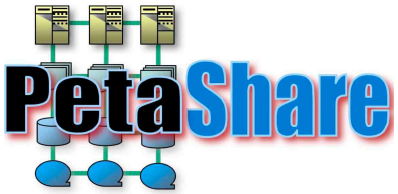


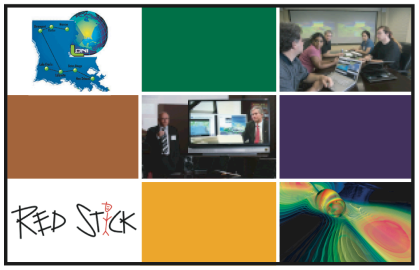
PetaShare:

Enabling Data Intensive Collaborative Science in Louisiana

Tevfik Kosar

Center for Computation & Technology
Louisiana State University

November 14, 2007



CENTER FOR COMPUTATION
& TECHNOLOGY

PetaShare – the Genesis

LONI brings:

+ fat pipes (40Gb/s) and vast computing resources (100 Tflops)

- missing a distributed data management and storage infrastructure

Our Goals:

- Bring additional storage to LONI
- Provide a CI for easy and efficient storage, access, and management of data.

"Let scientists focus on their science rather than dealing with low level data issues. The CI should take care of that."

GRID today

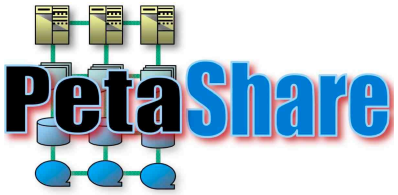
The Leading Source for Global News and Information from the evolving Grid ecosystem,
including Grid, SOA, Virtualization, Storage, Networking and Service-Oriented IT

"..The PetaShare system might become an important testbed for future Grids, and a leading site in next generation Peta-scale research."

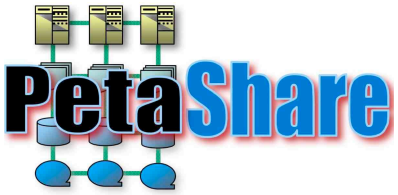
collaboration and sharing among the nation's education and research institutions. Simply purchasing high-capacity, high-performance storage systems and adding them to the existing infrastructure of the collaborating institutions does not solve the underlying and highly challenging data handling

".. has a potential to serve as a catalyst for coalescing researchers who might otherwise not develop the incentive to collaborate. ."

visualization on demand. PetaShare will enable scientists to focus on their primary research problem, assured that the underlying infrastructure will manage the low-level data handling issues.

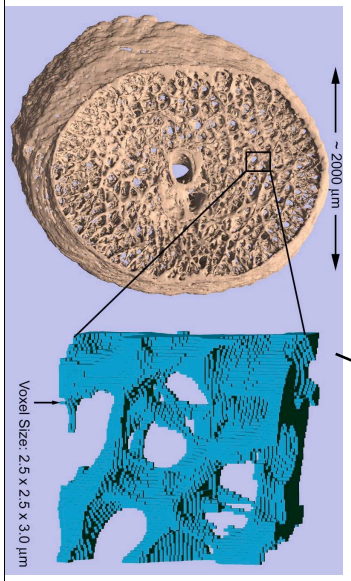


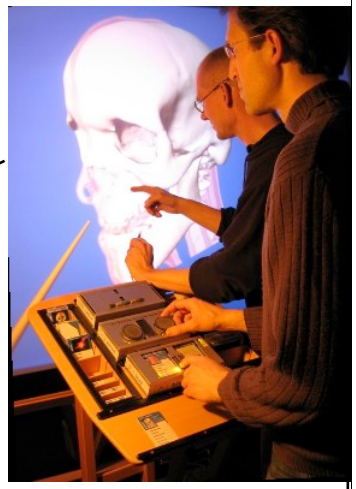
- Goal: enable domain scientists to focus on their primary research problem, assured that the underlying infrastructure will manage the low-level data handling issues.
- Novel approach: treat data storage resources and the tasks related to data access as first class entities just like computational resources and compute tasks.
- Key technologies being developed: data-aware storage systems, data-aware schedulers (i.e. Stork), and cross-domai meta-data scheme.
- Provides and additional 250TB disk, and 400TB tape storage (and access to national storage facilities)



