

# Table of Contents

<b>User guide for T3_TW_NTU_HEP usage.....</b>	<b>1</b>
Grid Access.....	1
Get a Grid certificate and the registration to CMS VO.....	1
Setup local Environment and prepare user analysis code.....	1
CRAB setup.....	2
Locate the dataset and prepare CRAB submission.....	2
Data selection.....	2
CRAB configuration.....	2
Configuration parameters.....	2
Summit jobs as normal cms role.....	3
Copy the output to an SE.....	3
Publish your result in DBS.....	3
crab.cfg for this Tutorial.....	4
Run Crab.....	5
Job Creation.....	5
Job Submission.....	6
Job Status Check.....	7
Job Output Retrieval.....	8
Where to find more on CRAB.....	8
T3_TW_NTU_HEP specific info.....	8
UI.....	9
CE.....	9
SE.....	9
Useful resources.....	9
Useful links.....	9
Slides and scripts for usage tutorial.....	9

# User guide for T3\_TW\_NTU\_HEP usage

## Grid Access

Please refer to the CMS Twiki workbook: <https://twiki.cern.ch/twiki/bin/view/CMS/WorkBookRunningGrid>  
WorkBookRunningGrid (Following descriptions are shamelessly copied and re-written for  
T3\_TW\_NTUHEP)

## Get a Grid certificate and the registration to CMS VO

People who has "normal" CERN computer accounts should go to <https://ca.cern.ch> directly and get their certificates following the instructions on the web page. For remote users, one should go to ASGC CA: [http://ca.grid.sinica.edu.tw/certificate/request/request\\_user\\_cert.html](http://ca.grid.sinica.edu.tw/certificate/request/request_user_cert.html) and following the instructions. A signature from regional RA (NTU: Pao-ti Chang) is needed to complete the application.

To get a certificate from CERN CA and register to CMS VO, you can find detailed instruction here: <https://twiki.cern.ch/twiki/bin/view/CMS/WebHome?topic=SWGGuideLcgAccess>. If you get a certificate from another Certification Authority, the procedure to register to CMS VO with your certificate, should be the same.

## Setup local Environment and prepare user analysis code

Users of Taiwan can either use the login servers at CERN, lxplus, or make use of our T3 login servers: currently we have ntugrid1 and ntugrid3 are SLC5 and ntugrid5 is SLC6.

```
ssh ntugrid3.phys.ntu.edu.tw
```

Users who wants to login to T3 NTU UI should apply accounts with me or Kai-feng. Once login, the CMSSW environment is setup automatically. You don't need to source for LCG UI. To setup a basic CMSSW environment, one can source from CVMFS by:

```
source /cvmfs/cms.cern.ch/cmsset_default.sh
```

Install CMSSW project in a directory of your choice. In this case we create a "Tutorial" directory:

```
mkdir Tutorial
cd Tutorial
cmsrel CMSSW_4_1_9
cd CMSSW_4_1_9/src/
cmsenv
```

**\*CVS has been deprecated by CMS\*** Get from CVS the configuration file we are going to use for this tutorial:

```
# cvsroot CMSSW
# export CVSROOT=:gserver:cmssw.cvs.cern.ch:/local/repos/CMSSW
export CVSROOT=:pserver:anonymous@cmscvs.cern.ch:/cvs_server/repositories/CMSSW
cvs login      # '98passwd'
cvs co -r $CMSSW_VERSION PhysicsTools/PatAlgos/test/patLayer1_fromAOD_full.cfg.py
cd PhysicsTools/PatAlgos/test/
```

## CRAB setup

Setup on lxplus:

In order to setup and use CRAB from any directory, source the script `crab.(c)sh` located in `/afs/cern.ch/cms/ccs/wm/scripts/Crab/`, which always points to the latest version of CRAB. (this is the only script you need to source for running CMSSW) After the source of the script it's possible to use CRAB from any directory (typically use it on your CMSSW working directory).

```
source /afs/cern.ch/cms/ccs/wm/scripts/Crab/crab.sh
```

Alternatively, you can use a locally installed version: (maybe soon out-dated)

```
source /home/Crab/CRAB_2_8_1/crab.sh
```

## Locate the dataset and prepare CRAB submission

In order to run our analysis over a whole dataset, we have to find first the data name and then put it on the crab configuration file.

