# Perspective on Campus Usage and Needs: TTU Experience with Grid Computing

Alan Sill
Senior Scientist, TIGRE Project
High Performance Computing Center
Texas Tech University

# *Introduction, History*

- **TTU has had a working campus grid now for over 4 years**
  - *Based on Avaki technology initially (now Sybase)*
  - *Results from shared computations presented as far back as GGF8 (Kulish, Perez, Smith, etc.)*
  - *Expanded now to include campus Data Grid*
  - *Serves both shared lab and cycle-scavenging as well as cluster computer systems*
  - *SAS grid added last year (see talk this conf.)*

# *Characteristics of TechGrid (2002)*

- Conceived as a local cross-campus system for all of TTU
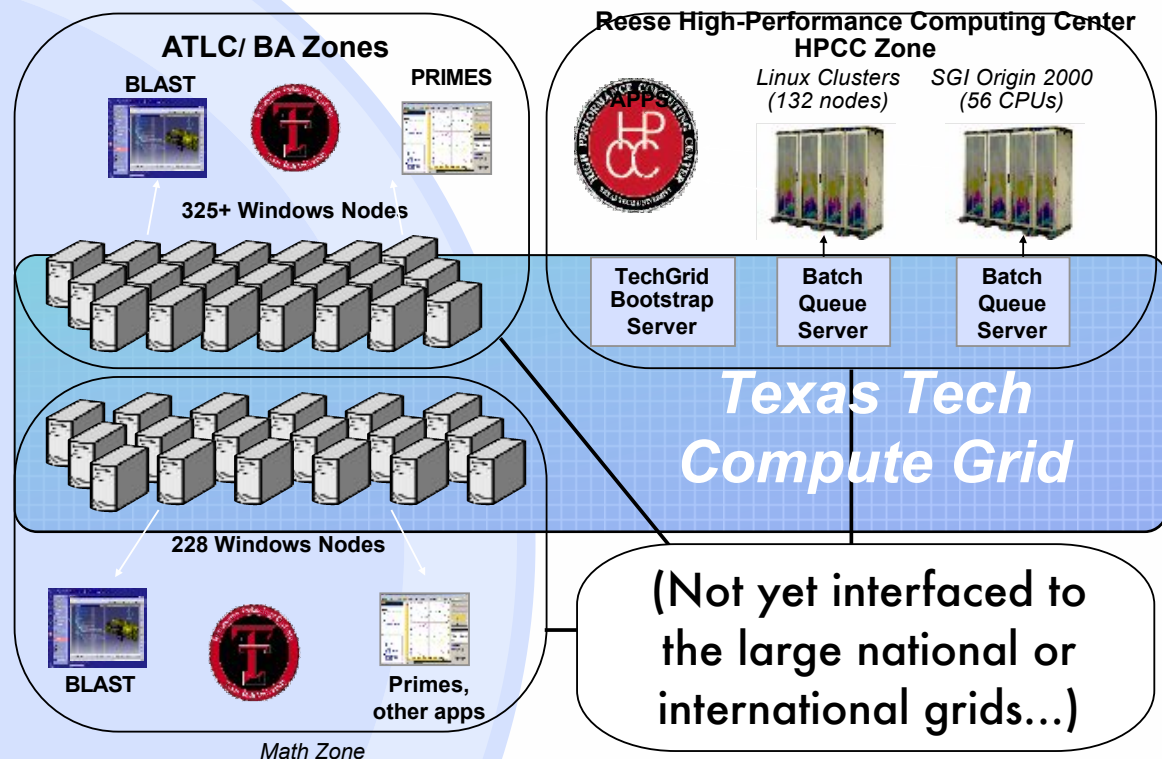- Both Linux and Windows resources included

