

High Performance Computing Center

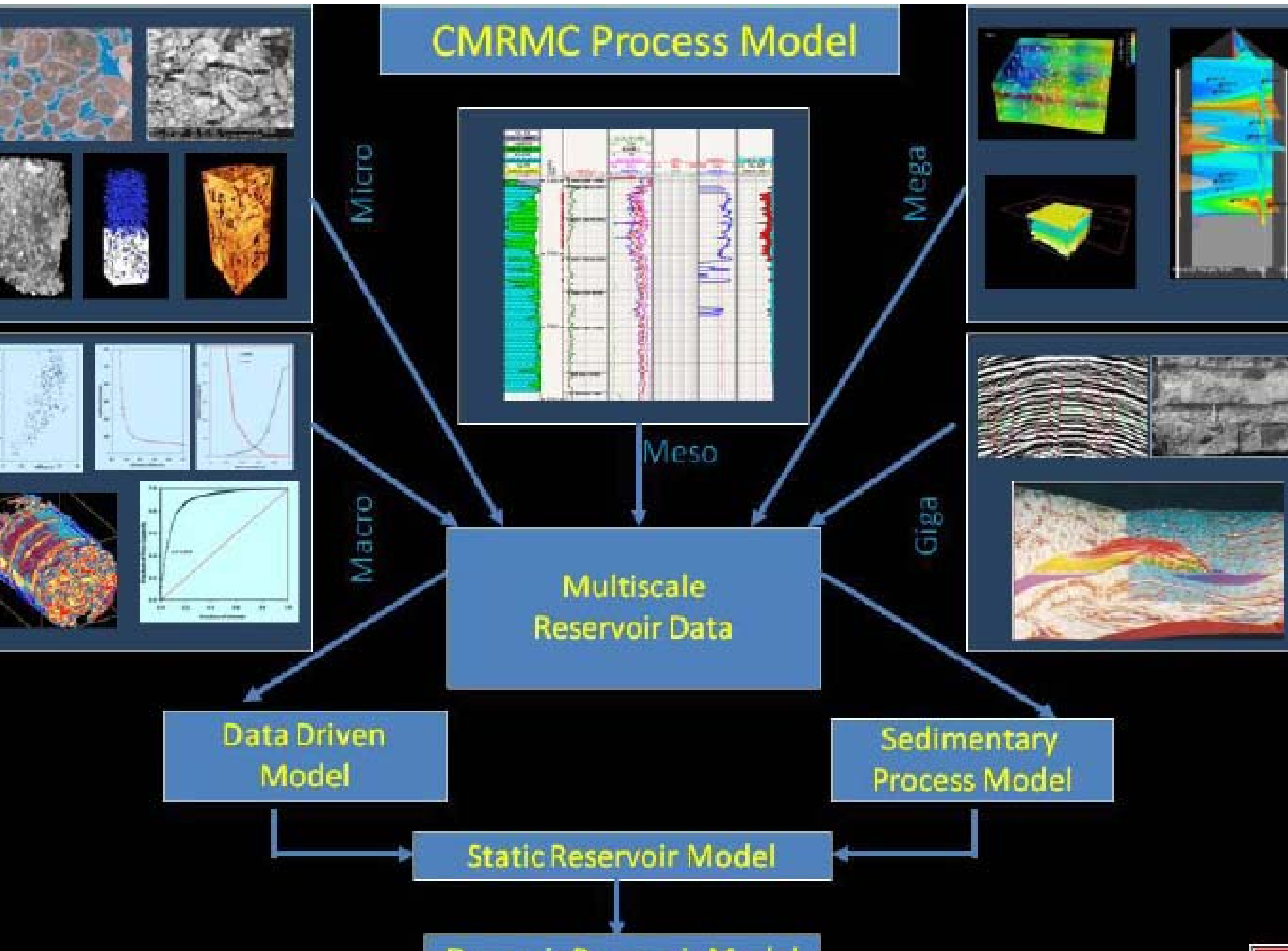


Emerging Technologies and Innovative Solutions for Petroleum Engineering

Ravi Vadapalli, Alan Sill and Shameem Siddiqui

Texas Tech University

Day of High Performance Computing Professionals, Houston, Nov 3,



Hydrocarbon reserves are depleting and effective management of existing reserves is necessary.

Partially untapped domestic reserves, and abundant ways.

Reservoir engineering is highly multidisciplinary and organization can provide all.

Industry is profit based and knowledge is the goal for higher educational systems.

Economy and environment are priorities for the government.

Knowing what tomorrow looks like helps all.

Data security is the “key” for cooperation.

Development of water, wind and geothermal depends on hydrocarbon resources and technologies.

Industry-academia-government partnerships is one solution fits all.

Students translate knowledge into expertise

Transferring training to schools saves industry time and helps opening new opportunities.

