Central CMS MC Requests: Overview

Internal: Hohlmann Research Group

Author: Jacob Chesslo

Table of Contents

Introduction

This document is an overview and guide on how to request central production requests, or MC Requests, i.e. those produced and run by CMS at CERN, of the Monte Carlo (MC) samples. In order to create MC Requests, certain "ingredients" for production are needed, such as Gridcards and Models, as well as the completion of certain steps, such as creating Gridpacks and emailing a specific spreadsheet. Specifically, this document goes over how to produce MadGraph5_aMCatNLO MC Requests, however the process for other Generator requests follow a similar process (see CERN's Twiki for instructions on how to produce MC Requests for other Generators).

Before moving forward with a MC Request, a few things are integral for success. One of which is a basic understanding of the Linux command line, basic programming in some of the commonly used languages, and a basic understanding of Github. Another is a CERN account, along with access to LXPLUS.

Make sure to read over this document and all of main the resources linked, especially the twiki pages https://twiki.cern.ch/twiki/bin/view/Main/ExoMCInstructions#Central_LHE_requests

Preliminary Steps and Knowledge

CERN Account

A CERN account is needed to access LXPLUS and some twiki pages, which are linked and referenced throughout this document and are incredibly helpful in troubleshooting issues. To create a CERN account, go to the following link: https://account.cern.ch/account/Management/NewAccount.aspx and request to make an account, filling out the necessary information. You will receive a username, usually your first-name initial and full last-name, and will then create your own password. This username and password is used to login to LXPLUS as well as some of the twiki pages.

Linux Command Line Basics

Command	Description
ssh [options] [user@hostname]	"Secure Shell" Access a server with a login
exit	Terminates server connection
mv [-f|-i|-n|-v] [-v] source target	"Move" Moves a file(s) to a target
mkdir [-pv] [-m] directory ...	"Make Directory" Makes a directory and names it
cd [-L|-P] [dir] cd ../ cd ~	"Change directory" Navigate through files and directories Moves to a parent directory Moves to home directory
ls [target]	"list" Lists files in current or targetted directory
nano [options] filename	Opens and allows edits to a file
pwd	"Print Working Directory"
rm [options] [target]	"Remove" Deletes directories and/or contents and/or files CAUTION: THERE IS NO "UNDO" FOR THIS
touch [target]	Creates a blank new file

Accessing and Using LXPLUS

LXPLUS Overview

LXPLUS (Linux Public Login User Service) is the interactive logon service to Linux for all CERN users. The cluster LXPLUS consists of public machines provided by the CERN IT Department for interactive work. LXPLUS mainly runs the CC7 (CERN CentOS 7) version of 64-bit Linux, although there are machines running C8 (CentOS8). The SLC6 (Scientific Linux CERN 6) version was retired in late 2020 and is no longer supported. Some of the programs used in creating Gridpacks are meant to run on SLC6 only, however there are updated versions of these programs that will run on CC7 or C8.

Accessing LXPLUS Using a Unix-Based Terminal

In a Unix-based Terminal (such as Linux or MacOS), type the following to access lxplus:

```
> ssh -Y username@lxplus.cern.ch
```

Followed by your password.

For specific login (such as to lxplus7 or lxplus8), use username@lxplus7.cern.ch or username@lxplus8.cern.ch instead. Do not specify a specific grid location however, as the

Accessing LXPLUS Using Windows

Some Windows machines requires the download of Putty, which can be downloaded for free from: Get Putty (Unofficial) - Microsoft Storewww.microsoft.com › en-us › putty-unofficial

To install the Windows 10's SSH Client, go to Settings > Apps and click "manage optional features" under Apps & Features. Click "Add a feature" and scroll down to the "OpenSSH Client (Beta)" option and click it, then clicking "Install". More information on this at the following link: https://www.howtogeek.com/336775/how-to-enable-and-use-windows-10s-built-in-ssh-commands/

MadGraph5_aMC@NLO

"MadGraph5_aMC@NLO is a framework that aims at providing all the elements necessary for Standard Model (SM) and Beyond Standard Model (BSM) phenomenology, such as the computations of cross sections, the generation of hard events and their matching with event generators, and the use of a variety of tools relevant to event manipulation and analysis. Processes can be simulated to LO accuracy for any user-defined Lagrangian, an the NLO accuracy in the case of models that support this kind of calculations -- prominent among these are QCD and EW corrections to SM processes. Matrix elements at the tree- and one-loop-level can also be obtained.

