RUHR UNIVERSITÄT BOCHUM

RUB

Astronomisches Institut
Ruhr-Universität Bochum

JÜLICH
FORSCHUNGSZENTRUM

Generic Pipeline

Lofar Pipeline Framework

17. September 2015 | Stefan Fröhlich

Mitglied der Helmholtz-Gemeinschaft

Pipelines

Stating the obvious:

- Single direction of workflow
- Black box to users

e.g. graphics pipeline

Mitglied der Helmholtz-Gemeinschaft

Lofar Pipeline Framework

- Written in Python2
- First implementation for WSRT

Features:

- Data distribution and tracking
- Job management
- Parallel execution

Provides standard set of pipelines

Mitglied der Helmholtz-Gemeinschaft

Lofar Pipeline Framework

- Configuration and control files

pipeline.cfg	tasks.cfg	mapfile
Where is stuff: Working directory Task configs Master/Node scripts Computing environment.	Possible steps to be used in the pipeline framework. A step consists of a pair of master/node scripts.	Control element. (host,input,skip) Runs each step for every input file (Measurement Set). Automatically created by plugins

Mitglied der Helmholtz-Gemeinschaft

Lofar Pipeline Framework

- Used like a fixed function pipeline
- Three layers of execution

Pipeline	Master Scripts	Node Scripts
Top level workflow script.	Head node version of functionality.	Worker version of master scripts.
Sets up master script calls (steps) and glues results together.	Extra data preparation parameter setup etc.	More step specific functions.
One for every pipeline.	One per functionality (diff. dppp version)	One per master

Mitglied der Helmholtz-Gemeinschaft

Lofar Pipeline Framework

- No clear distinction of pipeline description and framework
- Too many specific step functions are spread across layers
- Hard to create
- Harder to change
- Not end-user friendly (and not intended at the time of creation)

Mitglied der Helmholtz-Gemeinschaft

Generic Pipeline

What is the generic pipeline?

- Reorganized Lofar Pipeline

Pipeline Description	Pipeline Script	Master/Node Scripts
Workflow is described in a parset. For basic pipelines no programming is needed.	Former pipeline script now becomes a parser for the parset. Features of the pipeline are implemented here.	Only one master script necessary. Needs very few node scripts (e.g. casapy)

Mitglied der Helmholtz-Gemeinschaft

Generic Pipeline

Advantages

- Backwards compatible
- Uses the Lofar Pipeline Framework backend:
  - Job distribution and feedback
  - Tracking of data
  - Checkpoints after each step
- Use of standard steps is trivial (e.g. DPPP, AWImager, bbs-reducer)
- Creating own steps is easier than before

Mitglied der Helmholtz-Gemeinschaft

Generic Pipeline

Features

- Subpipelines (pipeline parset can be a step itself)
- Loops
- Plugins (for quick hacking, not tracked)
- Python steps (can ‘store’ values inside a pipeline run)
- Not Lofar bound. Can run anything.

Mitglied der Helmholtz-Gemeinschaft

Generic Pipeline

Set it up

- Load the Lofar environment (CEP3: use Loflm)
- Get a copy of the pipeline.cfg ($LOFARROOT/share/pipeline/pipeline.cfg)
- Configure your working directory and cluster setup
- Write a pipeline parset
- Run the pipeline:

```
genericpipeline.py mypipeline.parset –c mypipeline.cfg
```

Mitglied der Helmholtz-Gemeinschaft

Basic Example

```
pipeline.steps                  = [step1,step2,…]

step1.control.kind              = recipe/plugin
step1.control.type              = name
step1.control.executable        = /path/to/program
step1.argument.inputfile        = /your/mapfile
step1.argument.key              = value

step2.control.type              = taskX
step2.control.mapfile_in        = step1.output.mapfile
```

Mitglied der Helmholtz-Gemeinschaft

Generic Pipeline

Cookbook example: calibrate 3C 295

PreFacetCalibration pipeline: created as genericpipeline parset

Mitglied der Helmholtz-Gemeinschaft

Summary

- The Generic Pipeline is a Reorganized Lofar Pipeline
- The framework handles data tracking, parallel execution, checkpointing
- Users can define pipelines themselves and run them on every Lofar installation
- Functionality can be extended more easily than in the past
- Its possible to write pipelines without programming

Documentation:

http://www.astron.nl/citt/genericpipeline

Mitglied der Helmholtz-Gemeinschaft