Easy to deploy and easy to modify data reduction pipelines using KERN and CWL

packaging, containerization and pipelines

ADASS 2019 - Groningen

Gijs Molenaar

Pythonic.nl

Who am I

- pythonic.nl
- 8 years
- Research / Software Engineer
- Machine learning
- Contractor in science and industry
- Part-time remote PhD student South Africa

What did I work on?

- Large scale distributed pipeline deployment
- Packaging / improving radio astronomy software
- Transient Pipeline (TraP)

The problem

- SKA is coming
- Data volumes to high to transport data
- Processing on the spot
- Opinionated data processing

Pythonic.nl

But is it a problem?

- Why not let scientist process data on the spot
- Give them control over the data reduction
- Let them deliver a pipeline

What is required

- A technical solution
- Acceptation
- Procedures

What is required

- A technical solution
- Acceptation
- Procedures

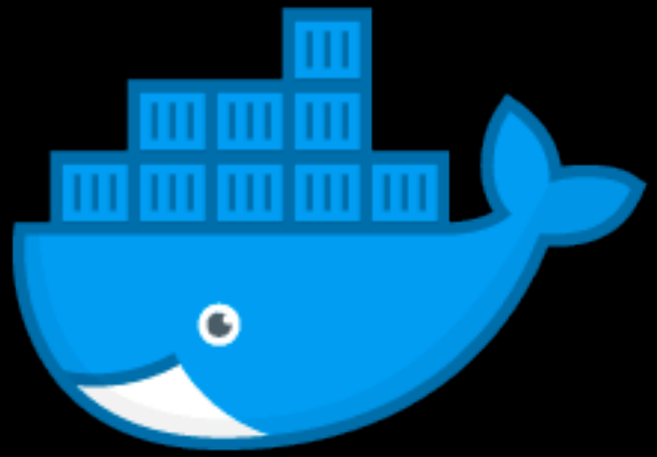
Scientific Software

- Often hard to install
- Locally and on cluster
- Issues like
  - Broken software
  - Compilation
  - Dependencies (versions)
  - Python 2 / 3

We want to

- Centralize agony
- Only compile once
- Compatible tools
- Make sure deployment are uniform between platforms

Solution?



docker

Solution

Containers

- Often leads to massive containers
- What is inside?
- How to combine containers?

Package management

- A boring old solution
- But does the job well
- Can be installed inside and outside container

Pythonic.nl

KERN

- Made for SKA South Africa
- Radio astronomy software packages
- basics, Imaging, pulsar
- A superset of Ubuntu LTS
- 75 packages and growing
- New release every +/- 6 months

Installation on Ubuntu 18.04

```
$ sudo apt-add-repository -s ppa:kernsuite/kern-5
$ sudo apt-get update
$ sudo apt-get install wsclean
```

More than packaging

- A platform for radio astron software improvement
- A community (reporting and fixing bugs)
- Upstream bug fixes

- Python-casacore binary wheel
- Presto python3

Debian

- Collaboration with Debian Astro blend
- Many KERN packages are now in Debian
- Ole Streicher is also here

