TIGRE Environment for Extending R & E Capabilities in Reservoir Simulations

Ravi Vadapalli
High Performance Computing Center
Texas Tech University, Lubbock

Outline

- Philosophy of the Ensemble Kalman Filter
- TIGRE Computing Environment
- Application Scope & Deployment Efforts
- Summary & Conclusions
- Extensions & Collaborations

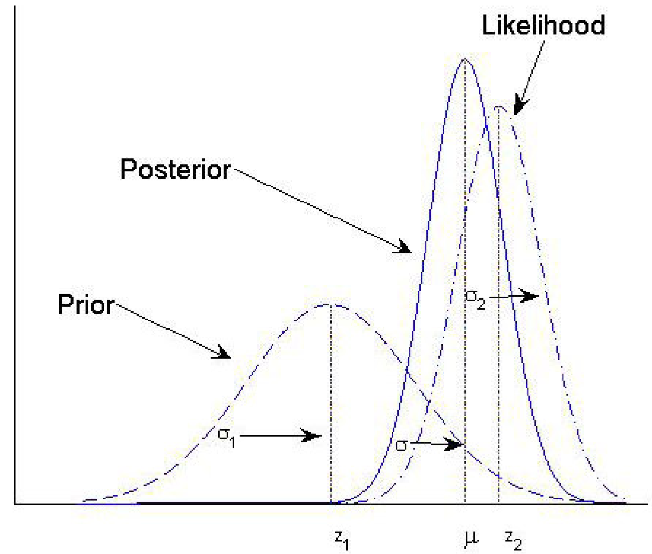
Ensemble Kalman Filter Approach

- A byproduct of Kalman filtering algorithm (Rudalph Kalman 1960).
- Introduced to handle non-linear ocean models
- Demonstrated for the solution of linear least squares in signal processing and control theory.
- Uses a set of equations to estimate the state of some process, with the possibility of assimilating new information as it arrives.
- Advantages
  - Don't have to start all over again
  - Incorporates all the information in a best possible way.

Kalman Filtering Algorithm Example

- You and your friend lost at night in a 1-D sea.
- Goal is to calculate your position with the help of star sighting.
- Assume: your friend has better knowledge of star sighting than you.

Use the observations in best possible way to reduce the position uncertainty.



$$\mu = \frac{\sigma_2^{\,2}}{(\sigma_1^{\,2} + \sigma_2^{\,2})} z_1 + \frac{\sigma_1^{\,2}}{(\sigma_1^{\,2} + \sigma_2^{\,2})} z_2$$

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^{\,2}} + \frac{1}{\sigma_2^{\,2}}$$

The variance of the “update step” decreases

Vazquez and Syversveen (2006)

Reduce the Uncertainty

- Assume that your friend is alone in the boat and the motion is governed by

$$\frac{dx}{dt} = u + w$$

- Discretize using the Euler time stepping:

$$x^t = x_0 + \Delta t u + \Delta t w_0$$

- t: model predicted "true state" of the process. x_0: position today and w_0: corresponding uncertainty.

$$x^f = z_0 + \Delta t u$$

$$(\sigma^f)^2 = \sigma_0{}^2 + \Delta t \sigma_w{}^2$$

- Project the equation "forward in time" (forecasting step) and replace your bad observation with it.

$$x^a = x^f + K_1[z_1 - x^f]$$

$$(\sigma^a)^2 = (1 - K_1)(\sigma^f)^2$$

- Let your friend take another look at stars tomorrow.

$$K_1 = \frac{(\sigma^f)^2}{(\sigma^f)^2 + \sigma^2}$$

f: existing info
a: new info
K: Kalman gain

The process is restarted with the moments x^a and σ^a

Kalman's Algorithm for Reservoir Simulation

- Data: Today's values of production variables (permeability, porosity) are available.
- Goal: To simulate gas/oil ratio in a well.
- Step 1: Forecast tomorrow's state of the reservoir w/ today's data (a-priori estimate) Simulation Step
- Step 2: Take new measurement of production data tomorrow and "update" the a-priori estimate.

History Matching

- The simulators are highly non-linear.
- Approximate the "true state" by the mean of an ensemble of model sates. (Evansen 1994)

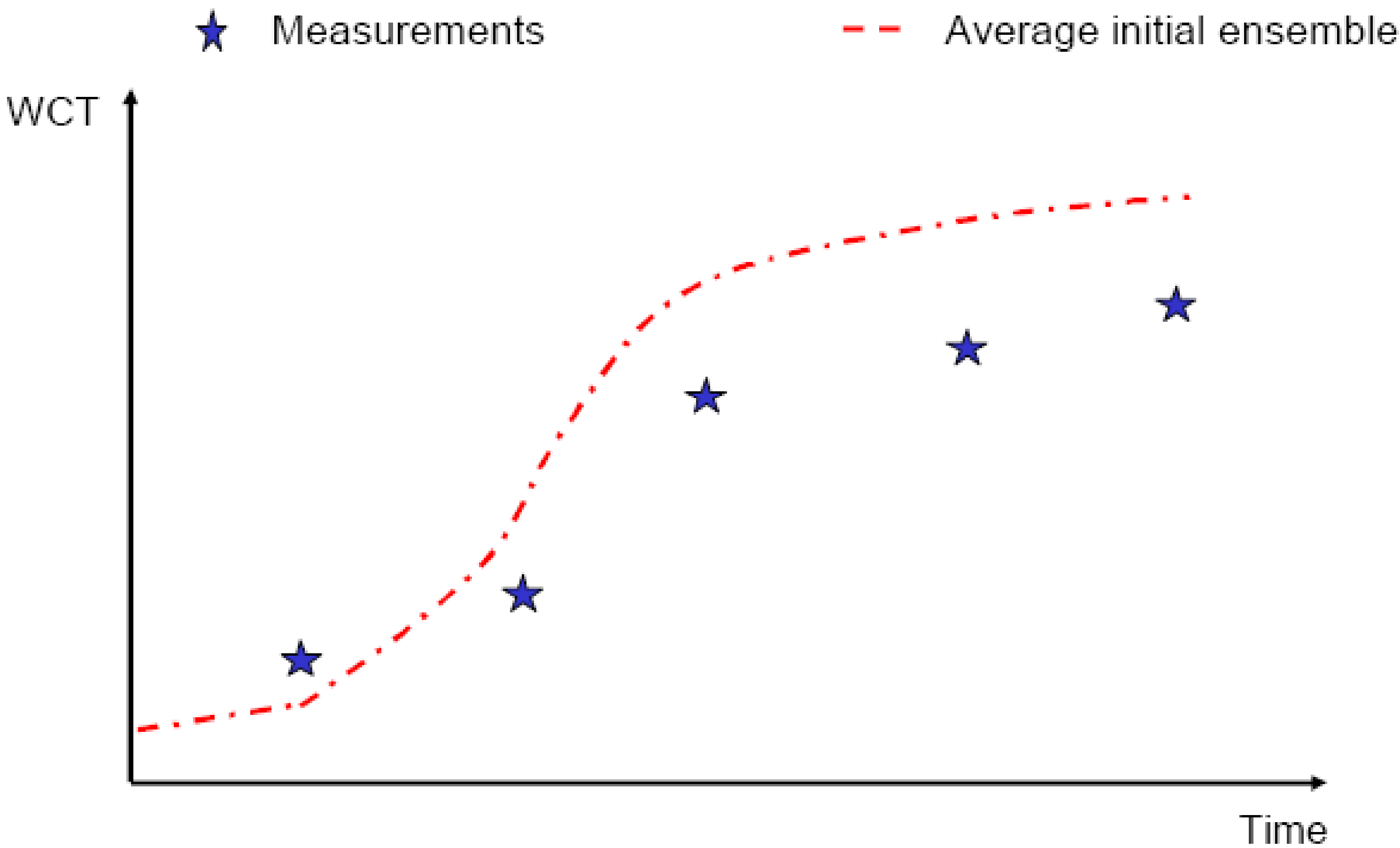
EnKF Algorithm

- **Initialize**: Define an initial ensemble X
- **Forecast**: Project forward in time the ensemble X using the equation $X^f = f(X) + W$, W is the noise matrix.
- **Analysis**: Generate observations Z by perturbing the real observations z. Update the ensemble X^f using the equation

