SnBDCV

Shake-and-Bake Distributed Collaborative Visualization

By: Amin Ghadersohi

CSE714 – Fall 2006 - Dr. Russ Miller.

Department of Computer Science and Engineering

Distributed Visualization

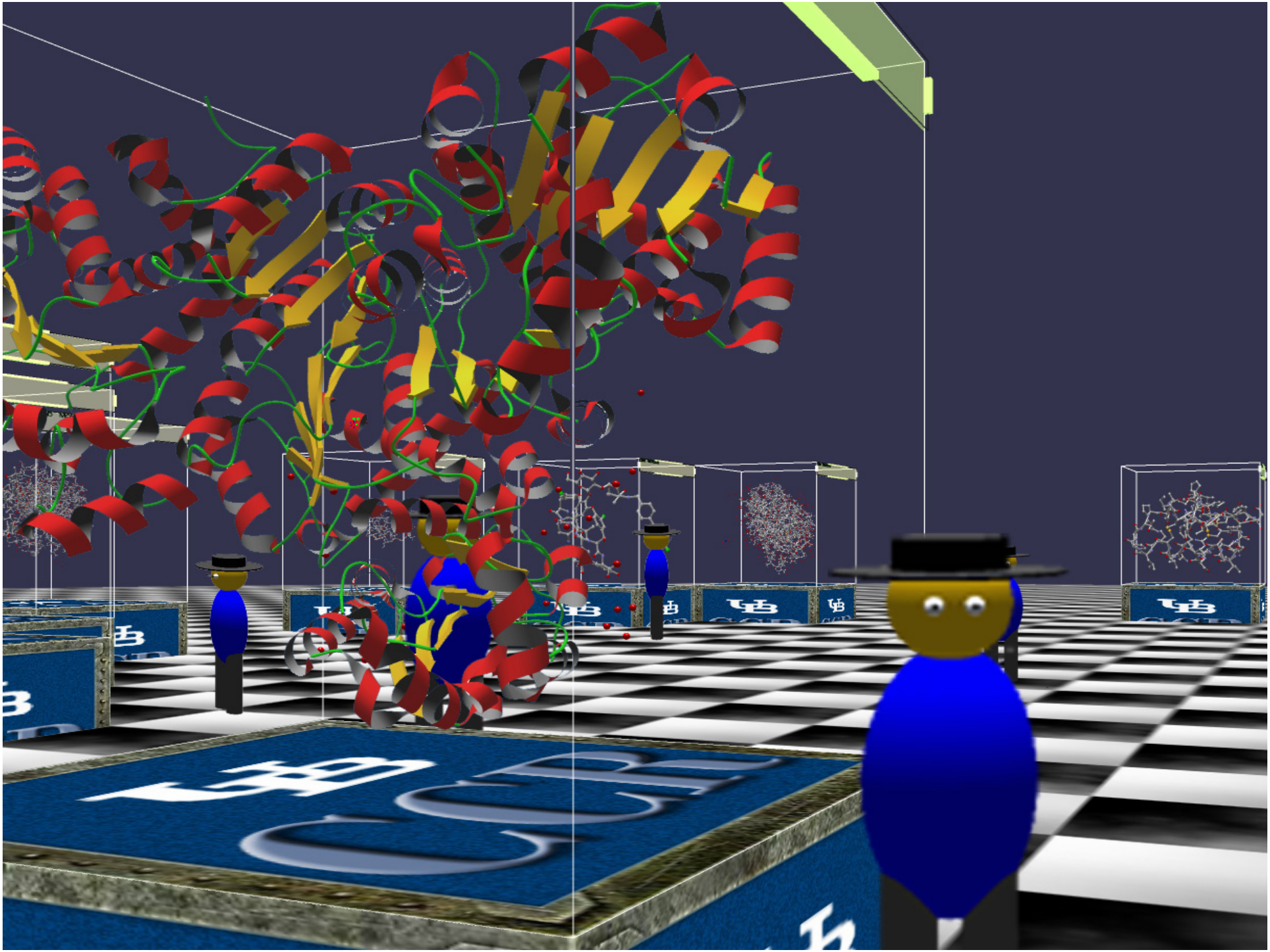
- Modules run on different machines.
- Distributed resources used in order to visualize large and computationally intensive datasets.
- Only the computation for the visuals is performed on distributed resources.

Collaborative Visualization

- Multiple human participants
- Joint interpretations of visualized data
- However, may still refer to a single view system with multiple people looking "over each other's shoulders"

Distributed Collaborative Visualization (DCV)

- Collaboration performed both at the system and human levels
- Sharing of resources and data from remote locations
- Common virtual environment
- Grid friendly!

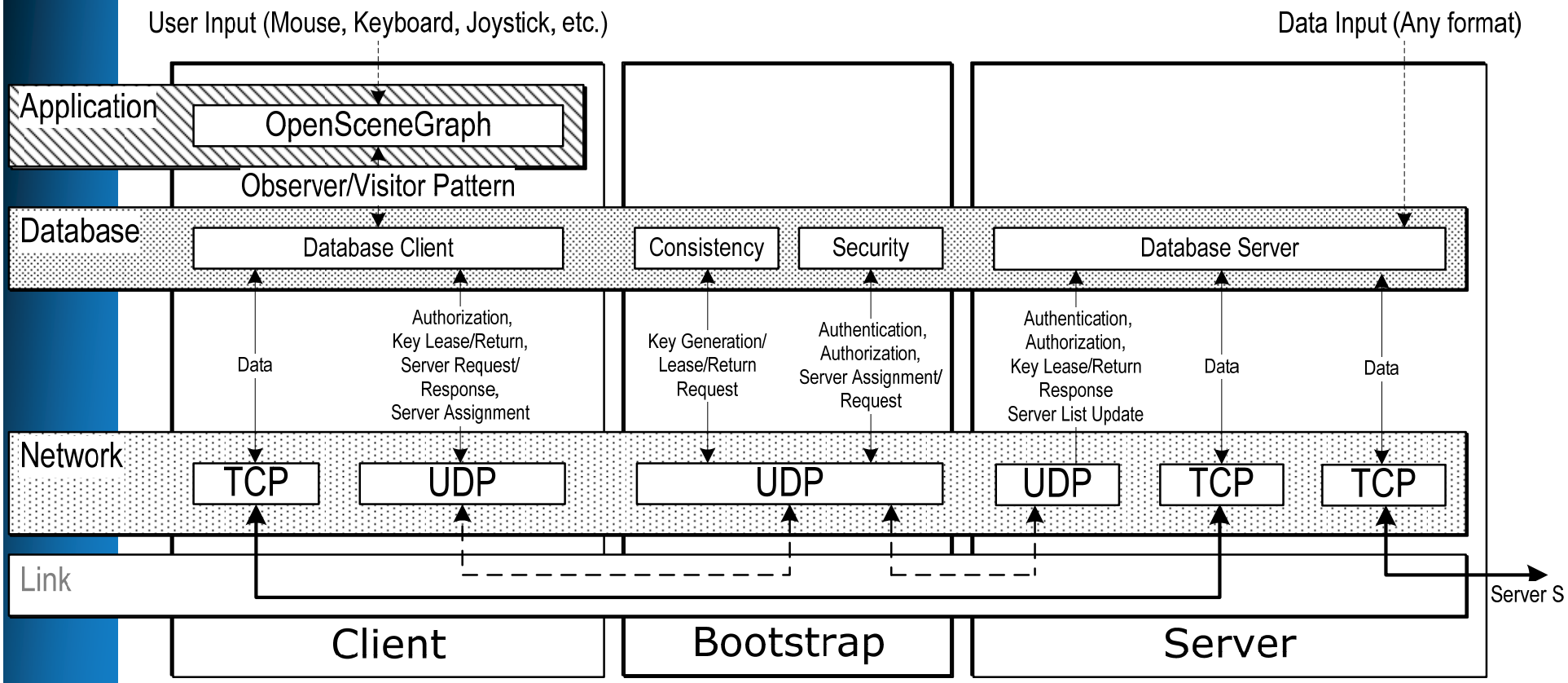


Requirements

- Networking
  - TCP and UDP
- Real-time Distributed Database
  - On top of network
- Data Consistency
  - Function of Real-time Distributed Database
- Scalability
  - Depends of Real-time Distributed Database
- Security
- Portability
- Graphics performance
  - OpenSceneGraph API
- Interaction performance
  - Tradeoff between consistency and interaction performance

