Lustre-WAN: Enabling Distributed Workflows in Astrophysics
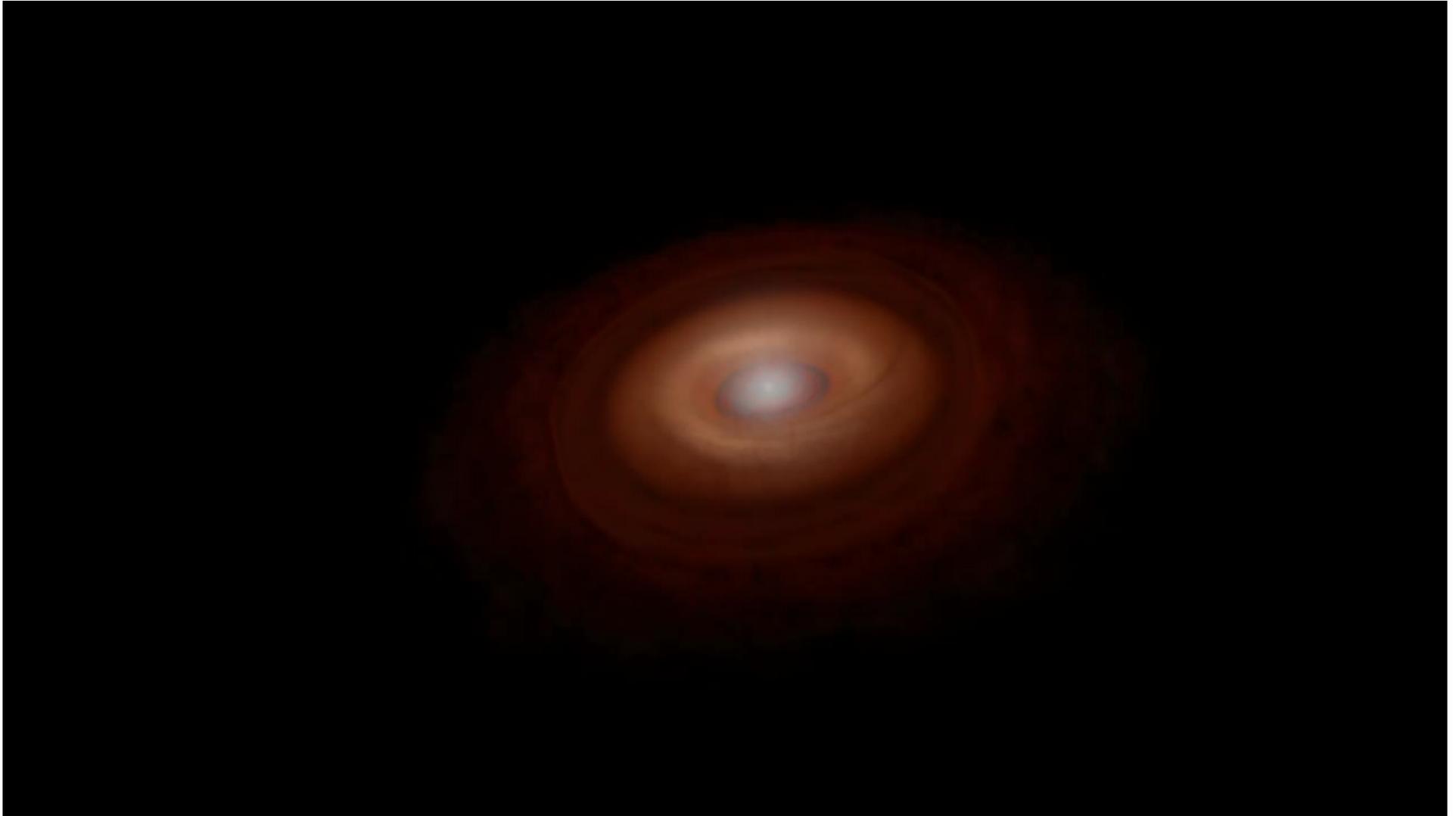
Scott Michael
scamicha@indiana.edu