$$X^a = X^f + P_e H'(HP_e H' + R_e)^{-1}(Z - HX^f)$$

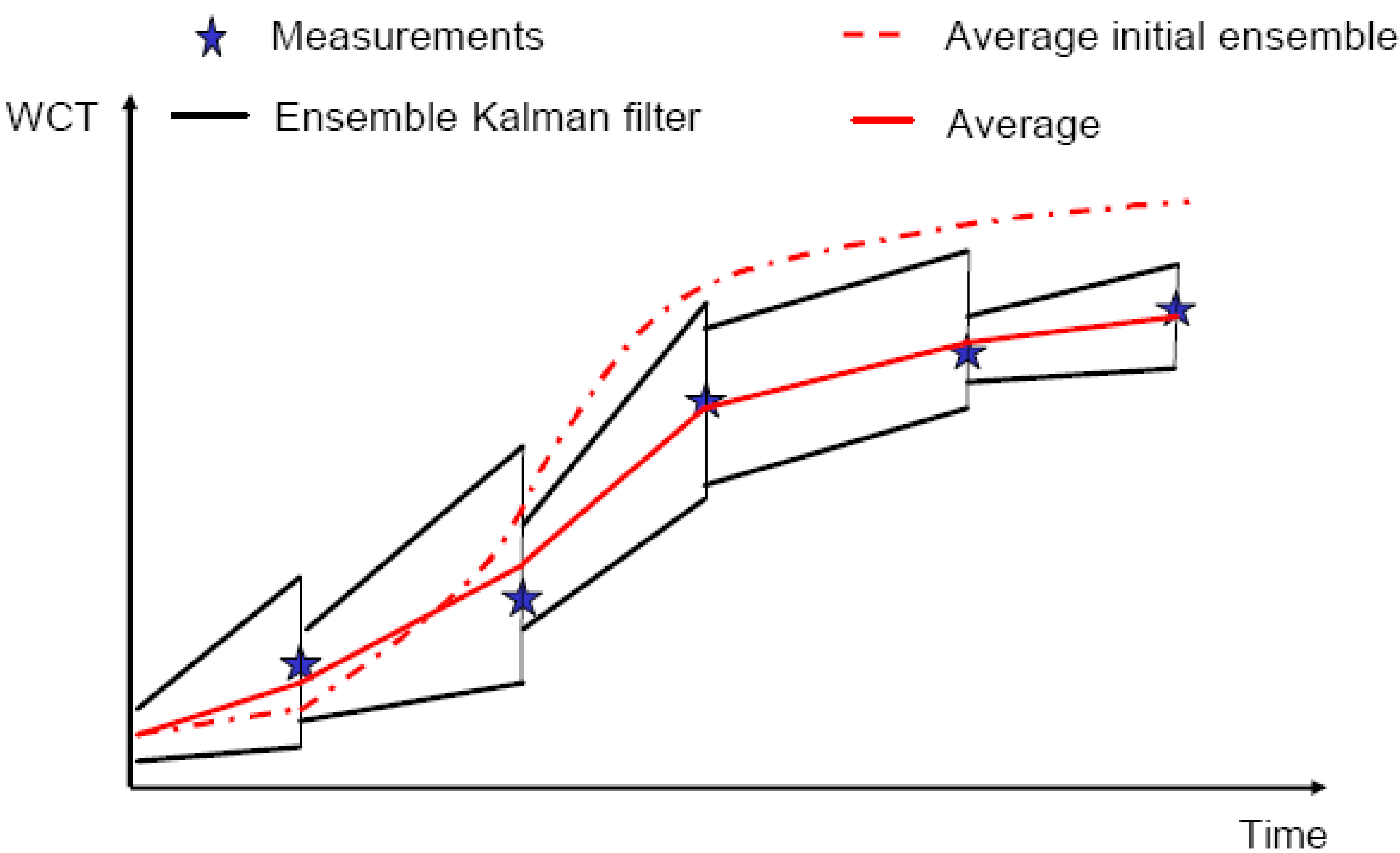
- Go to **Forecast** Step
- The costly part of the algorithm would be the Matrix inversion.

Ensemble Kalman Filter Idea



B. Vallés, et al. Rogaland Research

Ensemble Kalman Filter Idea



B. Vallés, et al. Rogaland Research

EnKF Data Assimilation Summary

- Based on Least Squares for continuous updating the reservoir model.
- Involves updating static and dynamic quantities such as pressure, saturations, permeability, porosity, etc.
- One flow simulation per ensemble member (suitable for parallel computing).
- Production data is assimilated "sequentially"
- Reservoir model is updated in real time.
- Uncertainty prediction is always up-to-date and directly computed from the ensemble members.
- Ensemble members are updated sequentially in time and reflect latest assimilated data.

Who is TIGRE?

- A project of the **Hi**gh **P**erformance **C**omputing **A**cross **T**exas (HiPCAT) consortium
  - HiPCAT started in 1998 with 5 universities in Texas, now 11 universities
- TIGRE has five partner universities

HiPCAT
High Performance Computing Across Texas

Houston

Rice

Texas A&M

Texas Tech

Univ. of Texas at Austin

What is TIGRE?

- Part of the State of Texas' plan to support Cyberinfrastructure
  - Funded by Texas Enterprise Fund [state fund to attract jobs]
  - Network - LEARN: $7.3M / 33 institutions
    - **L**onestar **E**ducation **A**nd **R**esearch **N**etwork - all optical
  - Software - TIGRE: $2.5M / 5 schools
    - People to build, validate and assess Grid software
- **Objectives**
  - Prototype a state-wide Cyberinfrastructure for
    - Supporting existing research and education
    - Attracting new industry
- **Goals**
  - Implement, deploy and validate grid software
  - Use three applications "of economic interest to the State of Texas" to demonstrate Grid capabilities
  - Document best practices, train potential users

Where is TIGRE?

TIGRE Milestones

- Year 1 (start 12/1/2005)
  - ✓ Q1
    - ✓ Certificate Authority, minimal testbed
  - ✓ Q2
    - ✓ F
    - ✓ h
  - ✓ Q3
    - ✓ Define server software stack
    - ✓ Simple application demonstration
  - ✓ Q4
    - ✓ Client software stack

  Annual Report

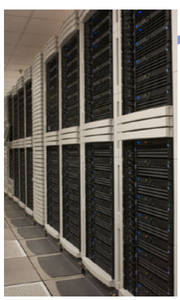
- Year 2
  - ✓ Q1
    - ✓ Customer management service (alpha)
  - ✓ Q2
    - duler
    - ✓ Stable software stack available
  - ✓ Q4
    - ✓ Hardening, documenting
    - ✓ Usage/participation policies
    - ✓ Demonstrate at SC07

  Annual Report

Deliberations on TIGRE Production Environment…
(in Progress)

