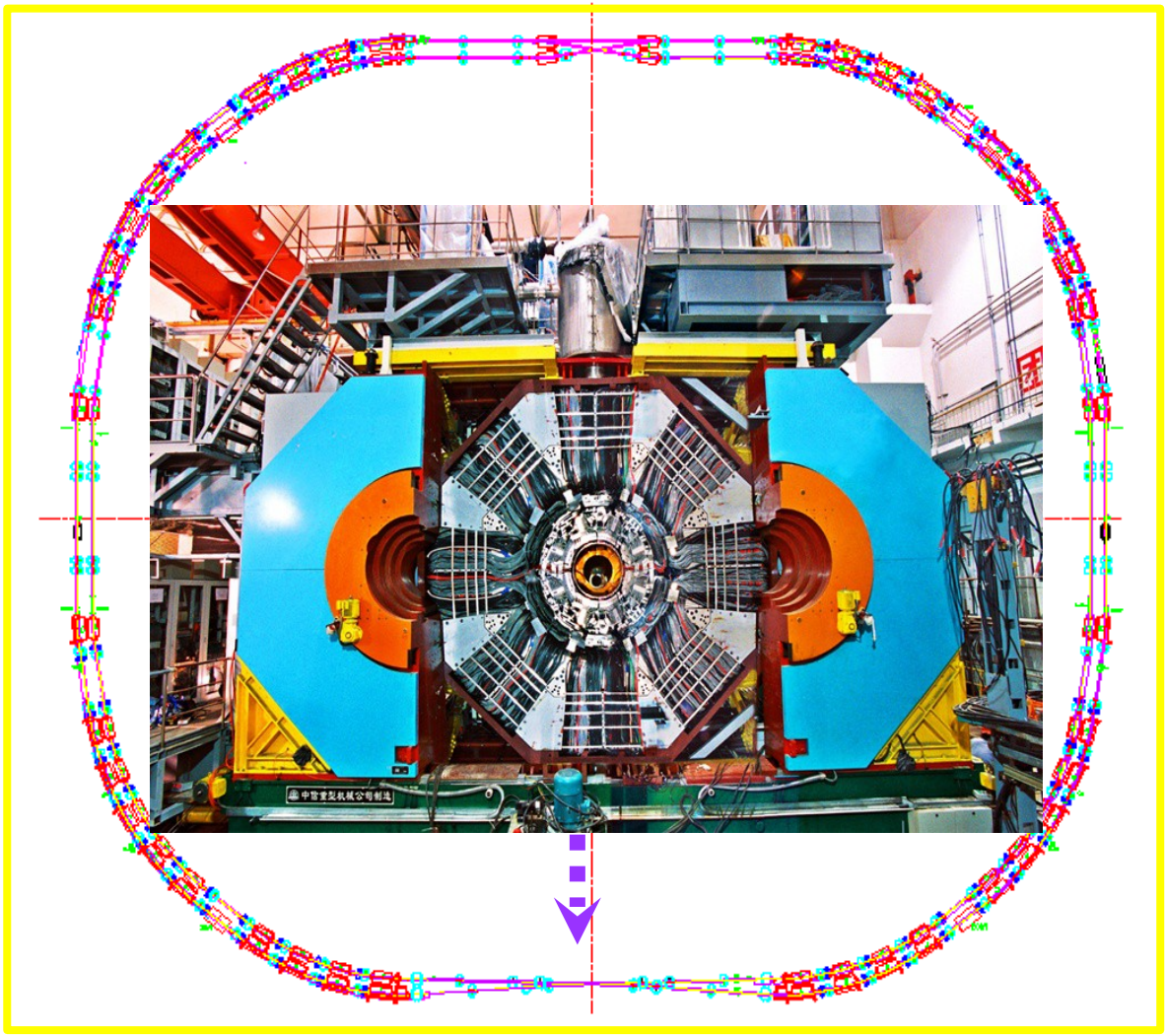
BES-III Distributed Computing

DENG Ziyan, LI Weidong, ZHANG Xiaomei *(IHEP CAS)*
LIN Lei *(Soochow Univ.)*
Caitriana NICHOLSON *(GUCAS)*
Alexey ZHEMCHUGOV *(JINR)*

The BES-III experiment

China, Germany, Italy, Japan, JINR, Korea, Netherlands, Pakistan, Russia, Sweden, Turkey, USA



Site:
IHEP CAS, Beijing, China

BEPC-II beam energy:
1.0-2.3 GeV

Design luminosity
1 x 10^{33}/cm^2/s @ψ(3770)