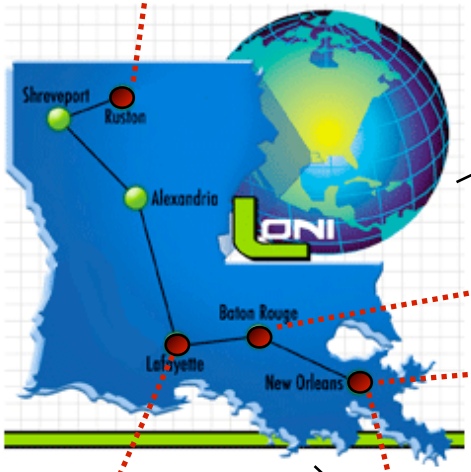
- PetaShare exploits 40 Gb/sec LONI connections between 5 LA institutions: LSU, LaTech, Tulane, ULL, and UNO.
- PetaShare links more than fifty senior researchers and two hundred graduate and undergraduate research students from ten different disciplines to perform multidisciplinary research.
- Application areas supported by PetaShare include coastal and environmental modeling, geospatial analysis, bioinformatics, medical imaging, computational fluid dynamics, petroleum engineering, numerical relativity, and high energy physics.

Participating institutions in the PetaShare project, connected through LONI. Sample research of the participating researchers pictured (i.e. biomechanics by Kodiyalam & Wischusen, tangible interaction by Ullmer, coastal studies by Walker, and molecular biology by Bishop).





Sensor Networking
High Energy Physics
Biomedical Data Mining
LaTech

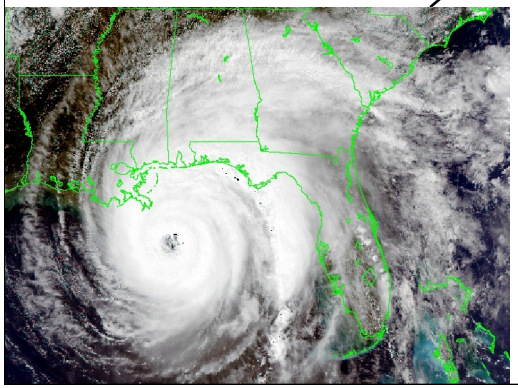


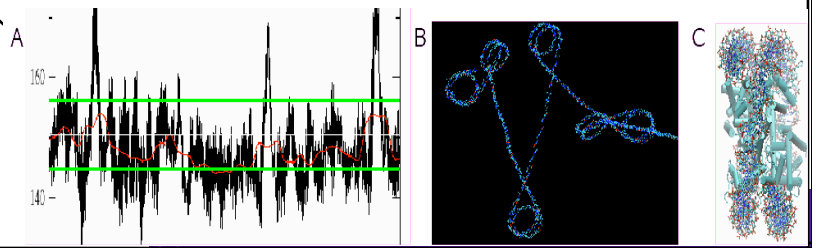
LSU
Coastal Modeling
Petroleum Engineering
Computational Fluid Dynamics
Computational Biology
Numerical Relativity