TIGRE Portal

http://tigreportal.hipcat.net/gridsphere/gridsphere?cid=resources



Grid Information Browser

Refresh

Parallel Computing Resources

Name	Institution	Department	System	CPUs	Peak GFlops	Memory GBytes	Disk GBytes	Status	Load	Jobs
Ada	Rice University	Computer and Information Technology Institute	Cray XD1	632	2939	1320	33736	↑		0R-154Q-0O
Alamo	University of Texas Health Science Center at San Antonio	Department of Biochemistry	Rocks i386 Linux Cluster	19	0	8.7	1126	↑		20R-76Q-0O
Cosmos	Texas A&M University	Texas A&M Supercomputing Facility	SGI Altix	128	666	256	4096	↑		29R-15Q-0O
Eldorado	University of Houston	Advanced Computing Research Laboratory	Eldorado Itanium2 Cluster	126	67	4	2232	↑		0R-0Q-1O
Jacinto	University of Texas Health Science Center at San Antonio	Department of Biochemistry	Microway Linux Cluster	44	66	85.1	8433.3	↑		0R-0Q-1O
Laredo	University of Texas Health Science Center at San Antonio	Department of Biochemistry	Dual Athlon Cluster	32	0	31.6	9509	↑		0R-0Q-0O
Lonestar	The University of Texas at Austin	Texas Advanced Computing Center	Dell PowerEdge Linux Cluster	5200	55000	10400	94900	↑		96R-1Q-392O
Minigar	Texas Tech University	High Performance Computing Center	Dell Linux Cluster	32	230	64	70	↑		8R-0Q-88O
RTC	Rice University	Computer and Information Technology Institute	HP Itanium II Linux Cluster	290	1044	596	7000	↑		0R-16Q-0O
TTU-Antaeus	Texas Tech University	High Performance Computing Center	Dell Xeon Cluster	192	2300	96	6000	↑		10R-0Q-17O
			Total:	**6695**	**62312**	**12861.4**	**167102.3**			

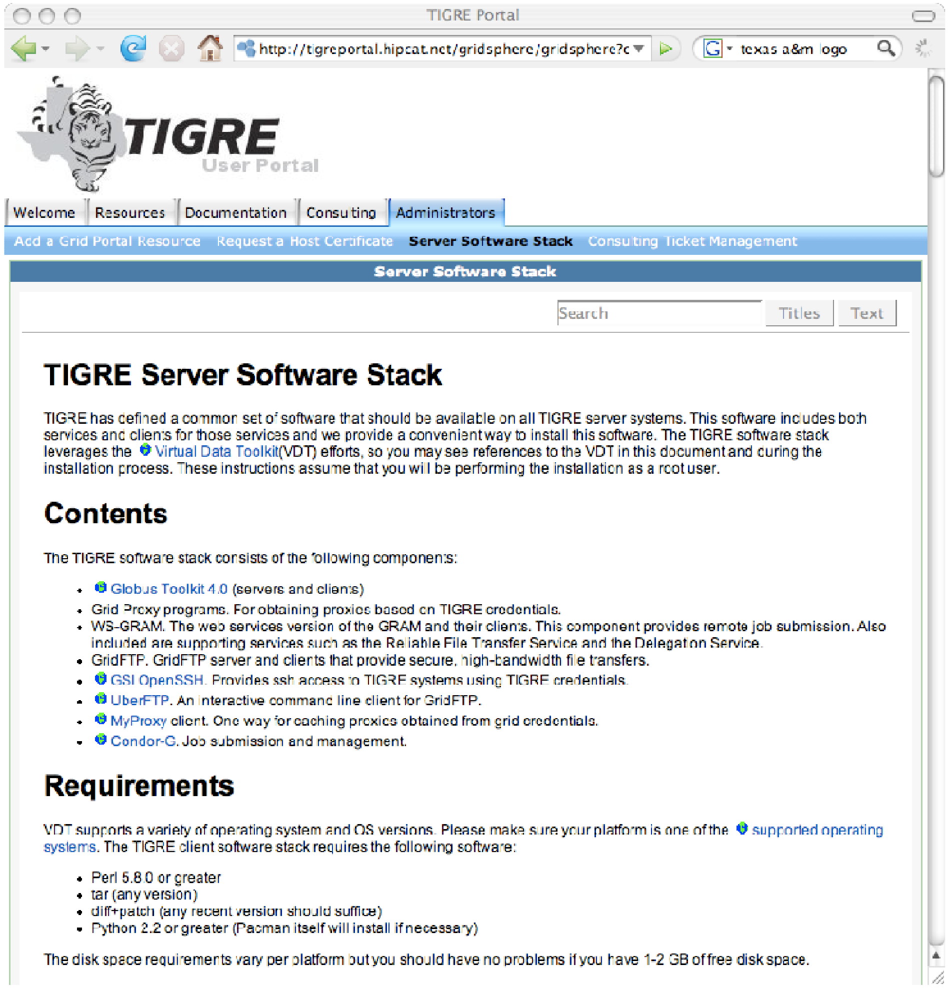
High Throughput Computing Resources

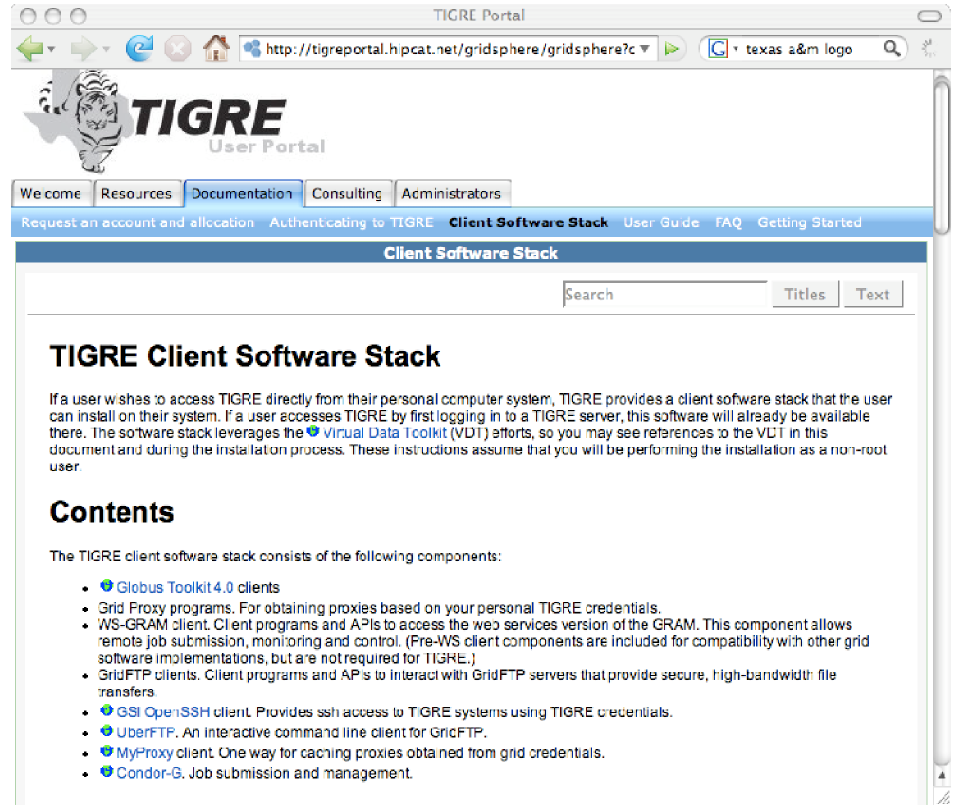
Name	Institution	Department	System	Active PCs	Active CPUs	Memory GBytes	Disk GBytes	Resource Details	Jobs
glb-test	Texas Tech University	High Performance Computing Center	Condor	**0** / 0	**68** / 68	14	1301	🔍	🔍
Rodeo	The University of Texas at Austin	Texas Advanced Computing Center	Condor	**31** / 33	**31** / 33	23	1189	🔍	🔍
			Total:	**31 / 33**	**99 / 101**	**37**	**2490**		

Jobs Key: **#R** - Number of Jobs Running, **#Q** - Number of Jobs Queued, **#O** - Number of Jobs in an Other State

