Condor Team 2010

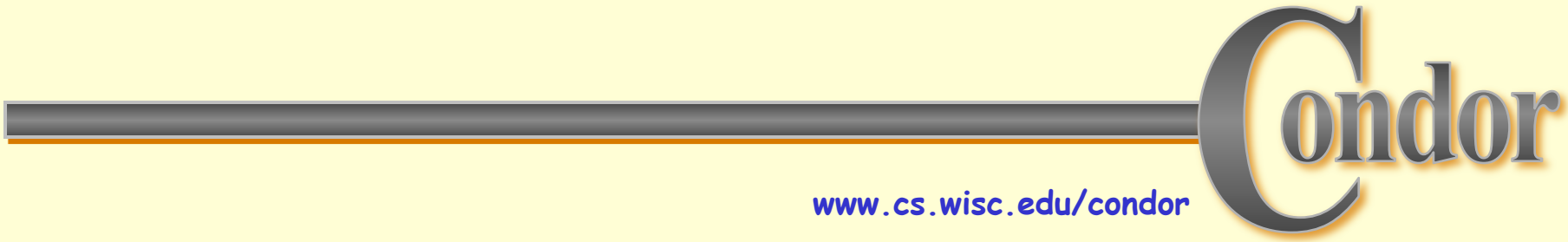
Established 1985

Welcome to Condor Week #12

(year #27 for our project)

Welcome to the 2nd Annual Condor meeting!!!

(Hope to see you all in the EuroGlobus+Condor meeting in southern Italy in mid June)



www.cs.wisc.edu/condor

Challenges Ahead

- Ride the "Grid Wave" without losing our balance
- Leverage the talent and expertise of our new faculty (Distributed I/O, Distributed Scheduling, Networking, Security)
- Expend our UW-Flock to a state-wide system (WiscNet?)
- Apply the Master Worker paradigm to domain decomposition problems (ongoing work with JPL)
- Scale our Master Worker framework to 10,000 workers.
- Open Source vs. Public Domain binaries vs. a Commercial version of Condor

Two new Institutes on the UW Campus - MIR & WID

Our Center for High Throughput Computing (CHTC) is an integral part of the vision and the operation of these two institutions

Providing Scientists with State Of The Art Cyber-Infrastructure Through Leadership in High Throughput Computing (HTC)

Cyber-Infrastructure = Hardware + Software + People

Leadership through novel concepts, frameworks and software technologies that are based on distributed computing principals (the what) and experimental Computer Science methodologies (the how)

06/27/97

This month, NCSA's (National Center for Supercomputing Applications) Advanced Computing Group (ACG) will begin testing Condor, a software system developed at the University of Wisconsin that promises to expand computing capabilities through efficient capture of cycles on idle machines. The software, operating within an HTC (High Throughput Computing) rather than a traditional HPC (High Performance Computing) paradigm, organizes machines into clusters, called pools, or collections of clusters called flocks, that can exchange resources. Condor then hunts for idle workstations to run jobs. When the owner resumes computing, Condor migrates the job to another machine.

To learn more about recent Condor developments, HPCwire interviewed Miron Livny, professor of Computer Science, University of Wisconsin at Madison and principal investigator for the Condor Project.

Why HTC?

For many experimental scientists, scientific progress and quality of research are strongly linked to computing throughput. In other words, they are less concerned about instantaneous computing power. Instead, what matters to them is the amount of computing they can harness over a month or a year --- they measure computing power in units of scenarios per day, wind patterns per week, instructions sets per month, or crystal configurations per year.

High Throughput Computing is a 24-7-365 activity

FLOPY ≠ *(60*60*24*7*52)*FLOPS*

The Exacycle Visiting Faculty Research Program is intended to give researchers access to very large amounts of CPU in a high-throughput computing environment. The program is focused on large-scale, CPU-bound batch computations in research areas such as biomedicine, energy, finance, entertainment, and agriculture, amongst others. For example, projects developing large-scale genomic search and alignment, massively scaled Monte Carlo simulations, and sky survey image analysis could be an ideal fit.

