

Astronomy ESFRI & Research Infrastructure Cluster ASTERICS - 653477



Integration of CASA with Jupyter for efficient remote processing

Aard Keimpema (keimpema@jive.eu)







H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).

Yearly archivable data



Near data processing

- SKA phase-1 will produce ~1 PB / day
- Data reduction will be done where the data is stored
- Possible solution: Remote interactive pipelines based on Jupyter notebooks
- Jupyter notebooks are displayed inside a web-browser
- Successor to IPython, CASA is based on IPython
- Not limited to python, bindings to 40+ languages exist

C C C Construction Likely Access Likely Construction Likely C Constructing Like	🧧 vla-cont-tutorial 🗙 🗌	Aard — D
<pre>#1 Apr 0 to the space back is a part with Physe is the twend to Tack is () part and recept the () part of the twend is () the context of the part of the twend is () the context of the part of the twend is () the part of the part</pre>	$\leftrightarrow \rightarrow \mathbb{C} \ \Delta$ 🛈 localhost:8888/notebooks/vla-cont-tutorial.ipynb	7 🛥 🔤 🗉 🗄
<pre>spin provide the control tuborial Las Checkpoint Geou.2017 (successed) provide the last Checkpoint Ge</pre>	🛗 Apps 🗋 🚺 Unity Game Develop 🧔 Spotify Web Player - 🌔 The Swedish Teache 🔍 Januari recept: Rund 🛐 South Indian-Style V 🔺 Bookmarks 🚳 15 Basic Stir Fry Sau 🔹	Other bookmarks
Image: Control of the control of th	JUpyter vla-cont-tutorial Last Checkpoint: 06/01/2017 (autosaved)	Logout
<pre>N * * * * * * * * * * * * * * * * * * *</pre>	File Edit View Insert Cell Kernel Help	Casa O
<pre>Definition of the state of</pre>	➡ ★ ↓ ▶ Run ■ C Code ▼	
<pre>source: https://canadides.neurodes.neurol.accontent</pre>	VLA Continuum Tutorial 3C391-CASA4.7	
Download the data from the above location and extract the dataset to the same directory as this notebook is located In [2]: [istobs(vis='3c301_ctm_mosaic_l0s_spw0.ms'] Theservertices: Name: /home/jupyter/data/casa/3c391_ctm_mosaic_l0s_spw0.ms MS Version 2 Theservertice: Names Miller-Jones Project: T.B.D. Theservertices: Name: /home/jupyter/data/casa/3c391_ctm_mosaic_l0s_spw0.ms MS Version 2 Theservertice: Names Miller-Jones Project: T.B.D. Theservertice: Names Miller-Jones Project: T.B.D. Theorem (UTC) Deservertice: Names Miller-Jones Project: T.B.D. Deservertice: Names Miller-Jones Project: Na	Source : https://casaguides.nrao.edu/index.php/VLA_Continuum_Tutorial_3C391-CASA4.7	
<pre>In [2]: listobs(vis='3c391_ctm_mosaic_l0s_spw0.ms') MeasurementSet Name: /home/jupyter/data/casa/3c391_ctm_mosaic_l0s_spw0.ms MS Version 2</pre>	Download the data from the above location and extract the dataset to the same directory as this notebook is located	
MeasurementSet Name: /home/jupyter/data/casa/3c391 ctm mosaic 10s spw0.ms MS Version 2 Observer: Dr. James Miller-Jones Project: T.B.D. Observer: Dr. James Miller-Jones Project: Dr.B. Observer: Dr. Jame	<pre>In [2]: listobs(vis='3c391_ctm_mosaic_10s_spw0.ms')</pre>	
Observer: Dr. James Miller-Jones Project: T.B.D. Observet: DVLA Data records: 845379 Total elapsed time = 28681.5 seconds Observet from 24-Apr-2010/08:02:10.0 to 24-Apr-2010/16:00:11.5 (UTC) Observet from 24-Apr-2010/08:02:10.0 to 24-Apr-2010/16:00:11.5 (UTC) Observet from 24-Apr-2010/08:02:10.0 to 24-Apr-2010/16:00:11.5 (UTC) Observet from 24-Apr-2010/08:02:10.0 to 24-Apr-2010/16:00:11.5 (UTC) Observet from 24-Apr-2010/08:02:10.0 to 08:02:30.0 1 0 01:331+2030 Gos (D) [10] ScanIntent 08:03:20.0 to 08:16:22.8 0 1 0 01:02:30.0 1 2 0 11:331+2030 Gos (D) [10] 08:03:30.0 to 08:16:20.0 3 0 01:331+2030 13975 [0] [10] 08:10:33.0 10 08:10:33.0 to 08:16:20.0 5 2 3:00 1 3 075910 [10] 08:23:48.0 to 08:19:44.8 0 5 2 3:301 (2 7821 [0] [10] 08:23:48.0 to 08:39:48.0 fo 3 3:391 (2 7821 [0] [10] 08:23:48.0 to 08:39:44.8 0 fo 8:39:48.0 fo 3 3:391 (2 7821 [0] [10] 08:33:8:0 to 08:39:48.0 fo 3 3:391 (2 7821 [0] [10] 08:33:8:0 to 08:39:48.0 fo 3 3:391 (2 7821 [0] [10] 08:33:8:0 to 08:39:48.0 for 08:39:48.0 for 08:394.0 for 08:494.0 for 08:394.0 for 08:494.0 for 09:394.0 for 0	MeasurementSet Name: /home/jupyter/data/casa/3c391_ctm_mosaic_10s_spw0.ms MS Version 2	
