



ALMA Pipeline Infrastructure Deployment

Lindsey Davis

ALMA Pipeline
Architect

Atacama Large Millimeter/submillimeter Array
Expanded Very Large Array
Robert C. Byrd Green Bank Telescope
Very Long Baseline Array



Infrastructure: Concepts

- Science goals ->
 - factored into observing unit sets, scheduling blocks, and ASDM(s)
- Standardization ->
 - project organization, observing unit sets
 - observing modes / calibration strategy
 - directory structure / directory names
 - reduction procedures / scripts
 - data products / data product names

Infrastructure Triggering Conditions

- Observing unit set is fully observed
 - all data has been taken
- Data is ready
 - bulk data in Archive in Santiago / ARCs
- Manual trigger for now

Infrastructure: Technical Inputs

- The UID of the parent observing unit set status entity
 - e.g uid://A001/X74/X29
- The intents file
 - e.g \$SCIPPIPE_SCRIPTDIR/intents_if.xml
- The procedure file
 - e.g \$SCIPPIPE_SCRIPTDIR/procedure_if.xml

Infrastructure: Runtime Environment

- The root working directory for all runs
 - e.g \$SCPIPE_ROOTDIR,
/export/lustre/pipeline/root
- The infrastructure logs directory
 - e.g \$SCPIPE_LOGDIR,
/export/lustre/pipeline/logs
- The intents / procedures directory
- e.g \$SCPIPE_SCRIPTDIR,
/export/lustre/pipeline/scripts

Infrastructure: Software/ Account Setup

- pipeproc account on multivac22
 - 32 bit machine
- Similar to a standard ALMA STE account with pipeline specific environment configuration
- Requires some ACS software packages, the ACS build system, Archive software packages, data model support packages, pipeline infrastructure packages
- Code is in Java

Infrastructure: Operations

- Retrieves project and status XML documents from the archive
- Creates processing tree in memory
 - gathers user intent information
 - gathers procedure information
 - creates pipeline processing request(s)
- Creates standard directory tree on Lustre
 - saves XML documents, pipeline processing request
 - downloads ASDM(s)

Infrastructure: Commands

- Show processing request
 - pipelineShowRequest <ousstatus_uid>
<intents>.xml <procedure>.xml
- Make the processing request
 - pipelineMake Request <ousstatus_uid>
<intents>.xml <procedure>.xml true/false

PPR: Contents

- General information
 - e.g. project title
- Project structure
 - links to observing project, observing unit set, ...
- Processing request
 - resource constraints, intents, processing commands and parameters, data sets
- Results handling
 - user notification

PPR: Environment, Software, Accounts

- Location of the pipeline scripts
 - `$SCIPIPE_HEURISTICS`,
 - `/export/lustre/pipeline/PIPELINE/Heuristics/src`
- Account pipeproc on multivac21
 - 64 bit machine
- CASA installed
 - casapy-stable (February 2012, #18527, tarball)
- All user software in Python

PPR: Execution Operations

- Converts XML document to Python object
 - summary, intents, commands, ASDMs
- Checks for pathological conditions
- Loops through the commands
 - locates command in Python task dictionary
 - inserts parameters into current Pipeline context
 - executes, and generates report
 - merges results back into context

PPR: Execution Commands

- Interactively from casapy
 - casapy-stable
 - import pipeline.infrastructure.executeppr as eppr
 - eppr.executeppr (<pprfile>.xml)
- casapy batch mode
 - Casapy-stable –nogui –nologger -c
\$SCIPIPE_HEURISTICS/runpipeline.py
<pprfile>.xml

Issues

- Heterogeneous operational environment
 - 32 bit / 64 bit
 - connected by Lustre file system
- Heterogeneous software and build system
 - ACS, ALMA data models and archive
 - CASA
 - pipeline tasks
- Non-standard projects, missing information
- Pipeline code still being refactored