CRAB tutorial
Cms Remote Analysis Builder

Stefano Lacaprara, Federica Fanzago, Marco Corvo

Department of Physics
INFN and University of Padova

CRAB Tutorial, 4 February 2005
Outline

Intro
- What is this tutorial about
- Prerequisites
- Further information

CRAB usage
- Getting CRAB
- Configuration
- Running CRAB
- Troubleshooting
What is this tutorial about and what is not

- Tutorial about CRAB Cms Remote Analysis Builder
  - A tool developed within Workload Management group open the Grid to the masses!
  - CRAB is aimed to give access to CMS analysts to all Data produced and available everywhere, using GRID (LCG) middleware
  - CRAB should hide as much as possible Grid complexities and sublets to CMS user (if you could believe it …)

- This is not a tutorial about
  - How to do analysis
  - How to use CMS software (ORCA)
  - general Grid usage
What is this tutorial about and what is not

- Tutorial about CRAB Cms Remote Analysis Builder
  - A tool developed within Workload Management group open the Grid to the masses!
  - CRAB is aimed to give access to CMS analysts to all Data produced and available everywhere, using GRID (LCG) middleware
  - CRAB should hide as much as possible Grid complexities and sublets to CMS user (if you could believe it . . . )

- This is not a tutorial about
  - How to do analysis
  - How to use CMS software (ORCA)
  - general Grid usage
Prerequisites

- What you need to access the Grid and access CMS data:
  - Work from an **User Interface** (UI)
  - Have a valid Grid certificate
  - Have a active grid-proxy on UI (**grid-proxy-init**)
  - Your Virtual Organization (VO) must be (also) **CMS**
  - Have CMS sw (**ORCA**) installed on UI
  - Know how to use ORCA interactively!
More information sources

- CRAB web page
  http://cmsdoc.cern.ch/cms/ccs/wm/www/Crab
- CRAB mailing list, for feedback and user support (you can register via SIMBA)
  cms-wm-crab-feedback@cern.ch
- CRAB savannah web page: for bugs report and features request
  https://savannah.cern.ch/projects/crab/
- README (which comes with CRAB)
- on-line crab manual -help
- EIS support support-eis@cern.ch for GRID related problems
Get CRAB

At CERN

- Any lxplus[7] node becomes an UI sourcing:
  /afs/cern.ch/cms/LCG/LCG-2/UI/cms_ui_env.csh|sh
- From any lxplus[7] node, just source:
  $CMS_PATH/ccs/wm/script/Crab/crab.csh|sh
  (it sources UI script is not done yet)
- Keep updated for you!

Elsewhere

- Get it from CVS (working with a better distribution system...)
- cmscvsroot CRAB
- cvs co -r <tag> UserTools
- latest tag (04-Feb-2005) is CRAB_0_0_7
How to start

- From a UI (lxplus[7] is fine) write and test your code in the usual way, using an ORCA working area (scram project ...)
- Once you are happy, decide which remote dataset you want to access
- Which dataset are available?
- Look at CERN PubDB page
  
  http://cmsdoc.cern.ch/cms/production/www/PubDB/GetPublishedCollectionInfoFromRefDB.php
- Warning!!! the data location service is ramping up: not all site are today up to date and accessible via CRAB
- Site today working fine: CNAF, LNL, BA; PIC almost; FNAL, IN2P3, CERN are working...
- LCG Worker nodes are RedHat7.3, ⇒ must use a RH7.3 UI: lxplus7 at CERN. Migration to SLC3 should take place soon
CRAB configuration

- From user working area (e.g., ~/ORCA_8_7_1/src/Workspace), issue the usual command `eval 'scram runtime -sh|csh'`
- Move to your Crab working area:
  - UserTools/src if CRAB via CVS
  - not needed if at CERN and crab.csh sourced!
- modify configuration file crab.cfg
- run crab
CRAB configuration

- From user working area (e.g. ~/ORCA_8_7_1/src/Workspace), issue the usual command eval `scram runtime -sh|csh`
- Move to your Crab working area:
  - UserTools/src if CRAB via CVS
  - not needed if at CERN and crab.csh sourced!
- modify configuration file crab.cfg
- run crab
CRAB configuration

- From user working area (e.g. ~/ORCA_8_7_1/src/Workspace), issue the usual command eval ‘scram runtime -sh|csh’
- Move to your Crab working area:
  - UserTools/src if CRAB via CVS
    - not needed if at CERN and crab.csh sourced!
- modify configuration file crab.cfg
- run crab
CRAB configuration

- From user working area (e.g. ~/ORCA_8_7_1/src/Workspace), issue the usual command `eval ‘scram runtime -sh|csh’`
- Move to your Crab working area:
  - `UserTools/src` if CRAB via CVS
  - not needed if at CERN and `crab.csh` sourced!
- modify configuration file `crab.cfg`
- run `crab`
CRAB configuration