Architecture

SnB DCV
Shake-and-Bake
Distributed Collaborative Visualization



Data Consistency

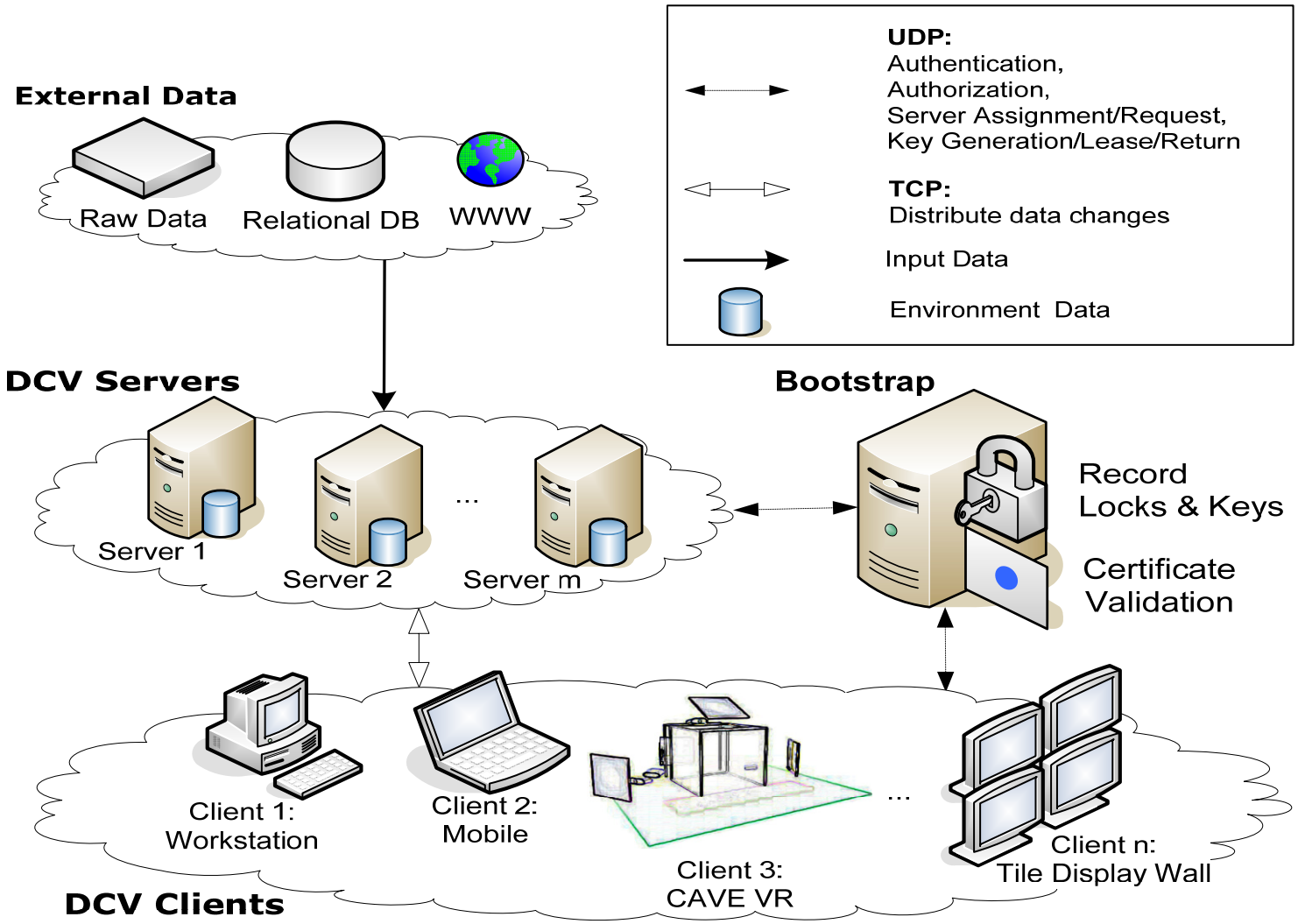
SnB DCV
Shake-and-Bake
Distributed Collaborative Visualization

- Central record locking mechanism.
- Performed by a third tier: the bootstrap
- Consistency guarantees
  - Performance tradeoff
    - A client needs to wait for permissions before changing the data
    - Decouple consistency from data distribution
    - The boostrap is a central, lightweight and fast protocol based on UDP
    - Scalable design

SnB DCV

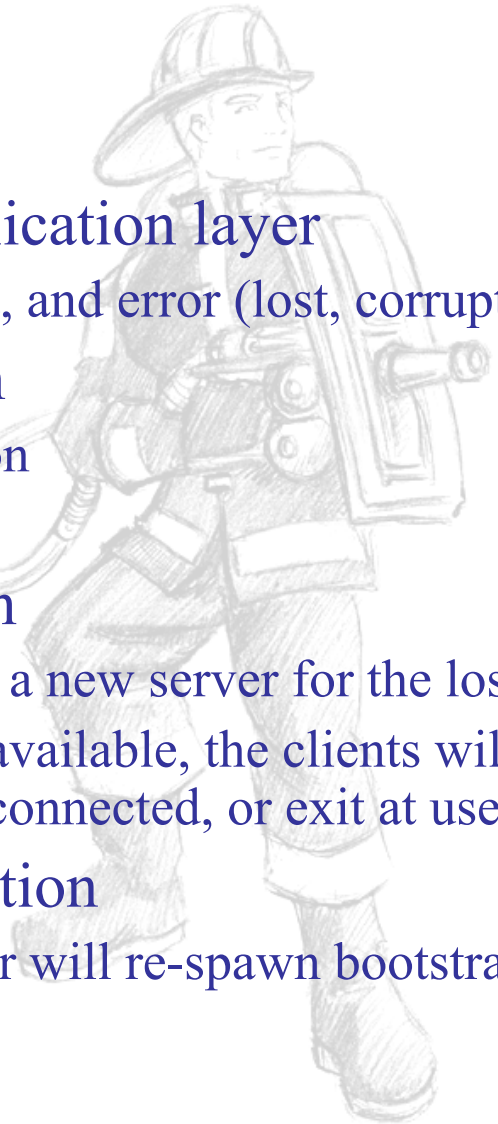
Shake-and-Bake

Distributed Collaborative Visualization



Failure Management

SnB DCV
Shake-and-Bake
Distributed Collaborative Visualization



- Network-aware application layer
  - Connection, dropout, and error (lost, corrupt packet) notification.
- Client disconnection
  - Wait for reconnection
  - Exit at user's will
- Server disconnection
  - Bootstrap re-assigns a new server for the lost server's clients.
  - If no others servers available, the clients will halt all operations until the server is reconnected, or exit at user's will.
- Bootstrap disconnection
  - Predetermined server will re-spawn bootstrap from backup data.