MadGraph5_aMC@NLO is the new version of both MadGraph5 and aMC@NLO that unifies the LO and NLO lines of development of automated tools within the MadGraph family. It therefore supersedes all the MadGraph5 1.5.x versions and all the beta versions of aMC@NLO."

GitHub

GitHub Account

Head to https://www.github.com and create an account, this is needed to access the cms-sw/genproductions repository and contribute to it.

Basics of Github

There

- git init
  - Creates a new or Reinitializes an existing Git Repository
  - https://git-scm.com/docs/git-init
- git clone
  - Clones a repository
  - https://git-scm.com/docs/git-clone
- git add
  - Add contents to the index
  - https://git-scm.com/docs/git-add
- git commit
  - Records changes to the repository
  - https://git-scm.com/docs/git-commit
- git status
  - Shows the current status of the working tree
  - https://git-scm.com/docs/git-status
- git branch
  - List, create, or delete branches (
  - https://git-scm.com/docs/git-branch
- git merge
  - Join multiple development histories together
  - https://git-scm.com/docs/git-merge
- git pull
  - Fetch from another repository or branch
  - https://git-scm.com/docs/git-pull
- git push
  - Update the remote refs with associated objects
  - https://git-scm.com/docs/git-push

Forking a Repository

This visual guide explains how to fork a Github repository:
https://docs.github.com/en/github/getting-started-with-github/fork-a-repo

Naming Conventions

The naming convention defines the primary dataset (PD) name, which allows for all MC sets submitted to CERN to be more organized. It's not possible to find a general rule for every single detail that can make up the physics contents of a sample and represent it as a general string, however the general rules are as follows (a more detailed and specific explanation is given in the link on the header of this section)

The convention is currently defined as
PROCESS_RANGETYPE-RANGELOWToRANGEHIGH_FILTER_TUNE_COMMENT_COMENERGY-GENERATOR

Descriptions:

- **PROCESS** is the Physics process in the sample. If one or more specific decays are simulated, the convention is [ParticleName(s)]To[ParticleName(s)]To[ParticleName(s)]... for any length of specified decay. Charge and Anti-Particle-ness are only specified if relevant, such as Wplus for W+ if only used in a sample (but not WplusWminus for W pair-production). If in one part of the process more than one particle of the same kind use the convention [number][ParticleName]... for each specific particle rather than all of the same particles.
    - Particle Names: All stated particles start with an uppercase letter followed by lowercase letters
    - Examples of Processes: QCD, TT, ZdToSd, HToWWToL2Nu, ZdTo2Fd1
- **RANGETYPE** is the variable the sample is binned in; **RANGELOW** is the lower bound of that range in GeV; **RANGEHIGH** is the upper bound of that range in GeV
    - Example: Pt-100To200, Pt-50
- **FILTER** denotes information on additional filters applied, such as GEN level filters
    - Examples: EMEnriched, MuEnriched
- **TUNE** is the underlying-event tune
    - Example: TuneZ2Star
- **COMMENT** for additional comments
    - Example: "OnlyQCDProduction"
- **COMENERGY** center of mass collision energy (with appropriate units). DO NOT use floats
    - Example: 7TeV, 990GeV, 2360GeV
- **GENERATOR** is the generator used
    - Example: pythia6, pythia8, herwig6, sherpa, powheg

Decimal Places

If one were to use decimal places in the naming convention (such as MZd91.1876), use a “p” instead of a “.” (such as MZd91p1876)

The Ingredients

Gridcards

Gridcards are the .dat files that are used to create Gridpacks, which contain various parameters and foundations for producing samples. XXX is filled in for the naming convention used. The main Gridcards used to create the Gridpacks are:

XXX_proc_card.dat	Declares the process to be generated. MANDATORY. Includes what is being decayed (and its definition if needed), what is being produced, what model to use, and what output file to make.
XXX_run_card.dat	Declares particular options on how the generator will run and generate the process, as well as specific kinematic cut values. MANDATORY. Identifies specific physical parameters, how many events, etc
XXX_madspin_card.dat	Declares how MadSpin will decay specific particles. Optional.
XXX_customizecards.dat	Declares the various physical parameters to be set, such as mass of particles and decay widths. Use this instead of XXX_param_card.dat when submitting samples (the system will create the param_cards automatically). Each line generally follows as: *set GRIDCARD PARAMETER PDGID VALUE* Such as *set param_card mass 1023 500*
XXX_extramodels.dat	Declares the custom models used in the sample. Note: the models referenced in this Gridcard are to be uploaded on https://cms-project-generators.web.cern.ch/cms-project-generators/ prior to submitting the samples to CERN to produce