UNO
Biophysics

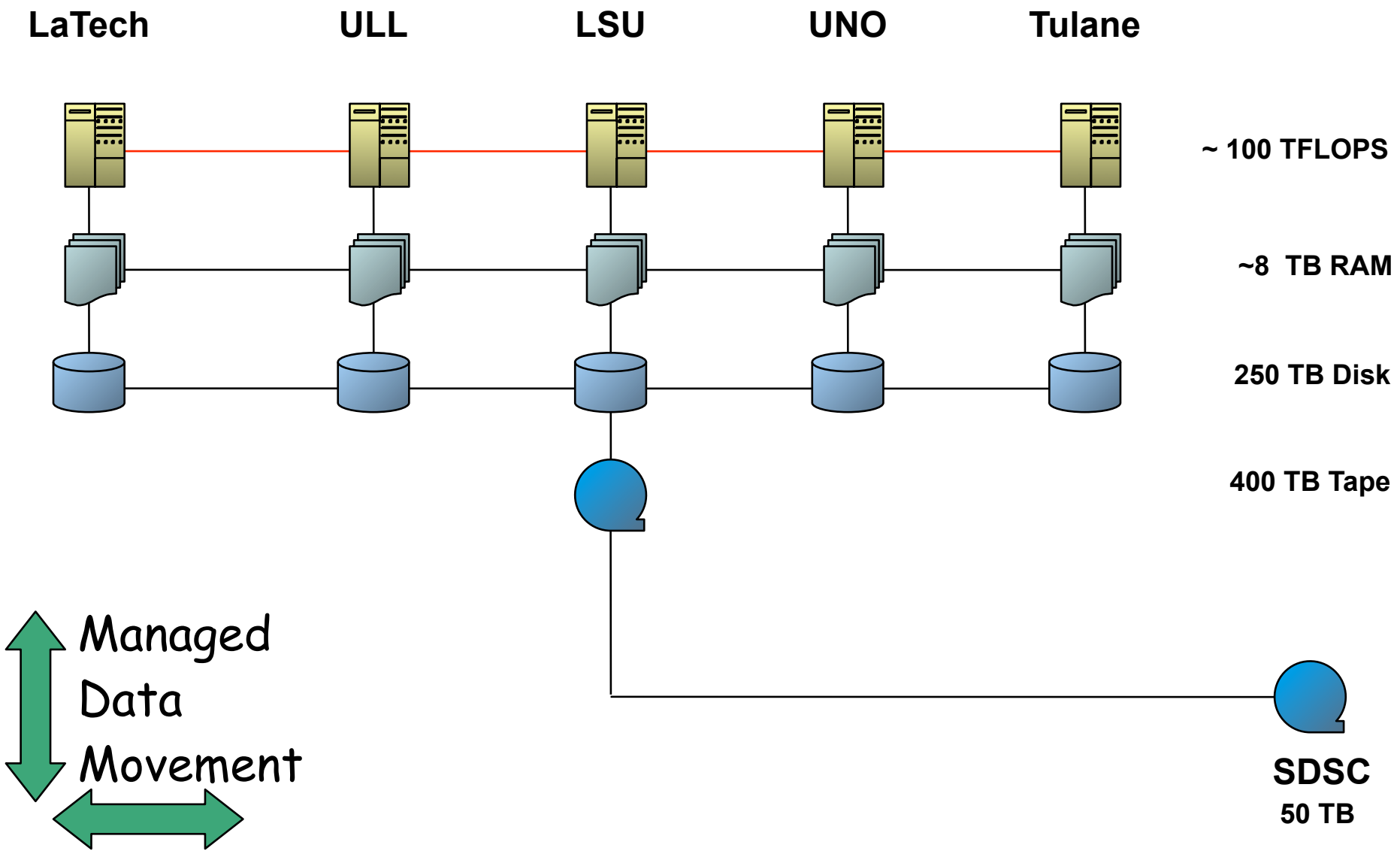
ULL
Geology
Petroleum Engineering
Sensor Networking