TIGRE Software Stacks

Both drawn from Virtual Data Toolkit (VDT)





Download from

http://tigreportal.hipcat.net/gridsphere/gridsphere?cid=serverstack

http://tigreportal.hipcat.net/gridsphere/gridsphere?cid=clientstack

TIGRE Software Stacks

- Server software stack
  - Globus 4.x
    - GRAM4 (web services servers) & pre-web services
    - GPIR monitoring
    - GSI OpenSSH server
    - Information service (MDS)
    - Resource discovery (MDS)
    - MyProxy client
  - UberFTP
  - Condor-G
  - Scheduling (GRMS/Gridway)
  - Job submission (GRMS/globusrun-ws)
  - Accounting (none)
  - Account management (GUMS/VOMRS OSG/TTU is using)
  - Privilege management (none)

- Client software stack
  - Globus 4.x (pre-web services and web services clients)
  - GSI OpenSSH client
  - UberFTP
  - MyProxy client
  - Condor-G

Both drawn from Virtual Data Toolkit (VDT)
Storage is still an unresolved issue for TIGRE

TIGRE Environment

- Grid middleware consists of “minimal” set of components derived from **VDT**.
- Additional components will be added as they become necessary.
- **PacMan** packaging and distribution mechanism is used for distribution of client/server software.
- TACC (IGTF accredited) issues X.509 user/resource certificates.
- TIGRE Institutions serve as the Registration Authorities for their respective user base.
- Visit http://tigreportal.hipcat.net for detailed instructions on getting user Grid certificates.

TIGRE Environment (contd.)

- Job Scheduling and Management
  - TIGRE Supports GRAM4-based job submission via web services.
  - Job submission scripts are written using XML.
  - GRAM translates XML into local batch scheduler script.
  - GridFTP is used for file staging
  - Login to remote clusters is accomplished through GSISSH service which requires Grid Certificates

Typical Job submission script

```
<job>
  <executable>binary</executable>
    <directory>${GLOBUS_USER_HOME}/DEST-DIR</directory>
    <argument>-f</argument>
    <argument>input file</argument>
    <argument>output file</argument>
  <fileStageIn>
    <transfer>
        <sourceUrl>gsiftp://submission-hostname:2811/input file</sourceUrl>
        <destinationUrl>file:///${GLOBUS_USER_HOME}/inputfile</destinationUrl>
    </transfer>
  </fileStageIn>
  <fileStageOut>
    <transfer>
        <sourceUrl>file:///${GLOBUS_USER_HOME}/output file</sourceUrl>
        <destinationUrl>gsiftp://hostname:2811/output file</destinationUrl>
    </transfer>
  </fileStageOut>
</job>
```

Customer Service System

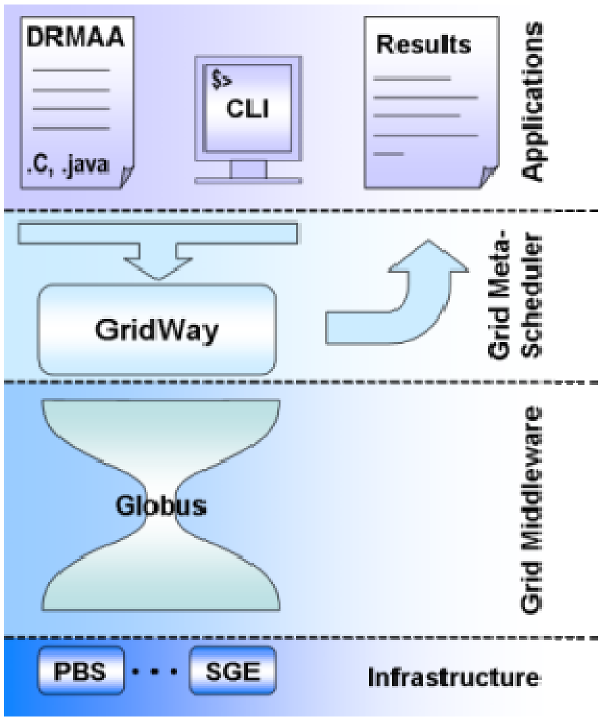
- The TIGRE Portal was designed and deployed using GridPort and maintained by TACC.
- TIGRE Portal http://tigreportal.hipcat.net provides one-stop-service point
  - Trouble tickets (OTRS)
  - MoinMoin Wiki for content and knowledge management.
  - Resource/user Certificate request and management
  - Tutorials, outreach and training materials

Global Scheduler

- Gridway
  - http://www.gridway.org/
  - Interface to local schedulers through GRAM
  - Resource management through Globus MDS
  - Dynamical job allocation and execution for efficient resource usage
  - Integrated in future releases of Globus Toolkit

- GridLab (Gridge) Resource Management System (GRMS)
  - http://www.gridge.org/content/view/30/66/, http://www.gridlab.org/WorkPackages/wp-9/
  - Interface to local scheduler (LSF, SGE, PBS) via GRAM
  - GridFTP for file staging
  - Use MDS querying for resource

Features of GridWay Metascheduler



http://www.gridway.org

- Interfaces with cluster specific schedulers such as LSF, PBS, etc.
- Simple to use command line interface for the end users.
- Provides dynamic scheduling and fault detection and recovery capabilities