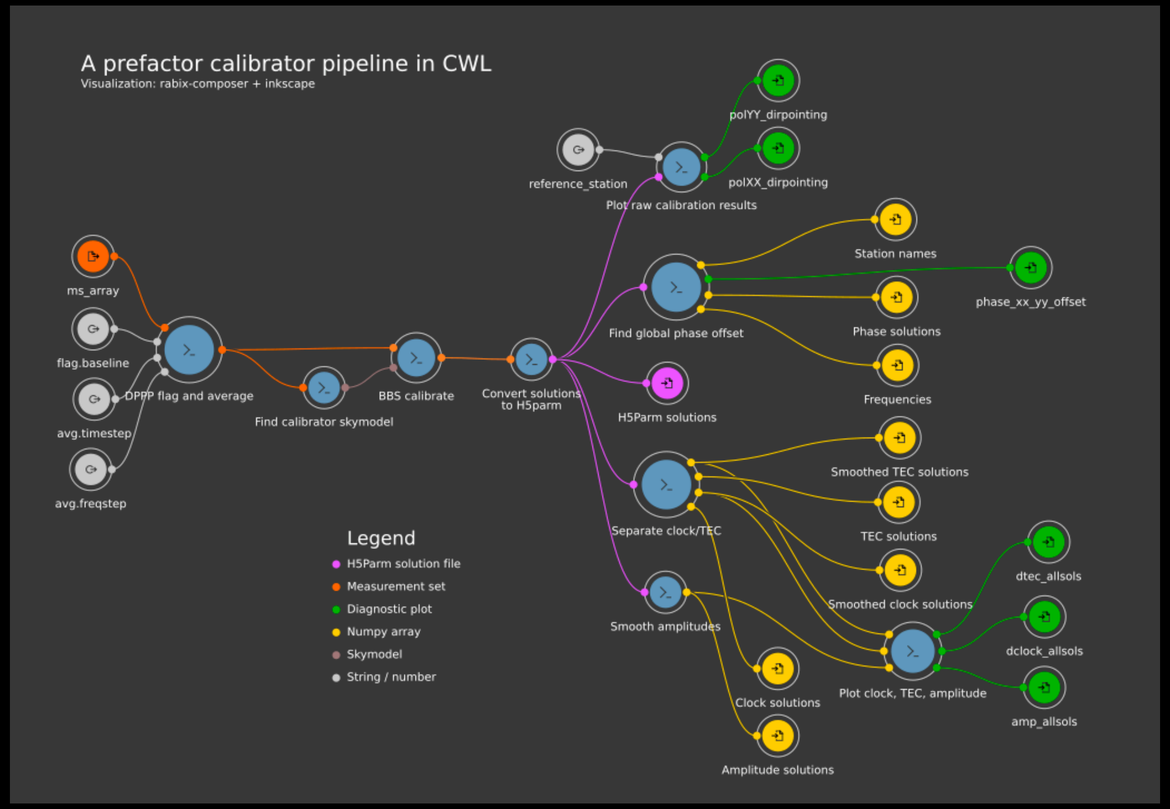
No KERN-6?

- No funding!
- If you use KERN and find it useful
- Please consider helping me find funding :)

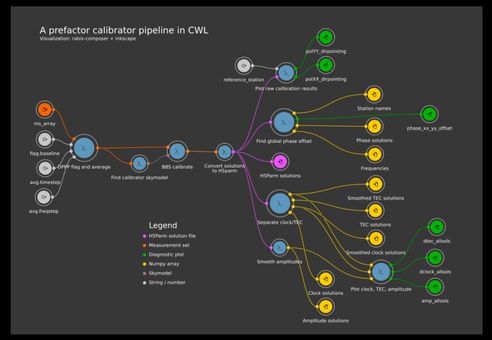
CWL

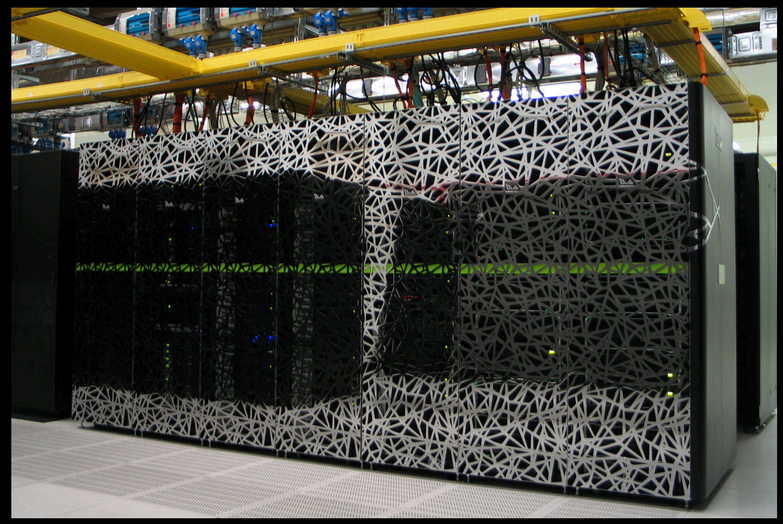
- Common Workflow Language
- The HTML of pipelines
- Method for describing your pipeline
- And how all components interact

