# CASA Details


Jeff Kern

CASA Team Lead


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

NRAO

# CASA 3.4

- Current release is CASA 3.3
- CASA 3.4 planned release is May 1, 2012
  - "Stable" updates are available from the CASA download page
    - http://casa.nrao.edu/casa_obtaining.shtml
- CASA 3.5 (around)

# Virtual Model Data

- Currently there are two scratch columns in the MS
  - MODEL_DATA
  - CORRECTED_DATA
- With 3.4 the Model Data column can be virtual
  - Reduced IO and Disk Consuption
  - Should be transparent
- Changes where time is spent in the

# New Flagging System

- New Flagging infrastructure introduced in 3.4
  - Replaces 'flagdata', 'flagdata2', 'flagcmd', 'testautoflag'
  - Plan to remove deprecated tasks in 3.5
- Underlying code re-written for
  - Performance Optimization
    - Asynchronous IO, Parallelization, pre-selection of data
  - Batch Modes (Many command in a single pass)
    - tflagdata: Single commands and batch commands from input text file
    - tflagcmd: Batch commands with the FLAG_CMD subtable
      - Similar to "flagcmd"

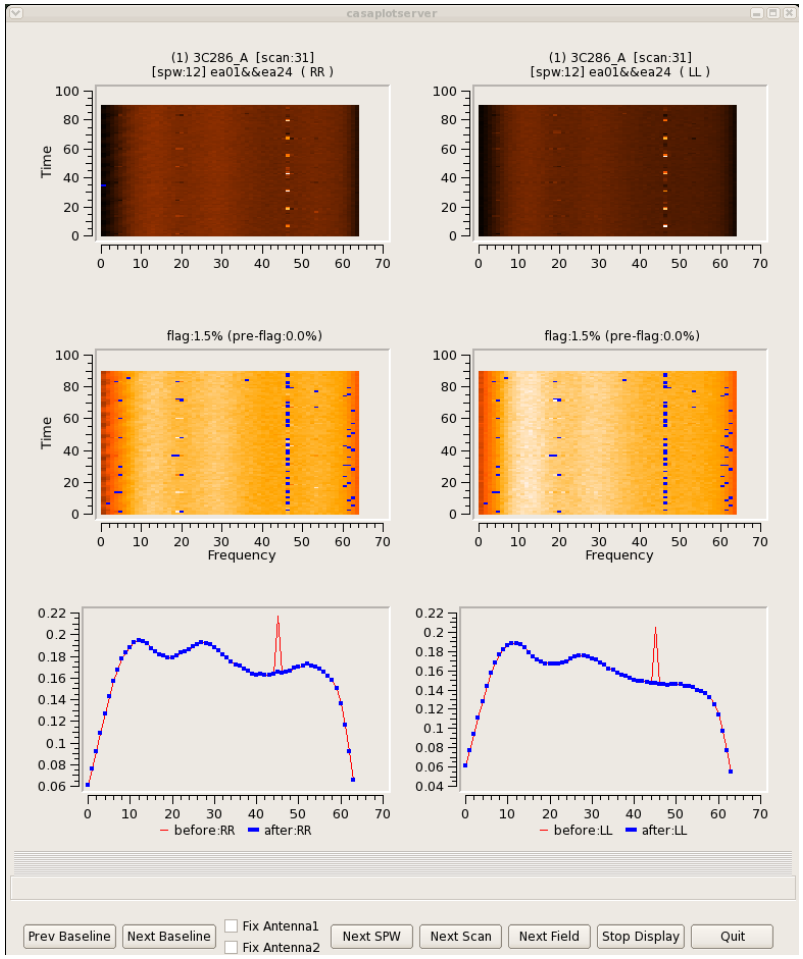
# New Flagging System

- New and Improved modes:
  - Automatic Flagging
    - "rflag" : based on E. Greisen's algorithm in AIPS
    - "tfcrop" : previously available in 'testautoflag'
  - Flag Extension
    - along time, frequency, polarization, baselines
  - Shadowing
    - Support for physically present antennas not listed in MS