**• Utilize spare cycles on 500 + CPUs across four sites, four administrative domains, firewalls.**

**• Variety of applications, including massively parallel apps.**

**• Capable of interfacing with a variety of queuing systems, including LSF (Platform), and PBS (via Rocks)**

**• Each Zone has different access control settings**

**• Each Zone has complete control over resources.**

**ATLC/ BA Zones**

BLAST  PRIMES

**325+ Windows Nodes**

**228 Windows Nodes**

BLAST  Primes, other apps

*Math Zone*

**Reese High-Performance Computing Center HPCC Zone**

*Linux Clusters (132 nodes)*  *SGI Origin 2000 (56 CPUs)*

APPSUTING

| TechGrid Bootstrap Server | Batch Queue Server | Batch Queue Server |

*Texas Tech Compute Grid*

(Not yet interfaced to the large national or international grids…)

(Perez et al., Internet2-2005 and Sybase 2005 conferences)

# *SAS Grid*

- Added 2 years ago and expanded greatly this past year for Business Statistics computation.

- Primarily based on Windows resources, but available also for Linux.

- Operating in a cycle-scavenging mode on BA computers.

- Requires a license for each machine; essentially a private grid.

- We do have it interfaced to the campus data grid – see next slide

# *Includes a Data Grid*

## Sybase technology – claims:

- **Provides access to a wide variety of data types cross hardware, operating systems, and system configurations**
- **Deploys on a domain by domain basis**
- **Expands and adapts to changing organizational needs by adding resources/services or interconnecting domains**
- **Provides for global sharing with local autonomy and control**
- **Continues to operate around point failures and provides automatic recovery**
- **Easy to deploy and maintain, highly reusable**

**TTU Campus Computers**

**HPCC at Reese Center**
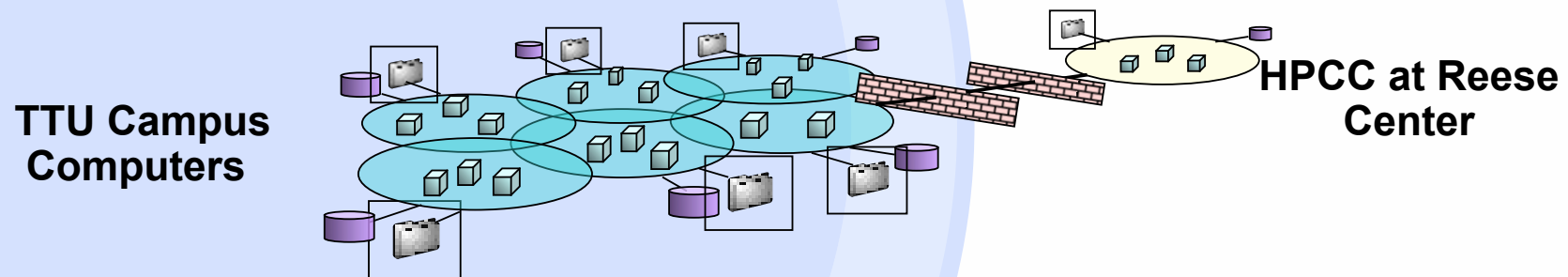
Our experience has been that this works within the cycle-scavenging grids, but we have not yet tried to extend it to the needs of our "data hungry" users.

# *Recent Large Scale Grid Work*

- **TTU also recently joined and helped to launch the Open Science Grid; & participate extensively in it**
  - *Different work than the previous campus grid, have not yet integrated these*
  - *More standards-based than some of our commercial grid deployments.*
  - *Designed to support some of our high-end, more data driven users*
  - *TTU participating in both Integration Test and Production grids for OSG*

- **Helped to form and charter member of The Americas Grid Policy Management Authority (TAG PMA).**

- **Participating in regional development (SURAGrid), state-wide grid building (TIGRE, THEGrid), and some international collaboration (NCHC/Taiwan, Shanghai, etc.) on specific topics of mutual interest.**

# *Open Science Grid (2005)*

- Standards-based.

- Open participation based on VOs.

- Interoperation activity with EGEE, TeraGrid.

- Oriented toward interconnection of large-scale clusters and resources (our interest).