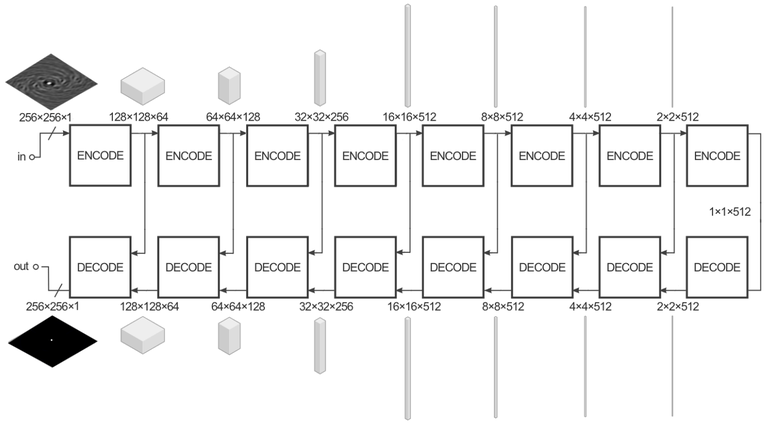
Prefactor

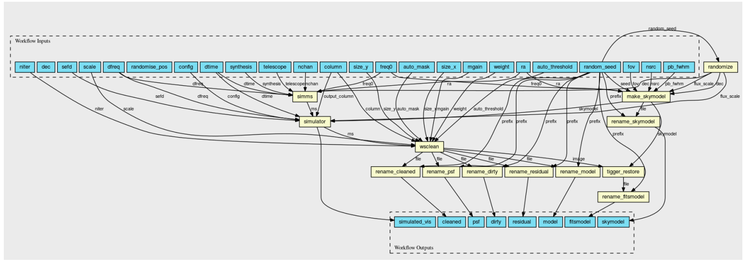

A prefactor calibrator pipeline in CWL
Visualization: rabix-composer + inkscape
ms_array
flag.baseline
avg.timestep
avg.freqstep
DPPP flag and average
Find calibrator skymodel
BBS calibrate
Convert solutions to H5parm
reference_station
Plot raw calibration results
polYY_dirpointing
polXX_dirpointing
Find global phase offset
Station names
phase_xx_yy_offset
Phase solutions
Frequencies
H5Parm solutions
Separate clock/TEC
Smoothed TEC solutions
TEC solutions
Smooth amplitudes
Smoothed clock solutions
dtec_allsols
dclock_allsols
Plot clock, TEC, amplitude
amp_allsols
Clock solutions
Amplitude solutions
Legend
H5Parm solution file
Measurement set
Diagnostic plot
Numpy array
Skymodel
String / number

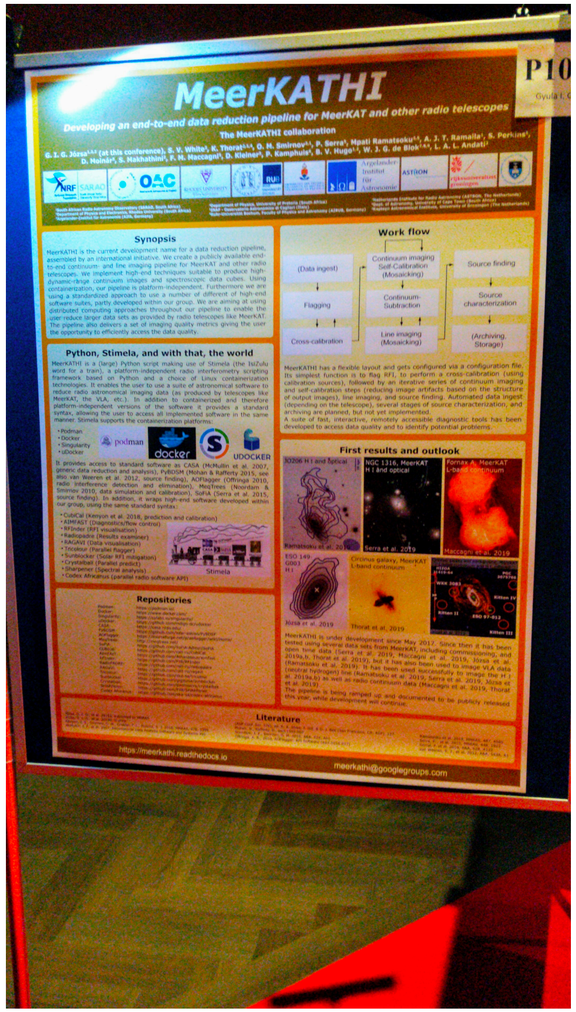
EOSCpilot 4lofar pilot











https://github.com/gijzelaerr/spiel

Want to know more?