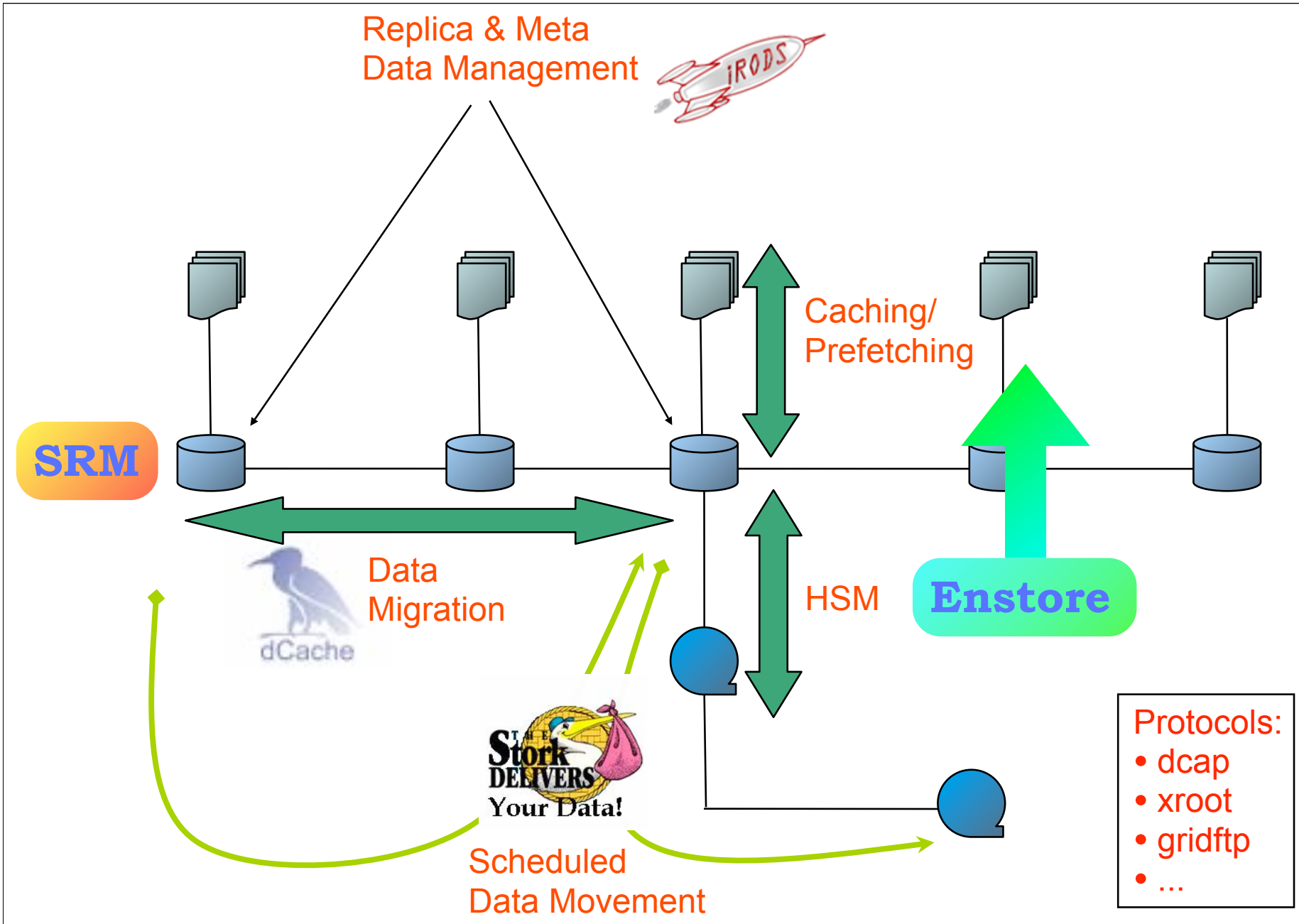
Tulane
Molecular Biology
Computational Electrophysiology



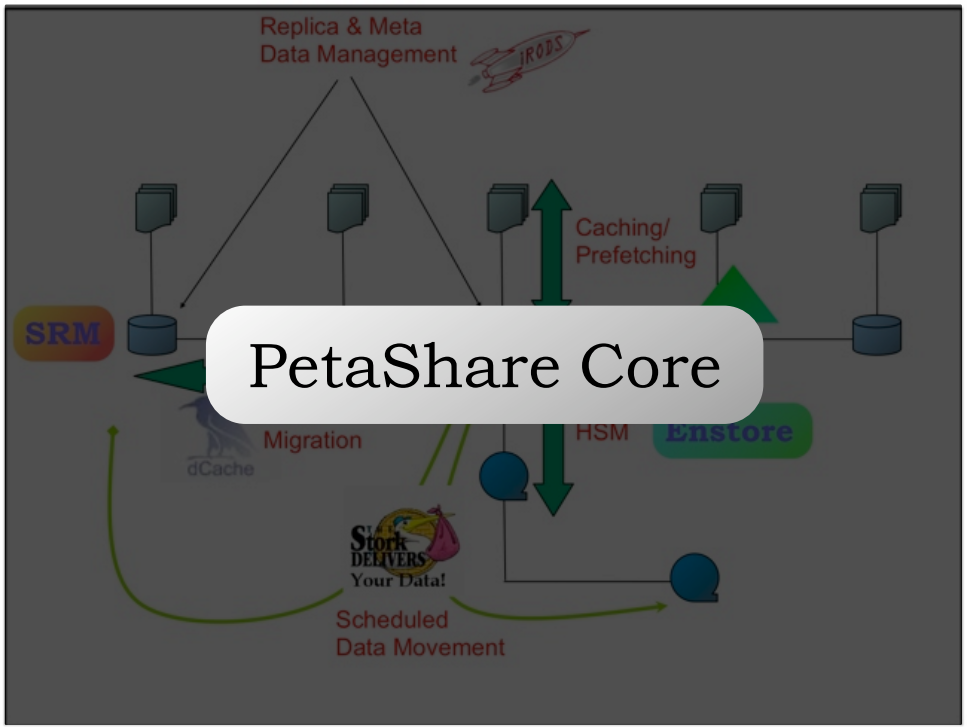


Infrastructure Overview

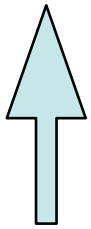



Replica & Meta
Data Management
iRODS
Caching/
Prefetching
SRM
Data
Migration
dCache
HSM
Enstore
The Stork DELIVERS Your Data!
Scheduled
Data Movement
Protocols:
• dcap
• xroot
• gridftp
• ...

