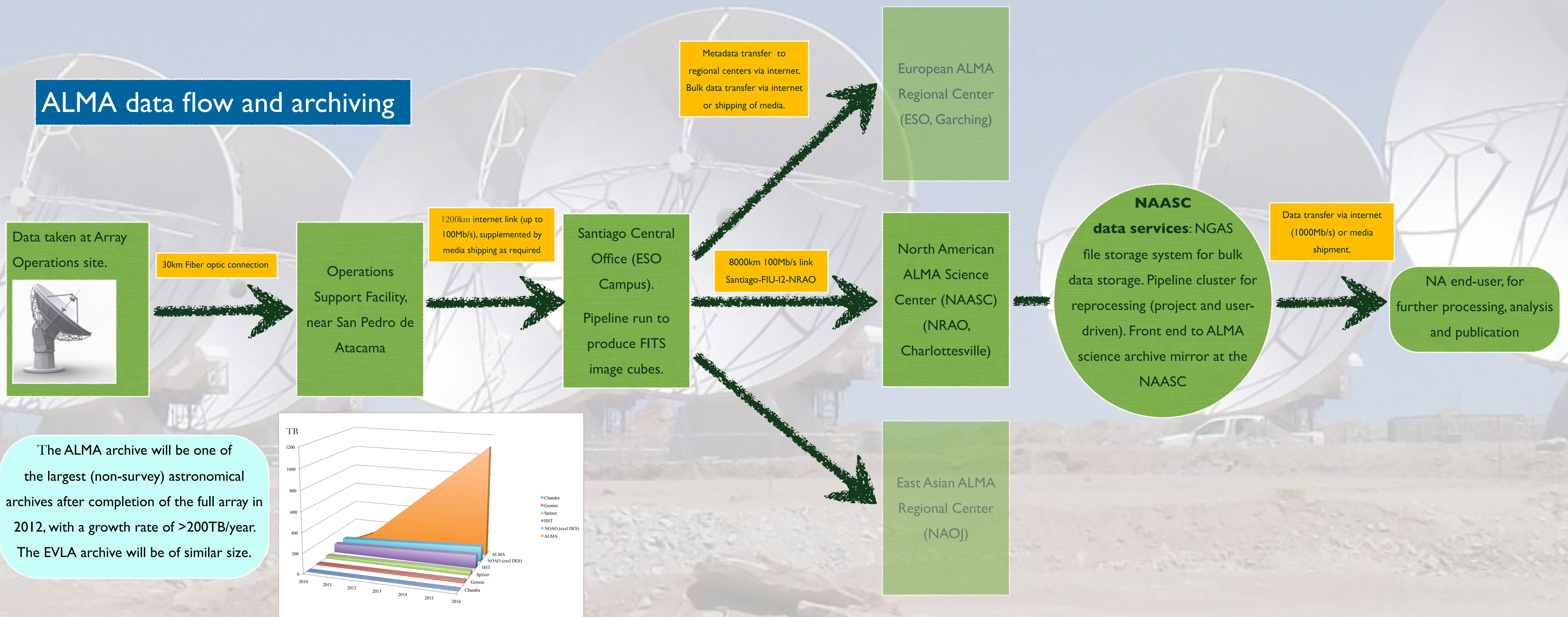


# Data processing and archiving at the North American ALMA Science Center

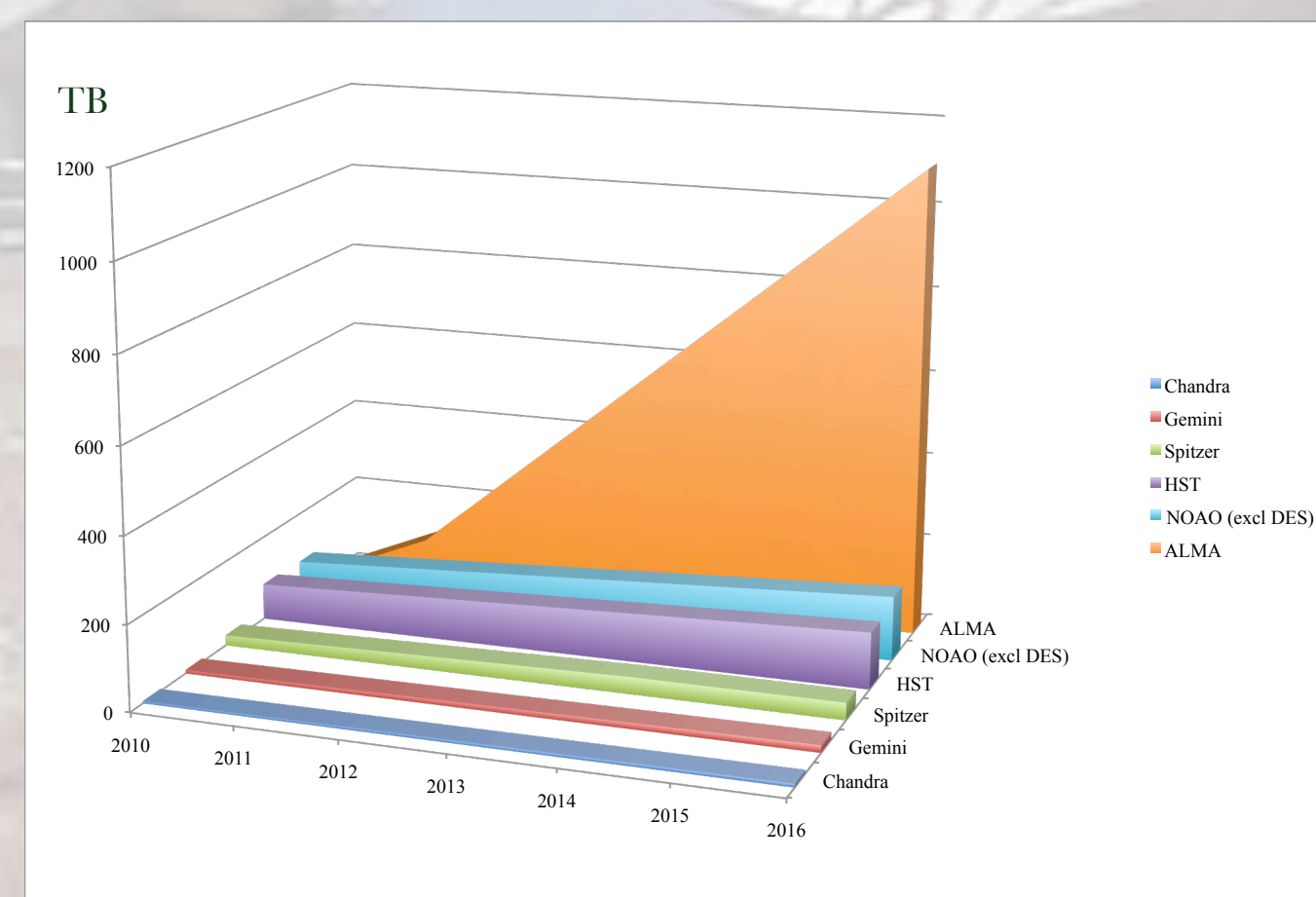
Mark Lacy, David Halstead, Mike Hatz  
North American ALMA Science Center, NRAO

The Atacama Large mm/submm Array (ALMA) is expected to begin Early Science operations in approximately one year. The data rate is expected to ramp up from ~20TB/yr (using the first 16 antennas) in 2011-2012 to the fully operational rate (with all 64 antennas) of ~200TB/yr in 2013. During this time, our data processing capabilities will also evolve, as we move from being able to process most datasets on desktop machines, to needing to use computing clusters to process a typical dataset. In this poster we present the data processing and archiving plans for the NAASC, and how these relate to the observatory pipeline processing and archiving activities at the Joint ALMA Observatory in Santiago.