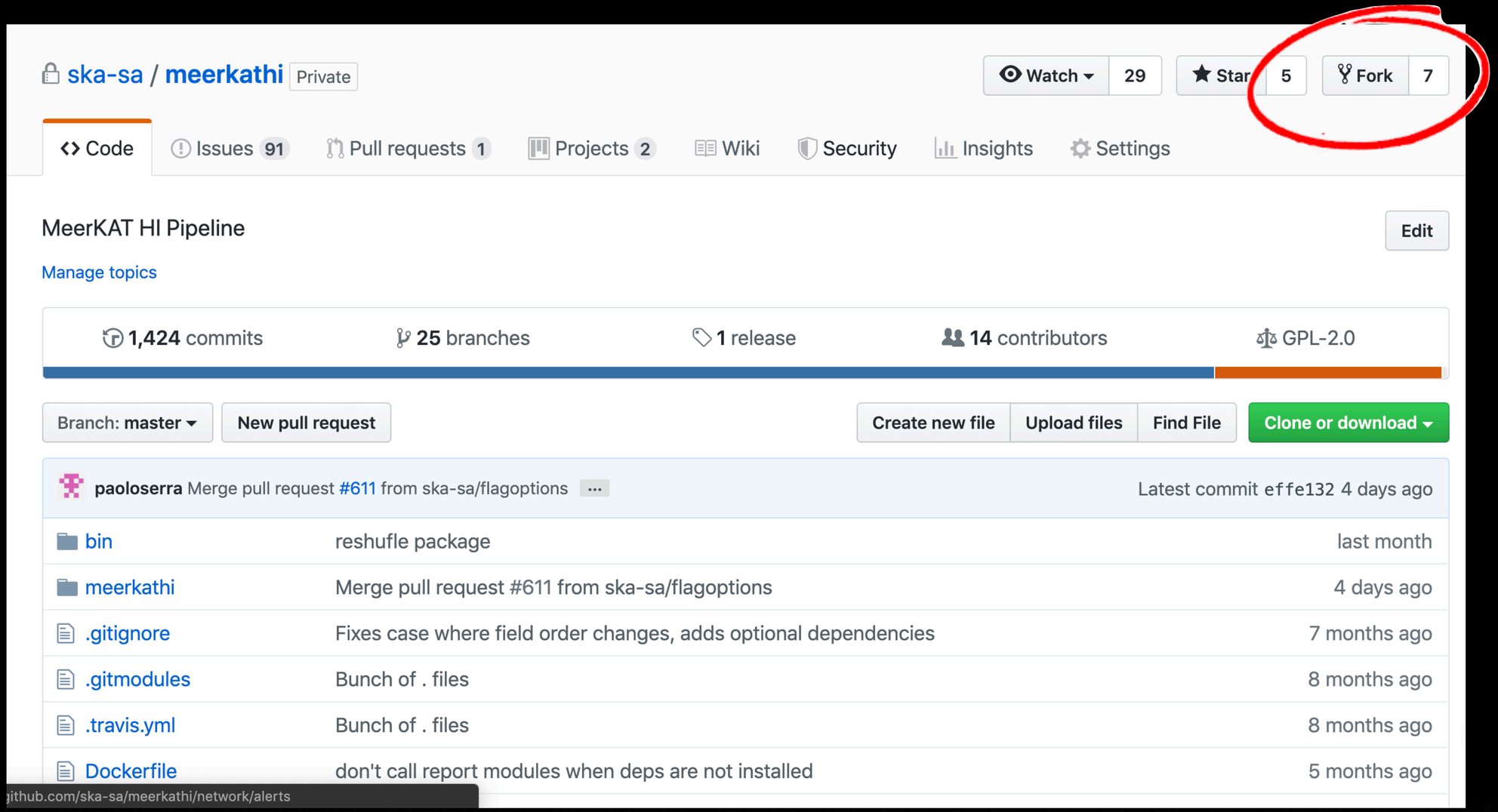
- Blog post with full story
- http://bit.do/radiopipelines

Pythonic.nl

In short

- I Believe astronomers can reduce their own data
- Temporarily hand over control of the data processing hardware
- Composable pipelines using packaging, containerization & CWL
- Remaining challenge: Organisational setup

Pythonic.nl


ska-sa / meerkathi Private
Watch 29
Star 5
Fork 7
Code
Issues 91
Pull requests 1
Projects 2
Wiki
Security
Insights
Settings
MeerKAT HI Pipeline
Edit
Manage topics
1,424 commits
25 branches
1 release
14 contributors
GPL-2.0
Branch: master
New pull request
Create new file
Upload files
Find File
Clone or download
paoloserra Merge pull request #611 from ska-sa/flagoptions
Latest commit effe132 4 days ago
bin reshufle package last month
meerkathi Merge pull request #611 from ska-sa/flagoptions 4 days ago
.gitignore Fixes case where field order changes, adds optional dependencies 7 months ago
.gitmodules Bunch of . files 8 months ago
.travis.yml Bunch of . files 8 months ago
Dockerfile don't call report modules when deps are not installed 5 months ago
github.com/ska-sa/meerkathi/network/alerts

Organisational setup

- Open source default pipeline
- Have the option to bundle a pipeline with your observation proposal

Questions?

http://bit.do/radiopipelines

http://kernsuite.info

Pythonic.nl