Load Balancing

SnB DCV
Shake-and-Bake
Distributed Collaborative Visualization

- One server can only serve so many clients
- More servers needed for extra clients
- Servers form a data backbone
- Bootstrap assigns servers to the clients dynamically based on system load statistics

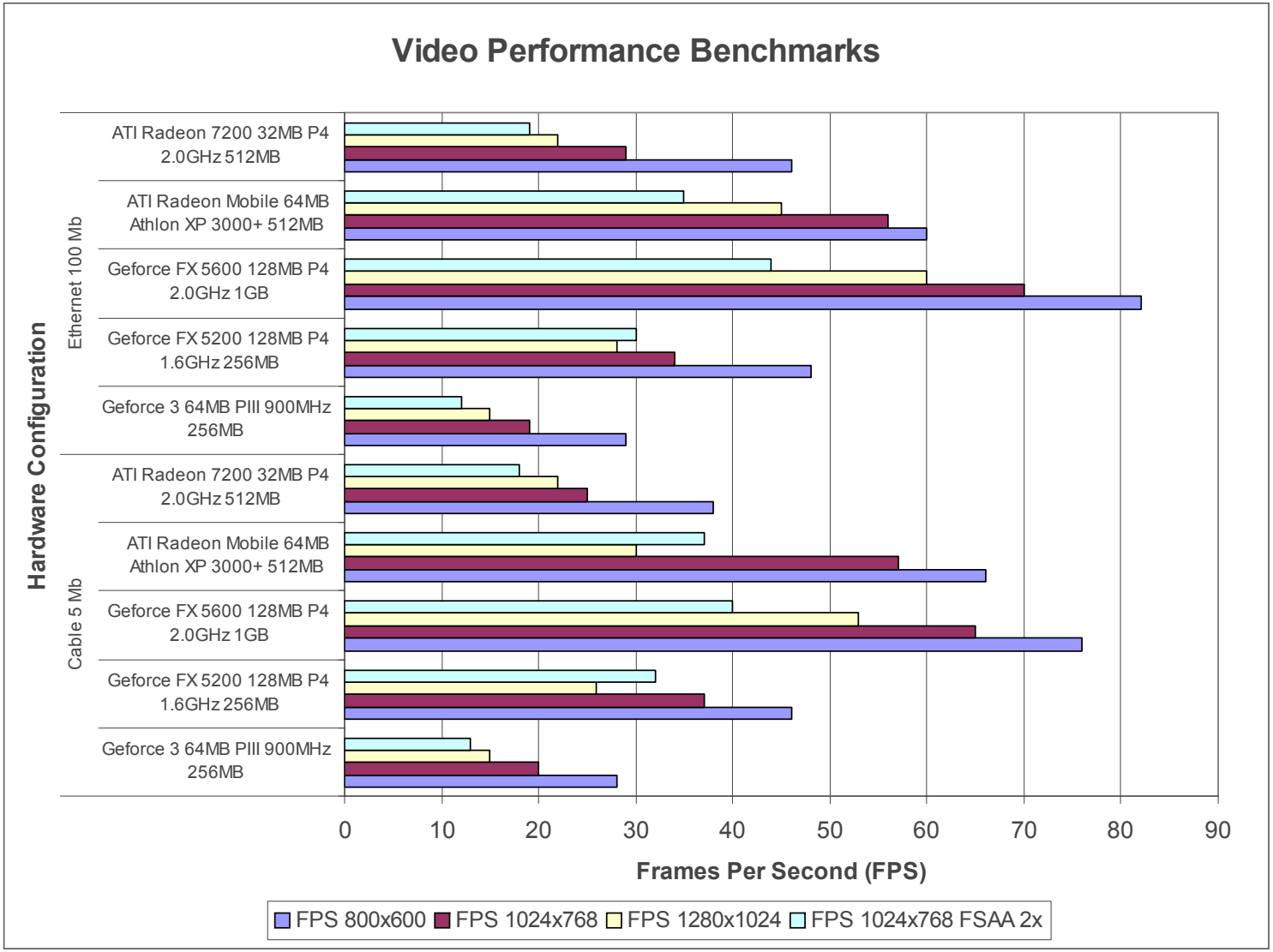
Security

SnB DCV
Shake-and-Bake
Distributed Collaborative Visualization

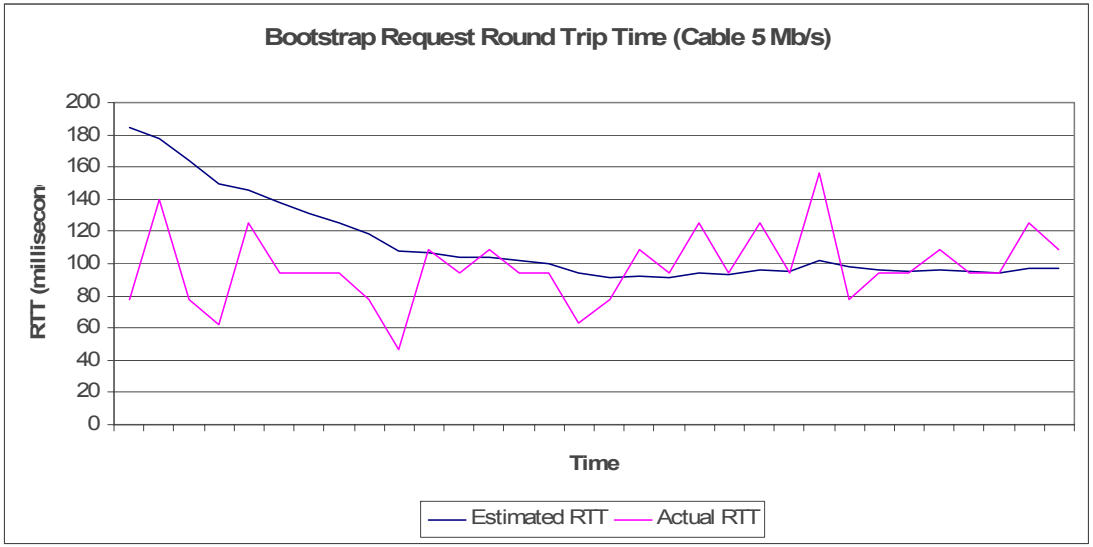
- Based on a certificate or proxy from a known and trusted Certificate Authority (CA) that can be verified by the bootstrap.
- Maintain a mapping of certificates to a list of valid IP addresses to block intruders with stolen or hijacked certificates.
  - Detect spoofing via the routing trace.