Led to the creation of

High performance computing operates in the context of resources provisioned for several related projects in which we are involved.

Local Resources - Grid and Local Users

- HPC clusters, department/researcher-owned resources and TTU campus grid

State-wide and Regional Resources

- TIGRE, SURAGrid, specialized regional clusters, etc.

National/International-Scale Cyberinfrastructure

- Open Science Grid, TeraGrid, and LHC Computing Grid

actively encourage our local scientists to make use of all



The TIGRE Project: an Example of a State-Wide Grid



A HiPCAT Project

TIGRE

Texas Internet Grid for
Research and Education

adapalli, TTU
TIGRE
pment Team



TIGRE project funded by legislation that also enabled

Created a grid infrastructure that enables integration of computing systems, storage systems, databases, visualization labs and displays, even instruments and sensors across Texas.

Facilitate new academic - government - private industry research partnerships by dramatically enhancing both computational capabilities and research infrastructure.

Address research areas of interest to the State of Texas in which substantial increase of computing power, data access, and collaboration are necessary.

Demonstrate new, preferred, enhanced or increased computing and storage handling capabilities offered by a statewide grid infrastructure.

**Project Completed
and Production-Ready**

Based originally on the Virtual Data Toolkit (VDT), working in close cooperation with VDT team members, Globus and the Open Science Grid (OSG). ***Greatly extended and enhanced VDT component services***

Uses a streamlined VDT set including GSI-OpenSSH, omitting much monitoring and accounting in favor of lightweight status reporting.

TIGRE implementation based on Web Services (GRAM4) **only**; pre-
available only upon request (no requests). ***Translation: a modern grid***

Client and server software stacks separately available.

Goal was “one page” installation instructions that can be implemented quickly by newcomers. ***(Achieved!)***

Authentication via X.509 (new TACC CA is now accredited by IGTF); authorization local, mostly via grid-mapfiles. ***(TTU hosts PEGrid CA)***

Installed on systems at all five primary TIGRE institutions; also running at other locations throughout the state. ***15-minute install process!***

Grid-enabled ensemble Kalman Filters approach for reservoir characterization.

Demonstrated the history matching application using Schlumberger's ECLIPSE® across the TIGRE sites.

Job scheduling and management was accomplished using GridWay metascheduler, and WS-GRAM4 based TIGRE stack.

Part of the work was used for PhD dissertation.

The grid deployment results were published in the 15th ACM Mardi Gras: Workshop on Grid-enabling Applications, Baton Rouge, LA.

Operate PEGrid with full data security while providing controlled access to researchers