The following cards are always needed: proc_card.dat and run_card.dat

Specific Noted Lines in Gridcards

proc_card.dat

- *import model MODEL_NAME*
    - Imports a model from that specified in extramodels.dat
    - Such as *import model unzipped_model_name*
- *generate interaction_and_decay_process*
    - Tells Generator what to create and decay
- *output XXX -nojpeg*
    - Tells how to name the output in creating LHE files

run_card.dat

- *### = nevents ! Number of unweighted events requested*
    - This number is used to create the Gridpack and is not necessarily the number of events produced (as seen in the spreadsheet).
    - The recommended number of events here is around 5,000 in order to produce a Gridpack
- For newer productions (2017 and more recent), in run_card.dat replace the following lines under # PDF CHOICE
    - *lhapdf = pdlabel ! PDF set*
      292000 = lhaid
    - With
    - *$DEFAULT_PDF_SETS = lhaid*
      $DEFAULT_PDF_MEMBERS = reweight_PDF

customizecards.dat

- *set GRIDCARD PARAMETER PDGID VALUE*
    - Sets specific parameters used in the
    - Such as: *set param_card mass 1023 500*

Extramodels.dat

- *model_name_as_listed_in_cms_project_generators_site*
    - Used to import the specific model from the cms-project-generators website

PDGID

The Particle Data Group (PDG) numbering scheme (PDGID or PDG ID) is the commonly accepted convention for numbering particles used in MC simulations, including hypothetical particles included beyond the Standard Model. All particle ID numbers need to be their respective PDGID numbers for MC Requests.

Gridpacks

MadGraph requests need to be generated Gridpacks, which are created from Gridcards using a premade automated script, found in the CMS GitHub.

- *./gridpack_generation.sh <name of process card without _proc_cards.dat> <folder containing cards relative to current location>*

For newer productions (2017 and more recent), use the gridcard_generation.sh script used in this specific branch: https://github.com/cms-sw/genproductions/tree/UL2019 (The one found in the main cms-sw/genproductions repository is for SLC6, which is no longer in service)

Models

Especially when using Beyond the Standard Model Models, the model generation files must be uploaded to the generators repositories.

Note: the models referenced in this Gridcard are to be uploaded on the CMS Project Generators website https://cms-project-generators.web.cern.ch/cms-project-generators/ prior to submitting the samples to CERN to produce. If they are not on the website, mention this to one of the moderators of the cms-sw/genproductions Github repository or email one of the MC&I conveners, who will take care of it.

Hadronizer Fragments

The generic Hadronizer Fragments are used found in the cms-sw/genproductions repository, however specialty Fragments can be made and uploaded to either your private Github repository or included in the original pull request.

The Process

MC&I Presentation

1. Email the MC&I Conveners and schedule a presentation on:
    a. Overview of Model
    b. Brief Summary of Analysis
    c. Generator Settings
    d. Distributions of Relevant Generator-Level Variables
    e. Specific Parameters of the Request
2. Wait for Approval from MC&I Conveners

Organize Gridcards, Script

1. Create Directories with the base Gridcards contained in them, naming them a generic PD name.
2. Create a Script which takes the Directories and its base Gridcards and creates all other Gridcards in their respective, specific PD directories.
    a. This is where basic programming skills come into play, especially shell scripting. For MC Requests that have more than a few different sets of Gridcards, the moderators at the cms-sw/genproductions github will want only the base Gridcards uploaded and a script to create the rest. An example that creates the Gridcards for a larger set, runs the script in LXPLUS that creates the Gridpacks, then cleans the directories in LXPLUS for each combination in the set is found here (Gridpack.sh)
    b. An example for the 2020 MC Request with all of the cards and scripts are found here: https://github.com/JacobChesslo/Research/tree/master/Gridpack%20Automation/LXPLUS%20Gridpack%20for%20MC%20Request%201.1
3. Verify that the Gridcards can be run in the Generator used and that the Script works successfully in LXPLUS.