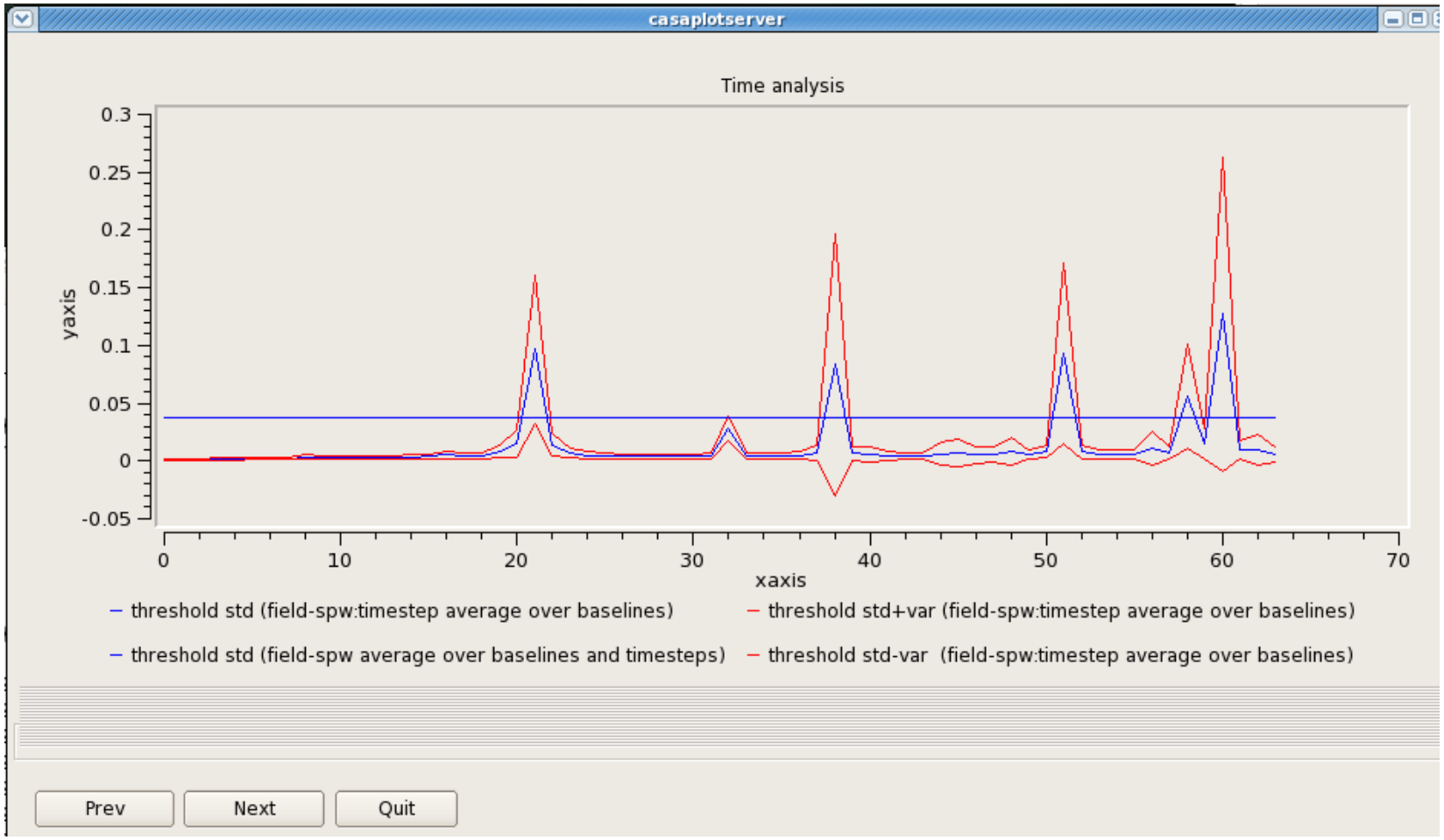
# New Flagging System

- Summary and Display
  - Integrated display option for single or batch modes to monitor flagging if desired.
    - Particularly useful for autoflagging
  - Summary reports
    - Generated by any mode either displayed or returned as python dictionaries.

# Interactive Display



# Flagging Report



# Flagging Output

| Origin | Message |
|---|---|
| FlagDataHandler::nextBuffer | ---------- |
| FlagDataHandler::nextBuffer | Chunk = 28, Observation = 0, Array = 0, Scan = 31, Field = 1 (3C286_A), Spw = 11, |
| Display::getChunkInfo | FlagAgentDisplay::getChunkInfo Field : 1 , 3C286_A Spw : 11 Scan : 31 : 31 |
| Manual_0::chunkSummary | => Data flagged in this chunk: 16.6667% |
| Quack_2::chunkSummary | => Data flagged in this chunk: 6.66667% |
| FlagDataHandler::nextBuffer | ---------- |
| FlagDataHandler::nextBuffer | Chunk = 29, Observation = 0, Array = 0, Scan = 31, Field = 1 (3C286_A), Spw = 12, |
| Display::getChunkInfo | FlagAgentDisplay::getChunkInfo Field : 1 , 3C286_A Spw : 12 Scan : 31 : 31 |
| Manual_0::chunkSummary | => Data flagged in this chunk: 16.6667% |
| Quack_2::chunkSummary | => Data flagged in this chunk: 6.66667% |
| FlagDataHandler::nextBuffer | ---------- |
| FlagDataHandler::nextBuffer | Chunk = 30, Observation = 0, Array = 0, Scan = 31, Field = 1 (3C286_A), Spw = 13, |
| Display::getChunkInfo | FlagAgentDisplay::getChunkInfo Field : 1 , 3C286_A Spw : 13 Scan : 31 : 31 |
| Manual_0::chunkSummary | => Data flagged in this chunk: 16.6667% |
| Quack_2::chunkSummary | => Data flagged in this chunk: 6.66667% |
| FlagDataHandler::nextBuffer | ---------- |
| FlagDataHandler::nextBuffer | Chunk = 31, Observation = 0, Array = 0, Scan = 31, Field = 1 (3C286_A), Spw = 14, |
| Display::getChunkInfo | FlagAgentDisplay::getChunkInfo Field : 1 , 3C286_A Spw : 14 Scan : 31 : 31 |
| Manual_0::chunkSummary | => Data flagged in this chunk: 16.6667% |
| Quack_2::chunkSummary | => Data flagged in this chunk: 6.66667% |
| FlagDataHandler::nextBuffer | ---------- |
| FlagDataHandler::nextBuffer | Chunk = 32, Observation = 0, Array = 0, Scan = 31, Field = 1 (3C286_A), Spw = 15, |
| Display::getChunkInfo | FlagAgentDisplay::getChunkInfo Field : 1 , 3C286_A Spw : 15 Scan : 31 : 31 |
| Manual_0::chunkSummary | => Data flagged in this chunk: 16.6667% |
| Quack_2::chunkSummary | => Data flagged in this chunk: 6.66667% |
| FlagDataHandler::nextChunk | ========== |
| Manual_0::msSummary | => Total data flagged in MS: 16.6667% |
| Clip_1_RR::msSummary | => Total data flagged in MS: 0.000977472% |
| Clip_1_RL::msSummary | => Total data flagged in MS: 0.000727421% |
| Clip_1_LR::msSummary | => Total data flagged in MS: 0.000591029% |
| Clip_1_LL::msSummary | => Total data flagged in MS: 0.000886544% |
| Quack_2::msSummary | => Total data flagged in MS: 6.14525% |
| TestFlagger::run | => Writing flags to the MS |