Data controls, strong authentication where needed

Execution controls for submission and retrieval of jobs

Full access for software manager

Provide large-scale scalable infrastructure

Full conformance to appropriate grid standards

WS-Trust, WS-Security, Secure Conversation

K.509 infrastructure for large-scale submission of jobs

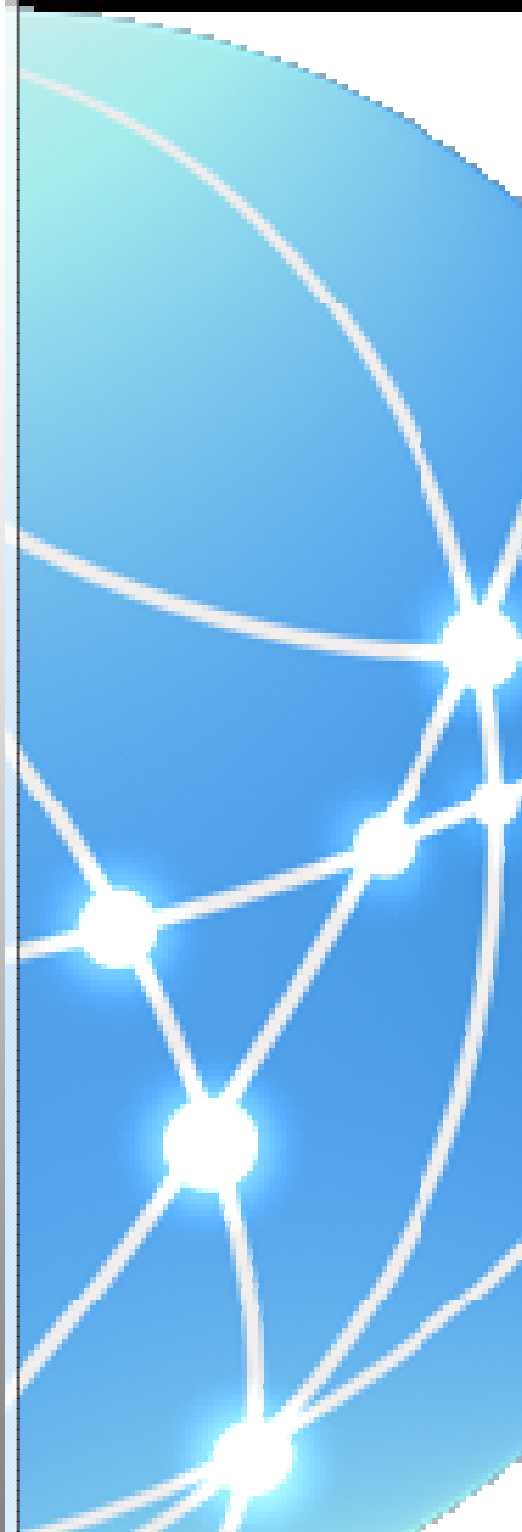
Web services with signed applications

Interface with .NET and web services portals

(EnginFrame etc.)

Create environment that is optimized for both commercial

and educational usage



creating the environment for full application access, we
controls in place for:

no has access to the applications

no can control and move job execution

specific controls on who can use and access data

weve this on a routine basis with existing grid controls by using:

personal credentials for all grid access

and user controls with individual account mapping

authentication and authorization for all job submission and data mover

on, we can impose any or all of the following:

Trust Secure Conversation, Transport Layer Security, Secure Messa

ate account mapping for all PEGrid users



es needed:

complete end-to-end security

“just in time” delivery as well as pre-staged data

specific controls on who can use and access data

security **from the outset!**

Integrate X.509 credentials into portal access system

Same host and user controls as used in job submission

Separate account mapping for all PEGrid users

Need very high performance data movement

Systems used in collaborating grids already handle petabytes

Must work across all centers and systems

Multi-stream GridFTP with SRM (Storage Resource Manager)

is the bill



FTP

Web Portal



X.509 AuthN/AuthZ



Read Config

Collect I/P conditions

Select Grid Resources

- Gather privileges

Select Model Space

- Checkout licenses
- Load environment

GridWay

Submit



Post-process

Master head and worker node based grid software:

Primarily obtained by standard X.509-secured grid methods - some of which we helped develop, produce and refine.

MS:

Controls mapping to accounts by X.509 certificates.

MS:

Support TIGRE, THEGrid, SURA (experimental) and PEGrid.

Proxy and squid:

Secured grid cache for high-performance database access.

Storage (SRM and GridFTP):

Collaboratively developed with Berkeley Storage Manager group

Center for Multiscale Reservoir Modeling and Characterization TO:

create research and educational opportunities through
industry-academia partnerships

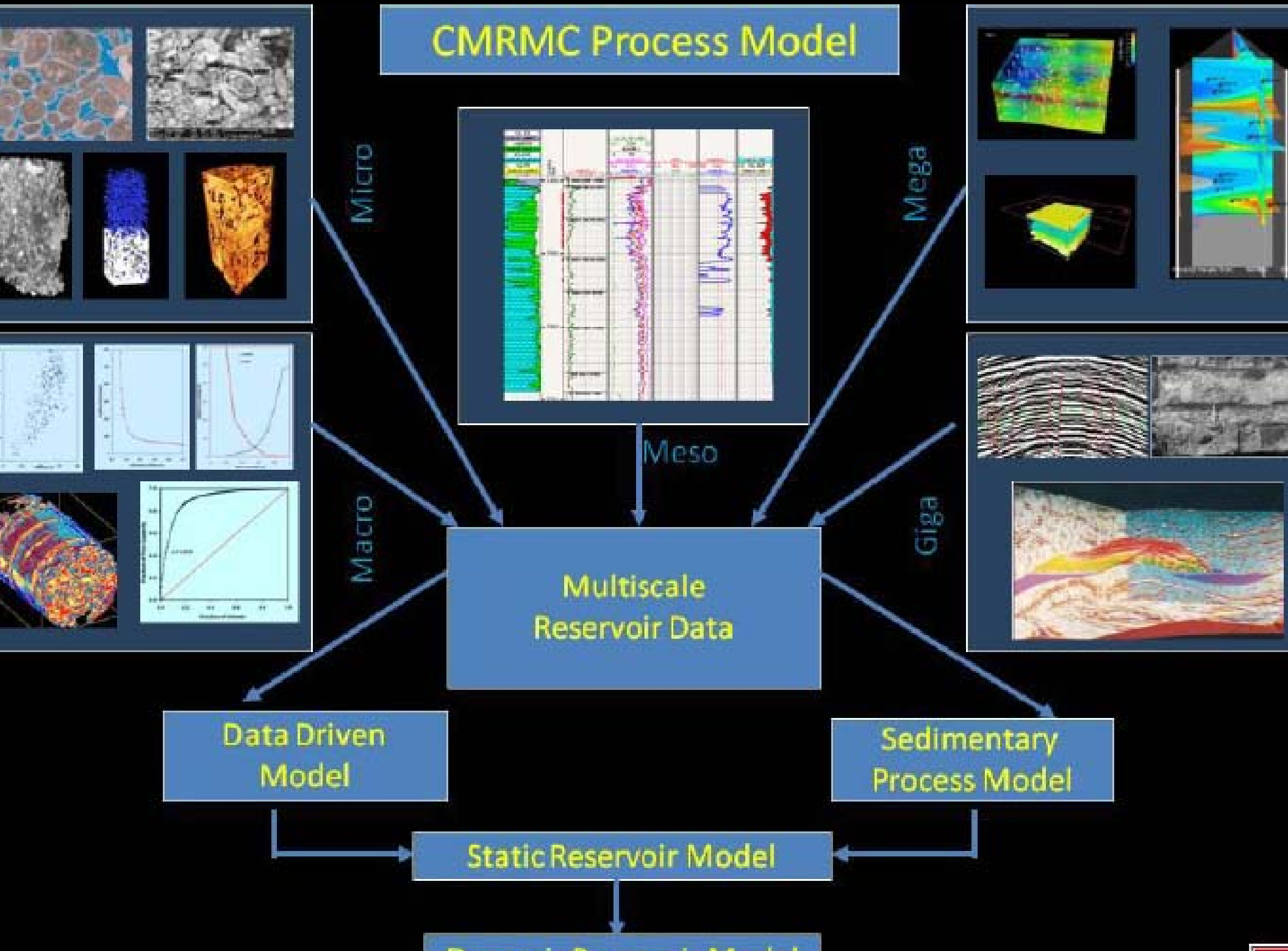
Create new partnerships with small scale oil/gas
operators