April 2011

INDIANA UNIVERSITY

Studying Planet Formation

Our Workflow Before Using Lustre-WAN

- Our workflow has three parts
  - Simulation – shared memory
  - Analysis – distributed memory
  - Visualization – proprietary software with interactivity
- In the past we have transferred data between HPC resources, stored the data locally, and performed analysis and visualization
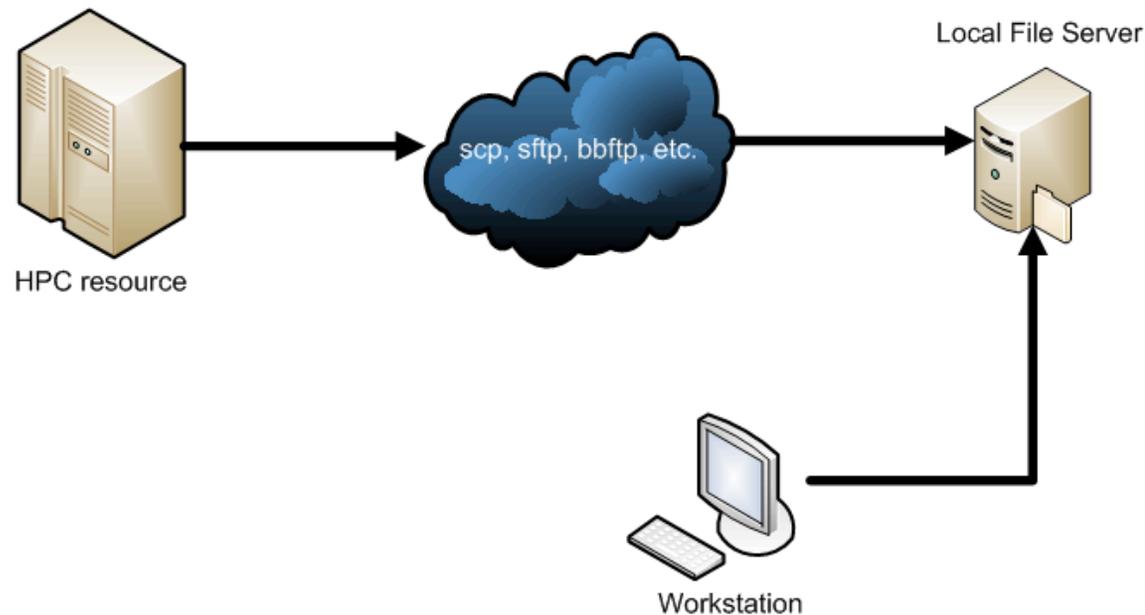
Using a Central Lustre File System

- Using Indiana University's Lustre based Data Capacitor WAN file system different systems in separate parts of the workflow can all access the same data
  - Data generated by SGI Altix
    - 4 simulations on NCSA's Cobalt
    - 6 simulations on PSC's Pople
  - Data analyzed by departmental resources and distributed memory machines
    - Indiana University's Big Red and Quarry
    - Mississippi State University's Raptor and Talon
  - Data visualization by departmental machines
    - Indiana University IDL license