<pre>ObservationID = 0 ArrayID = 0 Date Timerange (UTC) Scan FldId FieldName nRows SpwIds Average Interval(s) ScanIntent 24-Apr-2010/08:02:10.0 - 08:02:30.0 1 0 J1331+3830 550 [0] [10] 08:02:20.0 - 08:09:30.0 2 0 J1331+3830 13975 [0] [10] 08:02:20.0 - 08:20:20.0 1 0 J1331+3830 13975 [0] [10] 08:02:20.0 - 08:02:20.0 1 0 J1331+3830 13975 [0] [10] 08:02:20.0 - 08:20:20.0 1 0 J1331+3830 13975 [0] [10] 08:24:48(0 - 08:32+48,0 5 3 C391 C1 7590 [0] [10] 08:34:48(0 - 08:33+48,0 6 3 3 C391 C2 7821 [0] [10] 08:34:38(0 - 08:34:48,0 8 5 3 C391 C3 7821 [0] [10] 08:34:38(0 - 08:34:48,0 8 5 3 C391 C3 7821 [0] [10] 08:34:38(0 - 08:34:48,0 9 6 3 C391 C5 7843 [0] [10] 08:34:48(0 - 08:39:16,0 - 74 3 C391 C3 7821 [0] [10] 08:34:48(0 - 08:39:16,0 - 74 3 C391 C3 7821 [0] [10] 08:34:48(0 - 08:39:16,0 - 74 3 C391 C3 7821 [0] [10] 08:34:48(0 - 08:39:16,0 - 74 3 C391 C3 7821 [0] [10] 08:34:48(0 - 08:39:16,0 - 74 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 08:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 09:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 09:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 09:34:38(0 - 08:34:48,0 - 96 3 C391 C5 7843 [0] [10] 09:34:38(0 - 90 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</pre>	Observer: Dr. James Miller-Jones Project: T.B.D. Observation: EVLA Data records: 845379 Total elapsed time = 28681.5 seconds Observed from 24-Apr-2010/08:02:10.0 to 24-Apr-2010/16:00:11.5 (UTC)	
<pre>Label 1 2010 00:02:130 b 0 00:02:030 b 1 2 0 0 11331+3030 13975 [0] [10] 00:02:02 0 b 00:10:28.0 b 00:10:28.0 3 0 0 11331+3030 13975 [0] [10] 00:09:28.0 b 00:10:28.0 3 0 0 11331+3030 13975 [0] [10] 00:19:28.0 b 00:29:48.0 5 2 3331 [1 7500 [0] [10] 00:29:38.0 b 00:34:48.0 5 2 3331 [2 7621 [0] [10] 00:39:38.0 b 00:39:48.0 7 4 33312 2 7621 [0] [10] 00:39:38.0 b 00:49:48.0 9 6 5 3331 [2 7621 [0] [10] 00:39:38.0 b 00:49:48.0 9 6 5 3331 [2 7631 [0] [10] 00:39:38.0 b 00:49:48.0 9 6 5 3331 [2 7631 [0] [10] 00:39:38.0 b 00:49:48.0 9 6 5 3331 [2 7631 [0] [10] 00:39:38.0 b 00:49:48.0 9 6 5 3331 [2 7631 [0] [10] 00:49:38.0 b 00:49:48.0 9 6 5 3331 [2 7631 [0] [10] 00:49:39.0 b 00:49:48.0 9 6 5 3331 [2 7631 [0] [10] 00:49:39.0 b 00:49:48.0 9 [0] [10] 1n [3]: plotants(vis='3c391_ctm_mosaic_l0s_spw0.ms', figfile='plotants_3c391_antenna_layout.png') clearstat() # This removes the table lock generated by plotants in script mode</pre>	ObservationID = 0 ArrayID = 0 Date Timerange (UTC) Scan FldId FieldName nRows SpwIds Average Interval(s) ScanIntent 24 Apr 2010(08:02:10 0 08:02:20 0 1 0 11221:2020 650 [0] [10]	
In [3]: plotants(vis='3c391_ctm_mosaic_10s_spw0.ms', figfile='plotants_3c391_antenna_layout.png') clearstat() # This removes the table lock generated by plotants in script mode	08:02:10/08:02:10.0 06:02:30.0 1 0 01331+3030 13975 [0] [10] 08:02:20.0 08:16:28.0 3 0 J1331+3030 13975 [0] [10] 08:09:20.0 08:16:28.0 3 0 J1331+3030 13975 [0] [10] 08:19:38.0 08:29:48.0 5 2 3C391 C1 7590 [0] [10] 08:29:38.0 08:39:48.0 7 4 3C391 C3 7821 [0] [10] 08:39:38.0 08:44:48.0 8 5 3C391 C4 7821 [0] [10] 08:39:38.0 08:44:48.0 9 6 3C391 C5 7843 [0] [10]	*
$ \frac{3331_{ctm} mosaic_{105} spw0.ms}{400} \\ \frac{400}{400} \\ \frac{200}{400} \\ \frac{200}{40} $	<pre>In [3]: plotants(vis='3c391_ctm_mosaic_10s_spw0.ms',figfile='plotants_3c391_antenna_layout.png') clearstat() # This removes the table lock generated by plotants in script mode</pre>	
Show log	$\underbrace{\underbrace{600}_{0}}_{200} \underbrace{\underbrace{3c391}_{ctm} \underline{mosaic_{10s} spw0.ms}}_{a 28}}_{a 28}$	