### Data selection

To select data you want to access, use the **DAS** web page where available datasets are listed [DAS Data Discovery](#). For this tutorial we'll use :

```
/RelValZEE/CMSSW_2_1_9_STARTUP_V7_v2/GEN-SIM-DIGI-RAW-HLTDEBUG-RECO
```

## CRAB configuration

Modify the CRAB configuration file `crab.cfg` according to your needs: a fully documented template is available at `$(CRABPATH)/full_crab.cfg`, a template with essential parameters is available at `$(CRABPATH)/crab.cfg`. For guidance, see the list and description of configuration parameters. For this tutorial, the only relevant sections of the file are `[CRAB]`, `[CMSSW]` and `[USER]`.

### Configuration parameters

The list of the main parameters you need to specify on your `crab.cfg`:

- **pset**: the CMSSW configuration file name;
- **output\_file**: the output file name produced by your pset; if in the CMSSW pset the output is defined in `TFileService`, the file is automatically handled by CRAB, and there is no need to specify it on this parameter;
- **datasetpath**: the full dataset name you want to analyze;
- **total\_number\_of\_events**, **number\_of\_jobs**, **events\_per\_job**: you need to specify 2 of these parameters:
  - ◆ specify the *total\_number\_of\_events* and the *number\_of\_jobs*: this will assign to each job a number of events equal to  $total\_number\_of\_events/number\_of\_jobs$ ;
  - ◆ specify the *total\_number\_of\_events* and the *events\_per\_job*: will assign to each job the number *events\_per\_job* and will calculate the number of jobs by

*total\_number\_of\_events/events\_per\_job*;

◆ or you can specify the *number\_of\_jobs* and the *events\_per\_job*...;

- **return\_data**: this can be 0 or 1; if it is one you will retrieve your output files to your local working area;
- **server\_name**: the name of the server where you want to send your jobs;
- **scheduler**: the name of the scheduler you want to use;
- **jobtype**: the type of the jobs.

## Summit jobs as normal cms role

**Make sure the file access permission of your certificate under .globus/**

There's no special roles needed to run jobs on T3 NTU, but we reserve the rights to kill your jobs at any time.



```
$ voms-proxy-init -voms cms
```

(You can safely ignore the error about "Cannot find file or dir: /home/xxxxx/.globus/vomses")

To double-check the user role, one can issue the following command:

```
$ voms-proxy-info --all
```

## Copy the output to an SE

The change to copy the output to an existing **Storage Element** allows to bypass the output size limit (10MB) constraint.

## Publish your result in DBS

To process the private skims using CRAB, one need to publish the skim on DBS. A new set of private DBSs are set up recently with writing permissions. Details can be found here. For all registered CMS users, there are two DBS URLs available:

```
https://cmsdbprod.cern.ch:8443/cms_dbs_ph_analysis_01_writer/servlet/DBSServlet
https://cmsdbprod.cern.ch:8443/cms_dbs_ph_analysis_02_writer/servlet/DBSServlet
```

You need "/cms" role in your DN. You can check it with the following command:

```
$ voms-proxy-info -all
subject   : /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=yuanchao/CN=596728/CN=Yuan Chao/CN=proxy
issuer    : /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=yuanchao/CN=596728/CN=Yuan Chao
identity  : /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=yuanchao/CN=596728/CN=Yuan Chao
type      : proxy
strength  : 1024 bits
path      : /tmp/x509up_u13771
timeleft  : 11:59:56
=== VO cms extension information ===
VO        : cms
subject   : /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=yuanchao/CN=596728/CN=Yuan Chao
issuer    : /DC=ch/DC=cern/OU=computers/CN=voms.cern.ch
attribute : /cms/twcms/Role=NULL/Capability=NULL
attribute : /cms/Role=NULL/Capability=NULL
timeleft  : 11:59:55
uri       : voms.cern.ch:15002
```

Only the outputs of POOL output module can be publish. (not rootuples!) In your crab.cfg, the followings are needed:

```
[USER]
...
publish_data=1
publish_data_name=<publish_data_name>
dbs_url_for_publication = https://cmsdbsprod.cern.ch:8443/cms_dbs_ph_analysis_01_writer/servlet/D

publish_with_import_all_parents=0
```

Be careful on the publish name, it should not contain '/USER'. Or CRAB will rise an error. ([publish\_with\_import\_all\_parents] may need to be set as '0' to walk around a possible bug)

After the skim jobs are done, one can write the publish info by issuing:

```
$ crab -publish
```

