

CMS Analysis demo

Georgia Karapostoli (CERN/University of Athens)

LHCC comprehensive review

Demo session

Sep 25, 2006

CERN



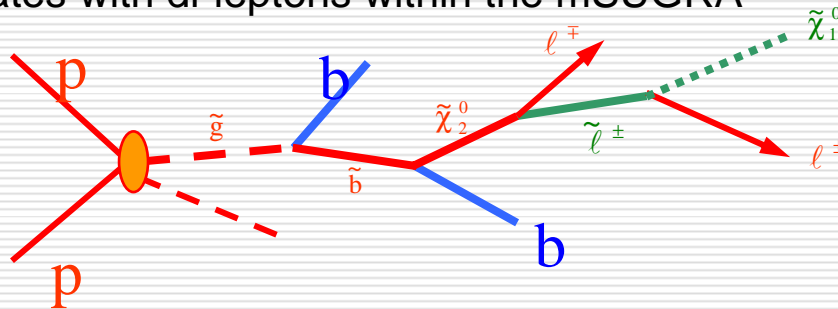
CMS Analysis from a physicist's view

User's workflow:

- Determine the physics process.
- Use CMS software to perform the analysis: ORCA (Object-oriented Reconstruction and Analysis software) with dedicated analysis application ExRootAnalysis.
- Determine the data set on which to run the analysis.

Example of a SUSY Analysis:

- Study signal observability of final states with di-leptons within the mSUGRA framework of Supersymmetry.



- Requires reduction of Standard Model background (ttbar, Z/W, di-bosons, QCD etc)
- **Needs large amount of simulated data and massive computational resources.**



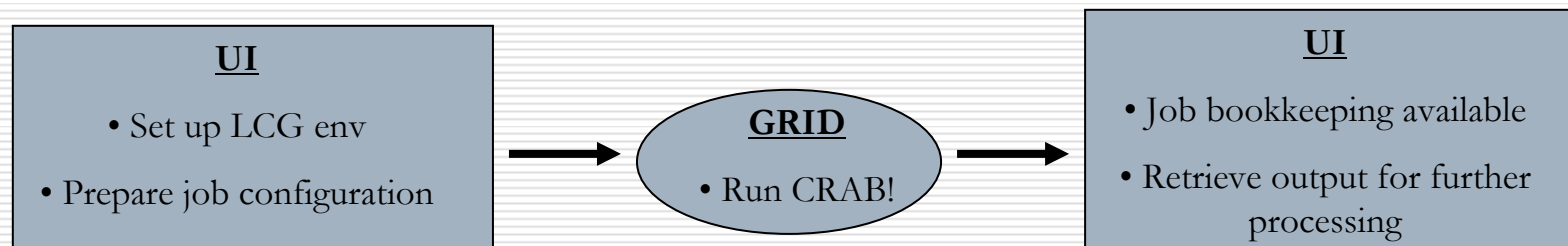
Introduction to the tool

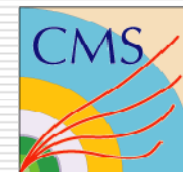
MOVING THE ANALYSIS TO THE GRID

- **Must be registered to the CMSVO server.**
- **Must be on a LCG User Interface (LCG UI)** (any machine that LCG middleware is installed)
- **CRAB (CMS Remote Analysis Builder) is an interface tool intended to simplify the process of creation and submission of CMS analysis jobs on the Grid environment.**

User's task to run analysis with CRAB:

- Test the analysis locally (analyze one file copied from dataset resident on a remote storage)
- Determine how to split jobs (must finish in reasonable time).
- Modify the CRAB configuration file.
- Run CRAB (submit and monitor job, retrieve output commands etc).