POSIX interface
- NO relinking
- NO recompiling
Based on Parrot



`petashell`



Web interface
& others..

Parrot

- Parrot makes a remote storage system appear as a local file system to an application.
- It does not require
  - any special privileges
  - any recompiling
  - or any change to existing programs
- Parrot interrupts I/O systems calls using *ptrace* and forwards them to remote storage
- Parrot is developed by Thain et al at University of Notre Dame

petashell

- a unix shell interface to PetaShare based on Parrot.

```
$ petashell

psh% cp /tmp/foo.txt /petashare/tulane/tmp/foo.txt
psh% vi /petashare/tulane/tmp/foo.txt
psh% cp /tmp/foo2.dat /petashare/anysite/tmp/foo2.dat

psh% genome_analysis genome_data -->
psh% genome_analysis /petashare/uno/genome_data

psh% exit
$
```

Data-Aware Scheduler: Stork

- Traditional schedulers not aware of characteristics and semantics of data placement jobs

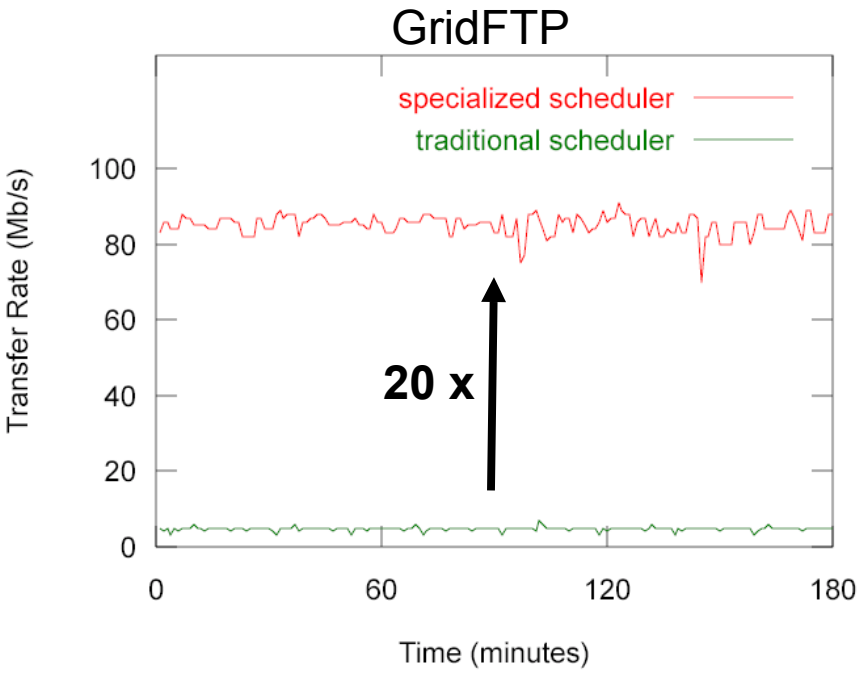
```
Executable = genome.exe
Arguments  = a b c d
```

```
Executable = globus-url-copy
Arguments  = gsiftp://host1/f1
             gsiftp://host2/f2
             -p 4 -tcp-bs 1024
```