Note the CRAB output about the published dataset name. Your user ID (hypernews ID) and a hash code will be added into the publish\_data\_name. You need this for the processing jobs.

```
...
2009-08-19 06:30:41,160 [INFO] --->>> End files publication
2009-08-19 06:30:41,174 [INFO] --->>> Check data publication: dataset /RelValMinBias/yuanchao-Mi
DBS url https://cmsdbsprod.cern.ch:8443/cms_dbs_ph_analysis_01_writer/servlet/DBSServlet

2009-08-19 06:30:41,930 [INFO] You can obtain more info about files of the data
set using: crab -checkPublication -USER.dataset_to_check=/RelValMinBias/yuanchao-MinBias_MC_31X_V
```

In this example, you need to put

```
"/RelValMinBias/yuanchao-MinBias_MC_31X_V3-v1_RAW-5ef010080a33c624ab2cff6acc1de0c6/USER"
as your [datasetpath] in crab.cfg. Also, you need to specify the [dbs_url] to
"https://cmsdbsprod.cern.ch:8443/cms_dbs_ph_analysis_01_writer/servlet/DBSServlet".
```

The publication of the produced data to a DBS allows one to re-run over the produced data that has been published. Below the instruction to follow and here the link to the how to. You have to add to the Crab configuration file more information specifying where you want to copy the results, the data name to publish and the DBS url instance where to register the output results.

```
[CMSSW]
datasetpath=<primarydataset>/<publish_data_name>/USER
### DBS/DLS options
dbs_url = <dbs_url_for_publication>
```

## crab.cfg for this Tutorial

You can find more details on this at the corresponding link on the Crab FAQ page.

The CRAB configuration file (crab.cfg) should be located at the same location as the CMSSW parameter-set to be used by CRAB with the following content:

```
[CMSSW]
total_number_of_events=3000
number_of_jobs=10
pset=patLayer1_fromAOD_full.cfg.py
datasetpath=/RelValZEE/CMSSW_2_1_9_STARTUP_V7_v2/GEN-SIM-DIGI-RAW-HLTDEBUG-RECO
output_file=PATLayer1_Output.fromAOD_full.root

[USER]
```

```
return_data=0
email=yourEmailAddressHere@cern.ch

copy_data = 1
storage_element = T3_TW_NTU_HEP

## To store the output at CERN Castor, comment out "storage_element" and un-comment the following
# storage_element = srm-cms.cern.ch
# storage_path = /srm/managerv2?SFN=/castor/cern.ch/user/(letter)/(UID)

user_remote_dir = MyFirstTutorialResults

#publish_data = 1
#publish_data_name = MyFirstTutorialResults
#dbs_url_for_publication = http://grid-dcms1.physik.rwth-aachen.de:8081/cms_dbs_prod_test/servlet

[CRAB]
scheduler=glite
jobtype=cmssw
```

## Run Crab

Once your `crab.cfg` is ready and the whole underlying environment is set up, you can start running CRAB. CRAB supports command line help which can be useful for the first time. You can get it via:

```
crab -h
```

in particular there is a [\\*HOW TO RUN CRAB FOR THE IMPATIENT USER\\*](#) section where the base commands are reported.

## Job Creation

The job creation checks the availability of the selected dataset and prepares **all** the jobs for submission according to the selected job splitting specified in the `crab.cfg`

The creation process creates a CRAB project directory (default: `crab_0_date_time`) in the current working directory, where the related crab configuration file is cached for further usage, avoiding interference with other (already created) projects

CRAB allows the user to chose the project name, so that it can be used later to distinguish multiple CRAB projects in the same directory.

```
crab -create
```

which first times could ask for proxy/myproxy passwords and that should produce a similar screen output like:

```
[ui02] /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test > crab -create
crab. crab (version 2.4.3) running on Thu Dec 18 13:38:01 2008
```

```
crab. Working options:
  scheduler          glite
  job type           CMSSW
  working directory  /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_
```

```
crab. Downloading config files for WMS: https://cmsweb.cern.ch/crabconf/files/glite_wms_CERN.conf
Cannot find file or dir: /home/yuanchao/.glite/vomses
Enter GRID pass phrase:
```

