PanDA in ASGC

22 Apr, 2016

Shao-Ting CHENG

larry.cheng@twgrid.org

https://dicos.grid.sinica.edu.tw

DiCOS

- Distributed Cloud Operating System.
- Based on PanDA, Rucio, pilot and GSI proxy.
- WebUI, HPC application support, automatic deployment, Restful API, DiCOSBox were developed and integrated
  - https://dicos.grid.sinica.edu.tw
- Supporting wider user communities beyond WLCG by federating distributed cloud resources close to the users
- Multiple VO by the same infrastructure: AMS, Bioinformatics, Physics, etc.
- Deployed at Institutes in Academia Sinica, and collaboration sites in Taiwan (NCU, NDHU, NCHC), CERN and U. Chicago
- 20K CPU cores, 20 PB disk

PanDA server

- Some our modifications
  - AdderAMSPlugin.py
  - SetupperAMSPlugin.py
  - Others in several scripts.
- MariaDB
  - We migrated from OracleDB to MariaDB on Aug, 2015
  - MariaDB is based on Mysql and PanDA also support Mysql. So everything goes fine.
  - But, there is “time out” problem when massive query to the DB.
    - We are still tuning the performance.
    - Plan to copy some tables to ElasticSearch for the searching, accounting and monitoring by the uses.

