HOW TO USE NAWAT_AGATA FOR DATA REPROCESSING ON THE GRID
EGAN SCHOOL 2012 GSI 3-7 DEC. 2012

This file is to describe how to use the nawat package to perform PSA and MERGING/TRACKING data processing for AGATA data on the Grid.

As the package is for the Grid use, it should be used/run from a User Interface machine. So login first into a user interface computer and copy the package there, in your working directory.

I. INTRODUCTION TO AGATA DATA PROCESSING ON THE GRID

1. The AGATA data processing on the Grid is performed in two steps.The first step consists in performing the PSA data processing on each input raw data file separately for the whole files of a given experiment. The second step consists in performing MERGING/TRACKING processing run by run, processing the whole PSA-processed files of a given run at once.

2. Convention:
A Grid file (for example SRM_AGATA_event_mezzdata.cdat.0001) is defined by its full pathname
srm://srmv2.ific.uv.es/lustre/ific.uv.es/grid/vo.agata.org/2011w29- \
test/run_00/Data/3R/SRM_AGATA_event_mezzdata.cdat.0001
The nawat package uses the following convention to separate the <InputStorage> from the <InputDataFilename>:
<InputStorage> is srm://srmv2.ific.uv.es/lustre/ific.uv.es/grid/vo.agata.org and
<InputDataFilename> is /2011w29-test/run_00/Data/3R/SRM_AGATA_event_mezzdata.cdat.0001
The <InputDataFilename> is defined as the full pathname of the file regarding the experiment

II. PREPARING THE PACKAGE

1. Login into a User Interface computer and go to your working directory

$> ssh -Y kaci@lyoserv.in2p3.fr
$> cd <your-working-directory>

In your working directory download the file nwt-agata.tar.bz2.

1. Unpack the file nwt-agata.tar.bz2 using

$> tar -xvjf nwt-agata.tar.bz2

2. Go inside the nwt-agata/ directory and unpack the file ConfExp.tar.bz2 :

$> cd nwt-agata
$> tar -xvjf ConfExp.tar.bz2

You will get the ConfExp/ directory which contains the configuration files and parameters for each run, for a given experiment. This ConfExp/ directory helps only as a template to show how to built your own ConfExp/ configuration directory. First, proceed with creating your own ConfExp/ directory that contains the right configuration files and parameters corresponding to your own experiment.

III. PREPARING THE PSA DATA PROCESSING

To run PSA data processing you will need basically preparing the three following
files: TaskConfig.nwt ; InputDataFilenames.nwt ; BasisFilenames.nwt

1. Copy the file TaskConfig.nwt-PSA into TaskConfig.nwt

$> cp TaskConfig.nwt-PSA TaskConfig.nwt

2. Edit the file TaskConfig.nwt and make the following changes to fit your needs:

Lines 1 to 5: no changes needed.

Line 6: replace /scratch by any temporal directory in the user interface (/tmp for example). This directory is used to receive the Output files from the Grid. You can create your own temporal directory (in the user interface machine) with

$> mkdir /scratch/my-user-name

Line 7: Maximum number of jobs you want to submit to the Grid. This number should not exceed the number of input raw data files to be processed (and which are listed in the file InputDataFilenames.nwt). See below, the comments on the content of the file InputDataFilenames.nwt (point 3.)

Lines 8 and 9: define the InputStorage (where the input data to process are located) and the OutputStorage (where the output obtained adf files will be stored). In order to maintain consistency, it is advised to use the same storage.
Note that a Grid file (for example SRM_AGATA_event_mezzdata.cdat.0001) is defined by its full pathname <InputStorage><InputDataFilename> where
<InputStorage> is srm://srmv2.ific.uv.es/lustre/ific.uv.es/grid/vo.agata.org and
<InputDataFilename> is /2011w29-test/run_00/Data/3R/SRM_AGATA_event_mezzdata.cdat.0001

Line 10: should be PSA for PSA data processing

Line 11: do not change, keep it as NONE

3. The file InputDataFilenames.nwt contains the list of names of the data files to be processed. Note that, as dicussed few lines above, the file InputDataFilenames.nwt will contain only the <InputDataFilename> part of the full filename path.

