



Analysis Using CRAB On The Grid

A practical introduction to

CRAB



Bockjoo Kim
U of Florida



Outline of Analysis Using CRAB

- **Certificate Processing: The first BIG steps**
- **Grid Tools**
- **CMSSW**
- **CRAB**
- **Example Configurations**
- **Example Analysis Using EDAnalyzer**
- **Putting Everything Together**



Certificate Processing



- **Create Certificate**

<https://pki1.doeagrids.org/ManUserEnroll.html>

- **Created Certificate**

usercert.pem* and **userkey.pem**

- **Register Certificate from RA's**

<http://www.doeagrids.org/>

- **Signed Certificate**

usercert.pem& and **userkey.pem**

- **Register yourself as a human at CERN**

http://cmsdoc.cern.ch/cms/aprom/www/top/CMS_VO.html#hrreg

- **Choose CMS as your VO**

<https://lcg-voms.cern.ch:8443/vo/cms/vomrs?path=/RootNode>

- **Choose your role as the cmsuser in /cms/uscms group.**

cmsfrontier, cmsphedex, cmsprod, cmssoft, cmst1admin, cmst1admin, cmsuser





Grid Tools: OSG Client+Condor-G

- **Available at some USCMS T2 Analysis Hosts**
Example : `alachua.phys.ufl.edu`
- **Install personal OSG client+Condor-G :**
<https://twiki.grid.iu.edu/twiki/bin/view/ReleaseDocumentation/ClientInstallationGuide>
- **OSG packaging : Pacman has to be used**
<https://twiki.grid.iu.edu/twiki/bin/view/ReleaseDocumentation/PacmanInfo>
- **Condor-G configuration maybe non-trivial**



Grid Tools: EGEE gLite UI

- Available at some USCMS T2 Analysis Hosts
- One can also install personal gLite UI :
<http://grid-deployment.web.cern.ch/grid-deployment/download/relocatable>
- Very non-trivial installation for first-timer
- No SL4 version yet (to be available soon)
- Important for interoperability (EGEE--> OSG)



CMSSW

- **CMSSW is required for CRAB job creation**
- **Available at OSG T2 sites**
- **One can also install CMSSW on {desk,lap}top**
 - https://twiki.cern.ch/twiki/bin/view/CMS/CMSSW_apinstaller : General install instruction
 - https://gdsuf.phys.ufl.edu:8080/cmssoft/cmssoft_install.tar.gz : install scripts on OSG
 - https://gdsuf.phys.ufl.edu:8080/cmssoft/install_cmssw_sl4.sh : Example SL4 CMSSW install
- **SL3 vs SL4 : CRAB not tested with SL4 yet**



CRAB

- **Available at some T2 sites**
UCSD, Caltech, Wisconsin, UFL
- **One can also install personal CRAB**
<https://cmsdoc.cern.ch/cms/ccs/wm/www/Crab>: Production Release
[/afs/cern.ch/cms/ccs/wm/scripts/Crab/CRAB_1_4_1_pre1.tgz](https://cmsdoc.cern.ch/cms/ccs/wm/scripts/Crab/CRAB_1_4_1_pre1.tgz) : Pre-release
- **Data Bookkeeping System (DBS) evolves:**
Something that one needs to watch out
- **CRAB Configuration :**
scheduler and data mostly



CRAB Configuration : scheduler

- **[CRAB] Section 'scheduler' configuration**

```
scheduler = condor_g # Condor-G Scheduler  
# scheduler = glitecoll # for gLite RB job submission
```

- **[EDG] Section comp. resource configuration
(Resource Broker or GridCat)**

```
ce_white_list = ufloridapg.phys.ufl.edu # one of USCMS T2 sites  
se_white_list = phys.ufl.edu # restrict to SE at Florida
```

- **[CONDORG] Section b-queue configuration**

```
globus_rsl = (queue=cms_grid)(maxWalltime=2880)(jobType=single)
```