## YuanChaoT3NTUHEPUsage < Main < TWiki

```
Your identity: /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=yuanchao/CN=596728/CN=Yuan Chao
Creating temporary proxy ..... Done
Contacting lcg-voms.cern.ch:15002 [/DC=ch/DC=cern/OU=computers/CN=lcg-voms.cern.ch] "cms" Done
Creating proxy ..... Done
Your proxy is valid until Fri Dec 26 13:44:47 2008
Your identity: /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=yuanchao/CN=596728/CN=Yuan Chao
Enter GRID pass phrase for this identity:
Creating proxy ..... Done
Proxy Verify OK
Your proxy is valid until: Thu Dec 25 13:45:25 2008
A proxy valid for 168 hours (7.0 days) for user /DC=ch/DC=cern/OU=Organic Units/OU=Users/CN=yuanchao
crab. Contacting Data Discovery Services ...
crab. Requested dataset: /RelValZEE/CMSSW_2_1_9_STARTUP_V7_v2/GEN-SIM-DIGI-RAW-HLTDEBUG-RECO has
crab. May not create the exact number_of_jobs requested.
crab. 11 job(s) can run on 3000 events.
crab. List of jobs and available destination sites:
Block      1: jobs                1-2: sites: srm-3.t2.ucsd.edu,srm.princeton.edu
Block      2: jobs                3-11: sites: srm-3.t2.ucsd.edu
crab. Creating 11 jobs, please wait...
crab. Total of 11 jobs created.
crab. Log-file is /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_0812
```

## Job Submission

With the submission command it's possible to specify a combination of jobs and job-ranges separated by comma (e.g.: =1,2,3-4), the default is all. To submit all jobs of the last created project with the default name, it's enough to execute the following command:

```
crab -submit
```

to submit a specific project:

```
crab -submit -c <dir name>
```

which should produce a similar screen output like:

```
[ui02] /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test > crab -submit
crab. crab (version 2.4.3) running on Thu Dec 18 14:06:34 2008
crab. Working options:
  scheduler          glite
  job type           CMSSW
  working directory  /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_0812
crab. Downloading config files for https://cmsweb.cern.ch/crabconf/files/server_bari.conf
crab. Registering credential to the server
crab. Registering a valid proxy to the server:
crab. Credential successfully delegated to the server.
crab. Starting sending the project to the storage dot1-prod-2.ba.infn.it...
crab. Task crab_0_081218_133622 successfully submitted to server dot1-prod-2.ba.infn.it
crab. Total of 11 jobs submitted
crab. Log-file is /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_0812
```

## Job Status Check

Check the status of the jobs in the latest CRAB project with the following command:

```
crab -status
```

to check a specific project:

```
crab -status -c <dir name>
```

which should produce a similar screen output like:

```
[ui02] /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test > crab -status
crab. crab (version 2.4.3) running on Thu Dec 18 14:52:36 2008
```

```
crab. Working options:
```

```
  scheduler          glite
  job type           CMSSW
  working directory  /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_081218_133622_306e8c4d-85ab-45d9-ab80-571420f834c7
```

ID	STATUS	E_HOST	EXE_EXIT_CODE	JOB_EXIT_STATUS	END
1	Submitting	N			
2	Submitting	N			
3	Submitting	N			
4	Submitting	N			
5	Submitting	N			
6	Submitting	N			
7	Submitting	N			
8	Submitting	N			
9	Submitting	N			
10	Submitting	N			
11	Submitting	N			

```
>>>>>>>> 11 Total Jobs
```

```
>>>>>>>> 11 Jobs with Wrapper Exit Code :
             List of jobs: 1-11
```

```
crab. Log-file is /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_081218_133622_306e8c4d-85ab-45d9-ab80-571420f834c7
```

Also, you can have a look at the web page of the server where you can see the status progress of you job.  
Simply, execute the command:

```
crab -printId
```

And you will get the unique id of your jobs:

```
[ui02] /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test > crab -printId
crab. crab (version 2.4.3) running on Thu Dec 18 14:25:43 2008
```

```
crab. Working options:
```

```
  scheduler          glite
  job type           CMSSW
  working directory  /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_081218_133622_306e8c4d-85ab-45d9-ab80-571420f834c7
```

```
Task Id = yuanchao_crab_0_081218_133622_306e8c4d-85ab-45d9-ab80-571420f834c7
```

```
crab. You can also check jobs status at: http://dot1-prod-2.ba.infn.it:8888/logginfo
      ( Your task name is: yuanchao_crab_0_081218_133622_306e8c4d-85ab-45d9-ab80-571420f834c7 )
```

```
crab. Log-file is /home/yuanchao/Tutorial/CMSSW_2_1_12/src/PhysicsTools/PatAlgos/test/crab_0_081218_133622_306e8c4d-85ab-45d9-ab80-571420f834c7
```



Copy the unique id of your task (in the above example: *mcinquil\_crab\_0\_081017\_011536\_72fd4e00-4d34-4862-893b-7057e0049f36*), go to the link of the Bari Server [Bari Server](#), paste the unique id in the text field and press the "Show" button.