- GridFTP for file staging
- Static resource discovery
- Dynamic scheduling

TIGRE Target Applications

Bioscience and Medicine

- UltraScan – Analysis of Macromolecular Assemblies
- Radiation Therapy modeling

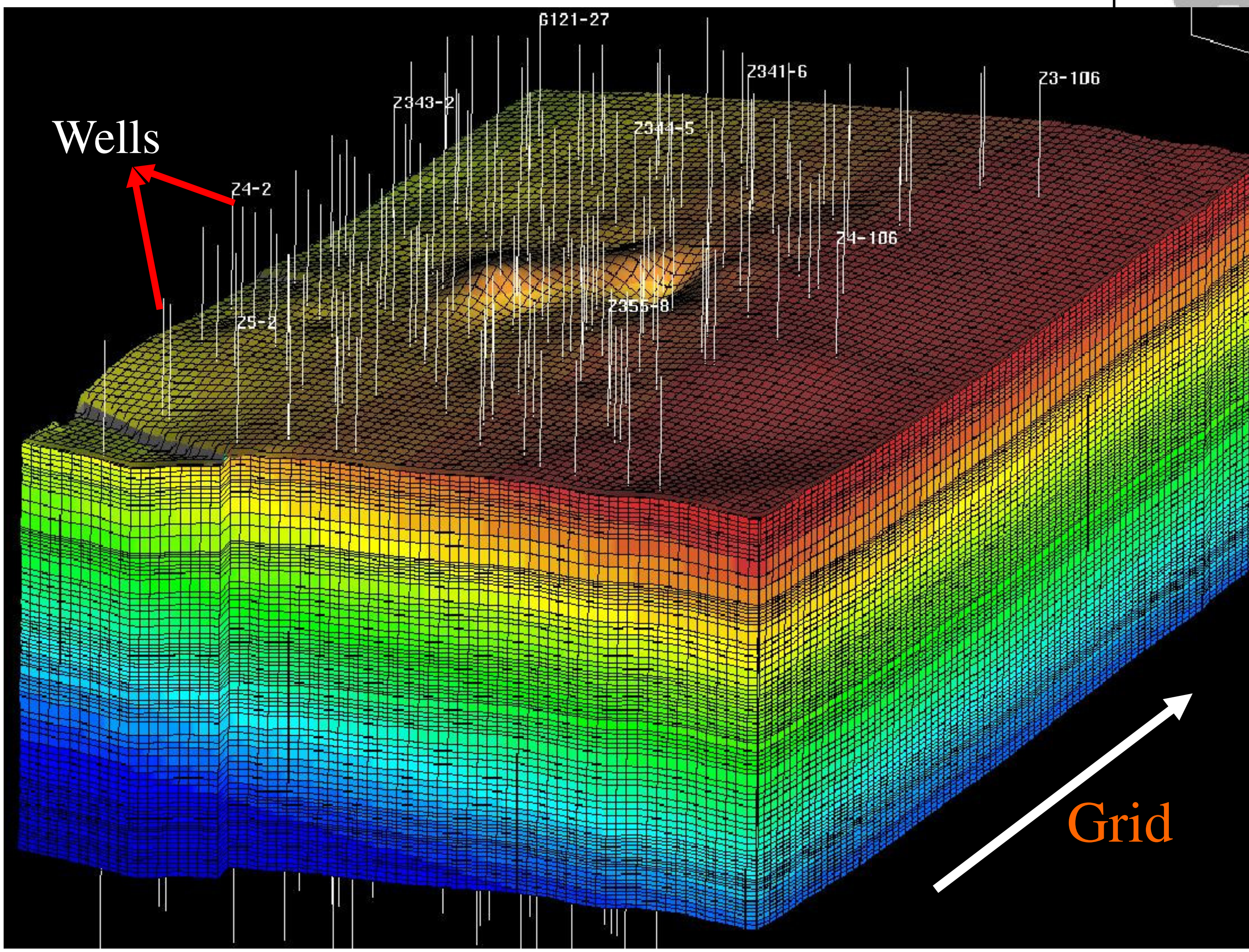
Air Quality Modeling

- MesoNet - meteorological information for Texas

Energy Exploration

- Ensemble Kalman Filter Data Assimilation Methodology using Schlumberger ECLIPSE

Production Field



Geological Models

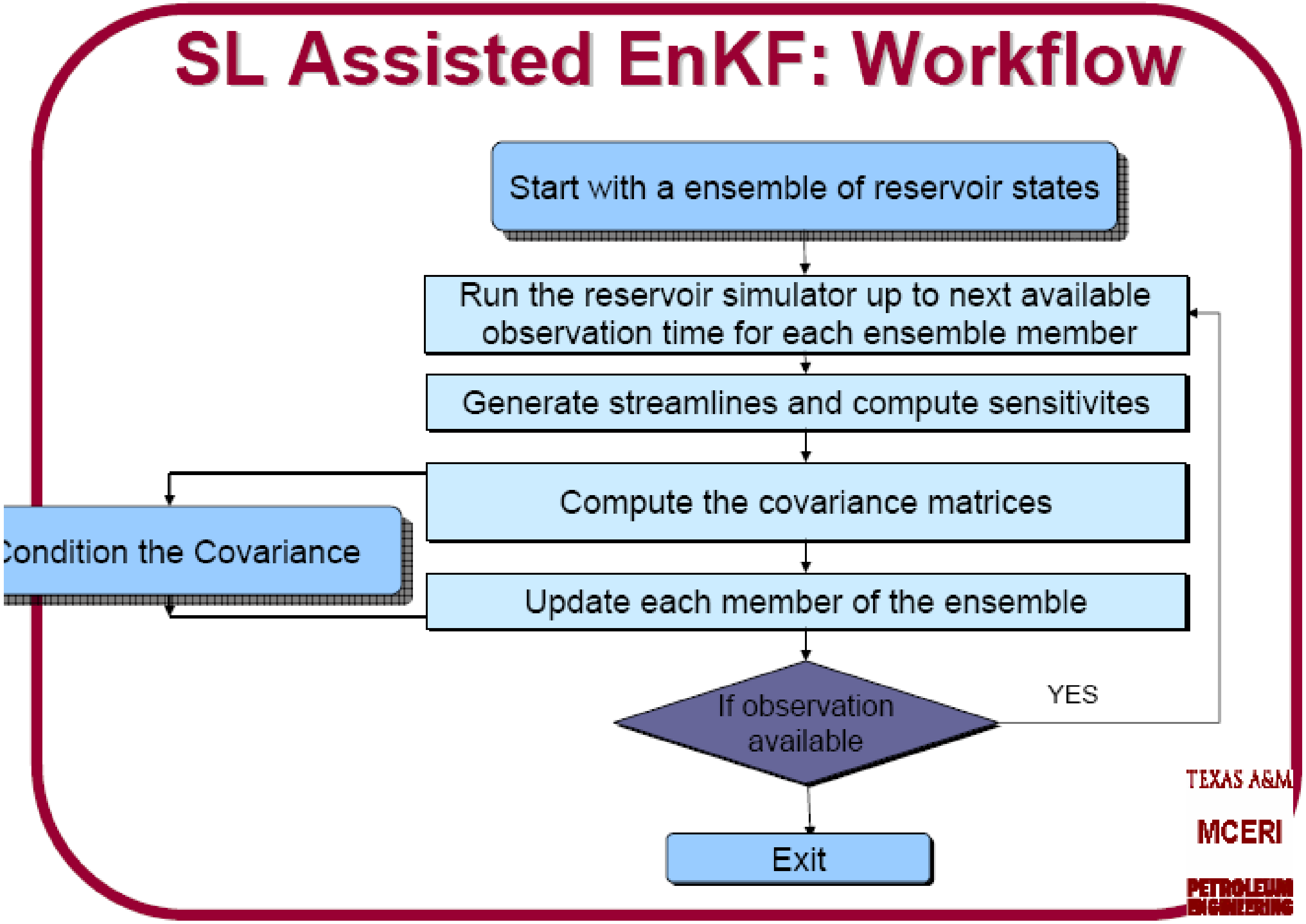
http://faculty.smu.edu/zchen/research.html

TAMU's EnKF Application

- Goal: Characterize reservoirs for optimal reservoir development and management.
- Complexity:
  - Hundreds of geological models, 50-60 years of real field production data
  - Integrates with Schlumberger ECLIPSE for model propagation
- TIGRE Scope: Supports industry standard data security requirements.
- Expected Outcomes:
  - Multi-institutional collaborations
  - Workforce development, education and outreach
- Collaborators (Current and in progress)
  - TTU, TAMU, UH and UT Austin

TAMU's EnKF

Streamline based covariant conditioning to reduce ensemble size



Simulation

- EnKF is Monte Carlo type [Parallelizable]
- TAMU EnKF is integrated with Schlumberger ECLIPSE
- Calculates cross-covariances between model-specific and field measurements (permeability, gas/oil ratio, etc) [Post-processing]
- Prototyped using MPI because it is straightforward.
- Deployed on Cosmos (TAMU's high demand compute cluster)
- **Runtime** depends on number of ECLIPSE licenses, compute nodes, ensemble size, and historical data.

MPI (VS) Grid Scheduling

Message-Passing

- Jobs won't **start** until requested nodes are available
- Takes longer to schedule N CPU job, if N is large.
- Licenses are node-locked until MPI_Finalize()
- Ideally, # of Licenses should be equal to N

Grid Scheduling

- Single CPU jobs
- Small/No wait time
- Licenses are node-locked and released.
- More jobs can be scheduled for the same # of licenses.
- Effective resource (CPU, and license) utilization

Cluster Scheduling Options

Heavily Loaded

- Takes longer to schedule N CPU job compared to N single processor jobs

Largely Available

- Parallelizability is restricted by the # of available licenses.