Open Science Grid
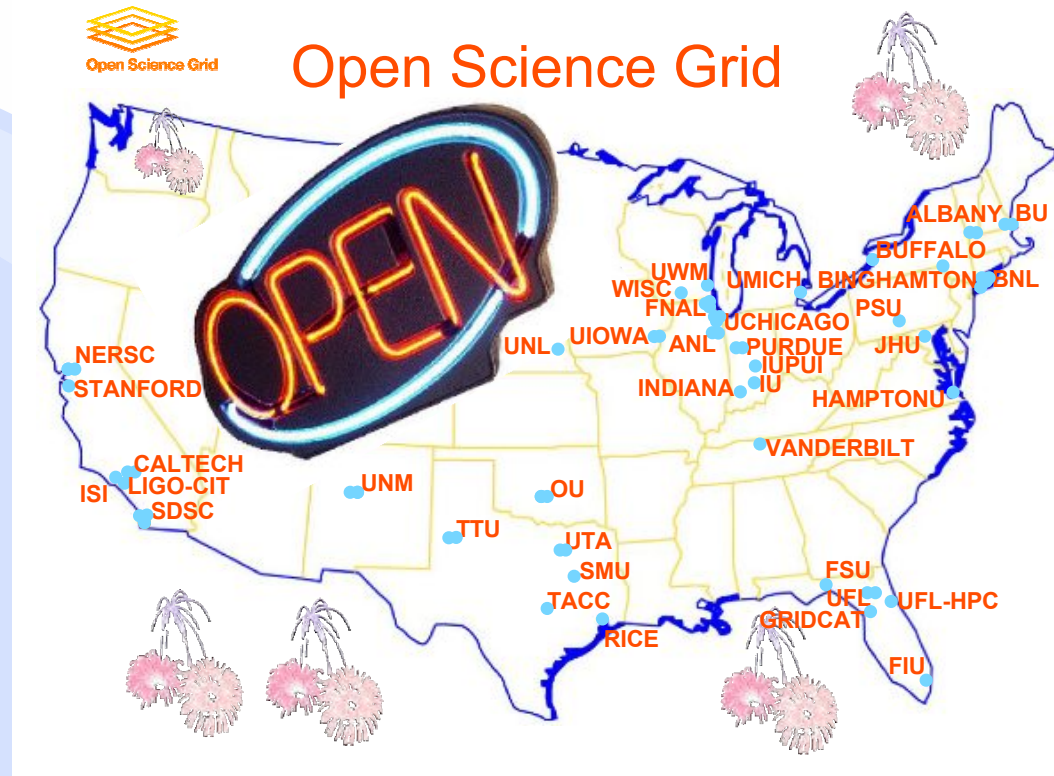
We think this is part of the solution for our high energy physicists, computational chemists, astronomers and other researchers for whom a large-scale virtual organization can be identified.

# High Performance Computing Across Texas (HiPCAT)

- Has been in operation for many years as a cooperative and collaborative organization for high performance computing centers throughout the state.

- Official projects started for special purposes:
  - *TIGRE - Texas Internet Grid for Research and Education - a cross-disciplinary, general purpose construction project to build a general grid*
  - *THEGrid - Texas High Energy Grid - complementary attempt to build a user community for HEP, nuclear physics, astrophysics and astronomy users*
  - *CDLT - Collaborative and Distance Learning Tools group*

- Partnering with cooperating LEARN project - Lone Star Education and Research Network
  - *Separately funded partner project to build high-speed fiber optic network between institutions.*

# *Some Observations*

- Although we have been following what we think is a well-organized and carefuly chosen path toward grid computing methods, a surprising diversity has developed.

- Have not covered here yet some very important considerations for authentication, authorization and interoperability that we think will play a major role in choosing our path to the future.

- Our users are just getting "good" at the grid => life promises to be very interesting in the near-term future as many more people join our efforts.

# *Some Conclusions*

- **Campus grids are diverse in user needs & must serve many masters.**
  - *What works for one set of users may not work for another - adapt and verify.*

- **While some expert communities are developing, not all users are there!**
  - *Must develop methods on many fronts.*
  - *Training still essential; also simplify software.*
  - *Advancement and diversity must continue.*