It is designed to match well to Google's compute infrastructure. To scale their applications, researchers must ensure their job can be partitioned into many small "work units" (typically tens of millions) each of which must fit in 1Gbyte of RAM, not exceed 1 CPU hour (wall clock time), and use very little disk IO (typically, no more than 5Gbytes of input and output data per work unit). If Exacycle provides support for external access through a well-defined API, Condor could support job submission to Exacycle as a backend.

David Konerding Senior Engineer and Tech Lead of Exacycle

GCC 2010 : The 9th International Conference on Grid and Cloud Computing

1-5 November 2010
Southeast University,
Nanjing, China.

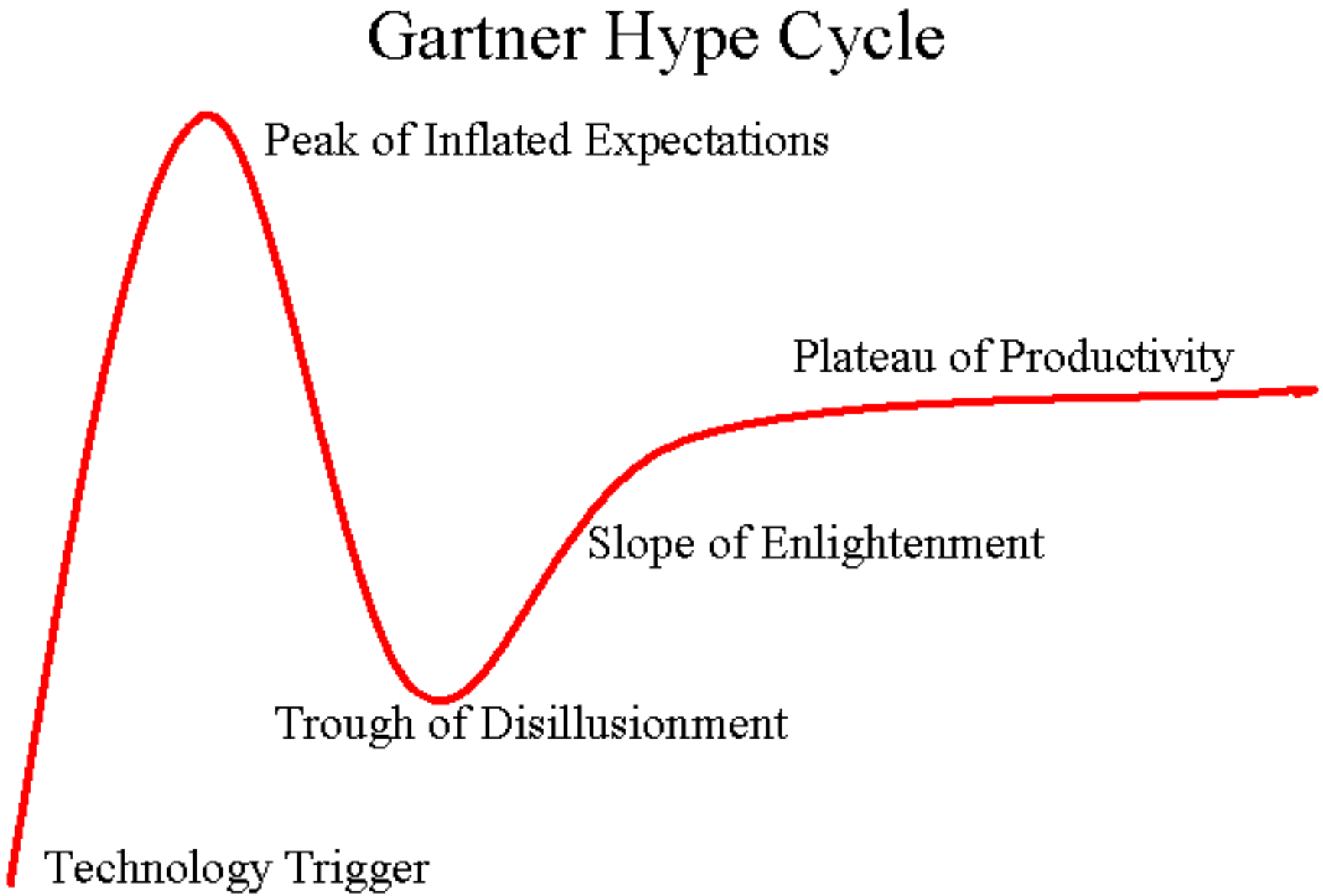
The words of Koheleth son of David, king in Jerusalem ~ 200 A.D.