- From user working area (e.g. ~/ORCA_8_7_1/src/Workspace), issue the usual command `eval ‘scram runtime -sh|csh’`
- Move to your Crab working area:
  - UserTools/src if CRAB via CVS
  - not needed if at CERN and crab.csh sourced!
- modify configuration file crab.cfg
- run crab
CRAB configuration

- From user working area (e.g. ~/ORCA_8_7_1/src/Workspace), issue the usual command `eval ‘scram runtime -sh|csh‘`
- Move to your Crab working area:
  - UserTools/src if CRAB via CVS
  - not needed if at CERN and `crab.csh` sourced!
- modify configuration file `crab.cfg`
- run `crab`
crab.cfg

- Mandatory keys:
  - dataset: The one you want to access
  - owner: ditto
  - executable: the one you used interactively! It is in your path since you did eval `scram runtime -csh`, didn’t you? If the executable uses some private libs, **crab** find them for you.
  - orcarc_file: the .orcarc file you used interactively: the very same! **crab** will change it for you
  - total_number_of_events: to be processed
  - job_number_of_events: number of event per job
  - output_file: your executable produces

Stefano Lacaprara, Federica Fanzago, Marco Corvo

CRAB
crab.cfg

- **data_tiers** you want to access with your executable: “DST,Digi,Hit”. Require parents to be published in the same site with primary owner. Default is just primary data tier.

- **PU not yet possible!** WARNING: ExDSTStatistics access PU!!!

- **additional_input_files** to be send to WorkerNode (WN) with the jobs: comma (,) separated list

- **output/log_dir** directory where crab will put output and log: if not set the default is crab_0_<date>_<time>/res
Running CRAB

`crab.py -help` and read it!!!
`crab.py -create N` to create N jobs (no submission)
`crab.py -submit N -continue` to submit the job you have created
(mind the `-continue`)
`crab.py -monitor -continue` to monitor (very primitive!) your jobs
and get automatically the output retrieved when the jobs are finished
`crab.py -create all -submit all -monitor` to do all in just one command

If all is fine, present your work at meeting and offer some wine/beers/etc to crab team!
if problems, see next slides (and no beer for us...)
Running CRAB

crab.py -help and read it!!!
crab.py -create N to create N jobs (no submission)
crab.py -submit N -continue to submit the job you have created (mind the -continue)
crab.py -monitor -continue to monitor (very primitive!) your jobs and get automatically the output retrieved when the jobs are finished
crab.py -create all -submit all -monitor to do all in just one command

If all is fine, present your work at meeting and offer some wine/beers/etc to crab team!
if problems, see next slides (and no beer for us...)

Stefano Lacaprara, Federica Fanzago, Marco Corvo
Running CRAB

```bash
crab.py -help and read it!!!
crab.py -create N to create N jobs (no submission)
crab.py -submit N -continue to submit the job you have created (mind the -continue)
crab.py -monitor -continue to monitor (very primitive!) your jobs and get automatically the output retrieved when the jobs are finished

crab.py -create all -submit all -monitor to do all in just one command
```

If all is fine, present your work at meeting and offer some wine/beers/etc to crab team!
if problems, see next slides (and no beer for us...)

Stefano Lacaprara, Federica Fanzago, Marco Corvo

CRAB
Troubleshooting

- New born system: please be patient and provide feedback!
- most common problems:
  - `ConfigParser.ParsingError` error in `crab.cfg`: typically space at beginning of line!
  - Cannot create job type "ORCA" for VO "CMS" You did not eval 'scram runtime -csh'
  - Attempt to submit 1 jobs but only 0 jobs were created
  - `crab.py -submit 1` but not -continue
  - Need to create grid-proxy Create one grid-proxy-init
  - dataset/owner not published with data tiers ... spelling
    could be that dataset/owner available but not its parents (e.g. only DST and not Digi,Hit)
Troubleshooting (II)

Job starts but ORCA crash: many possibilities:

- your fault: **double check your code!!!**
- you are trying to access other data tiers (typically PU)
- problem with local Pool catalog:
  - rerun with: CARFVerbosity=debug and PersistencyVerbosity=debug and give the output to Site PubDB maintainer or savannah or cms-wm-crab-feedback@cern.ch list
Troubleshooting (Job Aborted)

▶ You run `crab.py -mon -c` and you see **Aborted**. Many possibilities:

▶ **Do `edg-job-status <JodId>`**

  **Cannot plan: BrokerHelper: no compatible resources today**
  ORCA deployed at site is 8_7_1, if newer, no matching resources can be found

**Other messages** send output of `edg-job-get-logging-info -v 2 <jobID>` to support-eis@cern.ch