regional and small scale oil/gas operators

Cannot afford to keep up with advances in HPC

And, may not be able to afford consultancy costs

Willing to share datasets for academic purposes



GROUP
PO®

\$2.3 Million optimization software grant, and internships

{ Platform Computing

\$34,560 software grant
Collaborative project metaschedulers

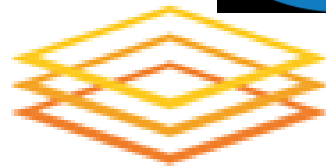
ICE

software grant
ships



Center for Multiscale Reservoir Modeling and Characterization

\$939,541



Open Science Grid



Schlumberger

\$42.5 Million software grant
training cost avoidance, in
and collaborations

Industries



onMobil

the world's toughest energy challenges.™

Academic/Research Collaborators



Amounts are for a 3 y

stands ready to bring high-performance methods into
proven applicability in the following areas:

Grid integration with a proven software stack

Built-in high transfer rate data access

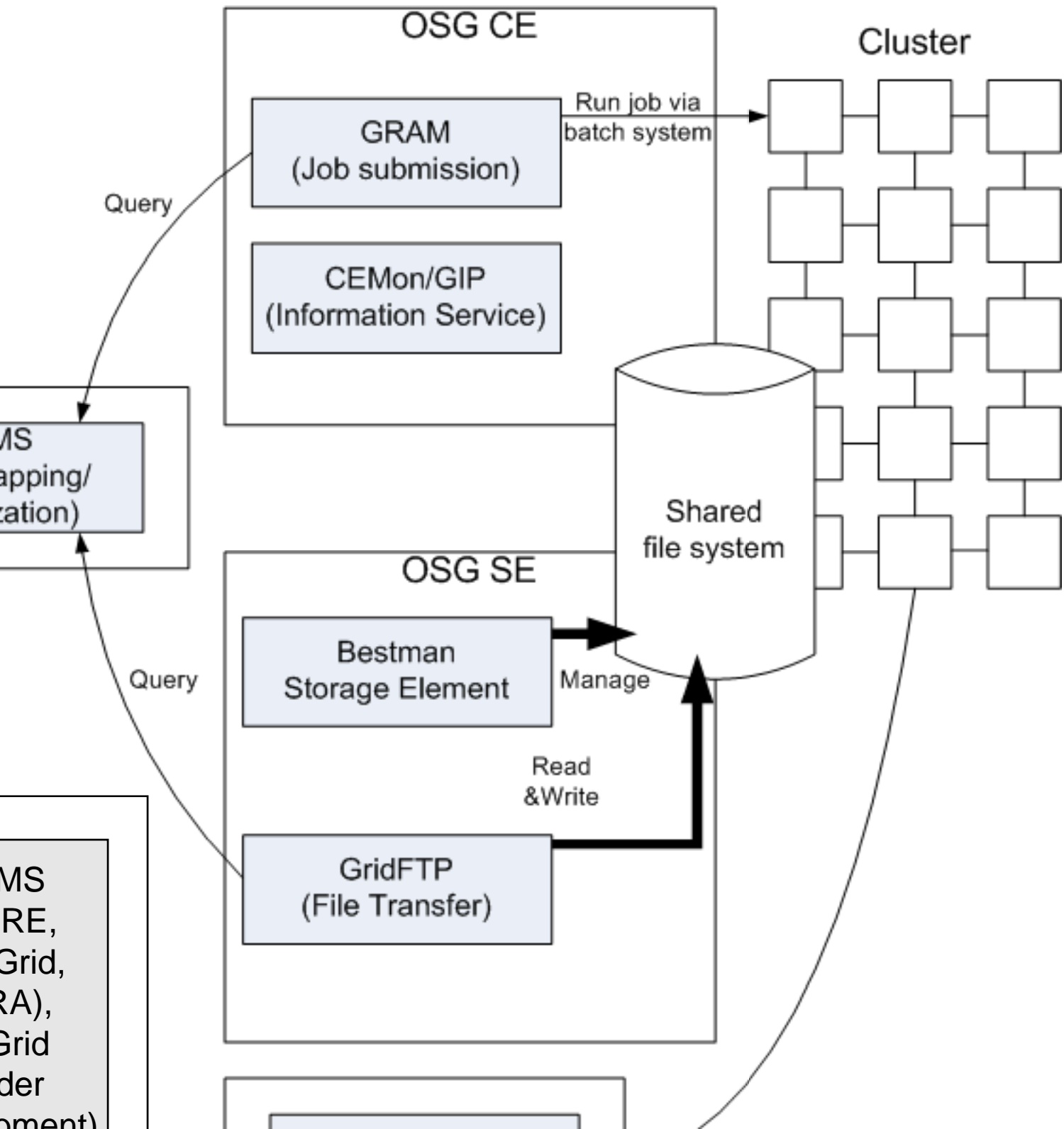
Integrated X.509 security that supports proprietary stan
data security

