ADMIT:
ALMA Data Mining Toolkit

Peter Teuben
University of Maryland

ADASS
Oct 19, 2016
ALMA

• A schizophrenic telescope

• Software: **CASA**
  - Calibration
  - Mapping or Imaging
  - Image/Cube Analysis
    - “align the phases; scale amplitudes”
    - “that fourier transform + deconvolution”
    - “separate the signal from the noise”

• Two telescope models (implemented as set of normalized Tables)
  - AlmaScienceDataModel (**ASDM**) for ALMA (and eVLA)
  - MeasurementSet (**MS**) for CASA

• ALMA Development Projects and Studies
  - **CARTA** (casaviewer replacement) and **ADMIT** (post-pipeline analysis) 2014-2016 - 2yr
  - **ASTUTE** (study - prepared for ADMIT) (2013 - 1yr)
  - **TP2VIS** (study how to combine single dish (“Total Power”) & interferometer data (2017 - 1yr)
ADMIT team

- PI: Lee Mundy (UMD)
- UMD: Peter Teuben, Marc Pound, Kevin Rauch
- UIUC: Leslie Looney, Doug Friedel, Lisa Xu, Robert Harris
- NRAO: Jeff Kern, Mark Lacy, John Hibbard (+...)
See also

- Artemix – ALMA RemoTE MIning eXperiment – P1.34
- Machine learning … spectral lines - P2.2
- ALMA Science Archive – P8.25
- ALMAWebQL v2 – D7


ADMIT:
Extract and add interesting science data products to the ALMA archive

- Python toolkit (generic – not CASA dependent)
- “flow” with reflow and dependencies (à la “make”)
  - **AT**  = ADMIT Task
  - **BDP** = Basic Data Product (xml wrapper)
- AT’s can be implemented using any module(s), e.g. CASA or radio-astro-tools or MIRIAD
  - Current version: using CASA tasks and tools + Numpy/Scipy
- BDP implement tables, but wrap images
- “MultiFlow” : combine flows
essential classes: 
AT and BDP

• AT:
  - BDP_in[]
  - BDP_out[]
  - Parameters

• BDP
  - AT
  - BDP_parents[]
  - BDP_children[]
ADMIT:
AT's and BDP's

- **BDP:**
  - Data wrapper (in XML)
  - Small data are inside the XML file, e.g. tables, numbers
  - Large data (fits files, png files) are XLINK'd
  - A BDP remembers how it was produces (AT w/ Parameters)

- **AT:**
  - Compute wrapper
    - can be implement in CASA, MIRIAD, NEMO etc.
    - can contain many calls to tasks in CASA/MIRIAD/NEMO etc.
  - Takes # input BDP(s)
  - Produces # output BDP(s)
    - an M to N mapping (M can be 0)
  - Has optional input parameters (“keyword=value”) to control the task
ADMIT flow diagram
AT and BDP’s

- FITS
- CasalImage
- CubeStats
- LineList
- LineCube
- LineCubeMoment0,...
A simple ADMIT flow:
cube statistics, spectra, line-id, line-cubes

Automagically generated and updated during a flow
A simple ADMIT flow:
with added continuum subtraction
**CubeStats_AT** computes image-plane robust statistics on datacubes. These statistics are particularly useful for identifying spectral lines in images where the noise varies as a function of frequency.

- **CASA image**: x.im
- **RMS method**: medabsdevmed
- **RMS value**: 1.131E-03
- **Dynamic range**: 1.361E+03
- **Data mean**: 8.325E-04

Emission characteristics as a function of channel, as derived by **CubeStats_AT** (cyan: global rms, green: noise per channel, blue: peak value per channel, red: peak/noise per channel).
<table>
<thead>
<tr>
<th>Task Name</th>
<th>Task ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingest_AT</td>
<td>taskid=0</td>
</tr>
<tr>
<td>CubeStats_AT</td>
<td>taskid=1</td>
</tr>
<tr>
<td>CubeSum_AT</td>
<td>taskid=2</td>
</tr>
<tr>
<td>SFind2D_AT</td>
<td>taskid=3</td>
</tr>
<tr>
<td>CubeSpectrum_AT</td>
<td>taskid=4</td>
</tr>
<tr>
<td>LineSegment_AT</td>
<td>taskid=5</td>
</tr>
<tr>
<td>PVSlice_AT</td>
<td>taskid=6</td>
</tr>
<tr>
<td>PVCorr_AT</td>
<td>taskid=7</td>
</tr>
<tr>
<td>LineID_AT</td>
<td>taskid=8</td>
</tr>
<tr>
<td>LineCube_AT</td>
<td>taskid=9</td>
</tr>
<tr>
<td>Moment_AT</td>
<td>taskid=10</td>
</tr>
<tr>
<td>CubeSpectrum_AT</td>
<td>taskid=11</td>
</tr>
<tr>
<td>Moment_AT</td>
<td>taskid=12</td>
</tr>
</tbody>
</table>

Update ADMIT flow state (dry run)  Re-run ADMIT flow
On-line documentation of LinELD_AT

Parameters

Input BDPs

<table>
<thead>
<tr>
<th>Index</th>
<th>Type</th>
<th>File name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>CubeSpectrum_BDP</td>
<td>x.csp</td>
</tr>
<tr>
<td>1</td>
<td>CubeStats_BDP</td>
<td>x.cst</td>
</tr>
<tr>
<td>2</td>
<td>PVOcorr_BDP</td>
<td>x.pvc</td>
</tr>
</tbody>
</table>

Output BDPs

<table>
<thead>
<tr>
<th>Index</th>
<th>Type</th>
<th>File name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>LineList_BDP</td>
<td>x.ll</td>
</tr>
</tbody>
</table>

Keywords

numsigma: 5.0
force:
tierwidth: 0.0
online: False
ADMIT Task Inputs for test0.admit

- mincman: 4
- pattern: AUTO
- smooth: [ ]
- recalcnoise: False
- vlsr: -999999.99
- maxgap: 3
- reject: [ ]
- method: (*PeakFinder*; {'thresh': 0.0})
- mode: ONE

LineCube_AT (taskid=9)
Moment_AT (taskid=10)
CubeSpectrum_AT (taskid=11)
Moment_AT (taskid=12)

[Update ADMIT flow state (dry run)] [Re-run ADMIT flow]
Moment_AT creates moment maps using custom clip levels.

Moment_AT output for x.CO_115.27120/lc.im
Technical Details

• CVS → git: https://github.com/astroumd/admit
• Quasi-Agile
  – Prototyped (ASTUTE, BDP centric → AT centric)
  – Rapid development cycle
• Tests: unit, integration, regression (buildbot)
• Documentation: sphinx
• Runtime reporting: bootstrap
• Integration into CASA
# example install
1% git clone https://github.com/astroumd/admit
2% cd admit
3% autoconf
4% ./configure --with-casa-root=/opt/casa-release-4.7.0-el6
5% source admit_start.sh

# example usage
6% runa1 test253_spw3.fits
7% aopen test253_spw3.admit

# example CASA
% casa
Import admit
p = admit.project('foo.admit')
Timeline / Future

- ADMIT 1.0 “delivered” May 2016
  - ADMIT 1.1 delivered Nov 2016 (handover to NRAO)
  - Acceptance Test TBD
  - ADMIT 2.0 in a next development proposal?

- Future Features
  - Desktop vs. Pipeline usage
  - Expand from the current 2 (or 5) Recipes
  - Expand from the current ~20 AT’s
  - Ingest_AT uses “noise-flat” cubes