Table of Contents

5.1 Chapter Overview -- Getting Started

Goals of this page:

This page is intended to provide you with an overview of this entire Chapter, pointing out which parts are required reading to get physics analysis done on the CMS distributed analysis infrastructure, and those that are meant to provide intellectual stimuli and broader context.

Contents

- Introduction
- Basic requirements for using the Grid
- Obtaining and Installing your Certificate
- Connecting your certificate to your account
- Using your grid certificate
- Accessing an LCG User Interface
    - ◆ Preinstalled
- Review status

Introduction

CMS uses a globally distributed computing system for data analysis. The present Chapter has two objectives:

1. Provide you with all the information required to use the global system for physics data analysis.
2. Provide you with background information, and context, so that you start gaining some appreciation of the complexity of this system.

Those who really don't care about how things work, and just want to get their analysis off the ground, may want to skip all the material provided in the interest of our second goal above. The present section is meant to make this easy for you by providing guidance on what to skip. However, let us warn you upfront that eventually, you will need that more detailed background knowledge in order to understand, and react to failures of the distributed system that you will invariably be exposed to, while using it. **The complexity of this global system guarantees that an educated and intelligent user will often be more effective in getting stuff done, than somebody who knows nothing but the basics.**

Roadmap for Chapter 5

As a new user, you should **read the "must read" chapters in the order listed**, as concepts introduced in one will often be used in the next. This is especially true for Chapters 5.4, 5.5, and 5.6.

- Chapter 5.1 is a **must read**. It not only provides this roadmap, but also a discussion of the requirements to get started.
- Chapter 5.2 "Grid Computing Context" can be skipped by the impatient. It provides a general introduction of "grid" computing terms.
- Chapter 5.3 "Analysis Workflow" can be skipped, except for the very beginning of it. It explains how CRAB works under the hood, at least conceptually.

- Chapter 5.4 "Locating Data" is a **must read**. It explains how to find the datasets to run on and how to pull a single file to your desktop, so you can try out your executable interactively and do the bulk of your debugging.
- Chapter 5.5 "Data Quality Monitor" can be skipped initially. It explains how to refine the Data Finding process to include Data Quality Information
- Chapter 5.6 "Data Analysis with CRAB" is a **must read**. It explains how to use CRAB, the tool to use for doing data analysis on the globally distributed CMS data analysis infrastructure.
- Chapter 5.7 "Data Analysis with CMS Connect" is a **must read**. It explains how to use CMS Connect, the complementary service to CRAB for user-defined scripts via condor for doing late-stage data analysis that don't depend on cmsRun (the CMSSW executable). E.g Making histograms, plots, analyzing trees, etc.
- Chapter 5.8 "Dashboard Job Monitor" is a **must read**. It explains how to monitor the status of your jobs.
- Chapter 5.9 "The role of the T2s" can be skipped initially. It provides essential background to understand the disk space organization at T2s in CMS. As T2s are the places where the vast majority of data analysis in CMS takes place, it will eventually be vital for you to read this chapter carefully.
- Chapter 5.10 "Transfering Data" can be skipped initially. Once you have read chapter 5.7, you will understand how disk space is managed, and can then graduate to using it in style. This Chapter explains how to request datasets to be moved to T2s and T3s. Anybody in CMS can make such requests.
- Chapter 5.11 "Data Organization Explained" can be skipped initially. It explains a variety of terms that CMS uses to describe how data is organized and managed.
- Chapter 5.12 "Processing by Physics Groups". It talks about priority users privileges and convenors responsibilty towards such features.
- Chapter 5.13 "cmssh tutorial". A very useful tool to easily find your favorite data from the command line, copy files transparently without knowing Physical File Name location, etc.

Basic requirements for using the Grid

The remainder of this page deals with the essentials you need before you can even start doing anything on the globally distributed CMS data analysis infrastructure.

Note that initial testing and workbook exercises can be done on an LXPLUS machine (or another machine, properly configured), but proper analysis jobs and Monte Carlo production should be submitted to the globally distributed CMS data analysis infrastructure. Note: We will sometimes use the word "Grid" as a synonym to "globally distributed CMS data analysis infrastructure" for obvious reasons of brevity.

The basic requirements for using the Grid resources are:

- a local computer - either personal computer or a workcluster;
- being a member of CMS (CMS registration�);
- a digital certificate from one of the Certification Authorities (CA) recognised by WLCG (for example the CERN CA�)
- being a member in the CMS Virtual Organisation (CMS VO);
- correct mapping of grid certificate to username, see below
- access to a grid User Interface;

Obtaining and installing your Certificate

To **obtain your certificate and join the CMS VO**, follow the steps on this page.
That same page also has pointers to troubleshooting help if needed.

Note that it can take a few days for the certificate to be issued. The CA will give you instructions on how to load your certificate into your browser.