Using a Central Lustre File System
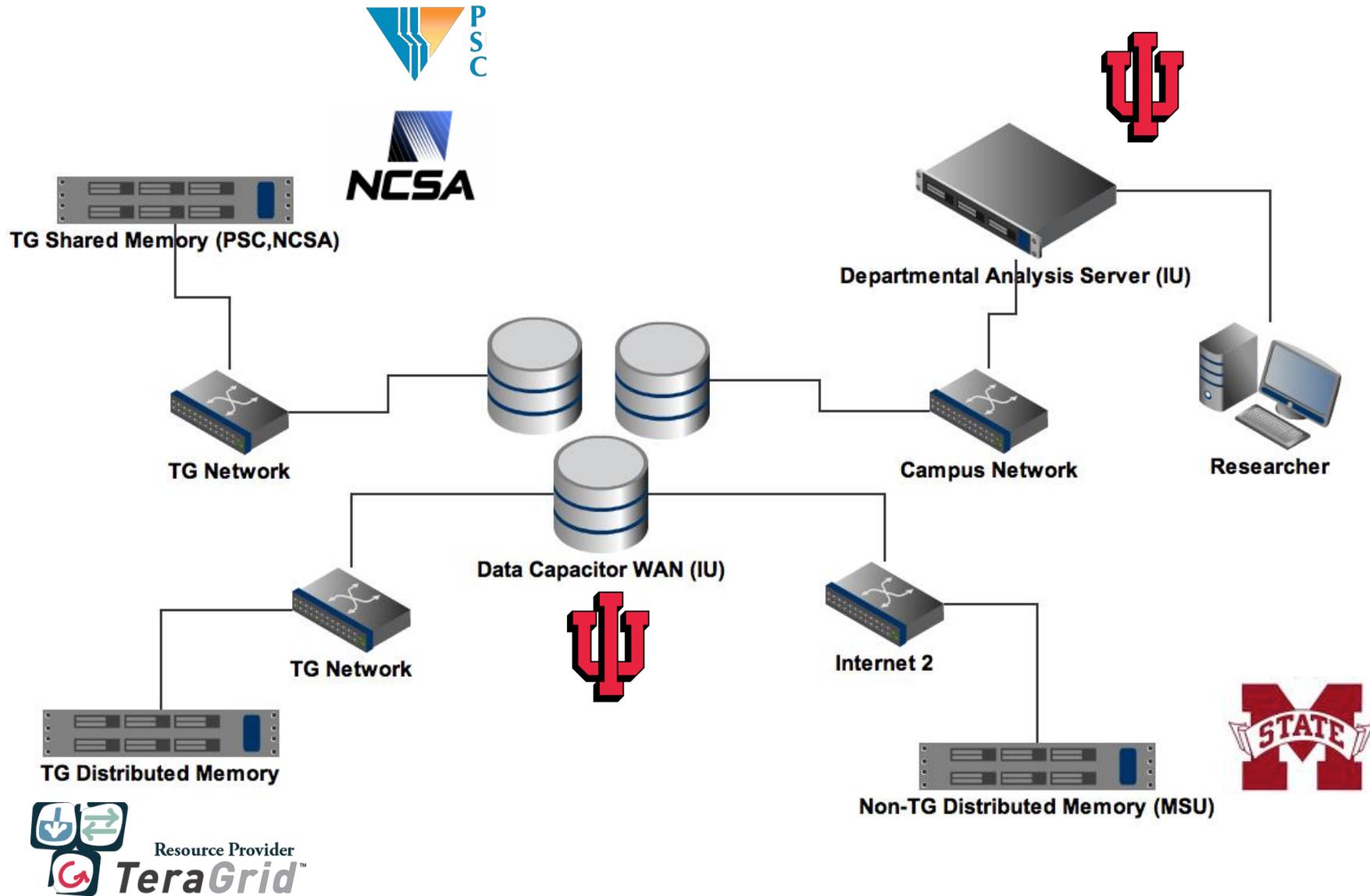
Use Cases for Lustre WAN

- Many cases where researcher needs resources outside a single data center
  - This is increasingly common in the TeraGrid
  - A user needs to use heterogeneous resources
    - Shared and Distributed Memory
    - GPUs
    - Visualization systems
  - A researcher needs to capture instrument data
  - A researcher needs to migrate to a different system
- Ideally every user can access all his data from any resource all the time

Thanks To

- DC Team
  - Steve Simms
  - Josh Walgenbach
  - Justin Miller
  - Nathan Heald
  - Eric Isaacson
- IU Research Technologies
  - Matt Link
  - Robert Henschel
  - Tom Johnson

- MSU Research
  - Trey Breckenridge
  - Roger Smith
  - Joey Jones
  - Vince Sanders
  - Greg Grimes
- PSC
- NCSA

INDIANA UNIVERSITY