EVLA Pipelines

February 24, 2012 – Socorro, NM USA

Bryan Butler

EVLA Computing Division Head

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

NRAO

Why all this *&!$ typing?!

We've shown you how to reduce this kind of data in general, and even provided some explicit guides (including all the commands), and you've done the reduction yourself.

You must be asking yourself: "Why can't this be automated? Why can't we just take all those commands from the casaguides page, and roll them up, and run them as a single script?"

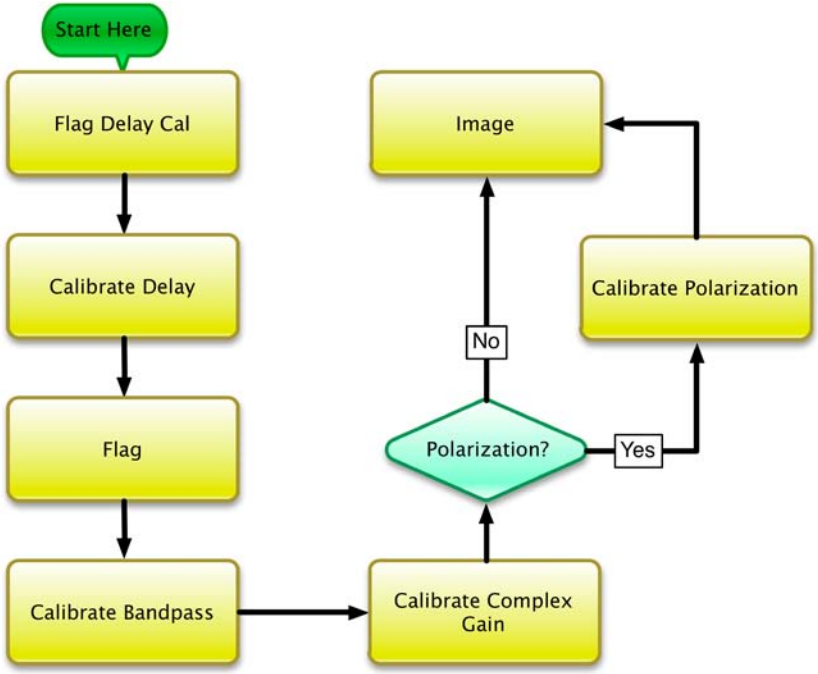
Well, you could, and it would work for specifically crafted datasets and scripts, but it turns out that it's not easy to do it so that it works for most types of observing without much human interaction - folks have been trying to make automated processing pipelines for radio astronomical data for decades and been only partly successful.

Nevertheless, we are going to do this for both EVLA and ALMA, and we believe we can do it for many observing modes, because we're being more careful about gathering the necessary data to make it viable.

Pipelines

What is a "pipeline"? A set of commands (perhaps as a single script, but not necessarily so) which is used to execute a series of post-processing steps, with little to no interaction.

What goes in is raw visibility data (maybe calibrated).
What comes out are "derived products" (the sequence of commands, logs, plots, tables, images, image cubes, etc.).



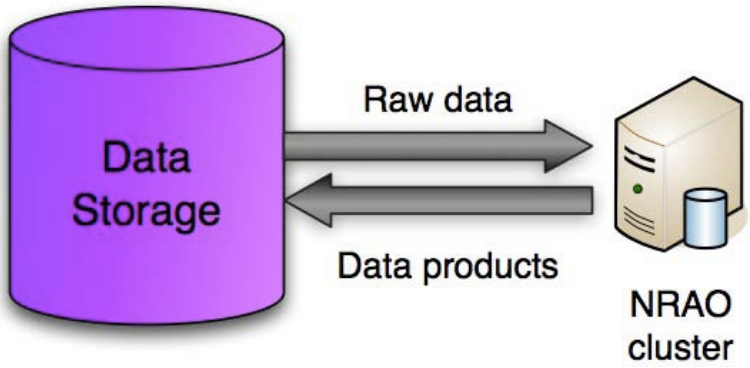
Pipeline Flavors

We intend to implement three different types of pipelines:

1. Standard pipeline
2. Batch reprocessing pipeline
3. Quasi-interactive reprocessing pipeline

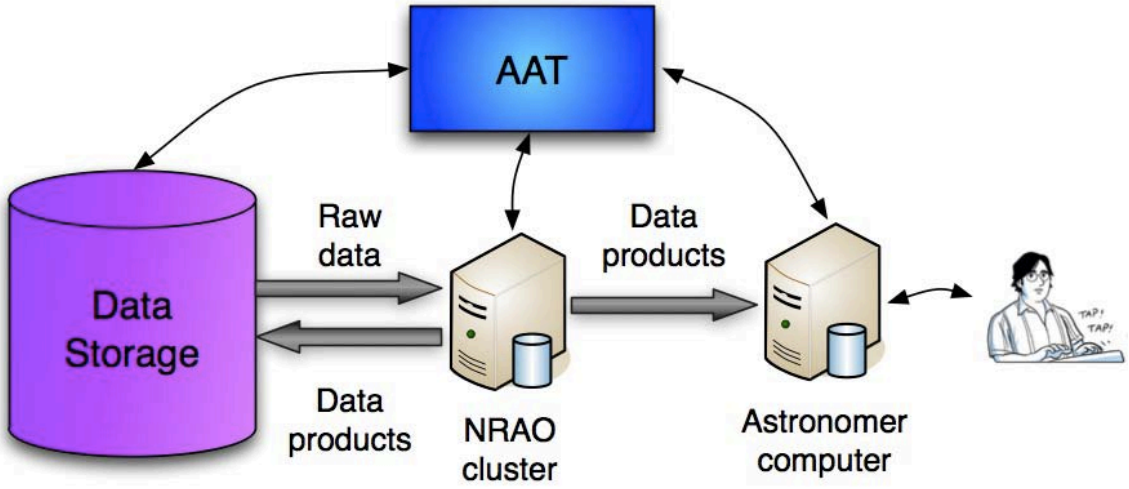
Standard Pipeline Reduction

- Happens to every "standard" SB as it is observed (successfully).
- No user-settable parameters – everything determined by "heuristics".
- Performed on NRAO cluster, with highest priority.
- 2 stages of QA – "level-1" is statistics from the data; "level-2" is human inspection.
- Data products are always stored back in the archive.



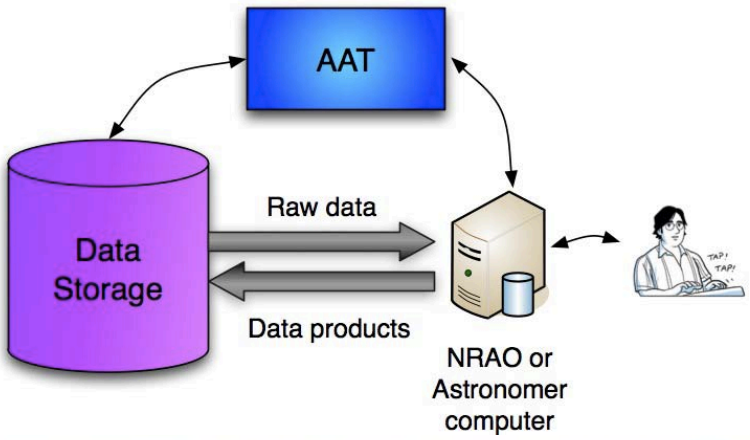
Batch Reprocessing

- Re-run of the standard pipeline, but user may set some of the parameters.
- Performed on NRAO cluster, with lowest priority.
- Only "level-1" QA done.
- Data products are only stored back in the archive if the user chooses.



Interactive Reprocessing

- More traditional post-processing (at terminal).
- Can be on NRAO cluster or workstation (with permission) or on astronomer home hardware.
- If on NRAO cluster, priority is intermediate.
- No QA done.
- Data products are only stored back in the archive if the user chooses.



Sounds Great, But, When???

The development of the pipeline infrastructure itself, much less the heuristics, is difficult, so it will take time to develop. We will have this all in place by the end of construction (Jan 2013), with early versions starting this spring into summer/fall.