CRAB Configuration : data

- **Dataset Location (DLS) : SE -> dataset**

`lfc_catalog=prod-lfc-cms-central.cern.ch/grid/cms/DLS/LFC`

`SE=srm1.phys.ufl.edu`

`dls-get-fileblock -i DLS_TYPE_LFC -e $lfc_catalog $SE # Lists all dataset at $SE`

- **[CMSSW] Section Variables**

- **Input Data Minbias CSA06, CMSSW_1_0_6**

`datasetpath = /CSA06-106-os-minbias12-0/RECO/CMSSW_1_0_6-
RECO_Hee5e8972ca2b9fe5859521c27772ba77`

- **No Input : e.g., starting from pythia evt. Gen.**

`datasetpath = None`

- **CMSSW configuration : pset = recoana.cfg**

- **Total events to be analyzed : total_number_of_events = -1**

- **Events/Job : events_job = 7000**

- **Output : output_file = myroot.root**



Example CRAB Configuration(crab.cfg)

[CRAB]

jobtype = cmssw

scheduler = condor_g

[CMSSW]

datasetpath = /CSA06-106-os-minbias12-0/RECO/CMSSW_1_0_6-RECO_Hee5e8972ca2b9fe5859521c277'

dbs_url = http://cmsdbs.cern.ch/cms/prod/comp/DBS/CGIServer/prodquery

pset = recoana.cfg

total_number_of_events = -1

events_per_job = 7000

output_file = myroot.root

[USER]

return_data = 1

use_central_bossDB = 0

use_boss_rt = 0

total_number_of_events = 1

[EDG]

lcg_version = 2

proxy_server = myproxy.cern.ch

ce_white_list = ufloridapg.phys.ufl.edu

se_white_list = phys.ufl.edu

virtual_organization = cms

retry_count = 2

lcg_catalog_type = lfc

lfc_host = lfc-cms-test.cern.ch

lfc_home = /grid/cms



Example CMSSW Configuration

-recoana.cfg == CMSSW configuration file contains :

```
process Everything = {
  service = Tracer {}
  include "Configuration/Examples/data/RECO.cff"
  include "FWCore/MessageLogger/data/MessageLogger.cfi"
  include "Configuration/CompatibilityFragments/data/CompatibilityWith081.cff"
  service = Timing {}
  service = SimpleMemoryCheck { untracked int32 ignoreTotal = 1 }

  source = PoolSource{
    untracked vstring fileNames ={
"/store/CSA06/2006/8/17/CSA06-082-os-TTbar/FE9DFC62-6238-DB11-9684-00E081339574.root"}
    untracked int32 maxEvents = 2
    untracked uint32 debugVerbosity = 10
    untracked bool  debugFlag    = true
  }

  module analysis = ExampleAnalysis
  {
    untracked string HistOutFile = "myroot.root"
  }
  path p = { reconstruction , analysis }
}
```



Example Analysis Using EDAnalyzer

- CMSSW_1_2_0/ExampleAnalysis/ExampleAnalysis/src/ExampleAnalysis.cc :

.
. .
. .
. .
. .

```
ExampleAnalysis::ExampleAnalysis(const ParameterSet& iConfig)
: fOutputFileName( iConfig.getUntrackedParameter<string>("HistOutFile",std::string("Exam
pleSmallAnalysisOutput.root")) ),
  fOutputFile(0)
```

```
{
.
.
.
.
.
}
```



Grid Environment Setup

- Initial Variables :

VDT_LOCATION, GLITE_LOCATION, CMSSOFT_DIR, CRABDIR

- Setup Grid Environment :

OSG : source \$VDT_LOCATION/setup.sh

gLite :

source \$GLITE_LOCATION/etc/profile.d/grid_env.sh ; source ~/.glite/slite_setenv.sh

On UF T2 analysis machines, this is given : try `which voms-proxy-init`

- Proxy Setup for Grid Credential :

