Working towards deterministic scientific pipelines

Gijs Molenaar
gijs@pythonic.nl
@gijzelaerr

Who am I

- for 2.5 years scientific software engineer @ API

- 3 months South Africa, Meqtrees
- September 2014 - April 2015 again

What I (try to) do

- Software engineering. AARTFAAC / TKP

- Improve current radio astronomy software landscape

What I think is important

- Write tests
- Run tests automated (jenkins)
- Documenation
- Communication with community
- Proper release management
- Software packaging
- Software distribution
- Usability & Installability for end user (astronomer)

Software packaging

Problem: scientist A want to run software B on some platform C

Specific per:

OS   Distribution   Library versions   compile flags

 system environment   paths   user preferences   sysadmin preferences

Solution STEP 1

Formalise and streamline installation procedure

package it up

Focus on one platform

Debian / Ubuntu

Ubuntu LTS specifically

https://launchpad.net/~ska-sa/+archive/main

Also there is

Debian astro

Working group to get often used packages in Debian

join mailinglist if interested

meqtrees / casacore / LUS (lofar)

for ubuntu 12.04 / 14.04

$ sudo apt-get install python-software-properties software-properties-common

$ sudo add-apt-repository ppa:ska-sa/main

$ sudo apt-get update

$ sudo apt-get install meqtrees

CASA

Was working on packaging CASA

hope to finish in 2015

Solution STEP 2

Don't adjust software to system

adjust system to software

Virtualisation & containment

Papino

- Radio astronomy springboard

- Virtual machine based
- Complete operating system
- contains various radio astronomy packages

- https://github.com/ska-sa/papino

Why?

- Solves many problems

- Easy to install anywhere (cluster, mac, PC)

- Easy to adjust (fork, change, build)

- Everything is scripted

Advantages

Brain dead installation procedure

Package once, install everywhere

Split interfering software

Old software / lib? not problem (take old distro)

Vagrant

install vagrant & virtualbox

$ git clone https://github.com/ska-sa/papino

$ cd papino

$vagrant up

Problem

Virtualisation overhead

memory barier (split of memory)

Solution

Docker!

Linux only

can run docker in vagrant for OS X

Usage

On Ubuntu 14.04:

$ sudo apt-get install docker.io

$ git clone https://github.com/ska-sa/papino

$ cd papino

$ ./docker.sh

Future plans

Use papino for reproducable science

Bundle software with paper?

Parallelisation & cluster deployment

shared memory between containers?

Data locality

Intensify communication with CASACORE developers. I want to contribute.