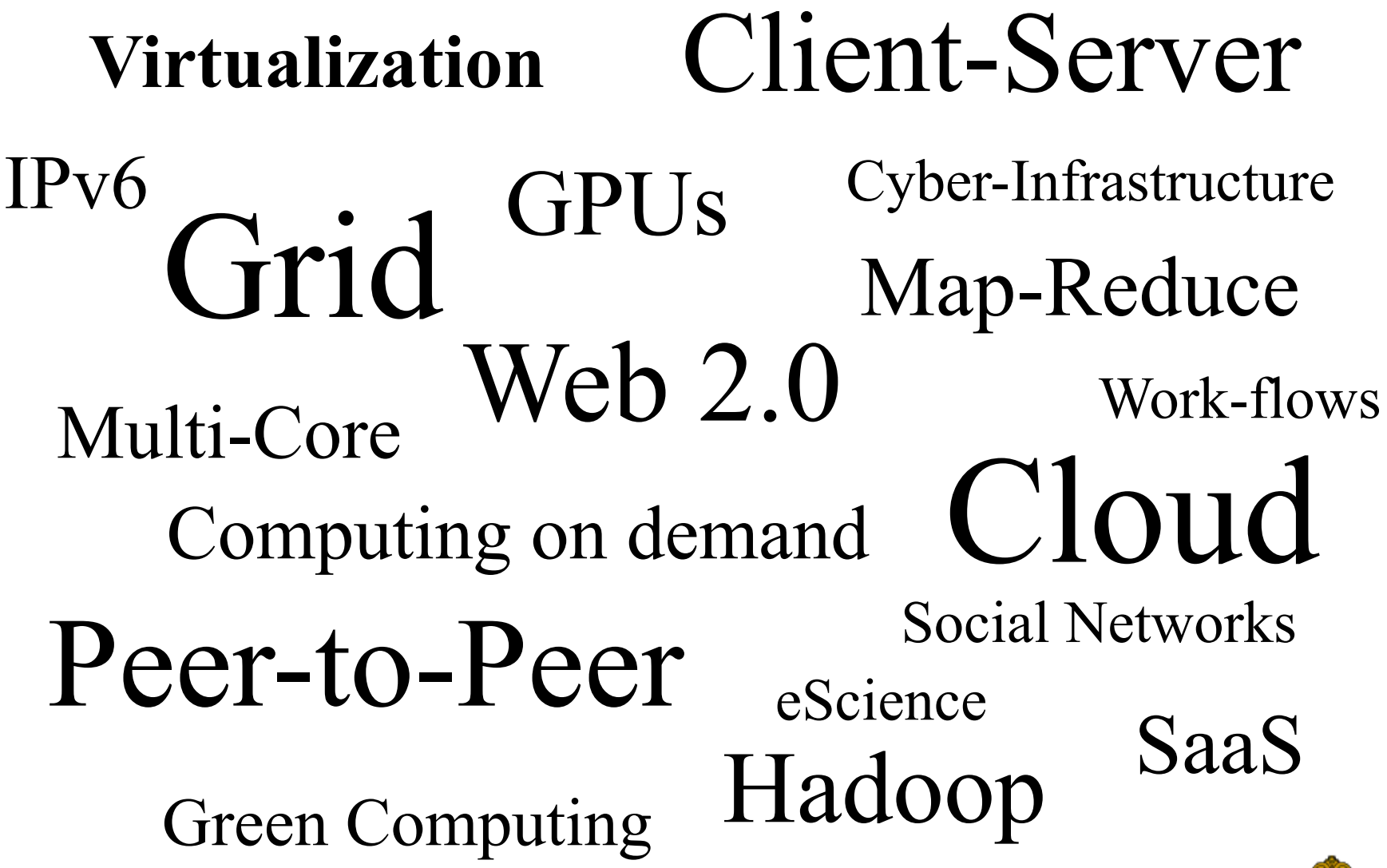
Only that shall happen
Which has happened,
Only that occur
Which has occurred;
There is nothing new
Beneath the sun!