Job preparation to run on the Grid

□ INTEGRATION WITH EXPERIMENT'S DATA MANAGEMENT

- CRAB uses a Data Discovery System to find the right data to access through DBS (Dataset Bookkeeping System) → *user must choose among existing named datasets.*

jm03b_TTbar_leptonic	jm_Hit245_2_g133	975959	FNAL
jm03b_TTbar_leptonic	jm_Hit245_2_g133	986000	CERN
jm03b_Wjets_0_20	jm_2x1033PU761_TkMu_2_g133_OSC	0	IN2P3
jm03b_Wjets_0_20	jm_2x1033PU761_TkMu_2_g133_OSC	371000	CERN

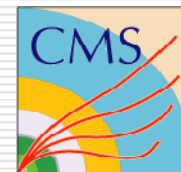
□ JOB PREPARATION (in CRAB cfg file)

Declare the input data (user consults a prototypal page which lists all data available with DBS). *The user does not need to know where the dataset is located, CRAB takes care for him/her.*

```
dataset = jm03b_Ttbar_inclusive
owner = jm_2x1033PU761_TkMu_2_g133_OSC
```

Specify splitting parameters (total number of events, events per job).
Analyze dataset with n parallel jobs.

```
total_number_of_events = -1 (all available)
events_per_job = 500
```



Job preparation (cont.)

Name of executable (specific to software application). *All software-specific libraries the job needs to run are transferred automatically by CRAB.*

```
executable = ExRootAnalysis
```

Output file list (names of the output files the application produces)

```
output_file = test.root
```

Option to retrieve output (output returned to UI or copied to a Storage Element SE). *For large outputs, the user should choose to store them in a SE (in this case handling of output gets asynchronous with job completion...)*

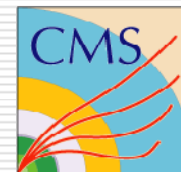
```
return_data = 1
```

```
copy_data = 0
```

Computing Element black/white list

ce_black_list (refuse access to CEs listed)

ce_white_list (allow access preferably to the CEs listed)

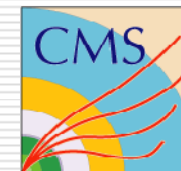


Job Submission

- ❑ **Set up LCG environment**
- ❑ **Set up software-specific environment**
- ❑ **Set up CRAB related environment**

- ❑ **RUN CRAB**

```
crab -create # Create all jobs. No submission!  
crab -submit 2 -continue [ui_working_dir] # Submit 2 jobs, the  
  ones already created (-continue)  
crab -create 2 -submit 2 # Create _and_ submit 2 jobs  
crab -status # Check the status of all jobs  
crab -getoutput # Get back the output of all jobs
```