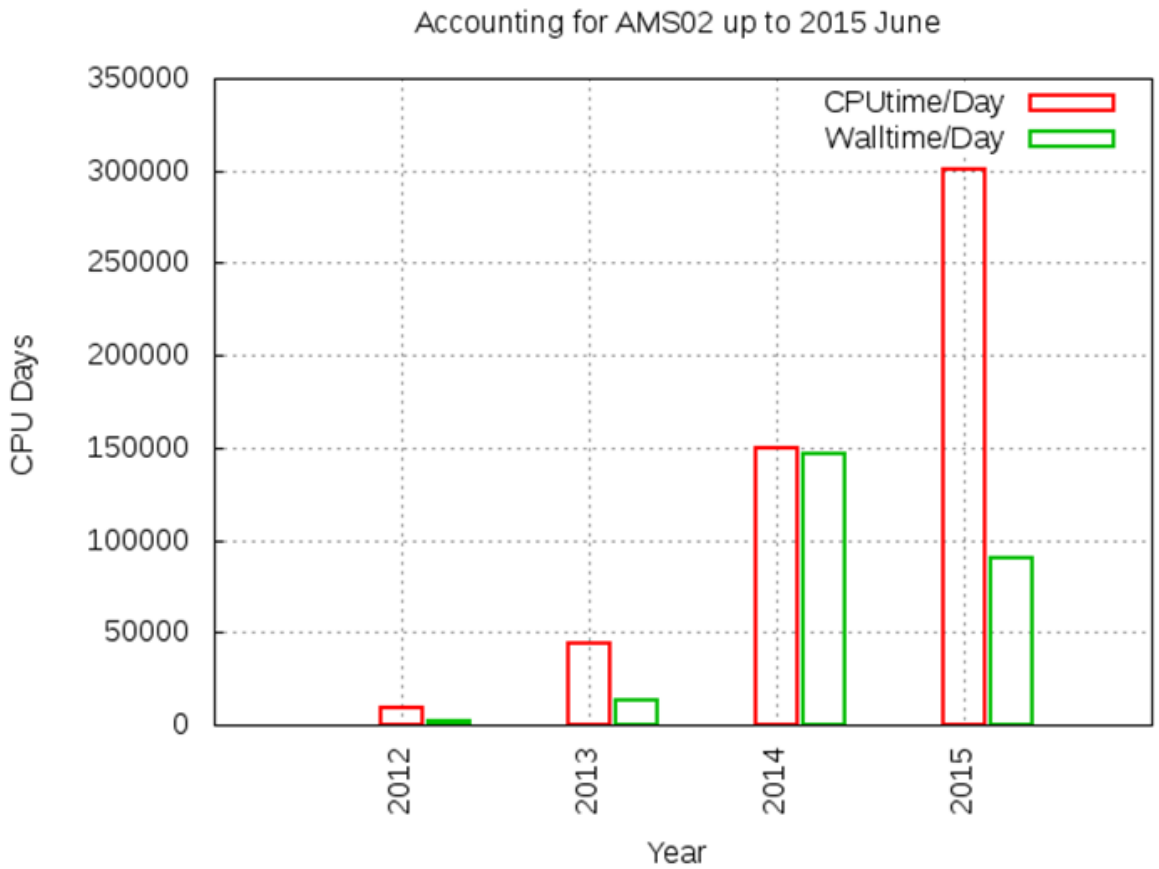
Pilot

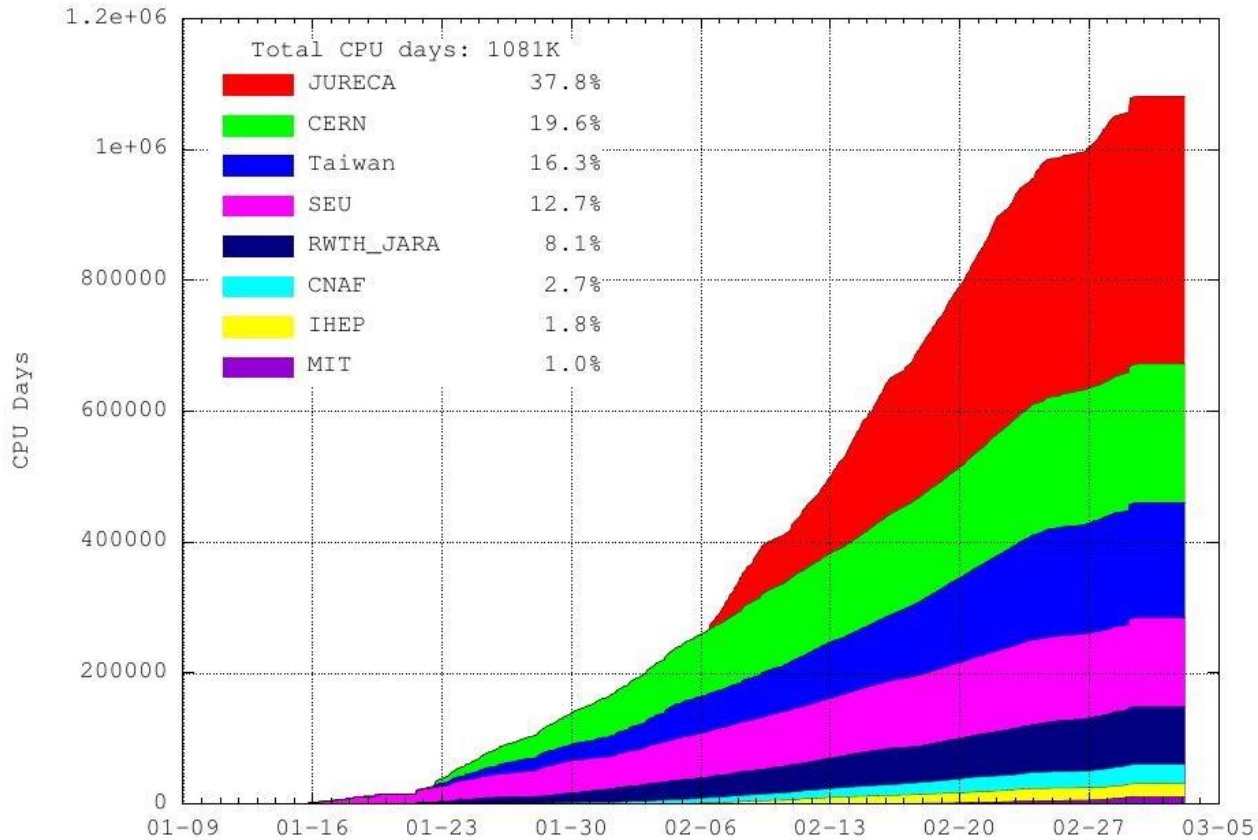
- We are now at version 64.0
- Some our modifications
  - AMSTaiwanExperiment.py
  - AMSTaiwanSiteInformation.py
  - Several TW specific statements in the script.
- I’m now re-organizing and revising the modification in more general way.
  - Hope to contribute to future Pilot release

AMS

- Alpha Magnetic Spectrometer (AMS) is the first user.
- ASGC acts as one of the primary analysis & data production centers of AMS since 2012.
- 600K+ CPU-Day running time and petabyte data in/out served by DiCOS.