4 / 22

Jupyter Architecture



- Notebook server knows nothing about target language
- All language specifics are in the kernel
 - Wrapper-kernel: written in python, easiest to implement
 - Native-kernel: written in target language, 5/22 much more work

Jupyter hub



Multi-user front-end for Jupyter

CASA Jupyter kernel

- Based on the generic python wrapper kernel
- Initialization:
 - Load needed python packages: casacore, casa tasks, matplotlib,
 - Setup environment: Config, logging, dbus, etc..
- CASA has python bindings for all tasks
- Many tasks open a C++ coded GUI, these are wrapped so that output goes to notebook.
- https://github.com/aardk/jupyter-casa

Example: plotms



plotms(vis='3c391_ctm_mosaic_10s_spw0.ms', field='0', correlation='RR', timerange='08:02:00~08:17:00', antenna='ea01&ea02', xaxis='channel', yaxis='amp', ydatacolumn='model')

Example: plotms



plotms(vis='3c391_ctm_mosaic_10s_spw0.ms', field='0', correlation='RR', ..., xaxis='channel', yaxis='amp', ydatacolumn='model', plotfile='plotms_temp.png', showgui=False)

Example: casaviewer



viewer('3c391_ctm_spw0v2_l.image')

Example: casaviewer



viewer('3c391_ctm_spw0v2_l.image', gui = False, 11/22 outformat = 'png', outfile = viewer_temp.png)