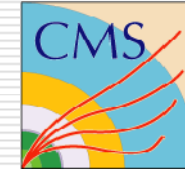


Example of real activity done with CRAB

• DATA ANALYZED

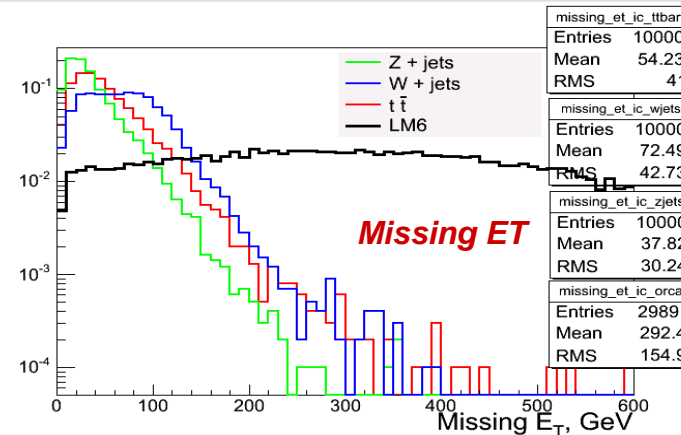
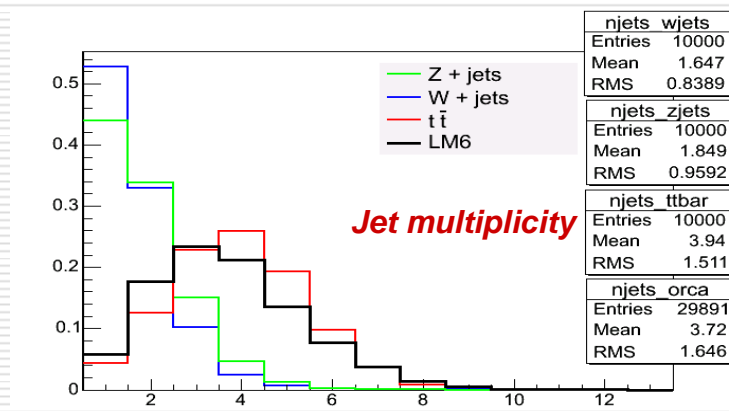
DATASET	DATASET SITE LOCATION	X-SEC (pb)	# EVENTS (L = 10 fb ⁻¹)	# EVENTS ANALYZED	NUMBER OF JOBS (= # output files)
su05_pyt_lm6	CERN	4	40 K	100 K	200
su05_pyt_LM1	CERN	41	410 K	400 K	800
jm03b_Ttbar_inclusive	FNAL, CERN, CNAF, DESY	492	5 M	3000 K	6000
jm03b_Zjets_85_150	FNAL, CERN, DESY	576	5800 K	290 K	580
jm03b_Zjets_150_250	FNAL, CERN, DESY	162	1600 K	150 K	300
jm03b_Wjets_85_150	FNAL, CERN, DESY	4300	43 M	570 K	1140
jm03b_Wjets_150_250	FNAL, CERN, DESY	1200	12 M	290 K	580
jm03b_WWjets_leptonic	FNAL, DESY	20	200 K	100 K	200
jm03b_WZjets_inclusive	FNAL, DESY	27	270 K	280 K	560
jm03b_ZZjets_inclusive	FNAL, DESY	11	110 K	400 K	800

- **Typical job properties:** takes a few hours (~ 3h) to run over 500 events (1job), ~1day to complete 200 jobs.
- **Grid problems may arise:** i) some jobs crash in periods of high throughput of jobs, ii) CEs down, have to consult BDII of sites for status of CEs.

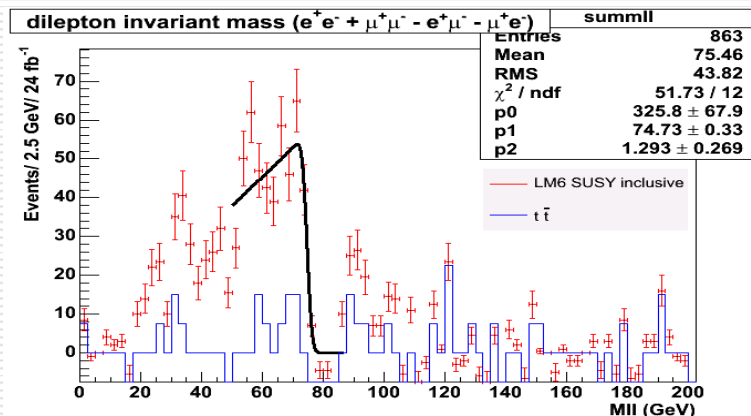


Some Physics Results

- **Output ROOT files contain collections of physics objects** (electrons, muons, Jets etc)
 - can be further processed with ROOT to plot p_T , invariant mass etc distributions.



- **Di-lepton invariant mass** distribution in supersymmetric events (mSUGRA LM6 benchmark point):



- Number of SUSY events used correspond to $\sim 24\text{fb}^{-1}$.
- Needed to rescale ttbar distribution (1600K events used only).
- **Needed to increase statistics of samples to account for large statistical uncertainties...**



Conclusion

□ **Advantages of the tool**

- user does not need to know where datasets reside (CRAB interacts with DBS to discover data)
- user does not need to interact with Grid directly (no worry about LCG commands, JDL files etc)
- user does not need to know which CEs match the requirements of the job (CPU power/time, CMS software installed etc)
- handles to check validity of user's proxy and renewal.
- handles to copy output to a SE in case of large output size.

□ **Currently, user must:**

- Check BDII of sites (status of CEs, versions of CMS software installed, publishing a Tag etc).
- Determine the output file size if returned to UI (maximum size limit when transferred by output sandbox).