Web-services based job submission

Novel industry-academia-government partnerships thro

Comprehensive research and training, creating a skilled
engineering workforce.

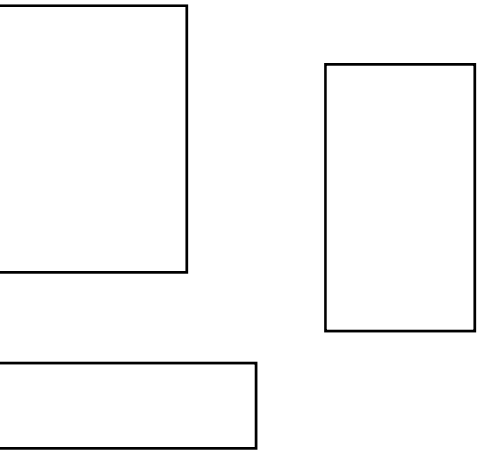




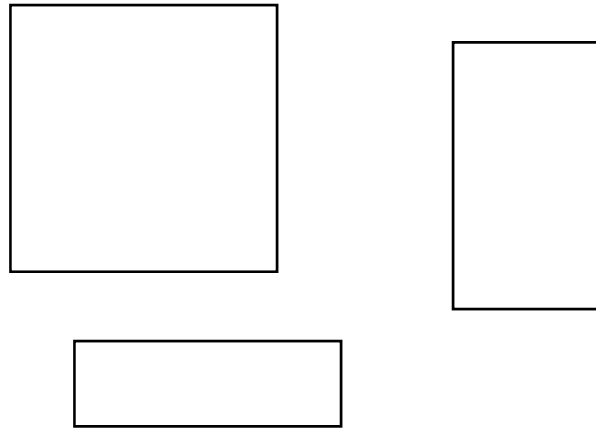
Clusters are configured with interfaces for job scheduling (Control Elements) and storage (Storage Elements), with additional services required to support define, control and schedule usage of virtual organizations (VOs).

TTU HPCC

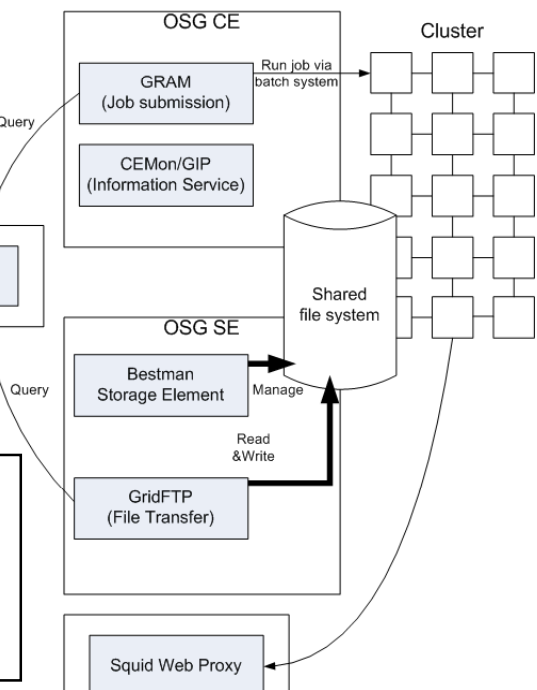
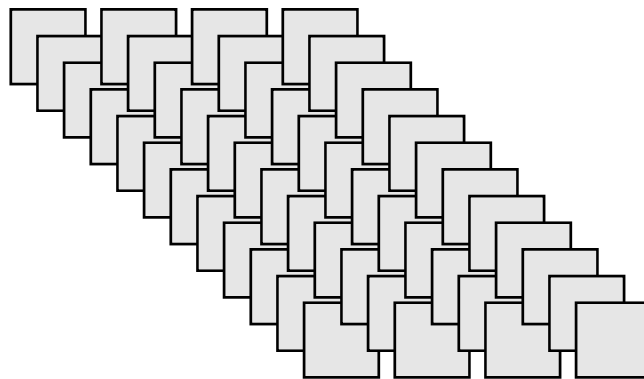
Other HPCC clusters