Achieved luminosity:
0.65 x 10^{33}/cm^2/s

Project timeline:

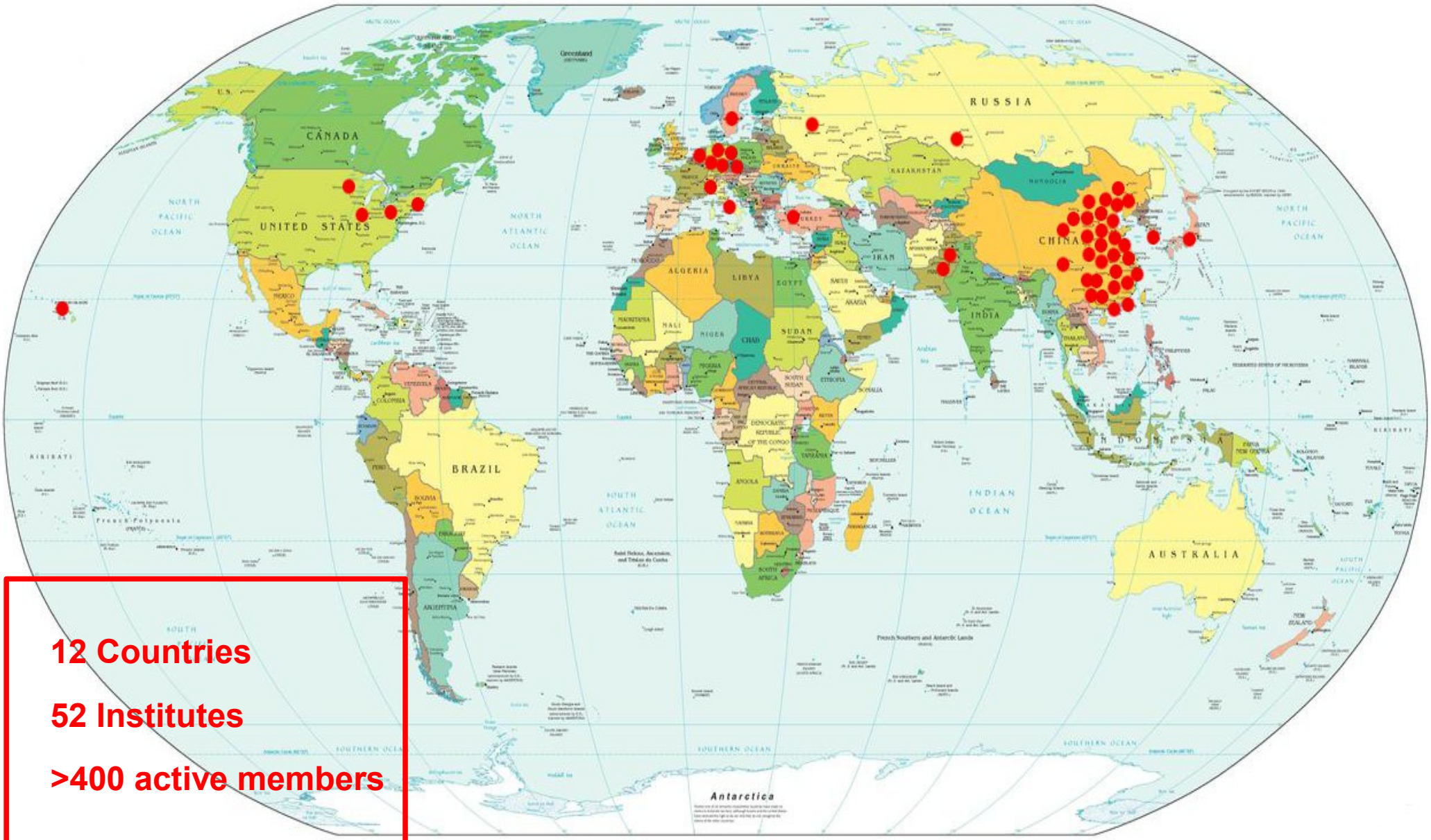
2004 – Start of BEPC upgrade

2006 - The detector installation

2007 - BEPCII/BESIII commissioning

2009 – Start of physics data taking

The BES-III Collaboration



12 Countries

52 Institutes

>400 active members

Data volume

Raw data		
J/Ψ	1.2×10^9 events	70 TB
Ψ'	$\sim 0.3 \times 10^9$ events	58 TB
Ψ''	2.9 fb^{-1}	169 TB
Ψ (4040)	0.5 fb^{-1}	15.5 TB
Total (with other data)	~ 0.4 PB	
DST	~100 TB	
MC	~20 TB	

World largest samples

Not much for LHC, but quite a lot for a single computing center

The problem

Grid could be a solution, but ...

