

TOTCSI

Configuration for Reconstruction

Igor Jurkowski & Radosław Lazarz

August 28, 2012

Contents

1	How to start configuration	2
2	Parameter explanations	2
2.1	General parameters	2
2.1.1	config.task_type	2
2.1.2	config.cmssw_dir	2
2.1.3	config.map_output	2
2.1.4	config.reduce_output	2
2.1.5	config.save_using_pool	3
2.2	TOTCSI workspace parameters	3
2.2.1	config.workspace.create	3
2.2.2	config.workspace.map_split_dir	3
2.2.3	config.workspace.reduce_split_dir	3
2.2.4	config.workspace.root_dir	3
2.2.5	config.workspace.shell_template	4
2.3	Input parameters	4
2.3.1	config.input_config.map_path	4
2.3.2	config.input_config.reduce_path	4
2.4	Reconstruction parameters	4
2.4.1	config.reconstruction.input_data	4
2.4.2	config.reconstruction.files_to_exclude	5
2.4.3	config.reconstruction.events_per_file_to_reconstruct	5
2.4.4	config.reconstruction.estimated_number_of_events_per_file	5
2.4.5	config.reconstruction.events_per_job	5
2.4.6	config.reconstruction.files_per_reduce	6
2.4.7	config.reconstruction.reduce_per_run	6
2.4.8	config.reconstruction.map_tuning_function	6
2.4.9	config.reconstruction.reduce_tuning_function	7
2.5	bsub parameters	7
2.5.1	Default bsub parameters for map and reduce	7
2.5.1.1	config.bsub.pool_size	7
2.5.1.2	config.bsub.swap_size	7
2.5.1.3	config.bsub.queue	7
2.5.2	bsub parameters for map phase	8
2.5.2.1	config.bsub_map.pool_size	8
2.5.2.2	config.bsub_map.swap_size	8
2.5.2.3	config.bsub_map.queue	8
2.5.3	bsub parameters for reduce phase	8
2.5.3.1	config.bsub_reduce.pool_size	8
2.5.3.2	config.bsub_reduce.swap_size	9
2.5.3.3	config.bsub_reduce.queue	9
3	Configuration example	9

1 How to start configuration

Use the following lines at the beginning of your configuration file in order to import configuration class with all default options.

```
from totcsi.totcsi_configuration_logic import *
config = TOTCSIConfiguration()
```

2 Parameter explanations

2.1 General parameters

2.1.1 config.task_type

Type of task to be run

Values:

- "Reconstruction"
- "Simulation"

```
# Default:
config.task_type = "Reconstruction"
```

2.1.2 config.cmssw_dir

CMSSW location (full path), beginning with "/"
Has to be given by user!

```
# Example:
config.cmssw_dir = "/afs/cern.ch/exp/totem/scratch... \
...Release/raw_data_reco/rev5679/CMSSW_4_2_4"
```

2.1.3 config.map_output

Output directory for files (logs, results, configurations) for map phase
Can be both CASTOR, EOS and AFS path
It has to be set by user!

```
# Example:
config.map_output = "/castor/cern.ch/user/.../MapOutputFolder"
```

2.1.4 config.reduce_output

Output directory for files (logs, results, configurations) for reduce phase
Can be both CASTOR, EOS and AFS path
It has to be set by user!

```
# Example:
config.reduce_output = "/castor/cern.ch/user/.../ReduceOutputFolder"
```

2.1.5 `config.save_using_pool`

Flag saying whether output should be saved first to pool and then copied to output location (might help with errors regarding CASTOR saving scripts) or directly to output location.