Logging

		Log Messages (:/jop92_1/keimpem	a/home/data/casa/casa-20171012-092720.log)
<u>F</u> ile <u>E</u> dit <u>V</u> iew			
昌 🗟 😤 📰 📈 💭 Search	Message:		🚜 💆 Filter: Time 🔽
Time	Priority	Origin	Message
2017-10-12 10:24:17	INFO	bandpass::Calibrater::	Initializing nominal selection to the whole MS.
2017-10-12 10:24:17	INFO	bandpass::calibrater::reset	Reseting solve/apply state
2017-10-12 10:24:17	INFO	bandpass::calibrater::setdata	Beginning selectvis(MSSelection version)
2017-10-12 10:24:17	INFO	bandpass::calibrater::reset	Reseting solve/apply state
2017-10-12 10:24:17	INFO	bandpass::Calibrater::selectvis	Performing selection on MeasurementSet
2017-10-12 10:24:17	INFO	bandpass::Calibrater::selectvis+	Selecting on field: 'J1331+3030'
2017-10-12 10:24:17	INFO	bandpass::Calibrater::selectvis	By selection 845379 rows are reduced to 31964
2017-10-12 10:24:17	INFO	bandpass::Calibrater::selectvis	Frequency selection: Selecting all channels in all spws.
2017-10-12 10:24:17	INFO	bandpass::calibrater::setdata	<pre>chanmode=none nchan=1 start=0 step=1 mStart='0km/s' mStep='0km/s' msSelect=''</pre>
2017-10-12 10:24:17	INFO	bandpass::calibrater::setapply	Beginning setapply(MSSelection version)
2017-10-12 10:24:17	INFO	<pre>Calibrater::setapply(type, applypar)</pre>	Arranging to APPLY:
2017-10-12 10:24:17	INFO	bandpass::::	(KAntPos Jones: Overriding with spwmap=[0] since KAntPos Jones is not spw-depe
2017-10-12 10:24:17	INFO	bandpass::::	(KAntPos Jones: Enforcing calWt()=False for phase/delay-like terms)
2017-10-12 10:24:17	WARN	bandpass::::	No VLATrDelCorr keyword in the antpos caltable; turning trop delay correction C
2017-10-12 10:24:17	INFO	<pre>Calibrater::setapply(type, applypar)</pre>	. KAntPos Jones: table=3c391_ctm_mosaic_10s_spw0.antpos select= interp=linear
2017-10-12 10:24:17	INFO	bandpass::calibrater::setapply	Beginning setapply(MSSelection version)
2017-10-12 10:24:17	INFO	<pre>Calibrater::setapply(type, applypar)</pre>	Arranging to APPLY:
2017-10-12 10:24:17	INFO	<pre>Calibrater::setapply(type, applypar)</pre>	. G Jones: table=3c391_ctm_mosaic_10s_spw0.G0 select= interp=linear spwmap=[-
2017-10-12 10:24:17	INFO	bandpass::calibrater::setapply	Beginning setapply(MSSelection version)
2017-10-12 10:24:17	INFO	<pre>Calibrater::setapply(type, applypar)</pre>	Arranging to APPLY:
2017-10-12 10:24:17	INFO	bandpass::::	(K Jones: Enforcing calWt()=False for phase/delay-like terms)
2017-10-12 10:24:17	INFO	<pre>Calibrater::setapply(type, applypar)</pre>	. K Jones: table=3c391_ctm_mosaic_10s_spw0.K0 select= interp=linear spwmap=[-
2017-10-12 10:24:17	INFO	bandpass::calibrater::setsolve	Beginning setsolve(MSSelection version)
2017-10-12 10:24:17	INFO	bandpass::Calibrater::setsolve	Arranging to SOLVE:
2017-10-12 10:24:17	INFO	bandpass::Calibrater::setsolve	. B Jones: table=3c391_ctm_mosaic_10s_spw0.B0 append=false solint=inf,none re
2017-10-12 10:24:17	INFO	bandpass::calibrater::solve	Beginning solve
2017-10-12 10:24:17	INFO	bandpass::Calibrater::solve	The following calibration terms are arranged for apply:
2017-10-12 10:24:17	INFO	bandpass::Calibrater::solve	. KAntPos Jones: table=3c391_ctm_mosaic_10s_spw0.antpos select= interp=linear
2017-10-12 10:24:17	INFO	bandpass::Calibrater::solve	. K Jones: table=3c391_ctm_mosaic_10s_spw0.K0 select= interp=linear spwmap=[-
2017-10-12 10:24:17	INFO	bandpass::Calibrater::solve	. G Jones: table=3c391_ctm_mosaic_10s_spw0.G0 select= interp=linear spwmap=[-
Negative Magazara			
Insert Message:			

CASA displays logging information inside *casalogger* task.