AMS


Accounting for AMS02 up to 2015 June
CPUtime/Day
Walltime/Day
CPU Days
350000
300000
250000
200000
150000
100000
50000
0
2012
2013
2014
2015
Year

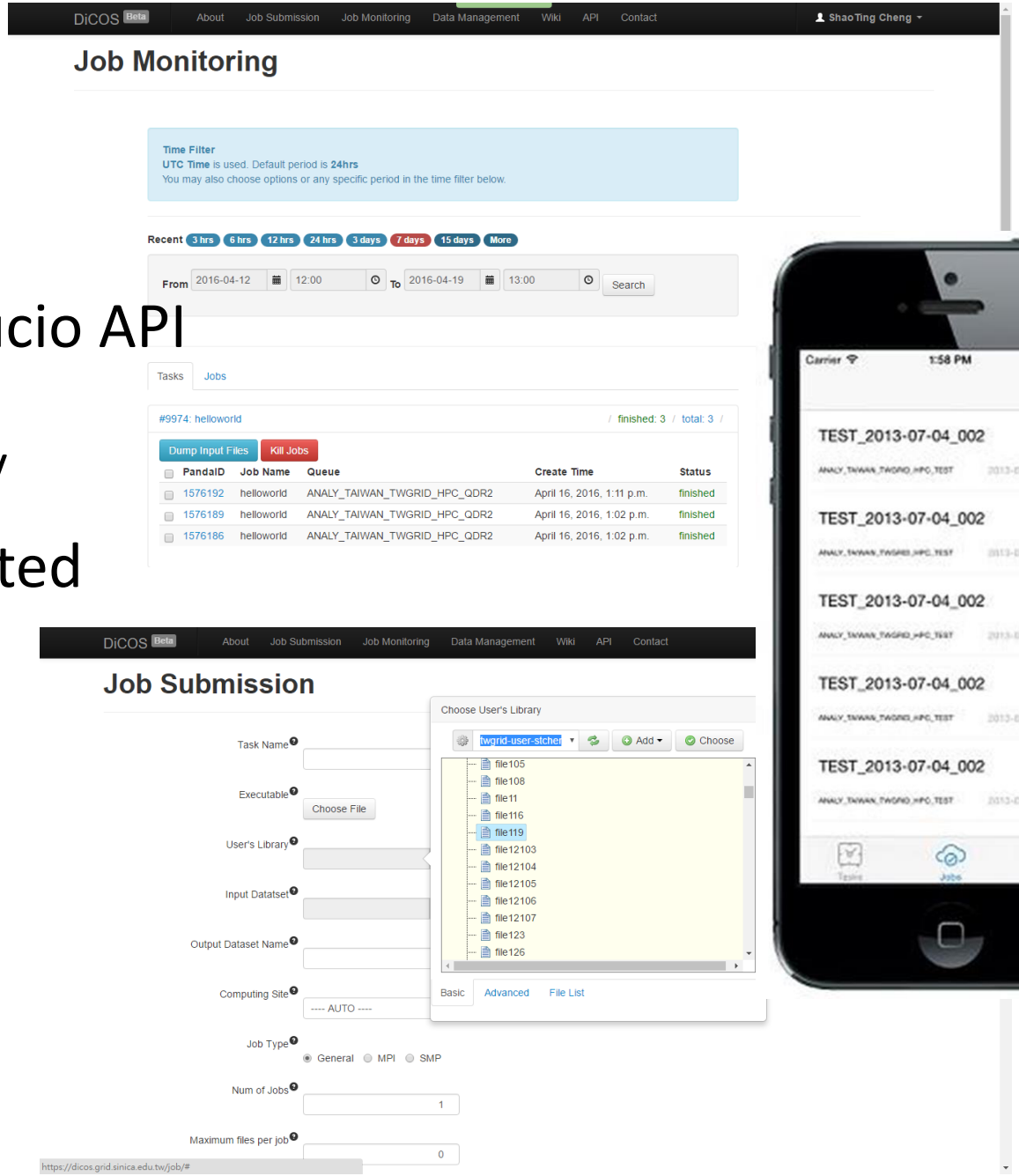

Total CPU days: 1081K
JURECA 37.8%
CERN 19.6%
Taiwan 16.3%
SEU 12.7%
RWTH_JARA 8.1%
CNAF 2.7%
IHEP 1.8%
MIT 1.0%
CPU Days
1.2e+06
1e+06
800000
600000
400000
200000
0
01-09
01-16
01-23
01-30
02-06
02-13
02-20
02-27
03-05

Bioinformatics Applications

- Cooperate with University of Chicago (on Biomedical Data Common) and Northwestern University from 2015.
- Started from RNA-seq data analysis and genome data analysis.
- Docker container is integrated with DiCOS for consistent and customized application environment.
- Until now, DiCOS contributes about half CPU time of the BDC-TP (Biomedical Data Commons-Technical Pilot) project since 2015.
- RNA information is the sensitive data, so the safety and the protected sharing rule is the big issue we are working on.