Values:

- True - save with a use of pool (Default)
- False - try to directly save to output location

```
# Default:  
config.save_using_pool = True
```

2.2 TOTCSI workspace parameters

2.2.1 `config.workspace.create`

Flag specifying whether workspace directories should be created if they do not exist.

```
# Default:  
config.workspace.create = False
```

2.2.2 `config.workspace.map_split_dir`

Folder with map splitted configurations

Relative to `root_dir`

It will contain the map configurations generated by TOTCSI

```
# Default:  
config.workspace.map_split_dir = "map_split"
```

2.2.3 `config.workspace.reduce_split_dir`

Folder with reduce splitted configurations

Relative to `root_dir`

It will contain the reduce configurations generated by TOTCSI

```
# Default:  
config.workspace.reduce_split_dir = "reduce_split"
```

2.2.4 `config.workspace.root_dir`

Base directory for all configuration files of TOTCSI

Can be both relative and full path

```
# Default:  
config.workspace.root_dir = "."
```

2.2.5 config.workspace.shell_template

Path to shell template. It is filled by TOTCSI and then submitted to computing cluster as job.

```
# Default:  
config.workspace.shell_template = "totcsi/templates/job_template.sh"
```

Default template looks like:

```
#!/bin/sh  
export RFIO_USE_CASTOR_V2=YES  
export STAGE_HOST={{stage_host}}  
export STAGE_SVCCLASS=default  
export SCRAM_ARCH=slc5_amd64_gcc434  
source /afs/cern.ch/cms/cmsset_default.sh  
cd {{CMSSW_path}}  
eval `scram runtime -sh`  
cd -  
{{rm_type}} {{file_to_remove}}  
{{mkdir_type}} {{dir_to_create}}  
cmsRun {{python_conf_file}}  
{{optional_copy_command}}
```

2.3 Input parameters

2.3.1 config.input_config.map_path

Path to map template (production)

Has to be given by user!

```
# Usage:  
config.input_config.map_path = "/afs/.../MapItTemplate.py"
```

2.3.2 config.input_config.reduce_path

Path to reduce template (validation)

Has to be given by user!

```
# Usage:  
config.input_config.reduce_path = "/afs/.../ReduceItTemplate.py"
```

2.4 Reconstruction parameters

2.4.1 config.reconstruction.input_data

Files to be taken into reconstruction.

Values:

- Path (relative or non-relative) to file containing in each line one path to input data
- List of paths to input data
- List of run numbers from which file paths will be extracted

Those values can be appended to each other, so for example user can specify some of the input data by direct paths and some by run numbers.

```
# Example with file:
config.reconstruction.input_data = paths_from_file("input.txt")
  # paths to files from input.txt file
  # (which is stored in main TOTCSI directory)
# Example with run numbers:
config.reconstruction.input_data = paths_from_run_numbers([6011, 6012] + range(6020, 6030))
```

2.4.2 config.reconstruction.files_to_exclude

Files that should not be reconstructed (they will be excluded from the input_data). Takes same format as input_data.

```
# Example (from list):
config.reconstruction.files_to_exclude = [""] # no files to exclude
# Example 2 (from file):
config.reconstruction.files_to_exclude = paths_from_file("exclude.txt")
```

2.4.3 config.reconstruction.events_per_file_to_reconstruct

Special option which gives user the ability to only reconstruct some of the events from each run file in order to get a peek at data stored in files.

Values:

- -1 - meaning all events (Default)
- Positive number - amount of events per file to reconstruct (from the beginning of the file)

```
# Default:
config.reconstruction.events_per_file_to_reconstruct = -1
```

2.4.4 config.reconstruction.estimated_number_of_events_per_file

How many events are in one input file (estimation made by user).

This information is needed for splitting one file into multiple jobs. If user wants to use the reconstruction.events_per_job option this has to be specified. It takes a positive integer as value.

```
# Example:
config.reconstruction.estimated_number_of_events_per_file = 50000
# Meaning that each file should have around 50000 events in it
# (last file of a run can have less)
```

2.4.5 config.reconstruction.events_per_job

How many events can be processed in one job at once.

Values:

- -1 - no maximum value for the amount of events per job (Default)
- Positive number - maximal amount of events processed in one job

```
# Default:
config.reconstruction.events_per_job = -1
```

2.4.6 `config.reconstruction.files_per_reduce`

How many files should be used in one reduce operation.