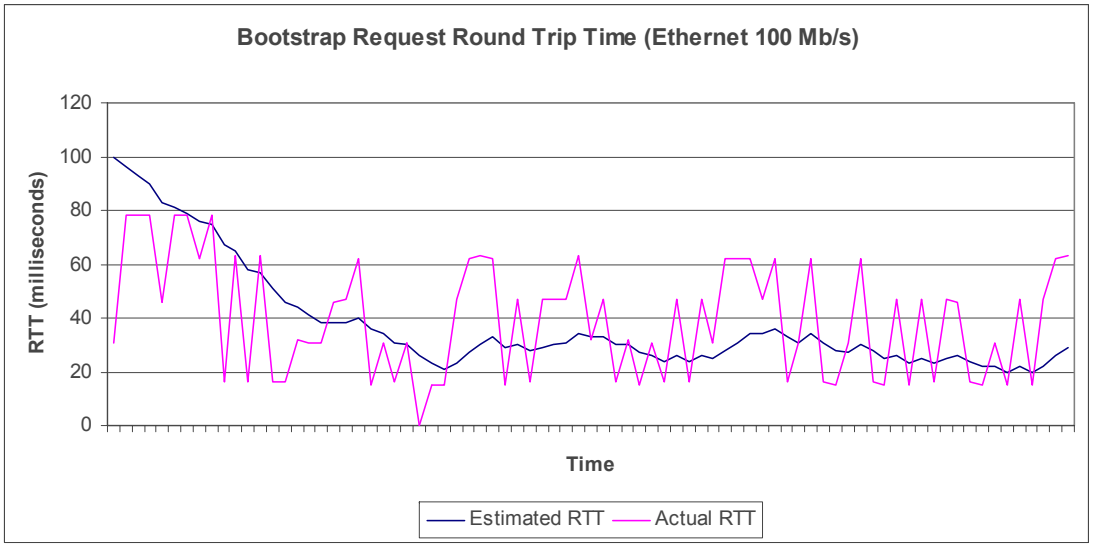
Graphics Performance

SnB DCV
Shake-and-Bake
Distributed Collaborative Visualization

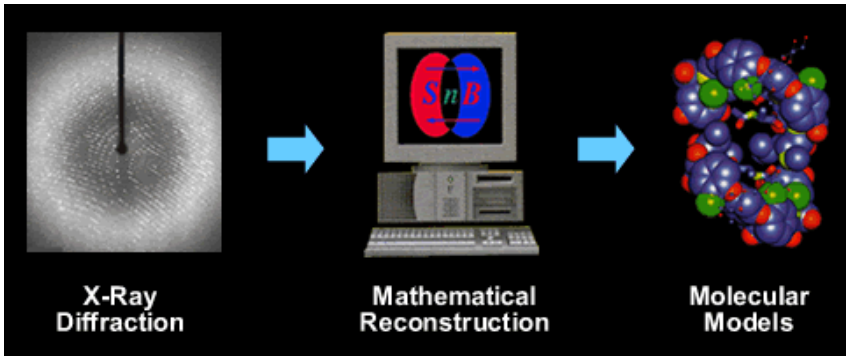


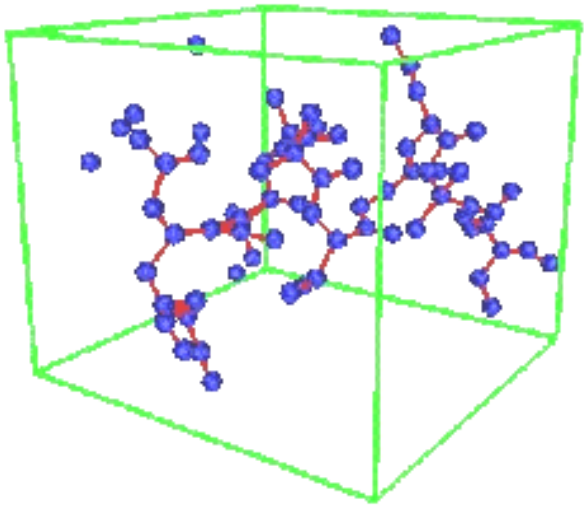
Interaction performance





Shake-and-Bake (SnB)



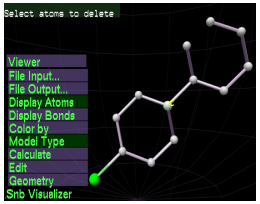


- *Shake-and-Bake* [Miller et al. 1994] is a state-of-the-art program for molecular crystal structure determination from X-ray crystallographic data.
- The *SnB* package:
  - Distributed to more than 500 laboratories worldwide.
  - Released in the mid-1990s.
    - primitive visualization.
- Advances in Grid Computing:
  - the *SnB* package no longer bound to one physical location.
  - Jobs queued/executed from anywhere in the world.
  - Need to monitor intermediate distributed results.

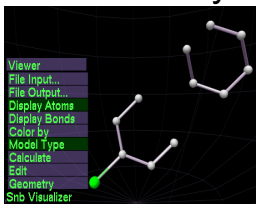
*SNB*VIS: *Shake-and-Bake* Visualizer

- Visualization of protein structures
- Integrate multiple data sources, such as *SnB* instances running on the Grid
- Monitor progress of structures
- Discuss, annotate, and edit structures as a team
- Used by chemists and the crystal determination community

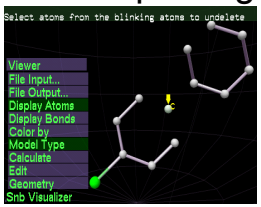
- Edit structures determined by the *SnB* package:



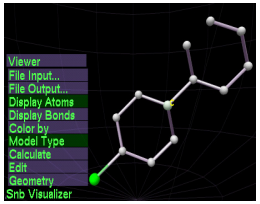
Select atom for deletion



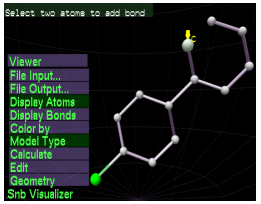
Atom deleted



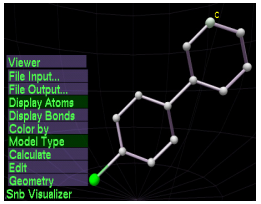
Undelete atom



Atom undeleted

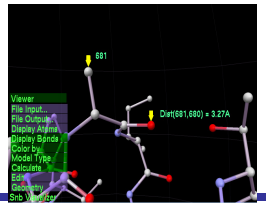


Select atoms to bond

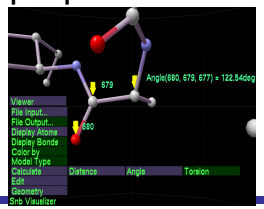


Atoms bonded

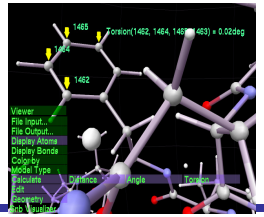
- Calculate structural properties:



Distances

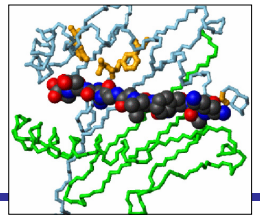


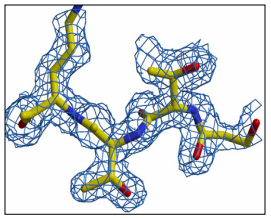
Angles

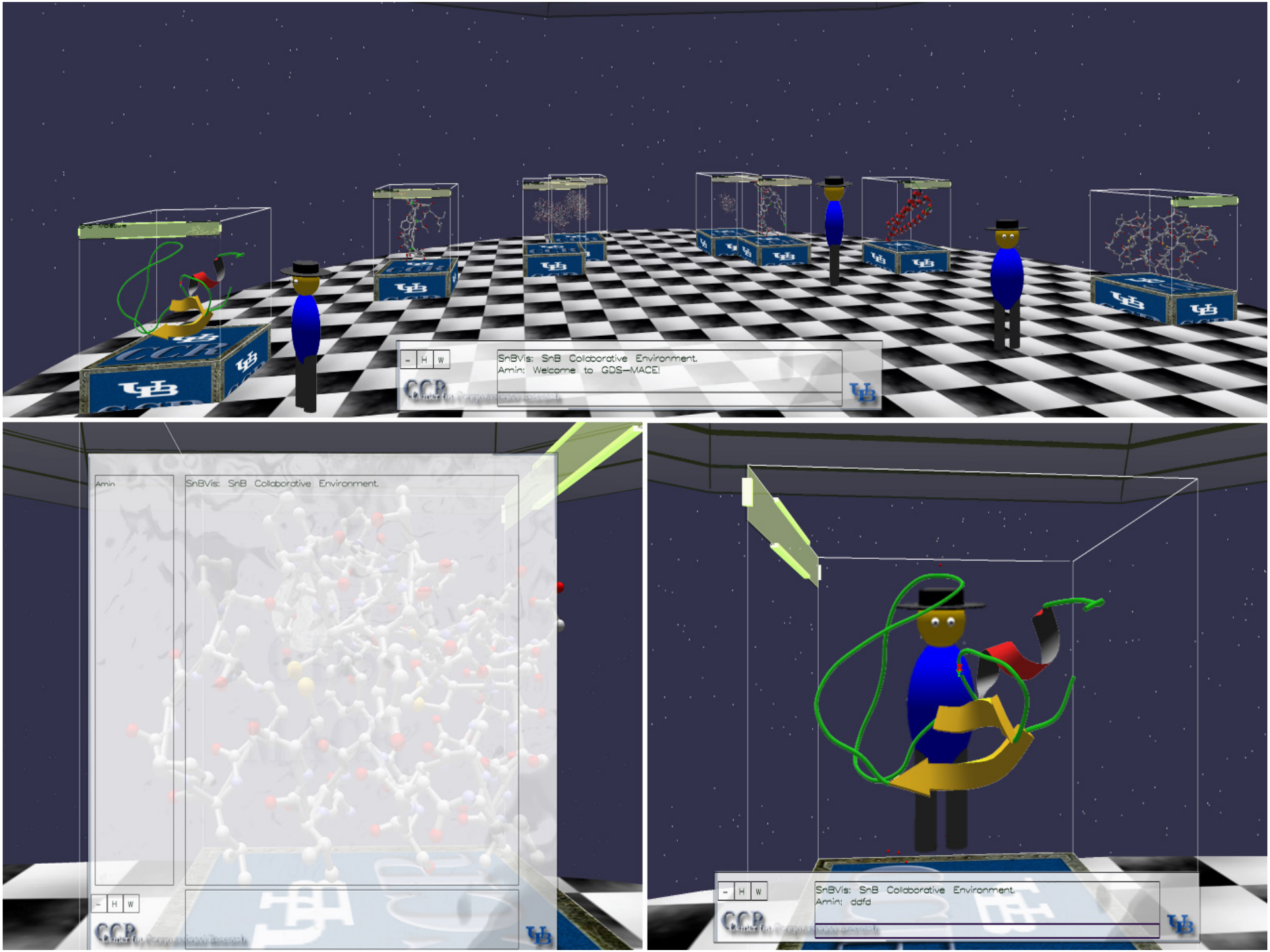


Torsion angles

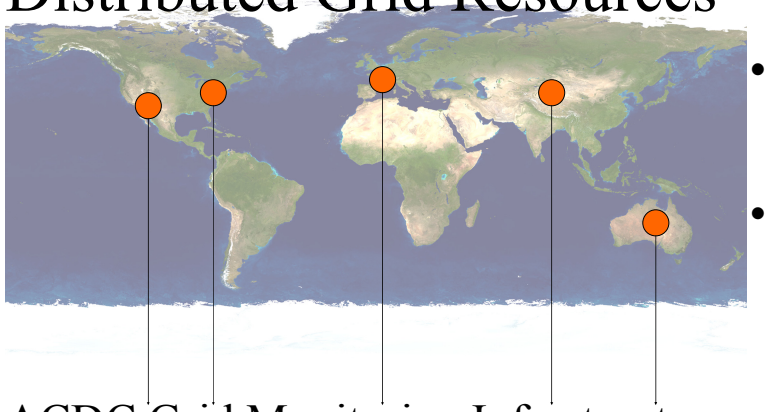
- Future:


Molecular Docking

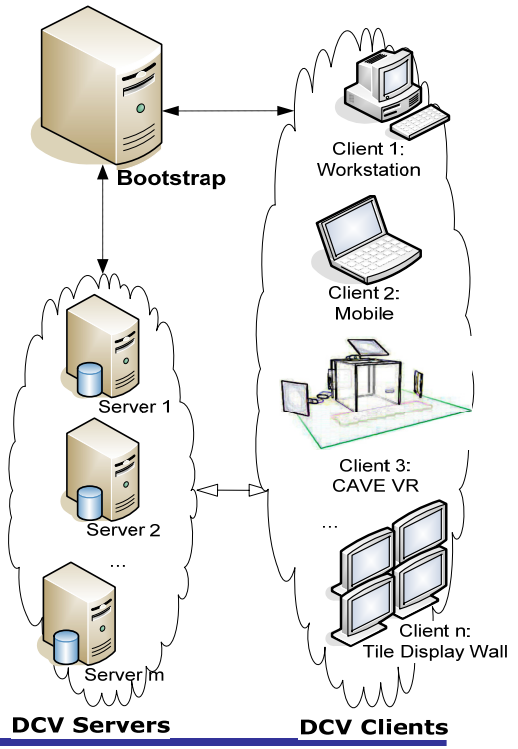

Electron Density Fitting

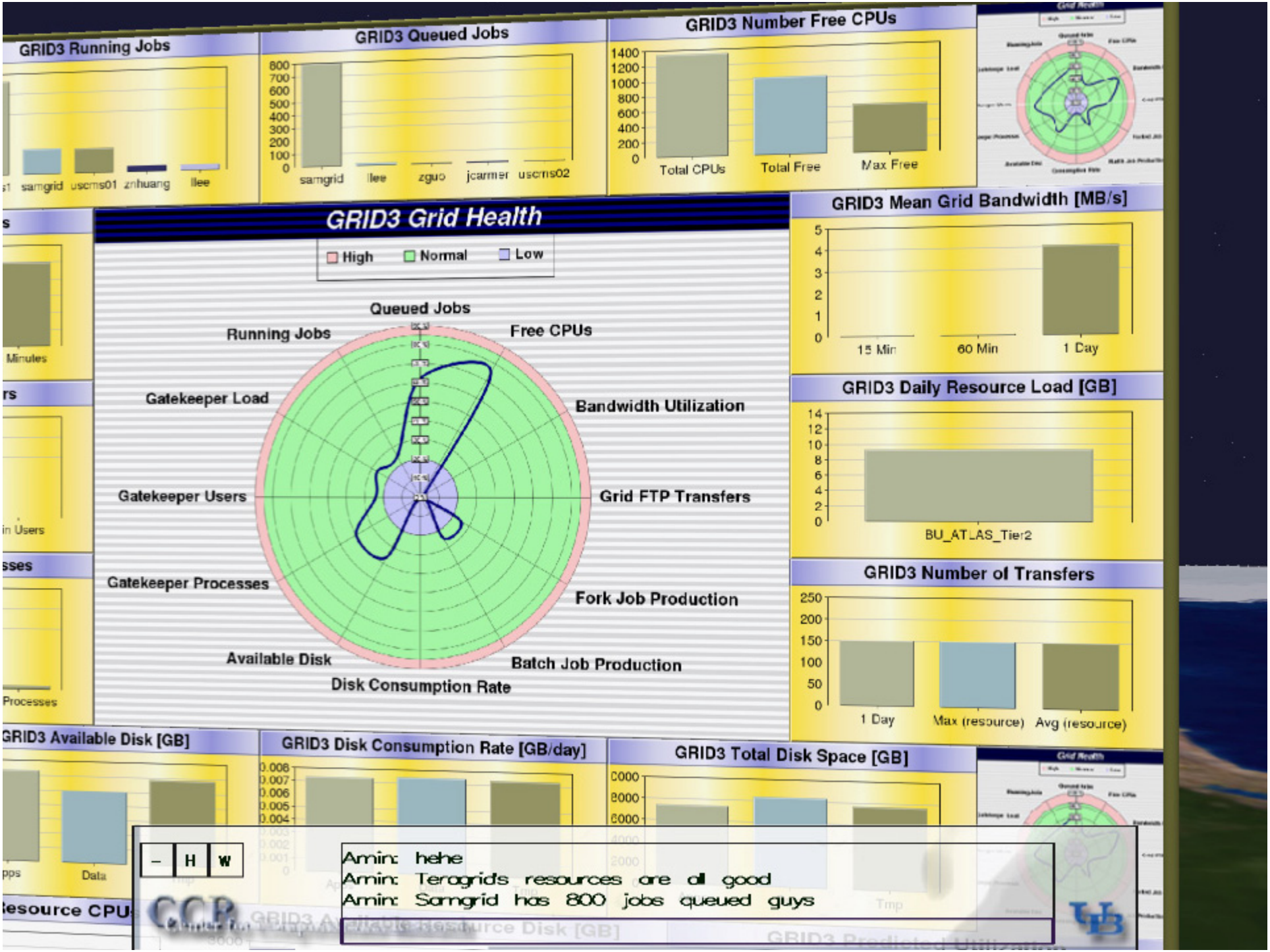

SnBVis: SnB Collaborative Environment.
Amin: Welcome to GDS—MACE!
Amin
SnBVis: SnB Collaborative Environment.
SnBVis: SnB Collaborative Environment.
Amin: ddfd