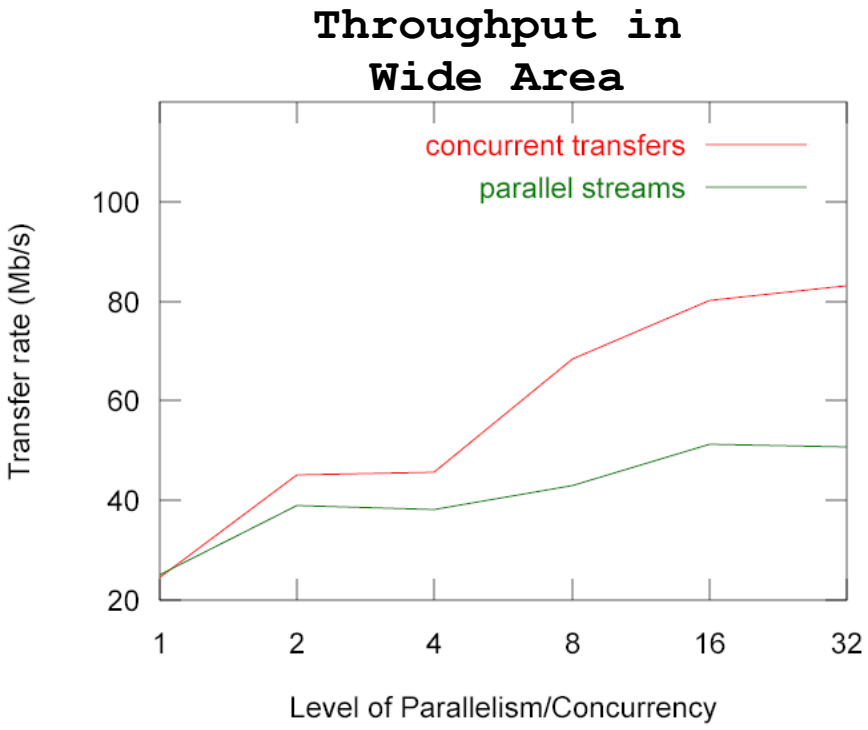
Any difference?

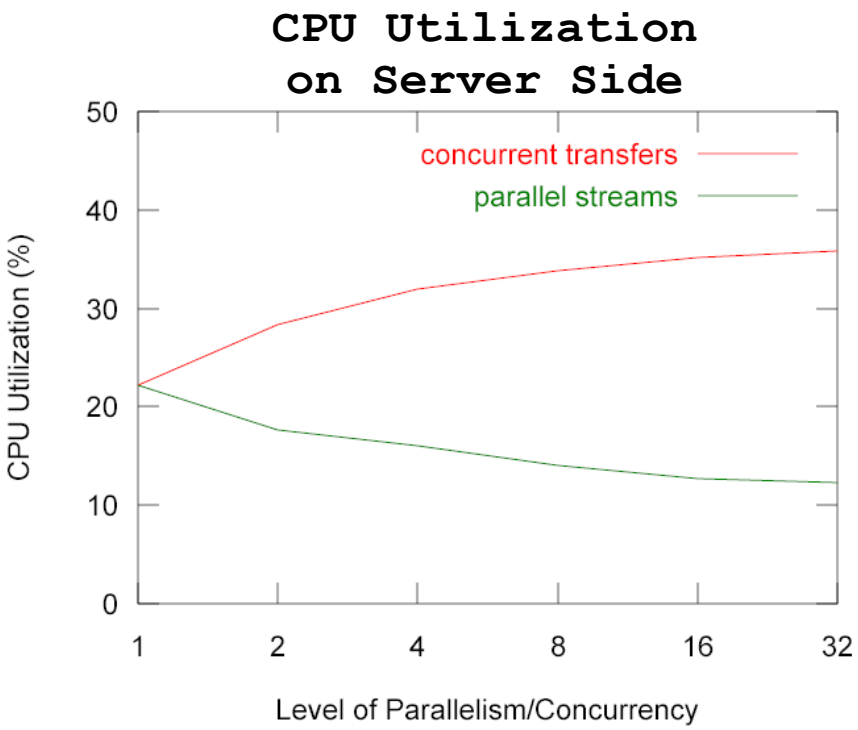
Data-Aware Scheduler: Stork

- What type of a job is it?
  - transfer, allocate, release, locate..
- What are the source and destination?
- Which protocols to use?
- What is available storage space?
- What is best concurrency level?
- What is the best route?
- What are the best network parameters?
  - tcp buffer size
  - I/O block size
  - # of parallel streams