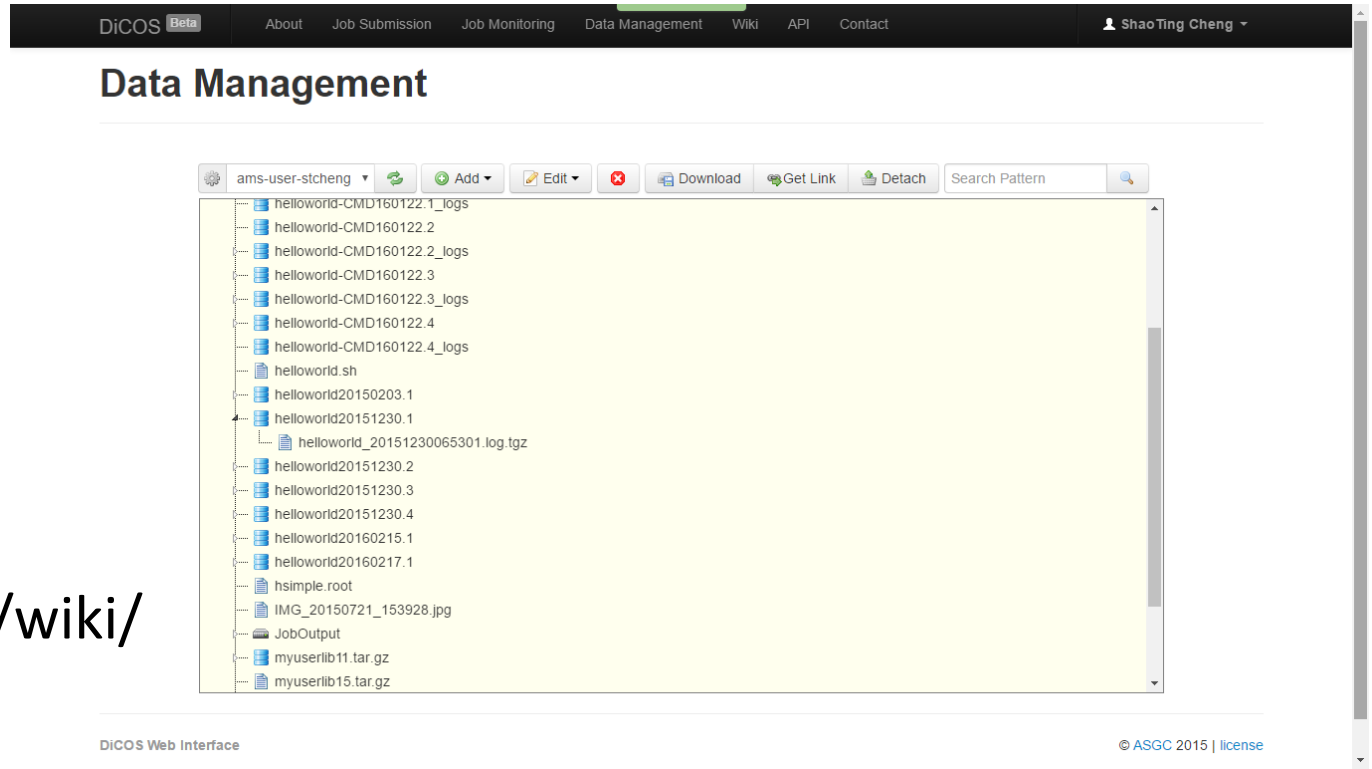
Web user interface & command-line tool

- Based on Panda client tool and Rucio API
  - Provide our own Restful API.
  - https://dicos.grid.sinica.edu.tw/api/
- Single sign-on & GSI proxy integrated
- Basic submission web page.
  - The easiest way to submit a job.
- Job monitoring
  - To see the job status
  - There is also smart phone interface



Web user interface & command-line tool

- Visualization of data management
  - Easily drag & drop
  - Upload & download
  - Root viewer
  - Online editor.
- Command-line tool
  - For the advantage usage,
    like massive jobs, variable input
  - https://dicos.grid.sinica.edu.tw/wiki/



Acknowledgement

- Thanks everyone who ever helped us: PanDA team, BigPanDA project, RUCIO team, etc.
- Especially thanks Paul and Tadashi for the instructions for PanDA server and pilot.