

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

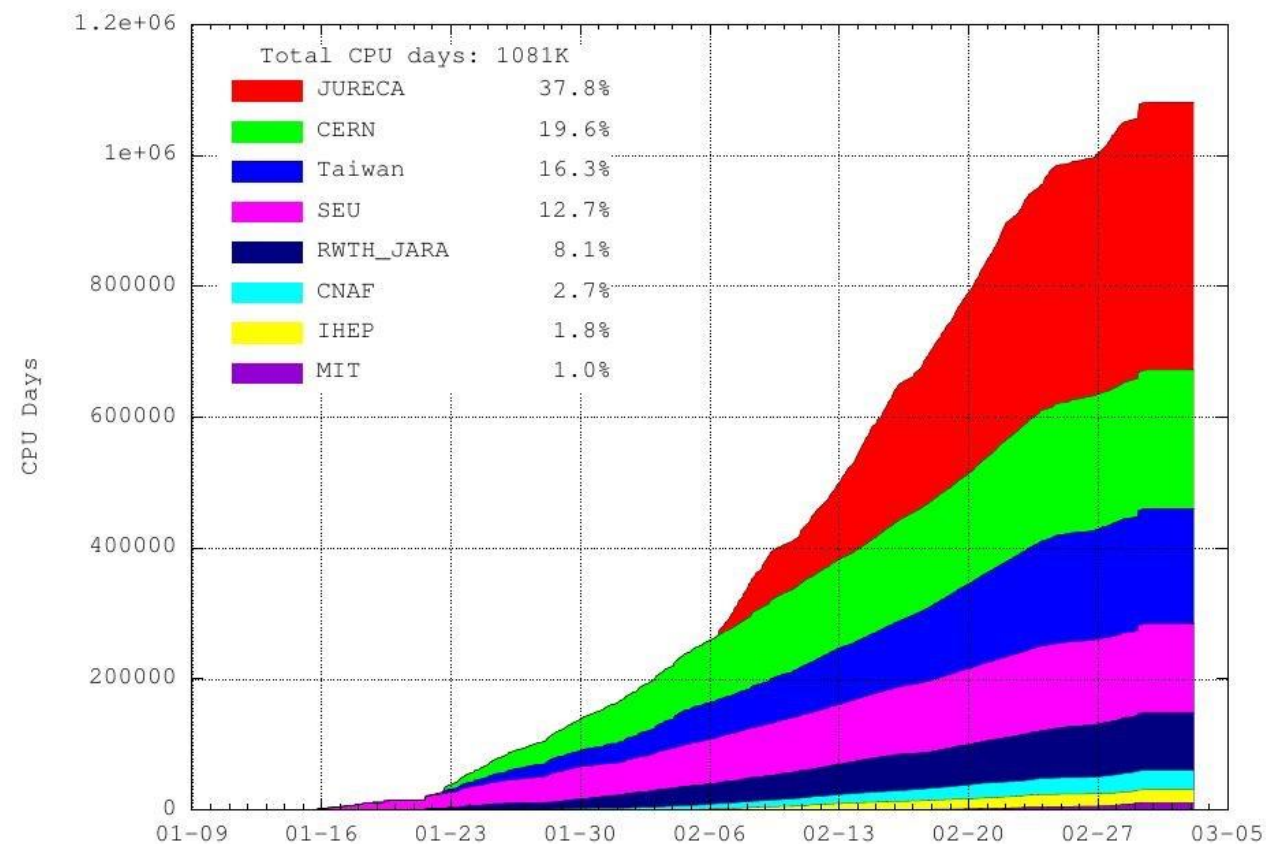
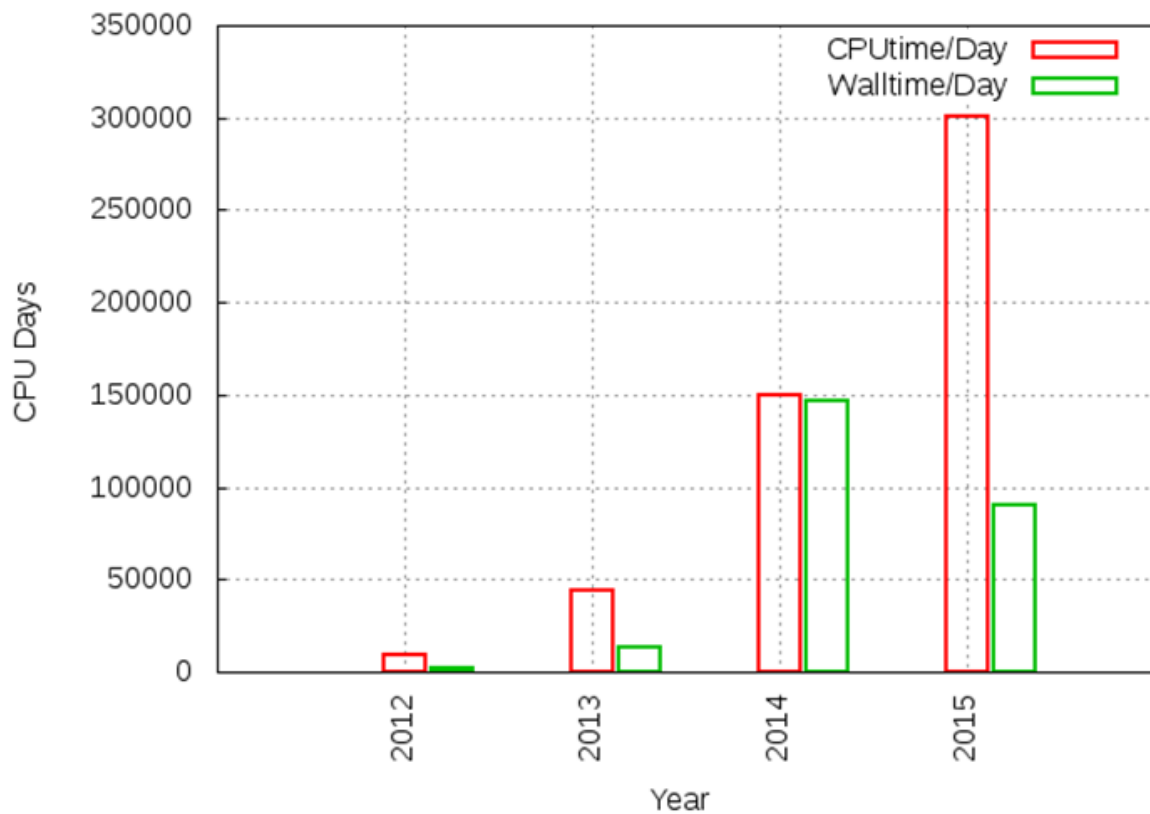
- We are now at version 64.0
- Some our modifications
 - AMSTaiwanExperiment.py
 - AMSTaiwanSiteInformation.py
 - Several TW specific statements in the script.
- I'm now re-organizing and revising the modification in more general way.
 - Hope to contribute to future Pilot release