## ALMA data flow and archiving

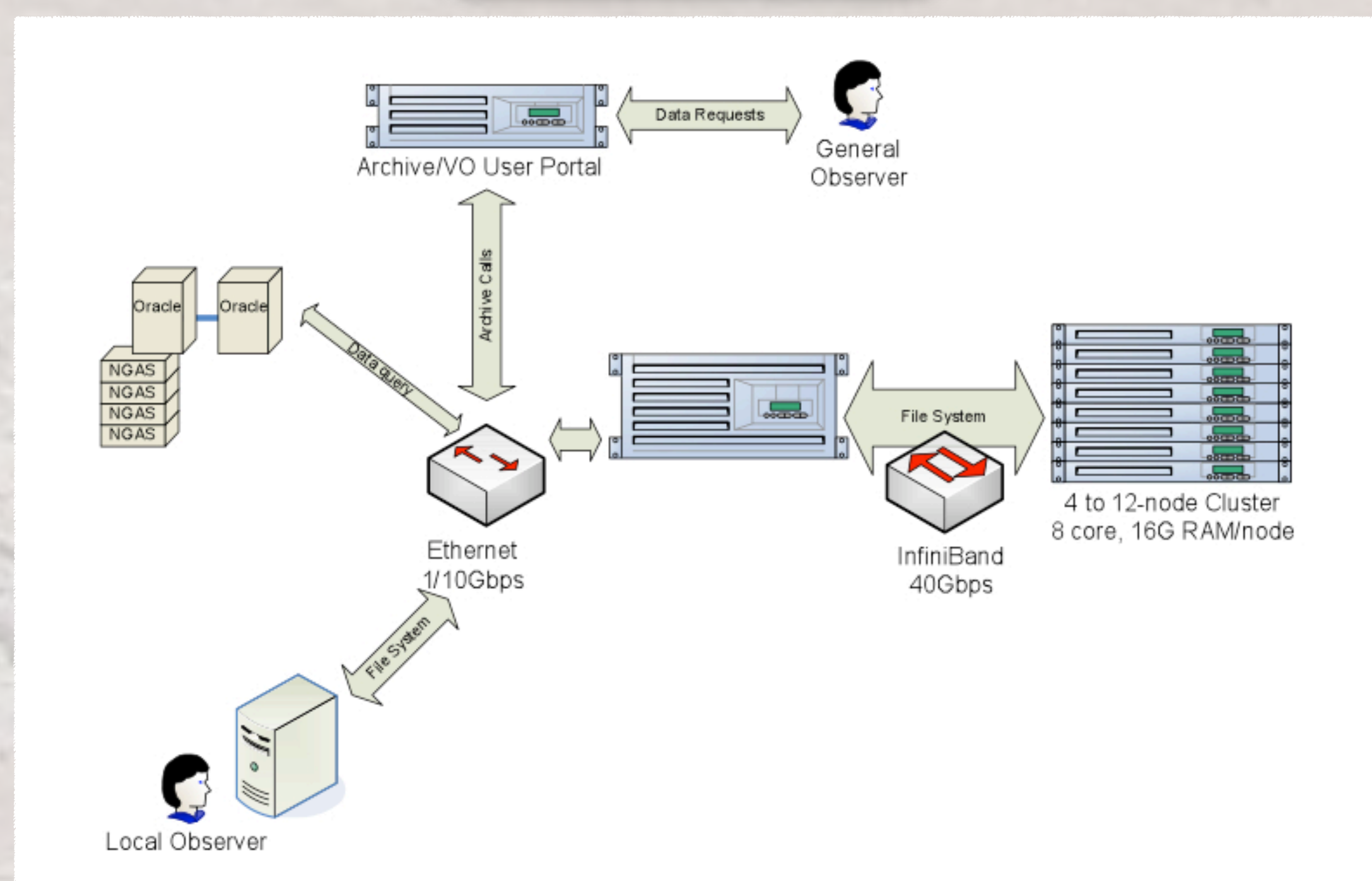


The ALMA archive will be one of the largest (non-survey) astronomical archives after completion of the full array in 2012, with a growth rate of >200TB/year. The EVLA archive will be of similar size.