Grid Scheduling is necessary

Grid-Enabling Effort

Step 1: Remove MPI calls from the code

- Convincing the user, making him do the necessary work
- Six weeks, ~2MB text messaging ☹

Step 2: Secure resources

- ECLIPSE, Code installation on Grid resources

Step 3: Setup License Sever environment

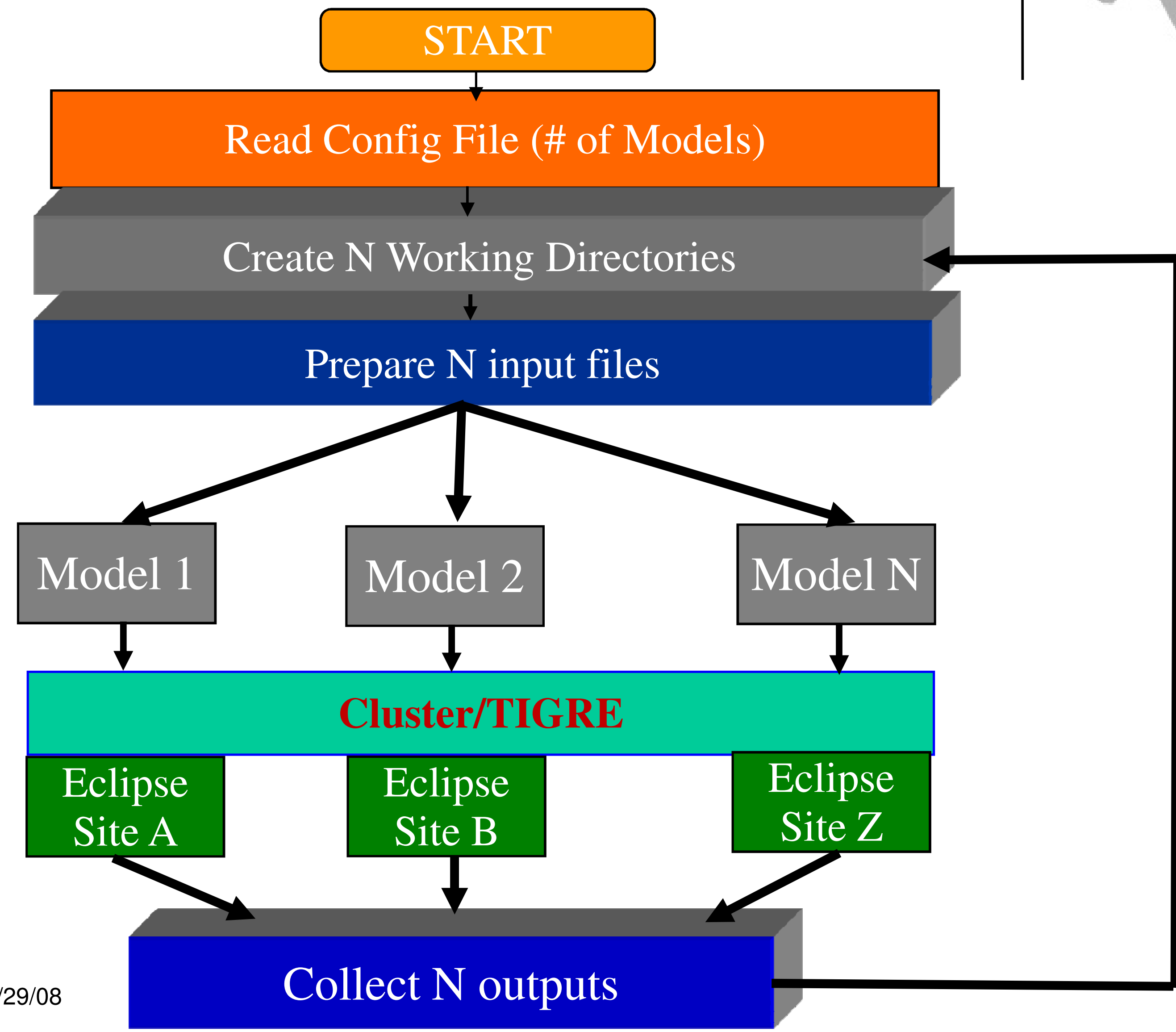
- Work with Schlumberger/Institutional contacts

Step 4: Demonstration of Services

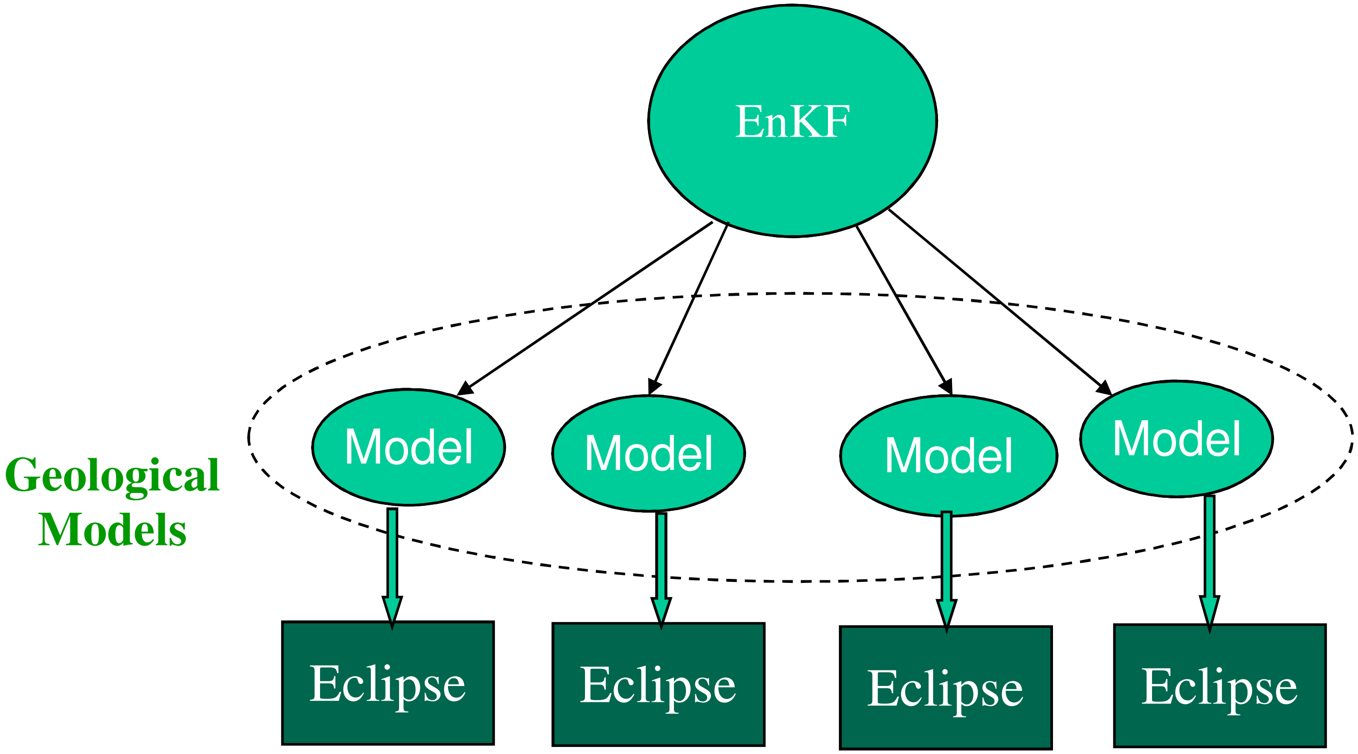
- Run Jobs across TIGRE, train user and return the token.