Values:

- -1 - as many files reduced in one operation as possible (all files with `reduce_per_run` set to `False`)
- Positive number - number of files that will be used in one reduce operation

Default:

```
config.reconstruction.files_per_reduce = -1
```

2.4.7 `config.reconstruction.reduce_per_run`

Should the reduce phase take as input all files or take into consideration run numbers and reduce input files for each run independently.

Values:

- `True` - take run numbers into consideration (Default)
- `False` - reduce different runs together

Default:

```
config.reconstruction.reduce_per_run = True
```

2.4.8 `config.reconstruction.map_tuning_function`

Function for managing runs' templates. User specifies function taking templates as input, which is a mapping with `key = run_number` and `value = configuration_file`. The function then makes changes in configuration so that each run has configuration tuned for it's reconstruction (map phase).

By default there are no changes in given templates.

Example:

```
def tune_map_configuration(map_templates):
    #appending
    for run in range(6908, 6928):
        map_templates[run] += '''
        toberemoved = []
        for xmlfile in process.XMLIdealGeometryESSource.geomXMLFiles:
            if xmlfile.endswith("RP_Dist_Beam_Cent.xml"):
                toberemoved.append(xmlfile)
        for xmlfile in toberemoved:
            process.XMLIdealGeometryESSource.geomXMLFiles.remove(xmlfile)
        process.XMLIdealGeometryESSource.geomXMLFiles.append("Geometry/TotemRPData... \
            ../data/2010_10_29-30_offsets/RP_Dist_Beam_Cent.xml")
        process.TotemRPGeometryESModule = cms.ESProducer("TotemRPGeometryESModule")
        '''
    #substitution
    for run in range(6190, 6213):
        map_templates[run] = Template(map_templates[run]).safe_substitute(GEOMETRY=... \
            ..."Geometry/TotemRPData/data/2010_10_29-30_offsets/RP_Dist_Beam_Cent.xml")

config.reconstruction.map_tuning_function = tune_map_configuration
```

2.4.9 config.reconstruction.reduce_tuning_function

Function for managing runs' templates. User specifies function taking templates as input, which is a mapping with key = run_number and value = configuration_file. The function then makes changes in configuration so that each run has configuration tuned for it's reconstruction (reduce phase).

By default there are no changes in given templates.

Example:

```
def tune_reduce_configuration(reduce_templates):
    #appending
    for run in range(6908, 6928):
        reduce_templates[run] += '''
        toberemoved = []
        for xmlfile in process.XMLIdealGeometryESSource.geomXMLFiles:
            if xmlfile.endswith("RP_Dist_Beam_Cent.xml"):
                toberemoved.append(xmlfile)
        for xmlfile in toberemoved:
            process.XMLIdealGeometryESSource.geomXMLFiles.remove(xmlfile)
        process.XMLIdealGeometryESSource.geomXMLFiles.append("Geometry/TotemRPData... \
        .../data/2010_10_29-30_offsets/RP_Dist_Beam_Cent.xml")
        process.TotemRPGeometryESModule = cms.ESProducer("TotemRPGeometryESModule")
        '''
    #substitution
    for run in range(6190, 6213):
        reduce_templates[run] = Template(reduce_templates[run]).safe_substitute(GEOMETRY=... \
        ..."Geometry/TotemRPData/data/2010_10_29-30_offsets/RP_Dist_Beam_Cent.xml")

config.reconstruction.reduce_tuning_function = tune_reduce_configuration
```

2.5 bsub parameters

2.5.1 Default bsub parameters for map and reduce

2.5.1.1 config.bsub.pool_size

Pool size for the job (integer)

```
# Default:
config.bsub.pool_size = 30000
```

2.5.1.2 config.bsub.swap_size

Swap size for the job (integer)

```
# Default:
config.bsub.swap_size = 2000
```

2.5.1.3 config.bsub.queue

Type of the queue to be used

Values:

- 1nh

- 8nh
- 1nd
- 2nd
- 1nw

```
# Default:  
config.bsub.queue = "1nh"
```

2.5.2 bsub parameters for map phase

2.5.2.1 config.bsub_map.pool_size

Pool size for the job (integer) for map phase (overrides config.bsub parameter)

```
# Default:  
config.bsub_map.pool_size = 30000
```