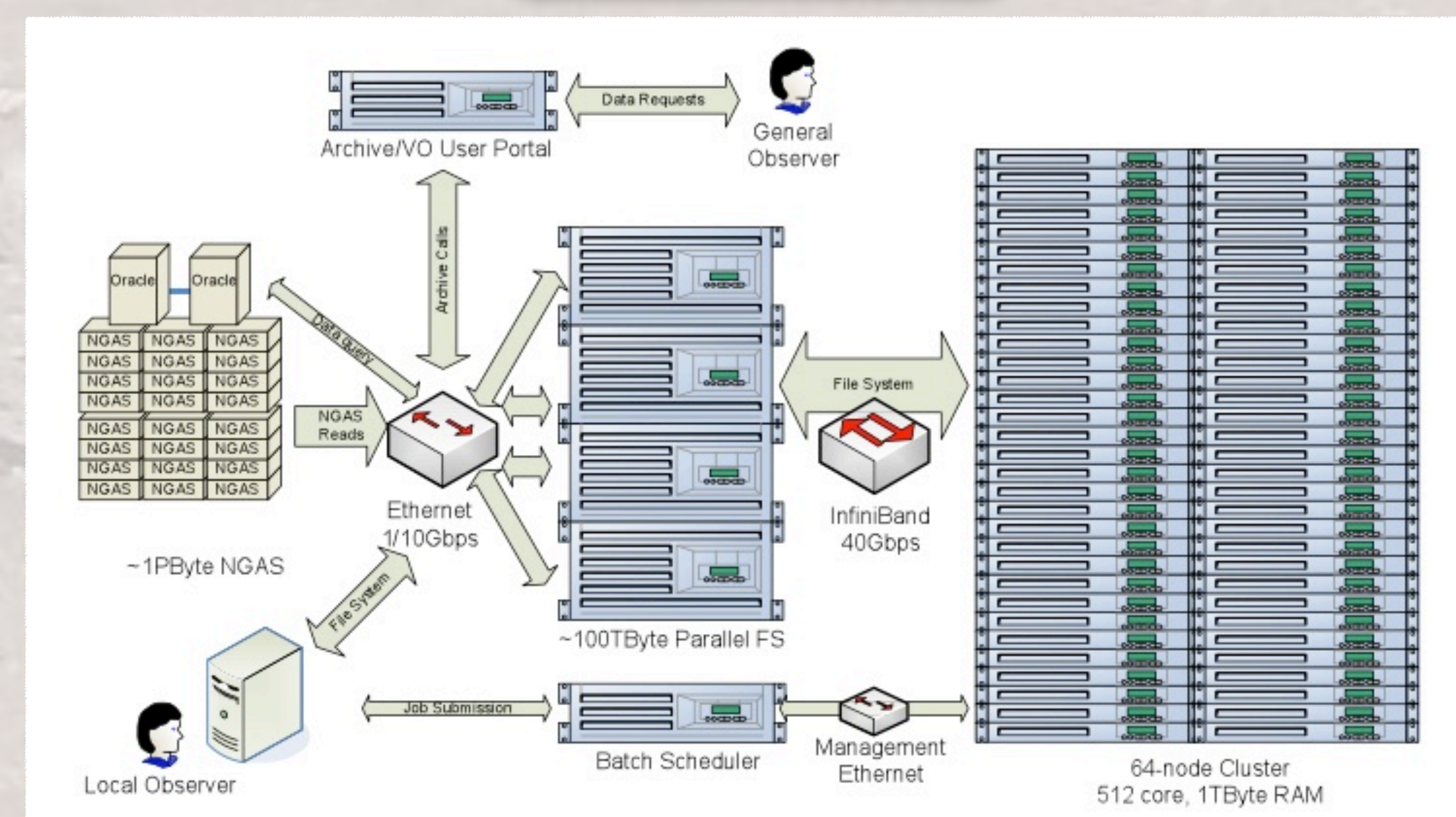


## Reprocessing at the NAASC

### Early Science



### Full operations



Our cluster will be grown in line with the processing requirements, starting with a single node early next year, to a full ~64 node machine early in 2012. The choice of cluster node will be governed by the cluster currently being constructed to process EVLA data at NRAO in Socorro, NM. Similar versions of this cluster will be installed at the ALMA Santiago Central Office and at the European and Japanese ALMA regional centers. Key features:

- Next Generation Archive System (NGAS, supplied by ESO).
- Fast file system (Lustre) for short-term storage while processing.
- Infiniband link between cluster nodes and the Lustre system.