Evaluation: We get a message from the user only when the Grid server is down ☺/☹

EnKF Workflow



EnKF Process Management



Degree of parallelism

Subject to availability of ECLIPSE licenses and CPUs

Compute Resources

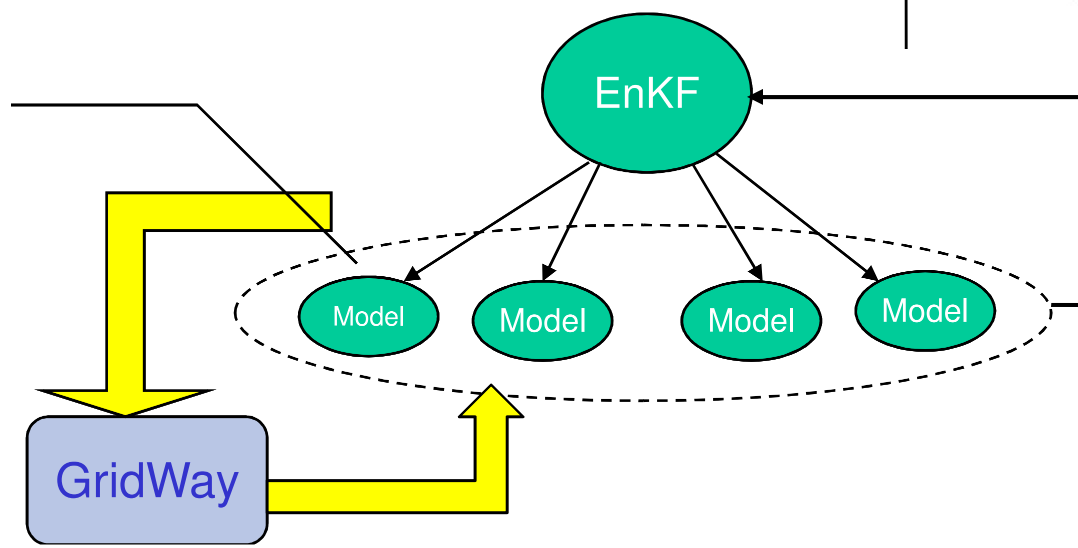
- Subset of TIGRE systems http://tigreportal.hipcat.net
- Simulation requires EnKF software and ECLIPSE licenses
- ECLIPSE licenses were pooled from TTU and TAMU
- GridWay Metascheduler for TIGRE wide job management

Name	Institution	System	OS	CPUs	Archi-tecture	Memory GBytes	Disk GBytes	Local Resource Manager
Cosmos	Texas A&M University	SGI Altix 3700	SUSE ES9	128	ia64	256	4096	**PBS**
TTU-Minigar	Texas Tech University	Dell Xeon Cluster	CentOS 4.4	32	x86_64	64	70	**LSF**
TTU-Antaeus	Texas Tech University	Dell Xeon Cluster	CentOS 4.4	192	x86_64	96	6000	**LSF**

Job Management using GridWay

Job template example

EXECUTABLE=runForward
REQUIREMENTS=HOSTNAME="cosmos.tamu.edu"|HOSTNAME="antaeus.hpcc.ttu.edu"|HOSTNAME="minigar.hpcc.ttu.edu"
ARGUMENTS=001
INPUT_FILES=001.tar
OUTPUT_FILES=001.out.tar



TIGRE Resources as seen by GridWay

% gwhost

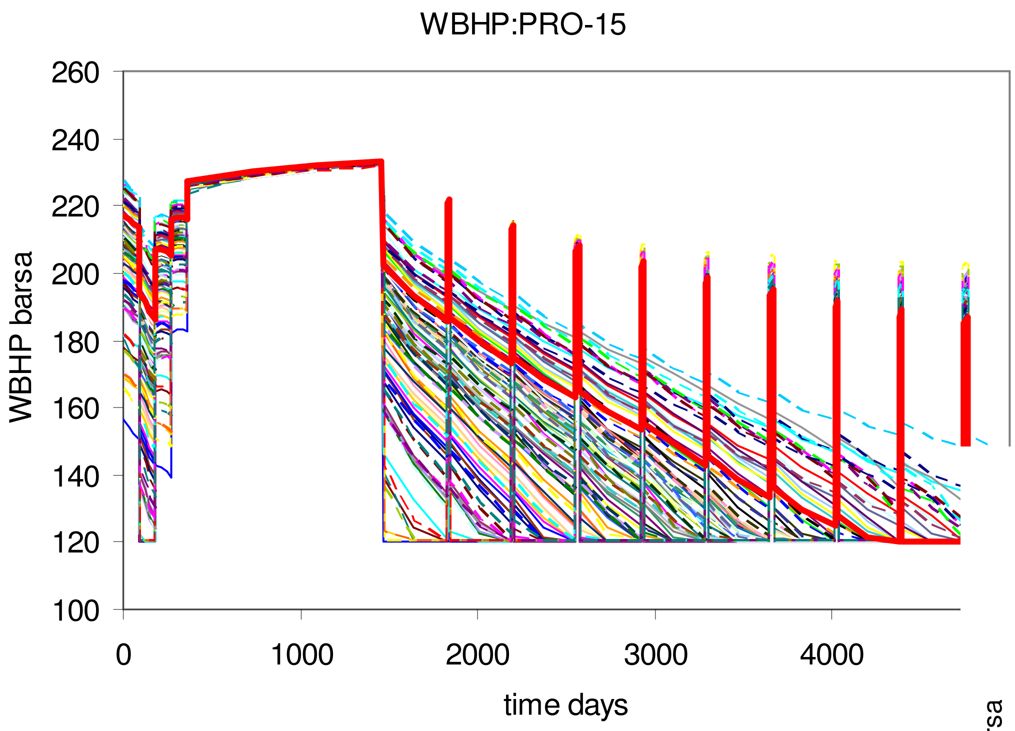
HID	PRIO	OS	ARCH	MHZ	%CPU	MEM(F/T)	DISK(F/T)	N(U/F/T)	LRMS	HOSTNAME
0	1	Linux2.6.9-42.0	x86_6	3591	1385	3783/4045	19688/56426	0/4/16	LSF	minigar.hpcc.ttu.edu
1	1	Linux2.6.9-42.0	x86_6	2000	385	3783/4045	19688/56426	0/4/4	LSF	antaeus.hpcc.ttu.edu
2	1	Linux2.4.21-sgi	ia64	1300	785	15000/16000	1221335/2231884	0/2/8	PBS	cosmos.tamu.edu
3	1	Linux2.6.9-42.0	x86_6	2660	385	11295/12303	150/200	0/4/4	Fork	lonestar.tacc.utexas.edu
4	1	Linux2.4.21-sgi	ia64	3201	185	180/431	40461/74312	0/2/2	Fork	eldorado.acrl.uh.edu
5	1	Linux2.6.9-42.E	ia64	999	185	3323/4124	509671/655157	0/2/2	Fork	gridgate.rtc.rice.edu

TIGRE Queue

USER	JID	DM	EM	NAME	HOST
pingluo	88	wrap	pend	enkf.jt	antaeus.hpcc.ttu.edu/LSF
pingluo	89	wrap	pend	enkf.jt	antaeus.hpcc.ttu.edu/LSF
pingluo	90	wrap	actv	enkf.jt	minigar.hpcc.ttu.edu/LSF
pingluo	91	wrap	pend	enkf.jt	minigar.hpcc.ttu.edu/LSF
pingluo	92	wrap	done	enkf.jt	cosmos.tamu.edu/PBS
pingluo	93	wrap	epil	enkf.jt	cosmos.tamu.edu/PBS

DM: Dispatch state
EM: Execution state
JID: Job Id and
HOST: Site specific cluster and its local batch scheduler.