Other HPCC clusters

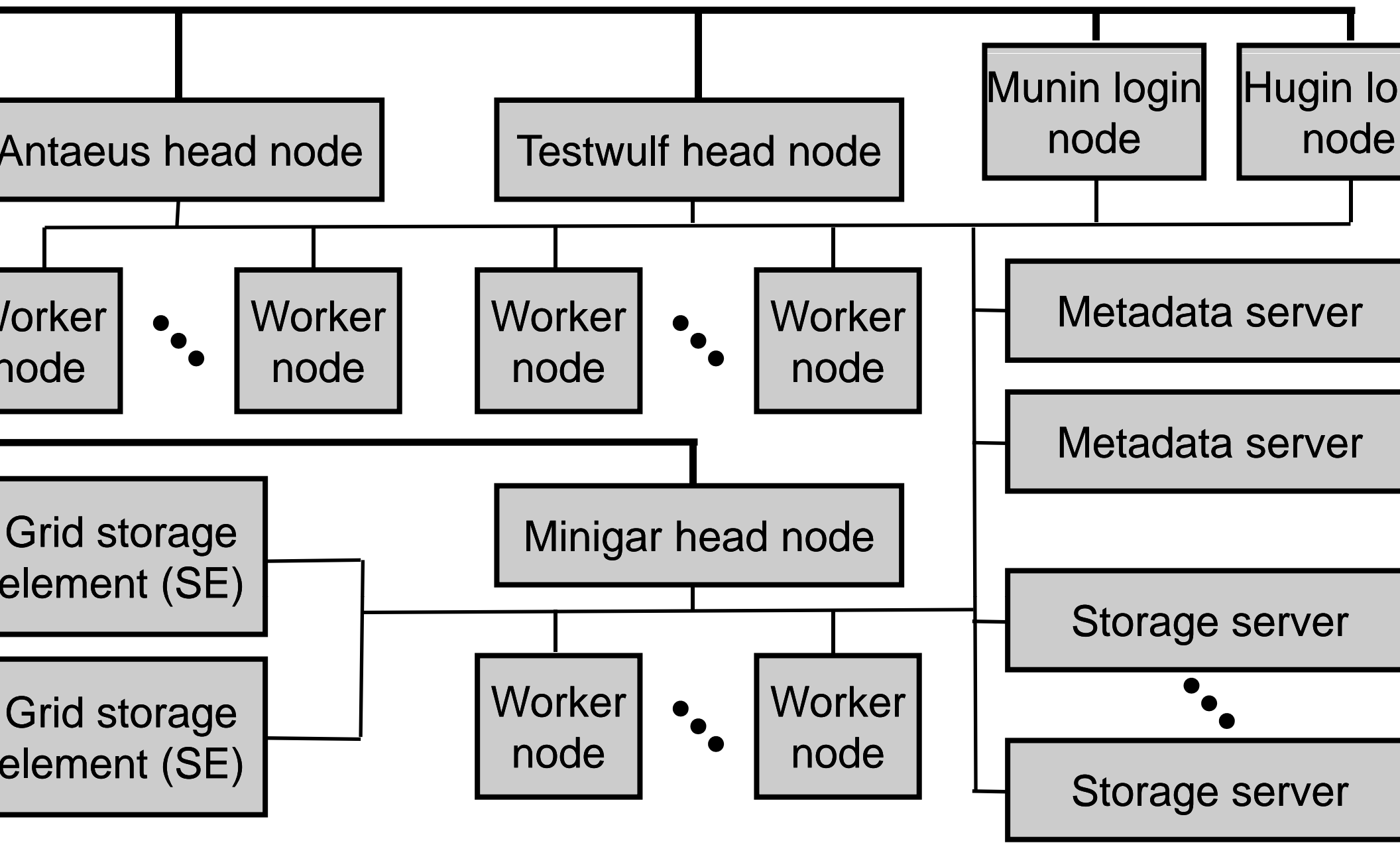


Campus Grid



Our resources supported in the context of the High Performance Computing Center PEGrid services are integrated clusters used in this and other related research and education

Public network



Private data network access through grid storage elements

lowest high
performance
computing resource.