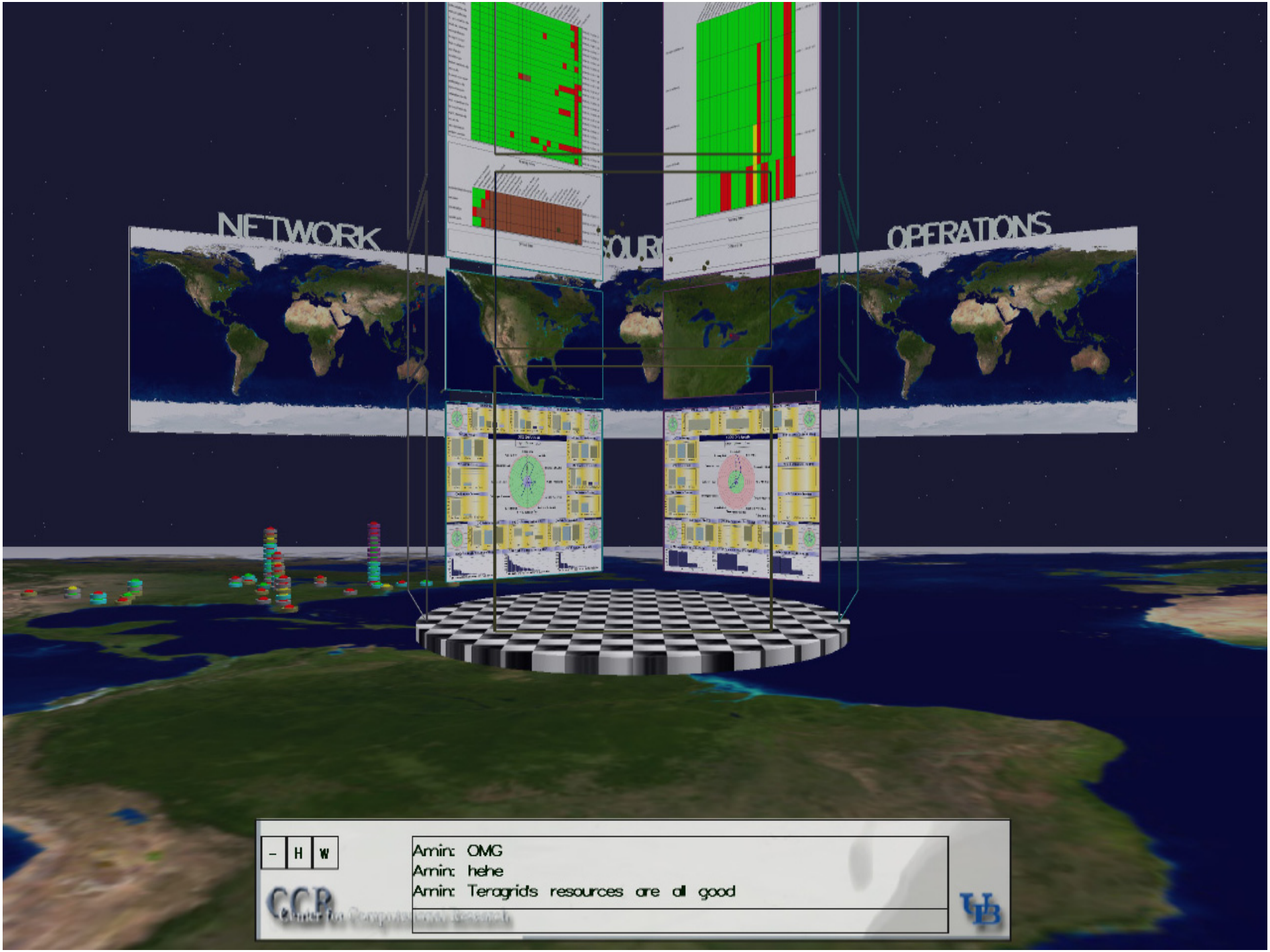
Grid *MACE: Grid Monitoring and Administration Collaborative Environment*