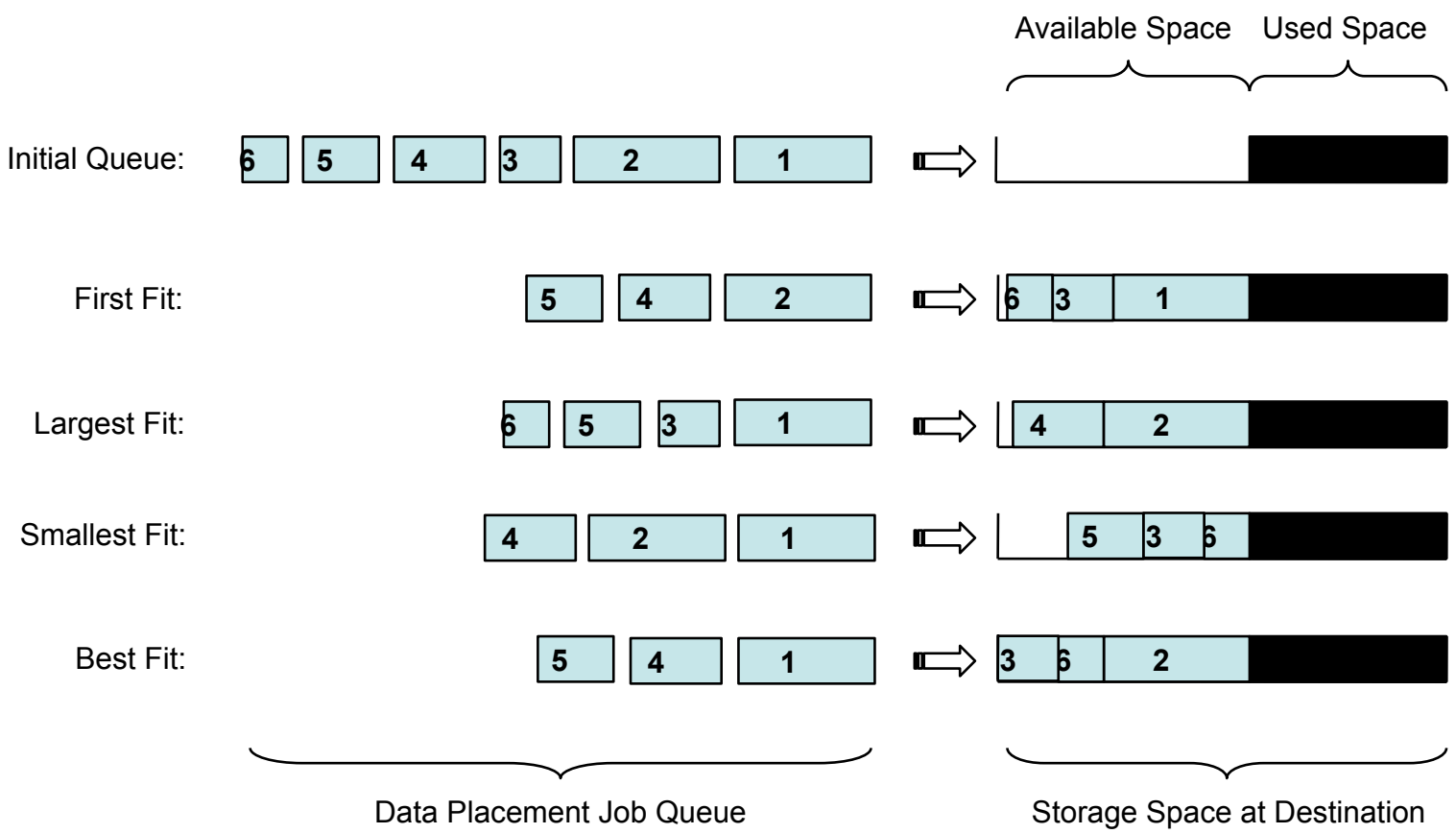
Optimizing Throughput and CPU Utilization at the same Time





- Definitions:
  - Concurrency: transfer n files at the same time
  - Parallelism: transfer 1 file using n parallel streams