Nodes x8=1680
of 3.0 GHz
Xeon with full
function band-
width Infiniband
interconnect.

teraflops

out of 500 on
2008 Top 500



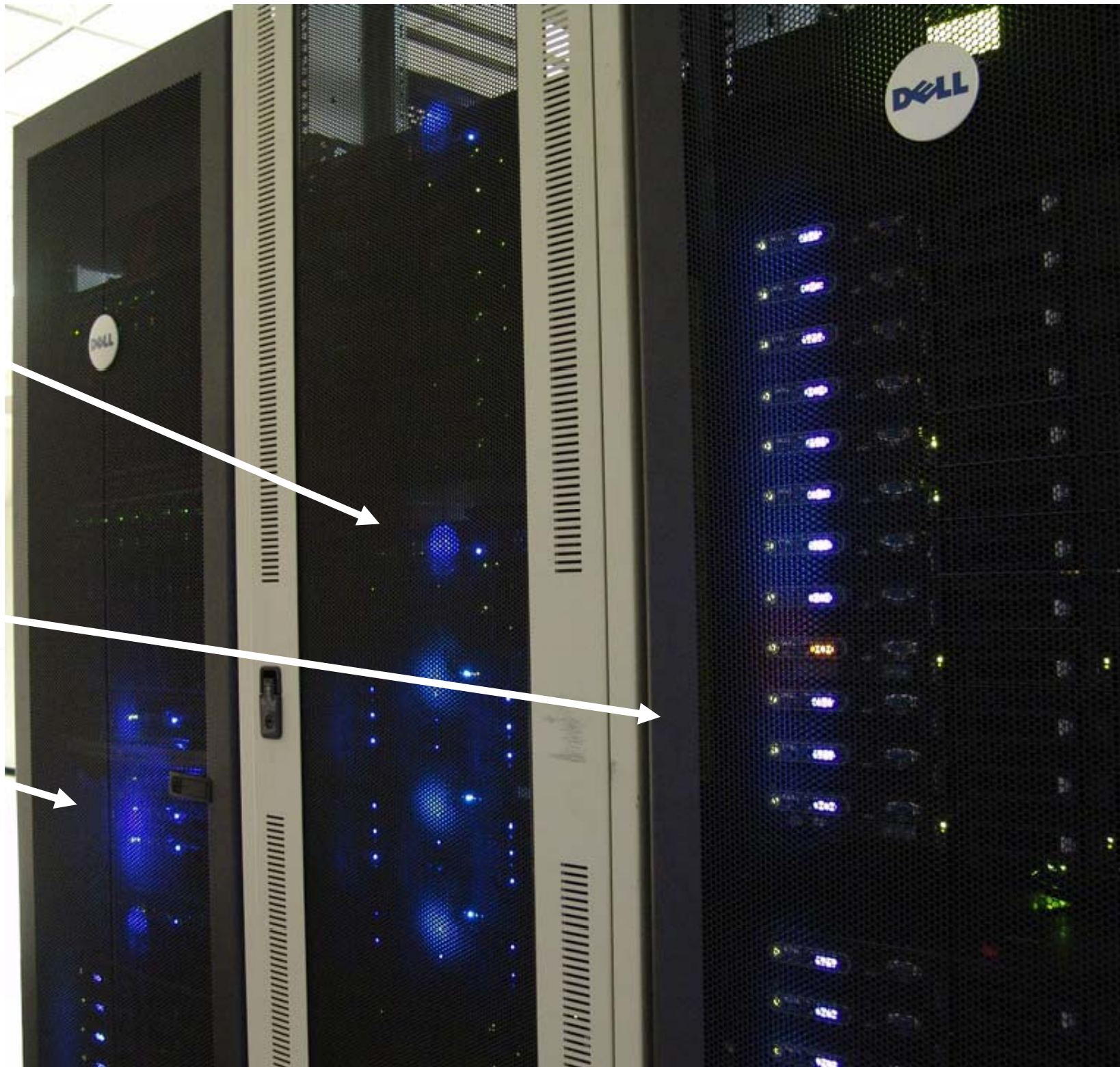
acks, mix of dual-dual and dual-quad-core Xeon cp
8 cpu cores total, separate Lustre file system used



and service nodes

marker nodes

and grid services





Uses Lustre

- Popular choice for storage at HPC centers
- Extremely high-speed delivery of data to worker nodes
- Multiple 2 TB servers with GbE fabric

PhEDEx - CMS Data Transfers

[Info](#) | [Activity](#) | [Data](#) | [Requests](#) | [Components](#) | [Reports](#)