## Job Output Retrieval

For the jobs which are in the "Done" state it is possible to retrieve the log files of the jobs (just the log files, because the output files are copied to the Storage Element associated to the T2 specified on the crab.cfg and in fact return\_data is 0). The following command retrieves the log files of all "Done" jobs of the last created CRAB project:

```
crab -getoutput
```

to get the output of a specific project:

```
crab -getoutput -c <dir name>
```

the job results will be copied in the `res` subdirectory of your crab project:

```
[lxplus231] ~/scratch0/Tutorial/CMSSW_2_1_9/src/PhysicsTools/PatAlgos/test > crab -get all -c
crab. crab (version 2.4.0) running on Thu Oct 16 16:32:18 2008
```

```
crab. Working options:
```

```
  scheduler           glite
  job type            CMSSW
  working directory   /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/PhysicsTools
```

```
crab. Starting retrieving output from server dot1-prod-2.ba.infn.it...
```

```
crab. Results of Jobs # 1 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 2 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 3 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 4 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 5 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 6 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 7 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 8 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 9 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Results of Jobs # 10 are in /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/P
crab. Log-file is /afs/cern.ch/user/m/mcinquil/scratch0/Tutorial/CMSSW_2_1_9/src/PhysicsTools/Pat
```

## Where to find more on CRAB

- [CRAB Home](#)
- [CRAB FAQ](#)
- [CRAB Man](#)
- [WorkbookGridJobDiagnosisTemplate](#): Steps to identify the problems you experience with your grid analysis jobs.
- [CRAB\\_2\\_4\\_2 Documentation page](#)

## T3\_TW\_NTU\_HEP specific info.

One can specify ASGC CE and/or SE in the white list of CRAB to force your jobs to run at T3\_TW\_NTU\_HEP.

## UI

- ntugrid1.phys.ntu.edu.tw, ntugrid3.phys.ntu.edu.tw (SLC5)
- ntugrid5.phys.ntu.edu.tw (SLC6)

## CE

- ntugrid2.phys.ntu.edu.tw
- ntugrid5.phys.ntu.edu.tw

## SE

- ntugrid4.phys.ntu.edu.tw (paired with ntugrid2)
- ntugrid6.phys.ntu.edu.tw (paired with ntugrid5)

## Useful resources

Recursive copying all files under a given directory for DPM or SRM. Don't forget to change the user name in the script. Rename the file as \*.py and give proper permissions.

- DPM copying script: dpm\_rec\_cp.py

Change 'rfdi' and 'rfcp' to '/opt/lcg/bin/rfdi' and '/opt/lcg/bin/rfcp' if you are using on ASGC UI.

- SRM copying script: srm\_rec\_cp.py

Don't forget to change to your SE store/user home in the script.

- SRM directory deletion scrip: srm\_rec\_rm.py

It can delete all files under a folder. For safety, it doesn't do recursive deletion. (only one level)

```
[yuanchao@pcntu01 ~]$ srm_rec_rm.py
Usage: srm_rec_rm.py srm_path [matching_pattern]
```

It picks up the user name and corresponding path, so you only need to specify a relative srm\_path to your own area. It also removes empty directories.

[matching\_pattern] is just a simple string comparison. So no '\*' or '?' allowed.

## Useful links

- The "starting page" Twiki:

<http://wiki.twgrid.org/apwiki/test?action=show>

## Slides and scripts for usage tutorial

- ychao\_20120830.odf: T3 usage tutorial slides

- dump\_AODSIM.py.txt: CMSSW script for event dump

-- YuanChao - 26 Sep 2014

---

This topic: Main > YuanChaoT3NTUHEPUsage

Topic revision: r10 - 2016-02-19 - YuanChao



Copyright &© 2008-2019 by the contributing authors. All material on this collaboration platform is the property of the contributing authors. Ideas, requests, problems regarding TWiki? Send feedback