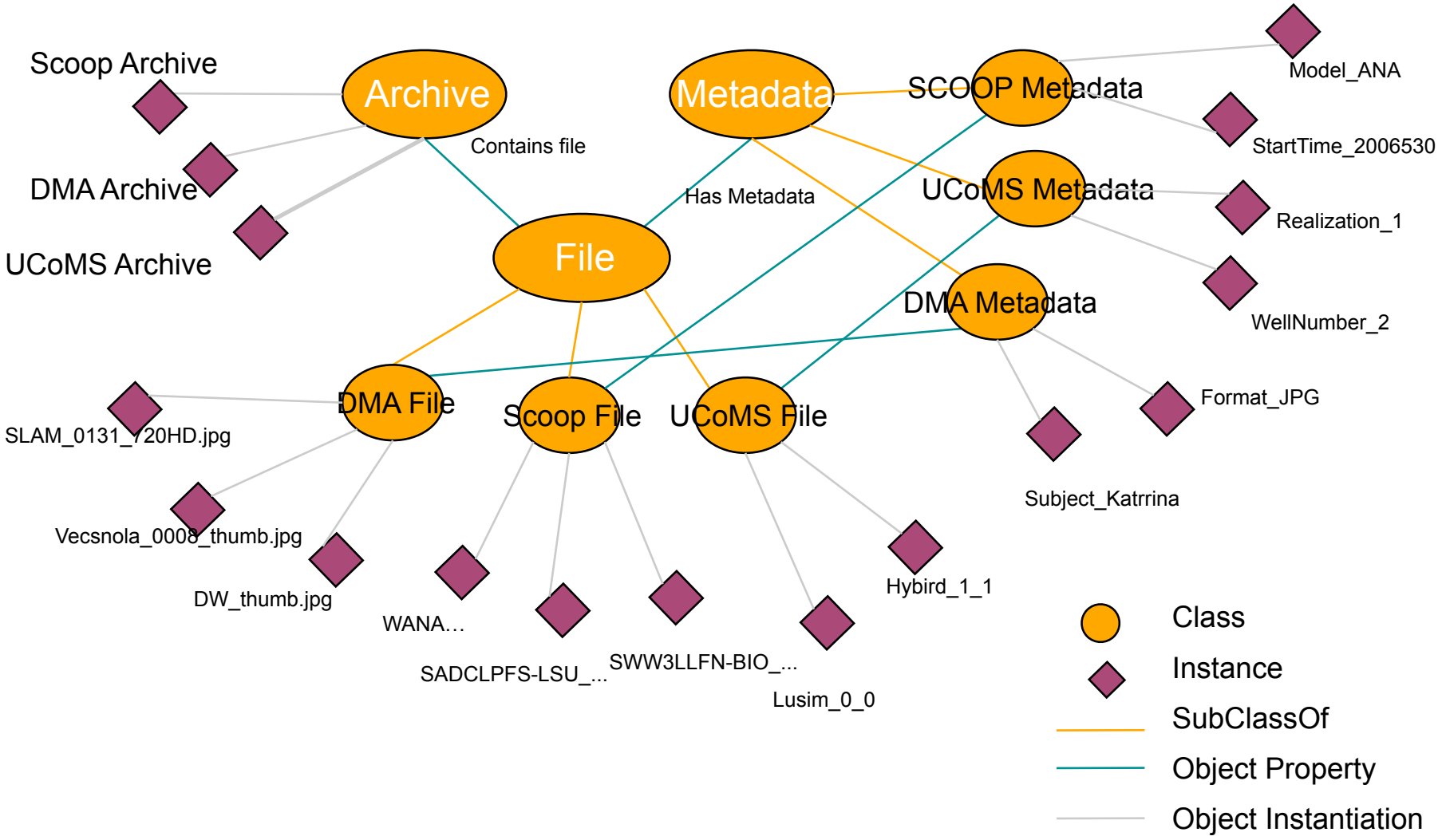
Storage Space Management



Cross-Domain Metadata

- SCOOP – Coastal Modeling
  - Model Type, Model Name, Institution Name, Model Init Time, Model Finish Type, File Type, Misc information …
- UCoMS – Petroleum Engineering
  - Simulator Name, Model Name , Number of realizations, Well number, output, scale, grid resolution …
- DMA – Scientific Visualization
  - Media Type, Media resolution, File Size, Media subject, Media Author, Intellectual property information, Camera Name …
- NumRel - Astrophysics
  - Run Name, Machine name, User Name, Parameter File Name, Thorn List, Thorn parameters …

Ontology definition



Summary

- PetaShare aims to enable data intensive collaborative science across state, by providing
  - Additional storage
  - Cyberinfrastructure to access, retrieve and share data
    - Data-aware storage
    - Data-aware schedulers
    - Cross-domain metadata

A system driven by the local needs (in LA), but has potential to be a generic solution for the broader community!

For more information on **PetaShare**: http://www.petashare.org
For more information on **Stork**: http://storkproject.org

Acknowledgment: This work was supported by NSF grant CNS-0619843.