# Calibration Table Refactor

- New Calibration tables come in 3.4
  - Display moves to plotMS
  - Improved interpolation (Freq. Dependent)
  - Can be applied to different MS
- Not backwards compatible
  - We will supply a task to convert old tables to new format

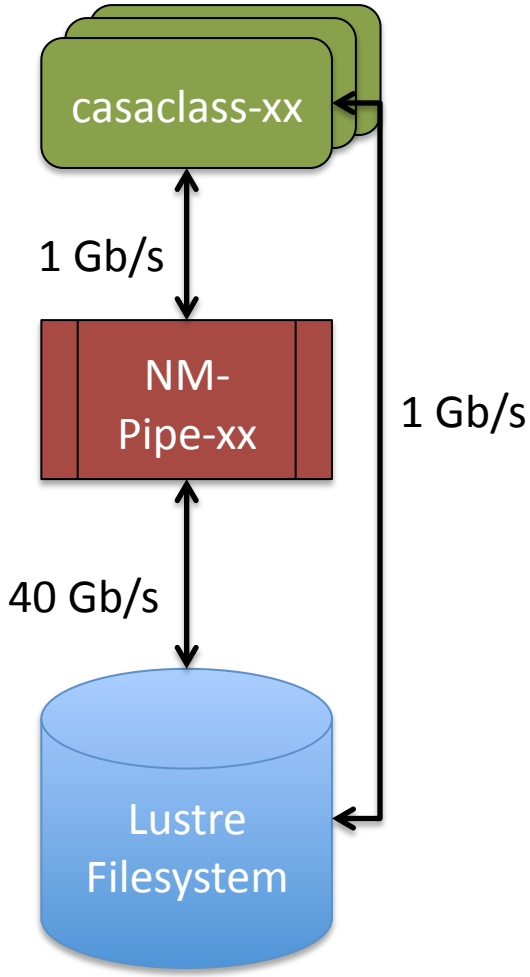
# CASA Performance

- CASA performance as experienced by a user is determined by three factors:
  - File I/O
  - Available memory
  - Processor speed

# Workshop Configuration

- Hex Core Intel Xeon
- 8 GB RAM

- Dual Hex Core Intel Xeon
- 24 GB RAM



# The processing performance problem

- Many tasks in CASA require traversing the entire data set and are data IO limited.

| | Peak Data Rate | OS Realized Data Rate | Time for 500 GB file |
|---|---|---|---|
| SATA Disk | 115 MB/s | 60 MB/s | 2h 22 m |
| Raid | 200-500 MB/s | 375 MB/s | 22 m |
| Lustre (10 GB) | 7+ GB/s | 700 MB/s | 12 m |

- Even imaging [in the simplest case] is IO limited, requiring about 65 MB/s per core to prevent the CPU from incurring wait states.

# Parallelization

- Note that this week, nearly all of the processing has been serialized (one core).
- Parallel CASA is coming
  - Already in the testing stages
  - Will make more efficient use of the available cores on a system.
  - Higher demand on memory
  - Will not help with the IO issues

# Conclusion

- As the capabilities of the EVLA continue to grow, so will the size of the data sets.
- Optimizing CASA's performance is not simply a matter of buying the fastest processor, or most memory.
- The CASA team is working to allow optimum performance on workstations as well as clusters.

https://science.nrao.edu/facilities/evla/data-processing/hardware-recommendations