Fork Github and Upload

1. Fork your Github to the cms-sw/genproductions GitHub
2. Add only the Base Gridcards (those used to create all others with the script) in their PD directory(ies) and the Script(s) into your forked repository under their specific Generator and subdirectories (such as under /bin/MadGraph5_aMCatNLO/cards/production/2017/13TeV/) by committing them to your private fork of genproductions
    a. It's important to put the Gridcards into a similar directory structure as you see in genproductions
3. Add Hadronizer Fragments to the Fragment Repository in a similar directory structure as you see in genproductions (such as in genproductions/python/ThirteenTeV) if they haven't already been added to the Fragment Repository.

    a. If the Hadronizer Fragments won't be included in the original pull request, upload them in your private repo and link this in the Spreadsheet later on.
4. Create a pull request for the Gridcards
    a. Be sure to include as much information in the description of this pull request as needed
5. Wait for the pull request to be accepted
    a. If the moderators of the cms-sw/genproductions repository see something that doesn't look quite right, you will be notified of the relevant changes to be made or asked questions about the request.

Gridpack Creation

1. Create the Gridpacks for each process
    a. Usually accomplished on LXPLUS7, however there are other options if desired (see
    b. An automatic script is easiest to produce many gridpacks, as each script will need to be submitted for production for each gridpacks. A copy of such a script for LXPLUS7 is found here, where Gridpack.sh should be ran inside of a copy of genproductions inside of your personal directory, where the Gridpack.sh and files should be placed in ~/genproductions/bin/MadGraph5_aMCatNLO/

Email the Spreadsheet

1. Create a copy of this spreadsheet, preferably in some accessible drive like Google Drive's Sheets or Dropbox so it can be shared, and fill it with the following information:
    a. Dataset name
        i. Will form part of the dataset name in DAS. See ProductionDataSetNames for proper name formatting
    b. Number of Events
        i. The number of events per each sample requested
        ii. This can be different from the number of events specified in the run_card.dat (the latter of which is used just to create the Gridpack, such as producing a Gridpack using a run_card specification of 5,000 however requesting 100,000 in the spreadsheet)
    c. Name of Hadronizer Fragment
        i. Either a URL (eg GitHub or Dropbox) or a file on lxplus
        ii. Generic Hadronizer Fragments are already on the genproductions repository
    d. Gridpack Location
        i. The complete path and file name of each Gridpack, can be put in one's the public repository on LXPLUS
    e. Gridpack's Card's URL
        i. The URL to the Gridpack's cards in the genproductions repository on GitHub.
        ii. Link to a specific commit rather than the master branch as the cards in the latter might change over time
        iii. Good Example Link

iv. Bad Example Link

f. Generator
   i. Set to “madgraph” or “powheg”
g. Cross Section (in picobarns [pb])
   i. Add this information if the samples requested are intended for use by multiple groups (eg backgrounds) or you have independent estimates of the cross section.
   ii. This can be left out for all other requests and the cross section will be set to 1 in McM.
h. Efficiency
   i. The ratio of events actually produced in the output ROOT file to the number you ran over
   ii. If the value is near 1, you do not need to include this information. Otherwise, add columns for efficiency and efficiency error to the spreadsheet

2. Send the editable spreadsheet to cms-exo-mcrequests@cernNOSPAMPLEASE.ch
   a. If the Hadronizer Fragment is not already in the genproductions repository, send the generator fragment as well.
   b. For priority samples (those requested to be produced before others), send the MC Conveners a list of those to be prioritized.

MC Request Completion

1. You will be assigned an MC Contact who will take over the MC Request once the Spreadsheet is complete.
2. Central Production takes at minimum months to complete.

Sources

CERN Account Creation
LXPLUS
SSHing on Windows, Putty Tutorial https://www.ssh.com/ssh/putty/windows/

Exotica MC Request Instructions twiki
https://twiki.cern.ch/twiki/bin/view/Main/ExoMCInstructions#Central_LHE_requests
MadGraph5_aMCatNLO Production Page
https://twiki.cern.ch/twiki/bin/viewauth/CMS/QuickGuideMadGraph5aMCatNLO#Software_MadGraph5_a_MCatNLO_Pro

Madgraph:
Homepage: http://madgraph.phys.ucl.ac.be
Download: https://launchpad.net/mg5amcnlo

cms-sw/genproductions Github Repository
https://github.com/cms-sw/genproductions

Copy of Hohlmann Group 2020 Central Production MC Request
https://github.com/JacobChesslo/Research/tree/master/Gridpack%20Automation/LXPLUS%20Gridpack%20for%20MC%20Request%201.1