Logging



Logging

🗢 PythonDataScienceHan 🗙 🥔 vi	a-cont-tutorial ×		ACIE		
$\langle \cdot ightarrow extbf{C}$ \bigcirc localhost:8888/	notebooks/vla-cont-tutorial.ipynb	☆	🕳 🚥		:
👯 Apps 🗋 😈 Unity Game Develop	😂 Spotify Web Player - 🌔 The Swedish Teache 🔍 Januari recept: Rund 🛐 South Indian-Style V 🌟 Bookmarks	»	Other b	ookma	irks
💭 Jupyter 🗤	/la-cont-tutorial Last Checkpoint: 39 minutes ago (autosaved)	Logout			
File Edit V	iew Insert Cell Kernel Help Not Trusted	Casa (
B + % 4					
In [27]:	<pre>bandpass(vis='3c391_ctm_mosaic_10s_spw0.ms',caltable='3c391_ctm_mosaic_10s_spw0.B0', field='J1331+3030',spw='',refant='ea21',combine='scan', solint='inf',bandtype='B', gaintable=['3c391_ctm_mosaic_10s_spw0.antpos', '3c391_ctm_mosaic_10s_spw0.G0', '3c391_ctm_mosaic_10s_spw0.K0'])</pre>				
	Show log 2017-10-12 10:24:17 INFO bandpass::Calibrater::open Opening MS: 3c391_ctm_mosaic_10s_spw0.ms for calibratic 2017-10-12 10:24:17 INFO bandpass::Calibrater:: Initializing nominal selection to the whole MS. 2017-10-12 10:24:17 INFO bandpass::Calibrater::reset Reseting solve/apply state 2017-10-12 10:24:17 INFO bandpass::Calibrater::reset Reseting solve/apply state 2017-10-12 10:24:17 INFO bandpass::Calibrater::reset Reseting solve/apply state 2017-10-12 10:24:17 INFO bandpass::Calibrater::selectvis Performing selection on MeasurementSet 2017-10-12 10:24:17 INFO bandpass::Calibrater::selectvis Fequency selection selecting all channels in all spws. 2017-10-12 10:24:17 INFO bandpass::Calibrater::selectvis Frequency selection: Selecting all channels in all spws. 2017-10-12 10:24:17 INFO bandpass::Calibrater::selectvis Frequency selection: Selecting all channels in all spws. 2017-10-12 10:24:17 INFO bandpass::Calibrater::setapply Beginning setapply(MSSelection version)2 2017-10-12 10:24:17 INFO bandpass::Calibrater::setapply Beginning setapply(MSSelection version)2 2017-10-12 10:24:17 INFO bandpass:::Calibrater::setapply(type, applypar) Arranging to APPLY: 2017-10-12 10:24:17 INFO bandpass::: (KAntPos Jones: Enforcing calWt()=False for phase/delay-like terms) 2017-10-12 10:24:17 INFO bandpass::: No VLATrDelCorr keyword in the antpos caltable; turning trop delay correction 2017-10-12 10:24:17 INFO bandpass:::Calibrater::setapply(type, applypar	pw- tion OFF.			

CASA for Jupyter

- NRAO CASA distribution is entirely selfcontained, it is essentially a linux distribution
- Too old for Jupyter, many conflicting packages, e.g. Matplotlib, IPython,
- We created a custom build of CASA using based on latest Python
- Only functional with Jupyter, no stand-alone CASA interpreter.
- Distributed as **Docker** and **Singularity** containers.

Docker containers

- Lightweight alternative to virtual machines
- Docker containers are isolated from host, resources have to be shared explicitly.
- Download Jupyter-CASA from dockerhub:
 - docker pull penngwyn/jupytercasa
- Run Jupyter-CASA:
 - docker run --rm -p 8888:8888 -i -t -v /tmp/.X11unix:/tmp/.X11-unix -e DISPLAY=\$DISPLAY penngwyn/jupytercasa /bin/sh -c "jupyter notebook"
- Docker not acceptable at many computing centres: Root privilege escalation

Singularity containers

- Can import directly from Docker
- Less isolated: /dev, /tmp, /proc, and \$HOME are shared with host
- Runs as in the user context of the user that executes: "singularity run"
- Download Jupyter-CASA from Singularity-hub
 - singularity pull shub://aardk/jupyter-casa
- Run Jupyter-CASA:
 - singularity run aardk-jupyter-casa-master.img
- Doesn't have root privilege escalation issues 17/22

Minimal re-computation pipelines



- Integrate minimal re-computation into Jupyter
- Original framework part of Hilado / Radionet
 - JIVE: Des Small and Mark Kettenis
 - U. Cambridge: Bojang Nikolic

Minimal re-computation pipelines



- Integrate minimal re-computation into Jupyter
- Original framework part of Hilado / Radionet
 - JIVE: Des Small and Mark Kettenis
 - U. Cambridge: Bojang Nikolic

Minimal re-computation pipelines



- Integrate minimal re-computation into Jupyter
- Original framework part of Hilado / Radionet
 - JIVE: Des Small and Mark Kettenis
 - U. Cambridge: Bojang Nikolic

Conclusions

- We have implemented Jupyter kernel for CASA suitable for pipelines
- We provide both Docker and Singularity images for easy deployment
- Future work includes:
 - Implement minimal re-computation
 - Jupyter hub integration
- https://github.com/aardk/jupyter-casa





Text for acknowledgement Slide

Acknowledgement

 H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).