

The University of New Mexico

DEFINING DATA INTENSIVE COMPUTING

BARNEY MACCABE

COMPUTER SCIENCE DEPARTMENT

&

CENTER FOR HIGH PERFORMANCE COMPUTING

HPC@UNM

THE CENTER FOR HIGH PERFORMANCE COMPUTING

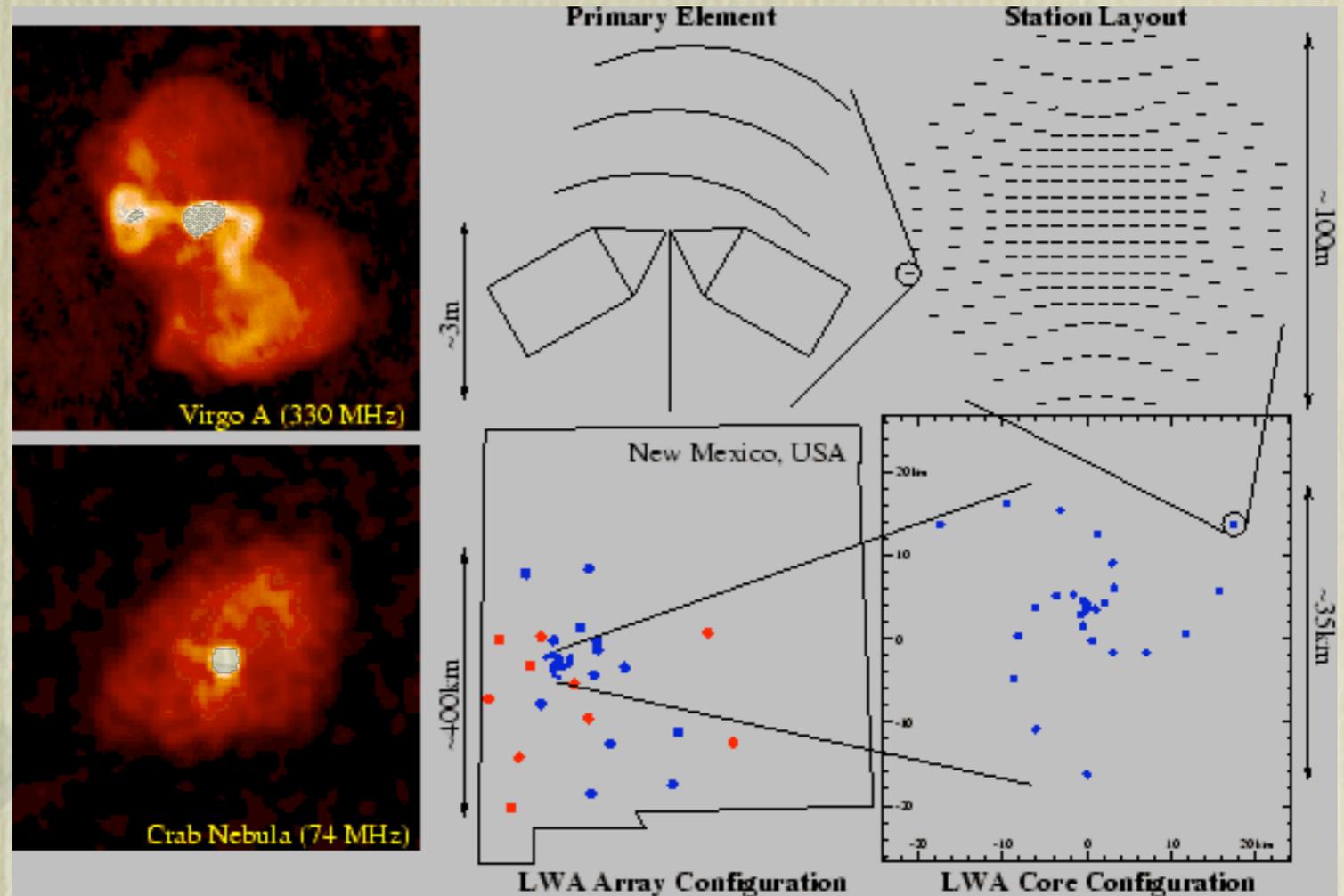
THE CENTER FOR HIGH PERFORMANCE COMPUTING

Driving Applications?

- ✿ What do you see as the driving applications for parallel file systems and high performance I/O over the next 5 to 10 years?
 - ✿ Remote Sensing (SBES Sim Based Eng Science)
 - ✿ Radio Astronomy
 - ✿ Medical applications
 - ✿ Social sciences
 - ✿ Fine arts & Humanities
- ✿ Does this represent a change over the last decade?
 - ✿ Yes.