To **setup the certificate** on the user interface from where you have to work you should:

- Export the certificate from your browser to a file in p12 format. How to export the certificate is very browser dependent. It will be something like Edit or Tools -> Preferences or (Internet) Options -> Advanced -> Security or Encryption -> View Certificates -> Your Certificates. In modern Firefox you should "backup" rather than "export" the certificate. You can find more instructions and hints for various browsers in this CERN CA help page☞. You can give any name to your p12 file (in the example below the name is `mycert.p12`).
- Place the p12 certificate file in the `.globus` directory of your home area. If the `.globus` directory doesn't exist, create it.

```
cd ~
mkdir .globus
cd ~/.globus
mv /path/to/mycert.p12 .
```

- Execute the following shell commands:

```
rm -f usercert.pem
rm -f userkey.pem
openssl pkcs12 -in mycert.p12 -clcerts -nokeys -out usercert.pem
openssl pkcs12 -in mycert.p12 -nocerts -out userkey.pem
chmod 400 userkey.pem
chmod 400 usercert.pem
```

- For openssl commands, you need to put the **same password** that you chose while importing the certificate in your browser, and you would also be asked for "Enter PEM pass phrase". One may choose to keep it same, so as to avoid password confusions ☺
- Verify that it all works by executing (n.b. you may need to setup a grid UI to execute this command, see below):

```
voms-proxy-init --rfc --voms cms -valid 192:00
```

- Ignore a (possible) message about not being able to find a `.glite/vomses` directory.

Some CAs provide the `usercert.pem` and `userkey.pem` files and then the user has to produce the p12 file to be imported to the browser. To convert the `usercert.pem` and `userkey.pem` files into a browser certificate `mycert.p12` do the following:

```
openssl pkcs12 -export -in usercert.pem -inkey userkey.pem -out mycert.p12 -name "my browser cert
```

To do CMS analysis on WLCG Grid resources, you will further require:

- A CMS analysis software environment setup on your local computer.
- Some sample datasets with local access (on a hard disk or other mass data storage system) so you can test your analysis code interactively before submitting your jobs on the grid. These local datasets are frequently subsets of one of the main CMS datasets resulting from a first-pass analysis job (RECO or AOD).
- To stage user data back to CERN with a non-CERN certificate you need to map it to your CERN account☞ (not yet enforced).

All CMS members using the Grid may benefit from subscribing to the Grid Annoucements CMS.HyperNews forum☞.

Connecting your certificate to your account

CMS Analysis with CRAB requires that the user's authentication credential is mapped to the a globally unique username. Currently authentication is based on grid certificate, where user is identified by the so called DN and we use the CERN primary computing account as username. If you use a certificated from CERN, this operation is fully transparent and you need to do no nothing but be aware of what you CERN username is. If you are using a grid certificate issued by a Certification Authority other than CERN CA, then read and follow the instructions in the Username for CRAB page to make sure your certificate is correctly mapped to your account.

Using your grid certificate

Each day you wish to use xrootd, CRAB, CMS Connect⬚, or similar technologies, you will need to authenticate your grid certificate with the command:

```
voms-proxy-init --rfc --voms cms -valid 192:00
```

Grid User Interface

The recommended way to submit jobs on the Grid is to use CRAB. It will allow you to access both EGI and OSG Grid resources in a fully transparent way. Minimal client as distributed by OSG or pre-installed on lxplus will do.

Preinstalled

- At CERN:
    - ♦ LXPLUS already has the grid commands needed for Crab, no need to issue any setup command.
- Other affiliated sites and institutions may provide generally available WLCG/OSG software for grid tools (see WorkBookRemoteSiteSpecifics to look for information for your institution).

Review status

Reviewer/Editor and Date (copy from screen)	Comments
CalebJamesSmith - 2021-07-06	Change filename "myCertificate.p12" to "mycert.p12" in commands for consistency.
MargueriteBeltTonjes - 2020-06019	remove GLiteUI/Install Your Own because it's SL5 and obsolete, added -valid to first use of voms-proxy-init
StefanoBelforte - 2020-02-28	remove references to SiteDB, remove some old obsolete things
StefanoBelforte - 20 Sep 2014	review and update grid documentation, remove duplications
StefanoBelforte - 14 Sep 2014	update reference to CERN Grid CA page
StefanoBelforte - 20 Aug 2014	remove reference to gLite UI
JohnStupak - Mar 2013	review with minor changes
NitishDhingra - 28-Mar-2012	See detailed comments below
StefanoBelforte - 22-Dec-2009	Complete Expert Review, minor changes
FrankWuerthwein - 04-Dec-2009	Complete Reorganization 1st draft ready for review

AndreaSciaba - 30 Nov 2009	Minor corrections (removed or replaced broken links)
SimonMetson - 30 Apr 2009	Updated the link to request a certificate (after a question from a user advice from Andrea Sciaba
MattiaCinquilli - 24 Nov 2008	added explicit commands to setup the certificate
AndreaSciaba - 24 Jan 2008	review with updated links and minor changes
StefanoLacaprara - 16 Nov 2006	review with minor changes
AnneHeavey - 03 Aug 2006	fairly substantial edits to Grid info

Detailed comments 28-Mar-2012 ▶ Hide ▼

Review with minor additions in the grid certificate set-up instructions. The page accomplishes its goal.

Responsible: StefanoBelforte
Last reviewed by: Main.David L Evans - fill in date when done -

This topic: CMSPublic > WorkBookStartingGrid
Topic revision: r77 - 2022-06-06 - MargueriteTonjes