- Medium-scale experiment
  - *WLCG tools are too complicated*
  - *Manpower is limited*
- Experiment specific middleware is unique and should be developed by ourselves
- Many sites do not participate in LHC => no experience with grid, no LCG infrastructure exists
- Network connectivity between sites is low

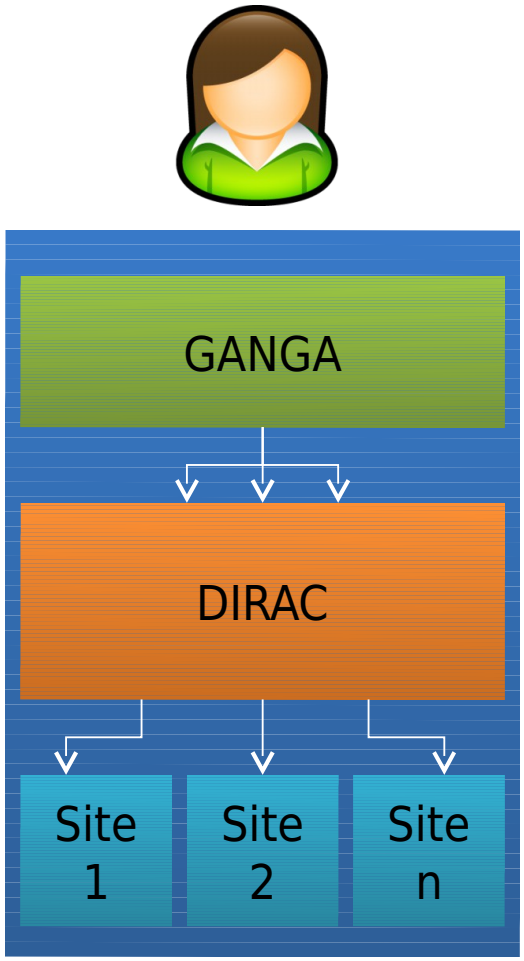
BES-III grid needs to be easy to set up, maintain and use, reliable and capable to work with low-bandwidth and unstable networks

The BES-III Distributed Computing Model

- All experimental data processed at IHEP. Remote sites participate in the MC production and data analysis only.
- Three operation models (depends on site capabilities)
  - *MC simulation at remote sites. Resulting data copied back to IHEP. MC reconstruction runs at IHEP (for sites with no SE or with only a small one)*
  - *MC production and reconstruction at remote sites. Resulting data copied back to IHEP.*
  - *Copying DST from IHEP and other sites and analyzing using local resources*
- Distributed analysis – maybe in future?

Job management

- **DiracGrid** is chosen as a solution
- **Components:**
  - GANGA for job submission and management
    - GangaBoss plug-in has been written to support BES software
  - DIRAC for running distributed computing jobs
    - DIRAC server is running at IHEP with clients at remote sites
    - Problem: Not all local resource management systems supported by DIRAC; new plugins may need to be developed
  - CVMFS (CERN VM File System) for deploying BOSS on target sites
    - Clients running at distributed sites can load BOSS version from server at IHEP



DIRAC Job Management

Data management

- DFC [DIRAC File Catalog] is used as the catalog solution
  - *metadata catalog*
  - *file catalog*
  - *dataset catalog (static or dynamic datasets)*
- FTS [File Transfer Service] is used for data transfer
  - *version 2.2.8 can handle both gsiftp and srm transfers*
  - *configured to work without BDII*
- BES-III Advanced Data manaGER takes care of the BES-III specific details
  - *User interface*
  - *Python API to be used in the job management system*
  - *Transfer of datasets between sites*

(noun) "large burrowing animals, with strong claws..."
(Collins English Dictionary)
so it can dig in and find that data ;-)

DFC vs AMGA

BES-III Metadata schema implemented in AMGA and DFC:

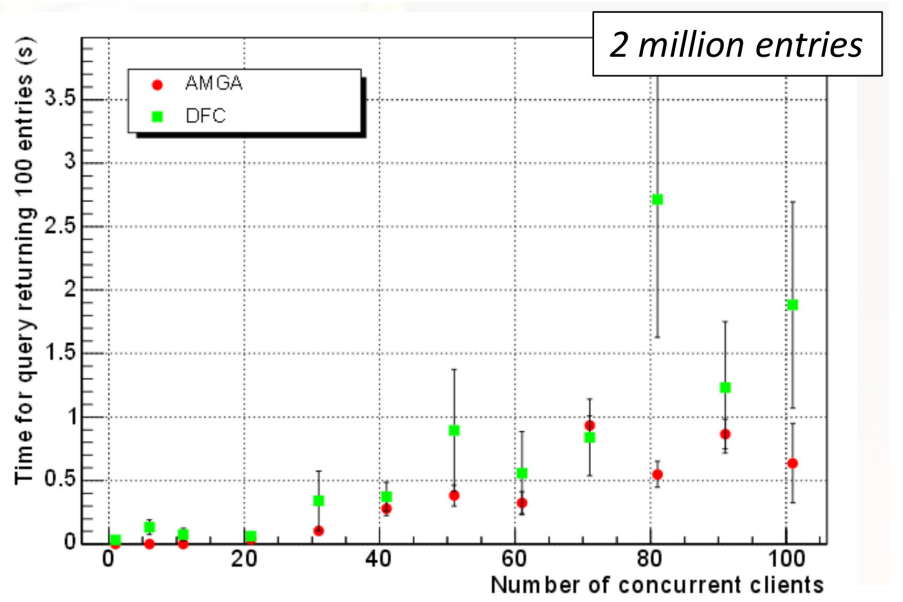
- MySQL backend Optimised configuration:
  - 8 DFC instances, max. 50 threads / instance
  - AMGA max. 140 processes
- Current BESIII data set used (~2x10^5 files)

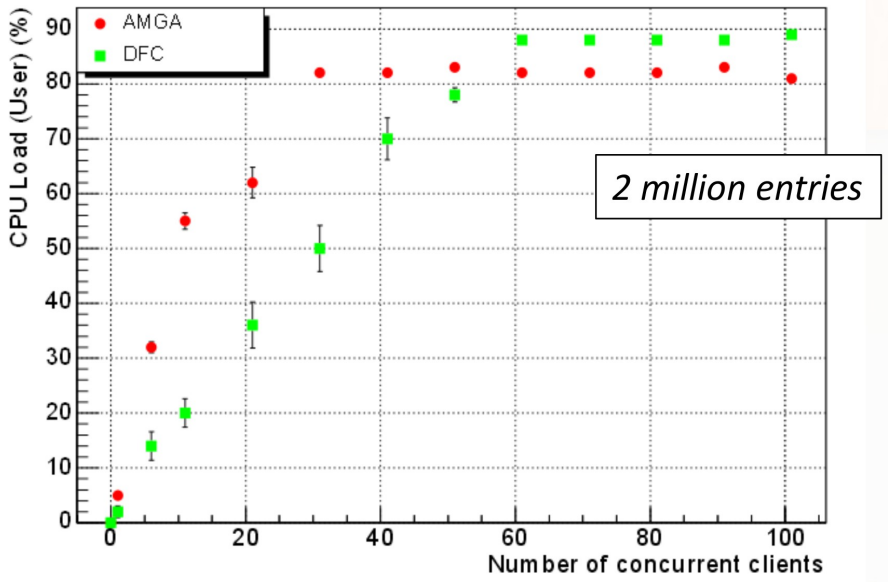
Results:

- With low number of clients, AMGA queries ~10x faster
- With high number of clients, query times approx. equal
- DFC CPU usage rises more slowly

Both give acceptable performance, but DFC meets more of BESIII requirements:

- file, metadata and dataset catalog functions at the same catalog;
- already part of DIRAC;
- easily integrated with GANGA.





Current status and plans

- Currently BES-III grid includes IHEP CAS, GUCAS and JINR.
- VO 'BES' is created and VOMS is set up at IHEP for user authorization
- CVMFS is used to distribute the experiment software
- DIRAC server is running at IHEP with clients at GUCAS and JINR sites. Tests of MC production chain are under way.
- FTS is set up at JINR. Now making tests and optimization with simple configuration of transfer channels
- BADGER prototype is being developed
- Five more sites (Univ. of Minnesota, USTC, Shandong Univ., Peking Univ., Wuhan Univ.) are expected to join BES-III grid before 2013

Summary

- The BES-III experiment is running since 2008 and currently it is the best source of data in τ-charm domain.
- Increasing data volume demands increasing computing power.
- Grid solution for medium scale experiments is wanted.
- BES-III grid is being constructed, based on the DIRAC infrastructure with experiment-specific data management
- Working prototype is set up already to join computing resources of IHEP CAS, GUCAS and JINR.
- Expect use of the BES-III grid in data production still this year.