4. The file BasisFilenames.nwt contains the list of names of the basis files to be used.
Note that in principle this file should not be changed, but please check if it fits your experiment.
Note also the location of the basis files: <InputStorage>/BaseFiles/LibTrap_XXXX.dat

IV. PREPARING THE MERGING/TRACKING DATA PROCESSING

To run MERGING/TRACKING data processing you will need basically preparing the two following files: TaskConfig.nwt ; PSADataFilenames.nwt

1. Copy the file TaskConfig.nwt-ANCTR-2 into TaskConfig.nwt

$> cp TaskConfig.nwt-ANCTR-2 TaskConfig.nwt

2. Edit the file TaskConfig.nwt and make the following changes to fit your needs:

Lines 1 to 5: no changes needed.

Line 6: replace /scratch by any temporal directory in the user interface (/tmp for example). This directory is used to receive the Output files from the Grid. You can create your own temporal directory (in the user interface machine) with

$> mkdir /scratch/my-user-name

Line 7: Maximum number of jobs you want to submit to the Grid. This number should not exceed the number of runs to be processed. See below, the comments on the content of the file PSADataFilenames.nwt (point 3.)

Lines 8 and 9: define the InputStorage (where the input data to process are located) and the OutputStorage (where the output obtained adf files will be stored). In order to maintain consistency, it is advised to use the same storage.
Note that a Grid file (for example SRM_AGATA_event_mezzdata.cdat.0001) is defined by its full pathname <InputStorage><InputDataFilename> where
<InputStorage> is srm://srmv2.ific.uv.es/lustre/ific.uv.es/grid/vo.agata.org and
<InputDataFilename> is /2011w29-test/run_00/Data/3R/SRM_AGATA_event_mezzdata.cdat.0001

Line 10: should be ANC+TR for MERGING/TRACKING data processing

Line 11: do not change, keep it as NONE

3. The file PSADataFilenames.nwt contains the list of names of the obtained PSA-processed (adf) data files for each run to be processed.
You should append below these files the event_vme data files of the ancillary detectors corresponding to the same runs as for the PSA-processed files.
Note that, as dicussed few lines above, the file PSADataFilenames.nwt will contain only the <InputDataFilename> part of the full filename path.

V. PREPARING THE ONLY TRACKING DATA PROCESSING

To run only TRACKING data processing you will need basically preparing the two following files: TaskConfig.nwt ; PSADataFilenames.nwt

1. Copy the file TaskConfig.nwt-ANCTR into TaskConfig.nwt

$> cp TaskConfig.nwt-ANCTR TaskConfig.nwt

2. Edit the file TaskConfig.nwt and make the following changes to fit your needs:

Lines 1 to 5: no changes needed.

Line 6: replace /scratch by any temporal directory in the user interface (/tmp for example). This directory is used to receive the Output files from the Grid. You can create your own temporal directory (in the user interface machine) with

$> mkdir /scratch/my-user-name

Line 7: Maximum number of jobs you want to submit to the Grid. This number should not exceed the number of runs to be processed. See below, the comments on the content of the file PSADataFilenames.nwt (point 3.)

Lines 8 and 9: define the InputStorage (where the input data to process are located) and the OutputStorage (where the output obtained adf files will be stored). In order to maintain consistency, it is advised to use the same storage.
Note that a Grid file (for example SRM_AGATA_event_mezzdata.cdat.0001) is defined by its full pathname <InputStorage><InputDataFilename> where
<InputStorage> is srm://srmv2.ific.uv.es/lustre/ific.uv.es/grid/vo.agata.org and
<InputDataFilename> is /2011w29-test/run_00/Data/3R/SRM_AGATA_event_mezzdata.cdat.0001

Line 10: should be TR for only TRACKING data processing

Line 11: do not change, keep it as NONE

3. The file PSADataFilenames.nwt contains the list of names of the obtained PSA-processed (adf) data files for each run to be processed.
Note that, as dicussed few lines above, the file PSADataFilenames.nwt will contain only the <InputDataFilename> part of the full filename path.

V. RUNNING THE DATA PROCESSING ON THE GRID

To run any of the above AGATA data processing, follow the following steps:

1. Create a valid proxy:

$> voms-proxy-init -voms vo.agata.org

2. Set your PATH environnement with your local working directory (.) :

$> export PATH=$PATH:.

3. Run nawat_agata

$> nawat_agata &

Click on the button CE-MATCH an choose from the dialog window the Computing Element where you want to run, and click OK.
Then click on the EXECUTE button to start the submission of jobs to the Grid.
You can modify the update status of the jobs by moving the slider to the value 10 minutes for example.
Wait untill the task finishes.
Exit nawat by cliking the button EXIT

Have a fun with nawat_agata !