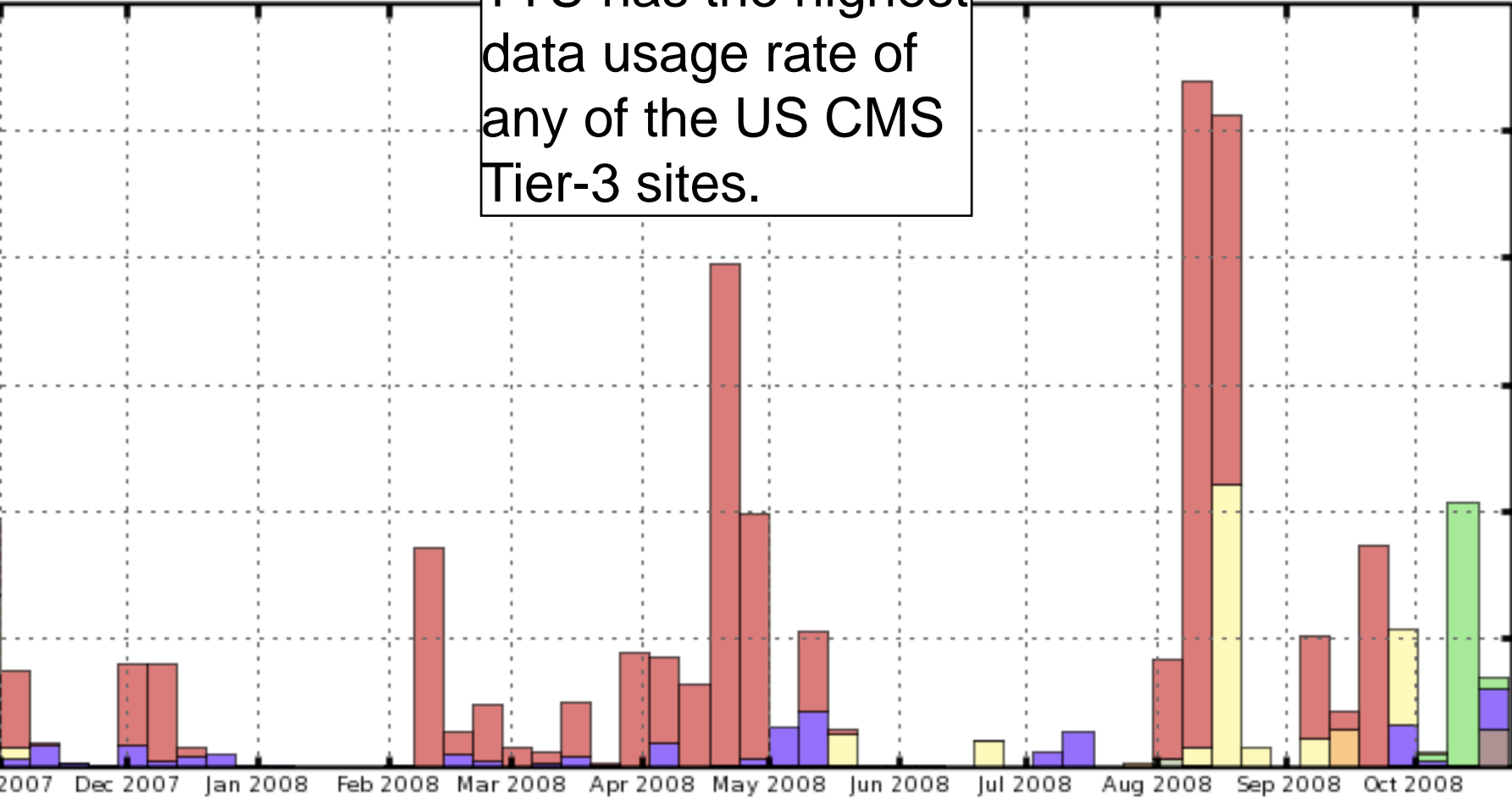
[Rate](#) | [Rate Plots](#) | [Queue Plots](#) | [Quality Plots](#) | [Routing](#) | [Transfer Details](#) | [Deletions](#) | [Recent Errors](#)

Rate by filter source destination

Weeks up to

CMS PhEDEx - Transfer Rate

52 Weeks Week 42 of 2008



TTU has the highest data usage rate of any of the US CMS Tier-3 sites.

Based
usage
of frequ
saturat
campu
Interne
the un
has ag
pursue
upgrad

stands ready to bring high-performance methods into play with *en* applicability in the following areas:

Grid integration with a proven software stack

Built-in high transfer rate data access

Integrated X.509 security

Web-services based job submission

stand ready to work with you on the following topics:

EnginFrame and web portal integration

.NET and WS-secured access via Petrel SDK

Other topics as needed by the community