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



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.

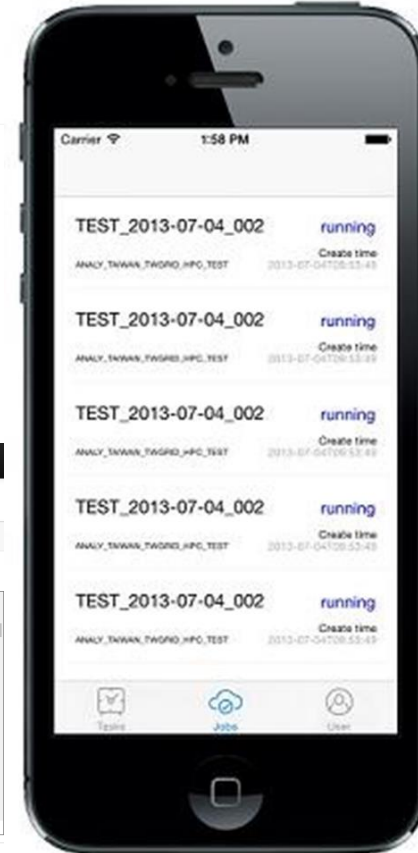
Web user interface & command-line tool

- 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

The screenshot shows the 'Job Monitoring' page of the DiCOS web interface. At the top, there is a navigation bar with links for 'About', 'Job Submission', 'Job Monitoring', 'Data Management', 'Wiki', 'API', and 'Contact'. The user 'ShaoTing Cheng' is logged in. Below the navigation bar, there is a 'Time Filter' section indicating that UTC Time is used and the default period is 24 hours. A 'Recent' filter bar shows options for 3 hrs, 6 hrs, 12 hrs, 24 hrs, 3 days, 7 days, 15 days, and More. A date range selector is set from 2016-04-12 12:00 to 2016-04-19 13:00. The main content area shows a table of jobs for the task '#9974: helloworld', with 3 finished jobs out of a total of 3. The table columns are PandaID, Job Name, Queue, Create Time, and Status. The jobs listed are:

PandaID	Job Name	Queue	Create Time	Status
1576192	helloworld	ANALY_TAIWAN_TWGRID_HPC_QDR2	April 16, 2016, 1:11 p.m.	finished
1576189	helloworld	ANALY_TAIWAN_TWGRID_HPC_QDR2	April 16, 2016, 1:02 p.m.	finished
1576186	helloworld	ANALY_TAIWAN_TWGRID_HPC_QDR2	April 16, 2016, 1:02 p.m.	finished

The screenshot shows the 'Job Submission' page of the DiCOS web interface. It features a form for submitting a job with fields for Task Name, Executable (with a 'Choose File' button), User's Library, Input Dataset, Output Dataset Name, Computing Site (set to '--- AUTO ---'), Job Type (General, MPI, or SMP), Num of Jobs (set to 1), and Maximum files per job (set to 0). A 'Choose Users Library' dialog box is open, showing a list of files from the 'twgrid-user-stcher' library, including file105, file108, file11, file116, file119, file12103, file12104, file12105, file12106, file12107, file123, and file126. The dialog also has 'Add' and 'Choose' buttons.



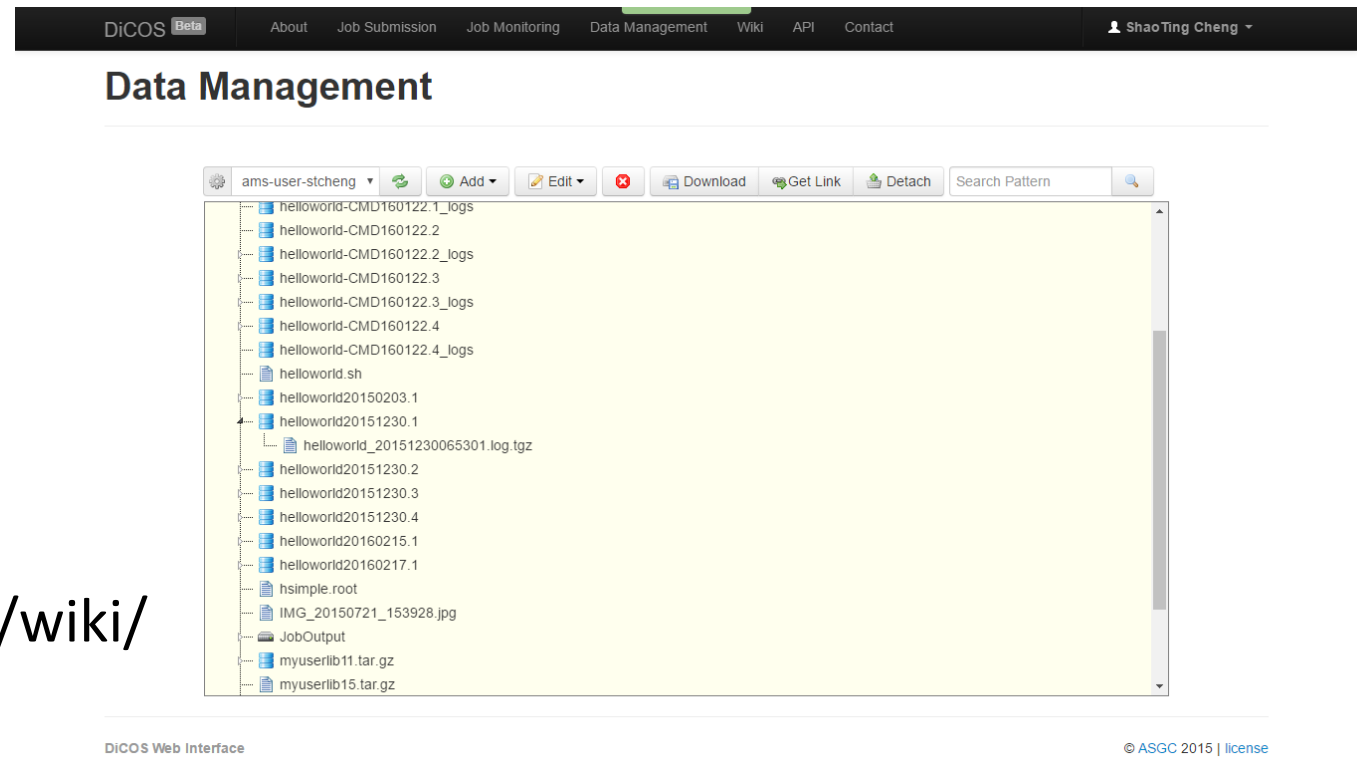
Web user interface & command-line tool

- 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.