On UF T2 analysis machines, you need to set up VOMS cert directory:

export X509_VOMS_DIR=/home/bockjoo/etc/grid-security/vomsdir

Make sure you have \$HOME/.globus/user{cert,key}.pem

voms-proxy-init --voms cms:/cms/uscms/Role=cmsuser # asks pass1

- Check Proxy : voms-proxy-info -all

- Proxy Delegation

export X509_USER_PROXY=`voms-proxy-info -path`

export X509_USER_CERT=\$X509_USER_PROXY

export X509_USER_KEY=\$X509_USER_PROXY

myproxy-init -v -d -s myproxy.cern.ch # asks pass2



Setup for CRAB

- Setup CMSSW

```
export SCRAM_ARCH=slc3_ia32_gcc323 ; source $CMSSOFT_DIR/cms/cmsset_default.sh
cd ; mkdir analysis_99 ; cd analysis_99 # creates analysis work directory
scramv1 -arch $SCRAM_ARCH project CMSSW CMSSW_1_2_0
cd CMSSW_1_2_0/src ; eval `scramv1 runtime -sh`
wget http://gdsuf.phys.ufl.edu:8080/cms_analysis/ExampleAnalysis.tar.gz
tar xzvf ExampleAnalysis.tar.gz ; scramv1 b # build EDAnalyzer
```

- CRAB Setup CMSSW

```
/home/coldfeet/scpaw/scripts/configure_crab # creates private CRAB installation and DB config.
source $CRABDIR/crab.sh
cd ; mkdir crab_99 ; cd crab_99 # this is going to be the crab work directory
- Site oriented directory
  mkdir -p Sites/Florida/{PG,IHEPA} ; cd Sites/Florida/PG ; vi $CRABDIR/python/crab.cfg
  cp /home/coldfeet/scpaw/scripts/recoana.cfg ./
  vi recoana.cfg
- Physics oriented directory
  mkdir -p Physics/DYZDimu/{PG,IHEPA}
  cd Physics/DYZDimu/PG ; vi crab-input-none.cfg ; vi dyzdimuana.cfg
```

- Test Driving CMSSW Before Going Wild

```
cd $HOME/crab_99/Physics/DYZDimu/PG
cmsRun dyzdimuana.cfg # pythin gen + SIM + DIGI + RECO + ExampleAnalysis
```



Executing CRAB and Retrieving Data

- Create Jobs

```
cd $HOME/crab_99/Sites/Florida/PG
crab -create # This creates a CRAB work directory with 'crab_0_yymmdd_hhmmss'
             # The CRAB work directory contains 'job' 'log' 'res' 'share'
             # 'job' : scripts
             # 'log' : log file
             # 'res' : stdout, stderr, root file, job log
             # 'share' : ASCII DB, BOSS job work area and binaries
```

- Submit One Job (Reports to Dashboard)

```
crab -submit 1 -continue d ; mkdir analysis_99 ; cd analysis_99
```

- Submit All Jobs (Reports to Dashboard)

```
crab -submit -continue
```

-Check Job Status (Reports to Dashboard)

```
crab -status -continue
```

- Retrieve Output (Reports to Dashboard)

```
crab -getoutput
```



Analyze Produced Plots and Data

- Merge root file

```
cd $HOME/crab_99/Sites/Florida/PG
root
root> TChain ch2("DimuTree");
root> ch2.Add("./crab_0_070131_121512/res/myroot_*.root");
root> ch2.Merge("./DimuTree.root");
```

- Create Plots

```
cd $HOME/crab_99/Sites/Florida/PG
root> new TBrower
root> .....
root> .....
```

- Today's Goal :

- Create CRAB Job
- Submit CRAB Job
- Retrieve CRAB Job
- Make plots



Variables on UF T2 Analysis Machine

- On `alachua.phys.ufl.edu` or another

`/home/coldfeet/scpaw/scripts/setup_crab.sh`