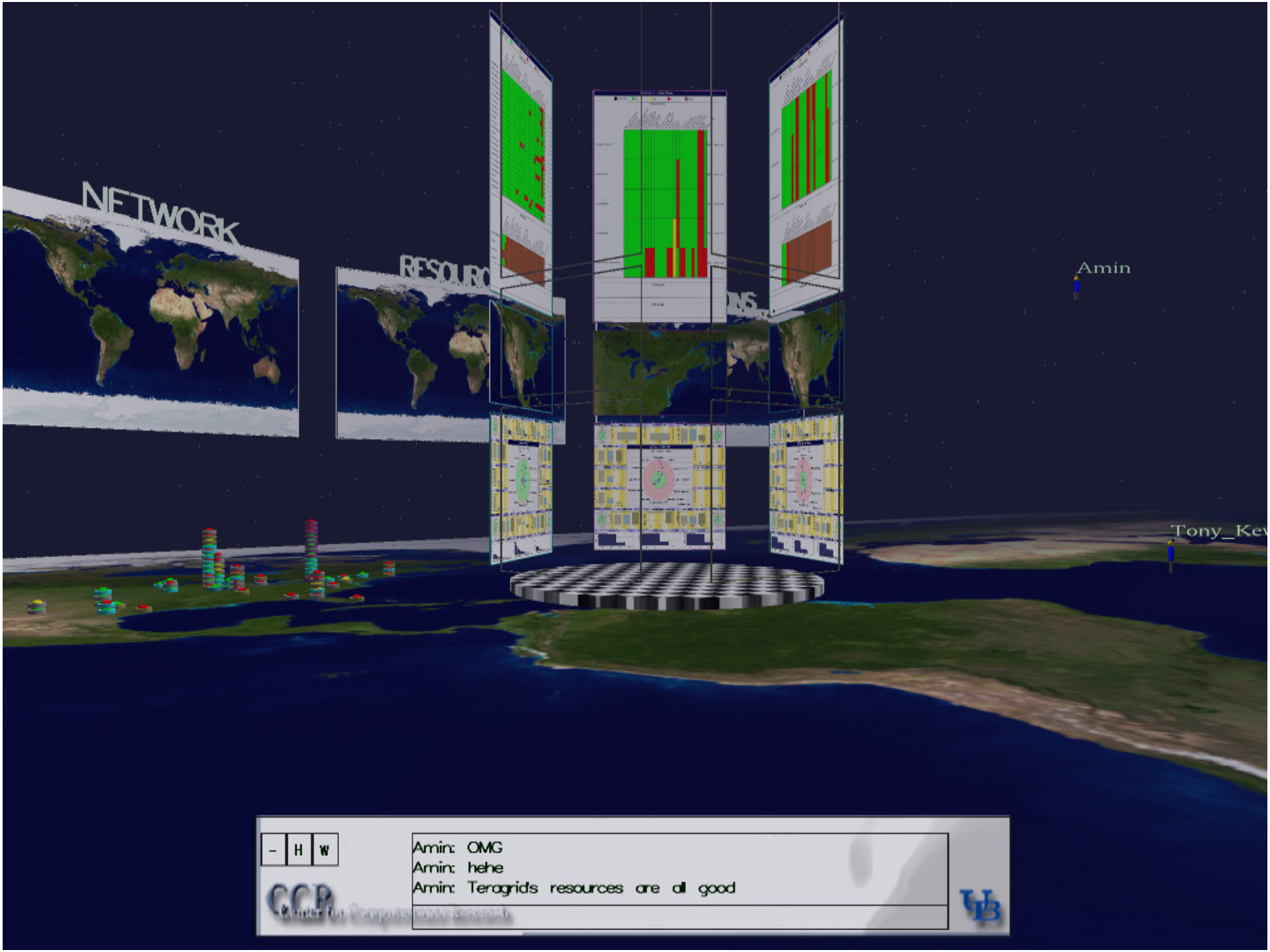


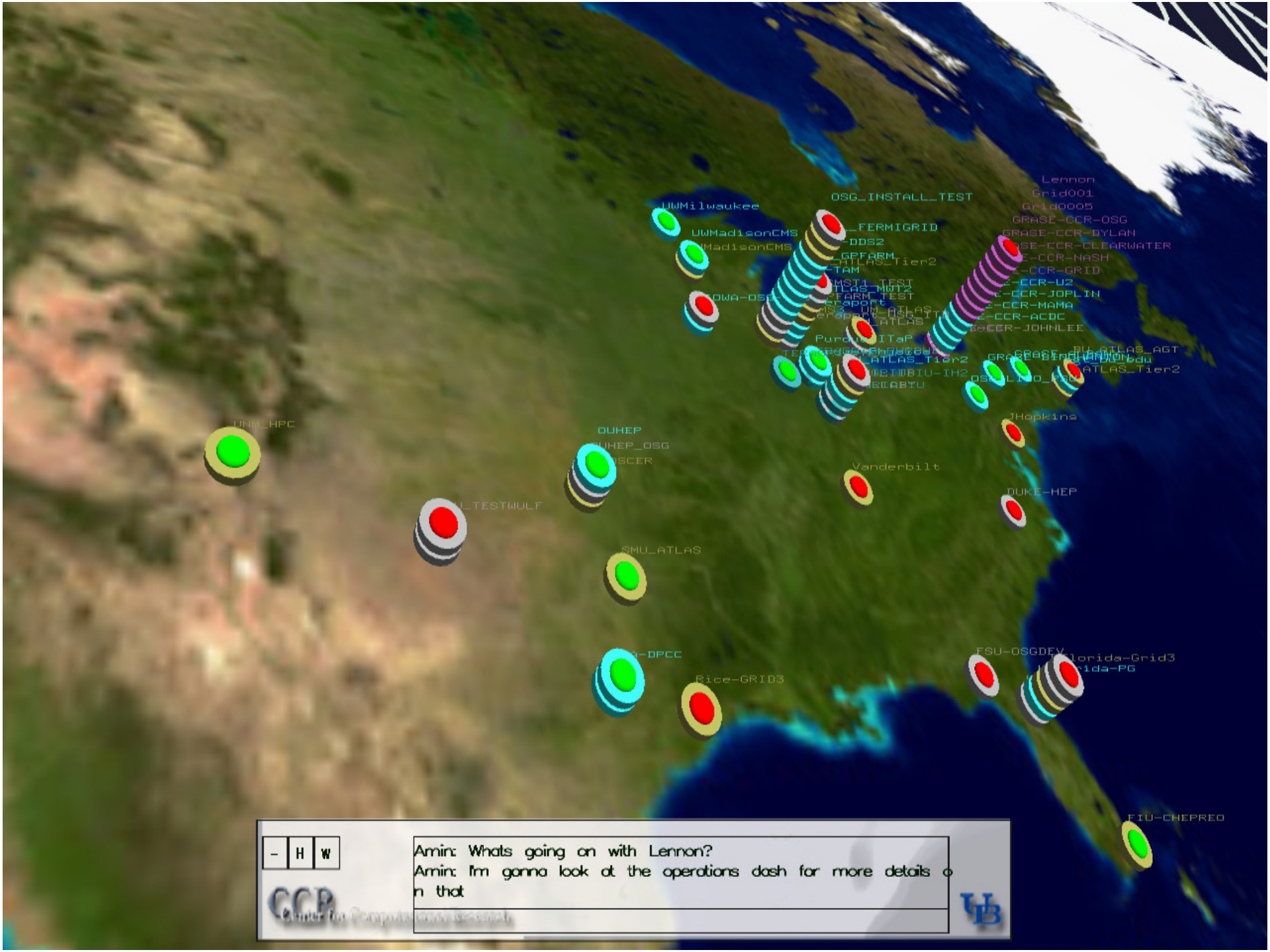
- Visualization of high volume, real-time Grid monitoring data
- Aid to administrators and users of the Grid




GRID3 Running Jobs
samgrid
uscms01
znhuang
llee
GRID3 Queued Jobs
800
700
600
500
400
300
200
100
0
samgrid
llee
zguo
jcarmer
uscms02
GRID3 Number Free CPUs
1400
1200
1000
800
600
400
200
0
Total CPUs
Total Free
Max Free
GRID3 Grid Health
High
Normal
Low
Queued Jobs
Running Jobs
Free CPUs
Gatekeeper Load
Bandwidth Utilization
Gatekeeper Users
Grid FTP Transfers
Gatekeeper Processes
Fork Job Production
Available Disk
Batch Job Production
Disk Consumption Rate
GRID3 Mean Grid Bandwidth [MB/s]
5
4
3
2
1
0
15 Min
60 Min
1 Day
GRID3 Daily Resource Load [GB]
14
12
10
8
6
4
2
0
BU_ATLAS_Tier2
GRID3 Number of Transfers
250
200
150
100
50
0
1 Day
Max (resource)
Avg (resource)
GRID3 Available Disk [GB]
Data
GRID3 Disk Consumption Rate [GB/day]
0.006
0.007
0.006
0.005
0.004
GRID3 Total Disk Space [GB]
8000
6000
4000
2000
Tmp
Minutes
Users
Processes
Resource CPU
GRID3 Predicted Utilization
Amin: hehe
Amin: Teragrids resources are all good
Amin: Samgrid has 800 jobs queued guys