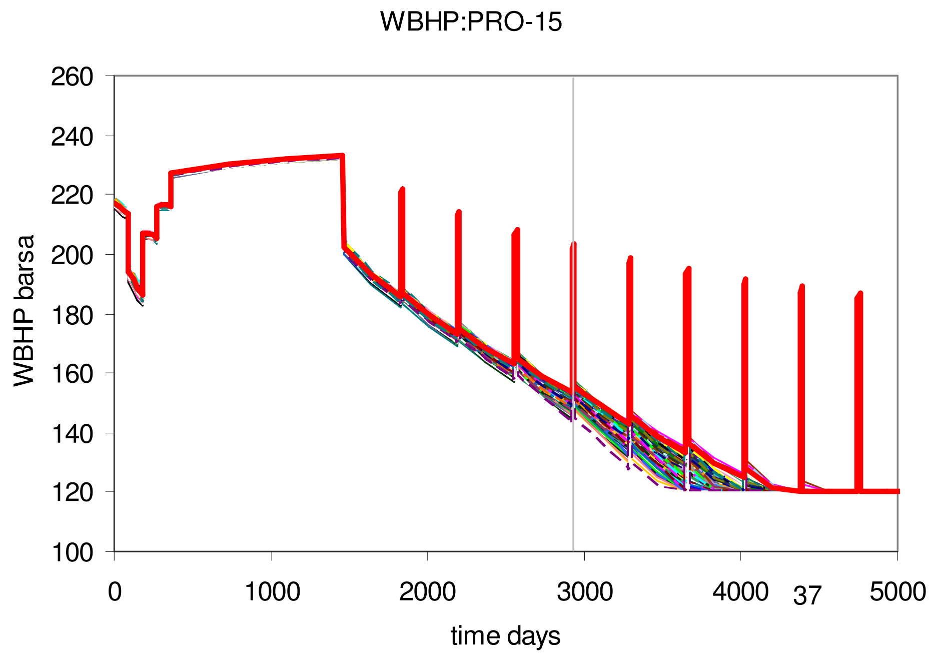
Test Results



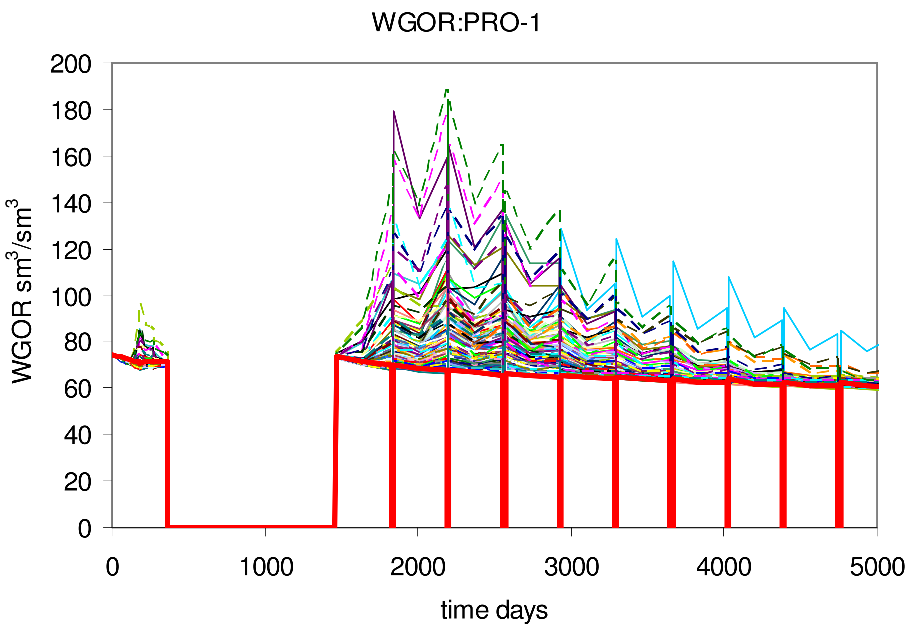
Individual Models

Bottom hole Pressure

History-matched models



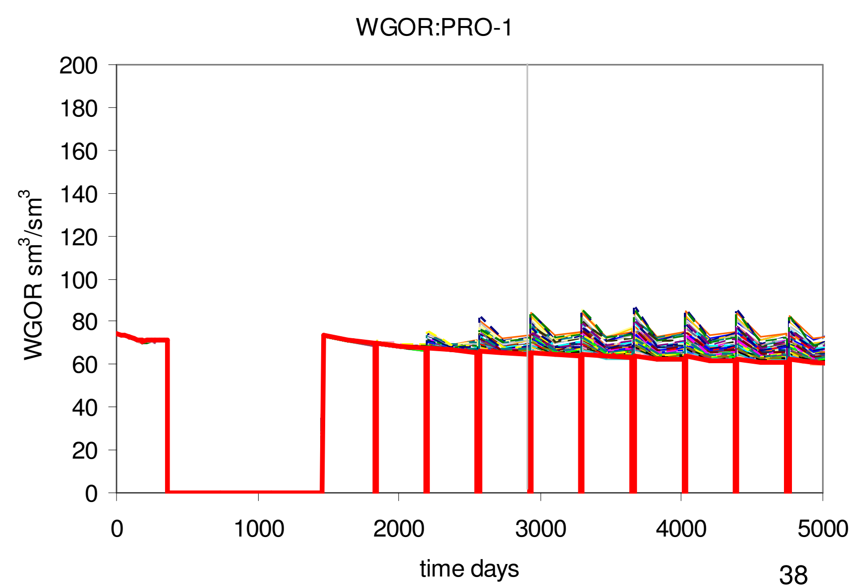
Test Results



Individual Models

Gas-to-Oil Ratio

History-matched models



Application Demo

- http://eldorado.acrl.uh.edu:9080/tigre-demo/demo2.pl
- The GWPS was iteratively SCP'ed from Cosmos (TAMU) to Eldorado (UH)
- Cosmos collects TIGRE runs from TTU, TAMU for post-processing.

Lessons Learned

- Cluster level QoS expected
- User recruitment and retaining
- Learning Curve.

Future Work

- Runtime analysis between MPI and Grid Scheduling
- Job-to-license Ratio Management

Summary & Conclusions

- The EnKF is a Monte Carlo type approach for real time analysis and characterization of reservoirs.
- The TAMU's EnKF implementation invokes Schlumberger ECLIPSE for model specific simulations.
- We have Grid enabled the MPI implementation and deployed on TIGRE using GridWay metascheduler.
- The Grid-enabling effort is necessary if a single institution can not support required number of licenses and computational resources.
- Efforts are underway to involve other TIGRE schools in this effort.
- 2 publications and 2 proposals (one submitted).

Extensions & Collaborations

- Extensions
  - Extension of the effort to involve other TIGRE Institutions
  - Environment for education & outreach
- Collaborations
  - Dr. Siddiqui, REF Proposal on creating an educational infrastructure through TIGRE.
  - Schlumberger for support with extensions

References

- The Kalman Filter (tutorials, references, etc.) http://www.cs.unc.edu/~welch/kalman
- A A-Vazquez and A R Syversveen, “The Ensemble Kalman filter – theory and applications in oil industry”, http://publications.nr.no/enkal.pdf
- Geir Evansen http://enkf.nersc.no
- B Vallés et al. Rogaland Research, “Reservoir Management by means of Data Assimilation” http://www.spe.no/bergen/doc
- Valestrand and Naevdal, IRIS & TAMU Petroleum, EnKF Workshop (2007) http://qp.rf.no/QuickPlace/enkfseminar
- TIGRE Project Documentation and Portal: http://tigreportal.hipcat.net
- R Vadapalli, P Luo, T S Kim, A Kumar and S Siddiqui, “Demonstration of Grid-enabled EnKF for reservoir characterization”, 15th ACM-MG08:Grid-enabling Applications (2008)

Credits

- TIGRE Development Team
  - Rice University: Chuck Koelbel, Kiran Thyagaraja
  - Texas A&M University: Taesung Kim, Ping Luo
  - Texas Tech University: Ravi Vadapalli, Alan Sill
  - University of Houston: T. Mark Huang
  - University of Texas at Austin: Rion Dooley, Marg Murray, Warren Smith
- Collaborators:
  - Akhil Datta-Gupta & Ajitabh Kumar (Texas A&M)
  - Lloyd Heinze & Shameem Siddiqui (TTU)

Thanks!!

Questions?