2.5.2.2 config.bsub_map.swap_size

Swap size for the job (integer) for map phase (overrides config.bsub parameter)

```
# Default:  
config.bsub_map.swap_size = 2000
```

2.5.2.3 config.bsub_map.queue

Type of the queue to be used for map phase (overrides config.bsub parameter)

Values:

- 1nh
- 8nh
- 1nd
- 2nd
- 1nw

```
# Default:  
config.bsub_map.queue = "1nh"
```

2.5.3 bsub parameters for reduce phase

2.5.3.1 config.bsub_reduce.pool_size

Pool size for the job (integer) for reduce phase (overrides config.bsub parameter)

```
# Default:  
config.bsub_reduce.pool_size = 30000
```


2.5.3.2 config.bsub_reduce.swap_size

Swap size for the job (integer) for reduce phase (overrides config.bsub parameter)

```
# Default:
config.bsub_reduce.swap_size = 2000
```

2.5.3.3 config.bsub_reduce.queue

Type of the queue to be used for reduce phase (overrides config.bsub parameter)

Values:

- 1nh
- 8nh
- 1nd
- 2nd
- 1nw

```
# Default:
config.bsub_reduce.queue = "1nh"
```

3 Configuration example

```
from totcsi.totcsi_configuration_logic import *

config = TOTCSIConfiguration()

config.task_type = "Reconstruction"

config.cmssw_dir = "/afs/cern.ch/i/ijurkows/cmssw/CMSSW_4_2_4"

config.map_output = "/castor/cern.ch/user/i/ijurkows/output"

config.save_using_pool = False #save right away

#setting the splitter workspace
config.workspace.create = True
config.workspace.root_dir = "/afs/cern.ch/i/ijurkows/private/MyTOTCSIWorkspace"
config.workspace.map_split_dir = "SplittedMaps"
config.workspace.reduce_split_dir = "SplittedReduces"

config.input_config.map_path = "/afs/cern.ch/user/i/ijurkows/public/\
    configurations_for_map/RPElasticbeta90energy3p5TeV2200october2011a3_cfg.py"
config.input_config.reduce_path = "/afs/cern.ch/user/i/ijurkows/public/\
    configurations_for_reduce/valRPElasticbeta90energy3p5TeV2200october2011a3_cfg.py"

config.bsub.pool_size = 20000
config.bsub_map.pool_size = 26000
```

```

# input.txt - file with path to data in each line
config.reconstruction.input_data = paths_from_file("input.txt")

config.reconstruction.files_to_exclude = \
["/castor/cern.ch/totem/LHCRawData/2012/Physics/run_0008.001.vmea", \
"/castor/cern.ch/totem/LHCRawData/2012/Physics/run_0008.002.vmea"]
def tune_map_configuration(map_templates):
    #appending
    for run in range(6908, 6928):
        map_templates[run] += '''
        toberemoved = []
        for xmlfile in process.XMLIdealGeometryESSource.geomXMLFiles:
            if xmlfile.endswith("RP_Dist_Beam_Cent.xml"):
                toberemoved.append(xmlfile)
        for xmlfile in toberemoved:
            process.XMLIdealGeometryESSource.geomXMLFiles.remove(xmlfile)
        process.XMLIdealGeometryESSource.geomXMLFiles.append("Geometry/\
        TotemRPData/data/2010_10_29-30_offsets/RP_Dist_Beam_Cent.xml")
        process.TotemRPGeometryESModule = cms.ESProducer("TotemRPGeometryESModule")
        '''
    #substitution
    for run in range(6190, 6213):
        map_templates[run] = Template(map_templates[run]).safe_substitute(GEOMETRY= \...
        ..."Geometry/TotemRPData/data/2010_10_29-30_offsets/RP_Dist_Beam_Cent.xml")

config.reconstruction.map_tuning_function = tune_map_configuration

```