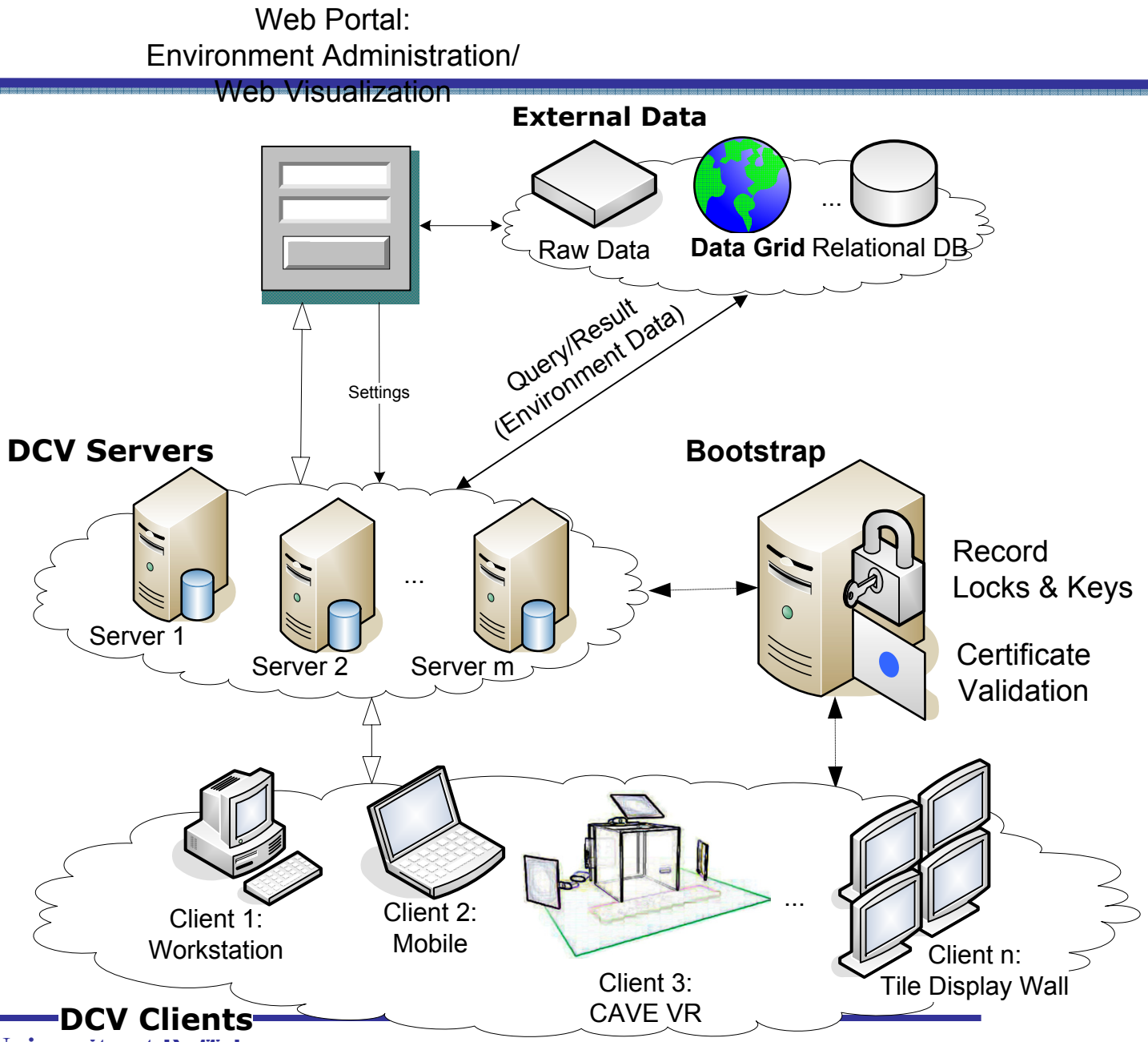

NETWORK
OPERATIONS
Amin: OMG
Amin: hehe
Amin: Teragrids resources are all good

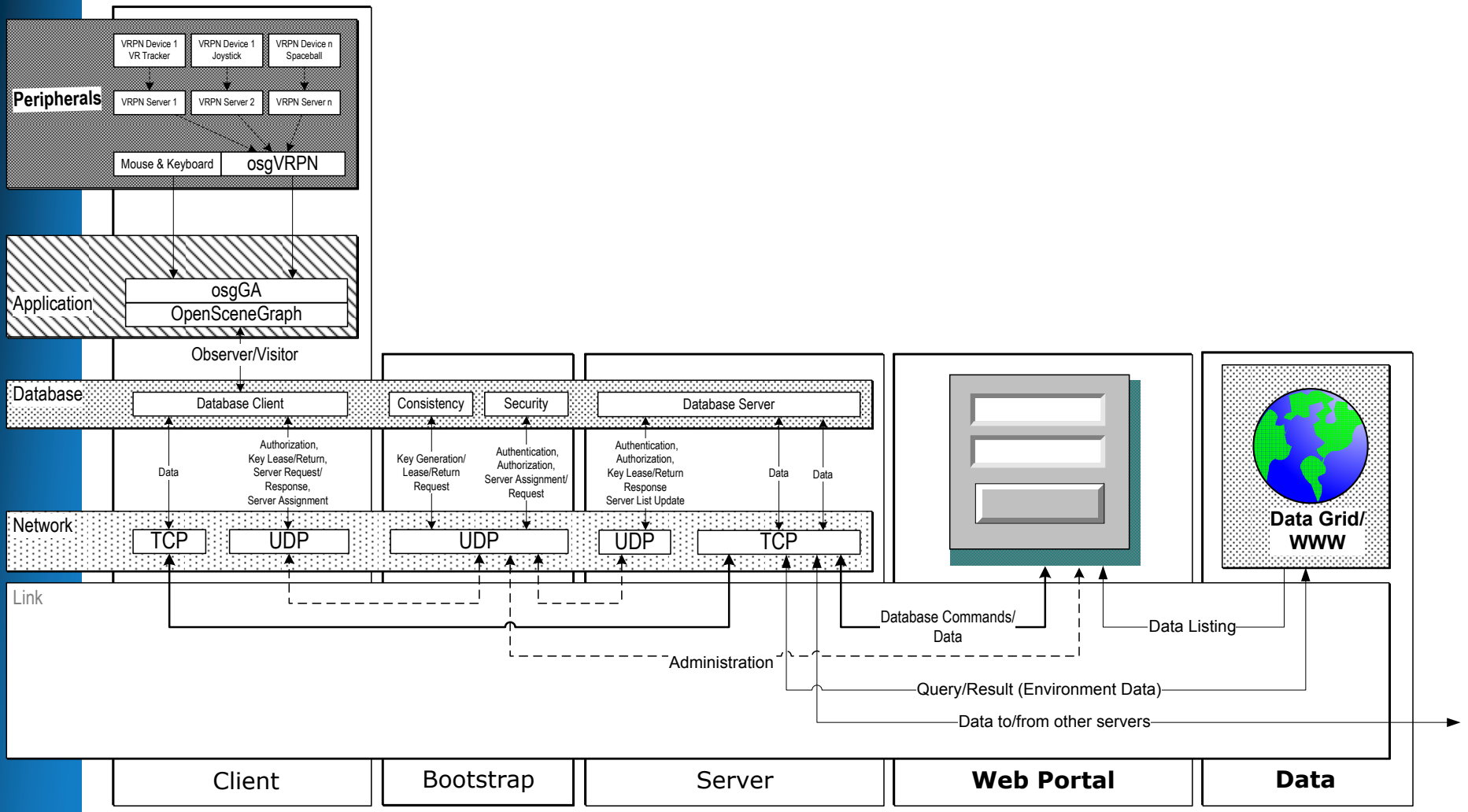

NETWORK
Amin
Tony_Kev
- H W
Amin: OMG
Amin: hehe
Amin: Teragrids resources are all good


UWMilwaukee
OSG_INSTALL_TEST
UWMadisonCMS
FERMIGRID
DDS2
GPFARM
TAM
Lennon
Grid001
Grid0005
GRASE-CCR-OSG
GRASE-CCR-DYLAN
GRASE-CCR-CLEARWATER
GRASE-CCR-NASH
GRASE-CCR-GRID
GRASE-CCR-U2
GRASE-CCR-JOPLIN
GRASE-CCR-MAMA
GRASE-CCR-ACDC
GRASE-CCR-JOHNLEE
ATLAS_Tier2
OUHEP
OUHEP_OSG
Vanderbilt
JHopkins
DUKE-HEP
SMU_ATLAS
Bice-GRID3
FSU-OSGDEV
FIU-CHEPREO
Amin: Whats going on with Lennon?
Amin: Im gonna look at the operations dash for more details o
n that

Summary:

- Enable distributed collaboration at both the system and user level via a three tier software architecture developed at CCR (Center for Computational Research http://www.ccr.buffalo.edu)
- Enable visualization and interaction with data across a geographically disparate network topology
- Integrate multiple data sources:
  - Scientific
  - Multimedia
  - Real-time
- Research and Development:
  - Real-time distributed databases
  - OpenGL 3D programming
  - 3D Modeling
  - Character animation
  - User interaction
  - Virtual Reality (CAVE VR)


Web Portal:
Environment Administration/
Web Visualization
External Data
Raw Data
Data Grid
Relational DB
...
Query/Result
(Environment Data)
Settings
DCV Servers
Server 1
Server 2
...
Server m
Bootstrap
Record
Locks & Keys
Certificate
Validation
Client 1:
Workstation
Client 2:
Mobile
Client 3:
CAVE VR
...
Client n:
Tile Display Wall
DCV Clients


VRPN Device 1
VR Tracker
VRPN Device 1
Joystick
VRPN Device n
Spaceball
Peripherals
VRPN Server 1
VRPN Server 2
VRPN Server n
Mouse & Keyboard
osgVRPN
Application
osgGA
OpenSceneGraph
Observer/Visitor
Database
Database Client
Consistency
Security
Database Server
Data
Authorization,
Key Lease/Return,
Server Request/
Response,
Server Assignment
Key Generation/
Lease/Return
Request
Authentication,
Authorization,
Server Assignment/
Request
Authentication,
Authorization,
Key Lease/Return
Response
Server List Update
Data
Data
Network
TCP
UDP
UDP
UDP
TCP
Link
Data Grid/
WWW
Database Commands/
Data
Data Listing
Administration
Query/Result (Environment Data)
Data to/from other servers
Client
Bootstrap
Server
Web Portal
Data