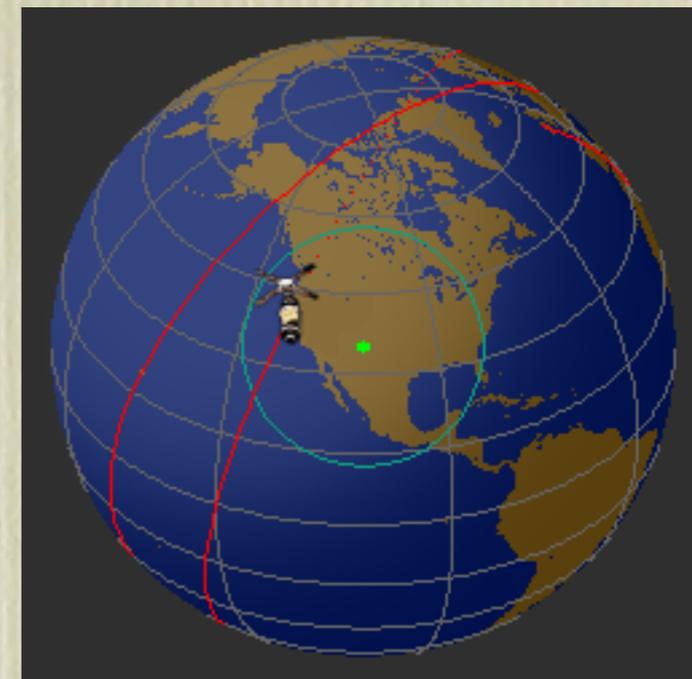
LWA—Radio Astronomy

- ☼ Long Wavelength Array
- ☼ Approximately 1TB/observation
- ☼ 3 observations/day
- ☼ 1 PB/year
- ☼ Processing requires 1 TOp/observation

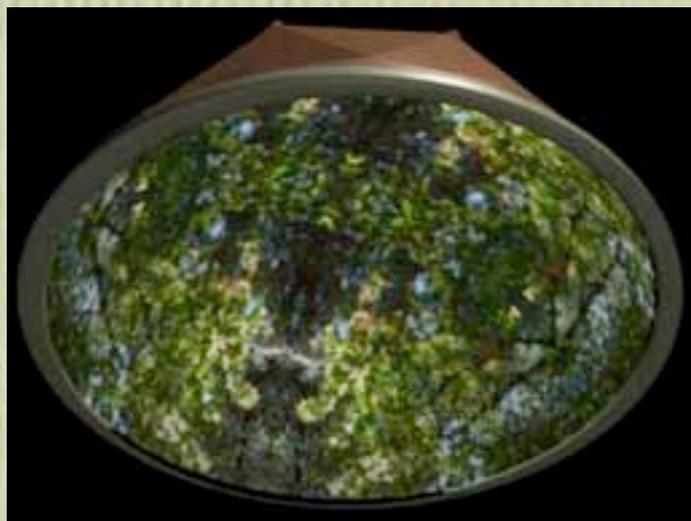


CREATE—Remote Sensing

- ✿ Center for Rapid Environmental Assessment and Terrain Evaluation
- ✿ Rapid assessment of satellite images
- ✿ 24–28 passes / day
- ✿ Current
 - ✿ Raw data 1.2 GB/pass–12.2 TB/year
 - ✿ Product 6 GB/pass–60 TB/year
- ✿ Future
 - ✿ 6–8 GB raw data/pass
 - ✿ couple to real-time modeling



ARTS Lab—Fine Arts



☀ Dome Theatre
(fulldome.org)

- ☀ take a planetarium, turn it 10–30° from horizontal
- ☀ play movies on the dome
- ☀ .75 TB/minute

Needy apps

- ☼ What do these applications need that current parallel files systems don't provide?
 - ☼ Fewer features.
 - ☼ Apps need “near physical” access to the I/O resources so they can manage these resources more effectively
 - ☼ Use libraries to add features
- ☼ Do we need a shift in thinking about high performance I/O?
 - ☼ Yes.

I/O and compute?

- ✿ Should we expect the same file systems to support applications that have both significant computation and I/O as well as those with only significant I/O?
- ✿ This is a resource management problem. We care about apps with significant computation and I/O because inefficient I/O wastes compute resources.
- ✿ If you don't have significant computation, I/O is less important. It may become the bottleneck for response time or require excessive buffer space.

Unix is dead, long live Unix

- ☼ The file systems from the 60's and 70's, designed for data processing, were quite different from those of the 80's and 90's, designed for interactive workstations. Have we lost any good ideas from those older systems that we should be reconsidering now?
- ☼ Channel programming?
- ☼ ISAM, HISAM, HIDAM?