Ecclesiastes, (קֹהֶלֶת, *Kohelet*, "son of David, and king in Jerusalem" alias Solomon, Wood engraving Gustave Doré (1832–1883)

Ecclesiastes Chapter 1 verse 9

The ups and down of Hype


Gartner Hype Cycle
Peak of Inflated Expectations
Plateau of Productivity
Slope of Enlightenment
Trough of Disillusionment
Technology Trigger


Virtualization
Client-Server
IPv6
Grid
GPUs
Cyber-Infrastructure
Map-Reduce
Web 2.0
Multi-Core
Work-flows
Cloud
Computing on demand
Social Networks
Peer-to-Peer
eScience
SaaS
Hadoop
Green Computing

But — and it's a big but — if we put those numbers on Gartner's own hype cycle, the industry will soon teeter at the "Peak of Inflated Expectations" (the highest point on Gartner's hype cycle new-technology adoption curve) And if the model proves true, 2015 looks like it may see a financial slide into the "Trough of Disillusionment" (the lowest point on the curve, directly following the high), **perhaps owing to persistent data breaches and the associated financial liability for interruptions in the cloud that prove beyond one's control**.

TotalCIO

A SearchCIO.com blog

Perspectives on Grid Computing

Uwe Schwiegelshohn Rosa M. Badia Marian Bubak Marco Danelutto
Schahram Dustdar Fabrizio Gagliardi Alfred Geiger Ladislav Hluchy
Dieter Kranzlmüller Erwin Laure Thierry Priol Alexander Reinefeld
Michael Resch Andreas Reuter Otto Rienhoff Thomas Rüter Peter Sloot
Domenico Talia Klaus Ullmann Ramin Yahyapour Gabriele von Voigt

We should not waste our time in redefining terms or key technologies: clusters, Grids, Clouds... What is in a name? Ian Foster recently quoted Miron Livny saying: **"I was doing Cloud computing way before people called it Grid computing"**, referring to the ground breaking Condor technology. It is the Grid scientific paradigm that counts!

Open Science Grid (OSG)
HTC at the National Level

The Institute for Distributed High Throughput Computing*

The proposed Institute for Distributed High Throughput Computing (InDHTC) brings together a diverse and accomplished group of computer and computational scientists who will *enhance and expand the impact of DHTC on DOE science through close interdisciplinary collaborations with the broader community that will research and formulate novel frameworks, develop advanced technologies, and build state-of-the-art software tools*. This effort will build upon the foundation established over the past 5 years by the Open Science Grid (OSG), expanding an existing network of interdisciplinary collaborations to cover the growing role that distributed computing across shared processing and storage resources plays in scientific discovery.

***submitted on 05/02/11 to DOE SciDAC-3**

Distributed High Throughput Computing

We define DHTC to be the shared utilization of autonomous resources toward a common goal, where all the elements are optimized for maximizing computational throughput. Sharing of such resources requires a framework of mutual trust and maximizing throughput requires dependable access to as much processing and storage capacity as possible. The inherent stress between the requirements for both trust and broad collaboration underpins the challenges that the DHTC community faces in developing frameworks and tools that translate the potential of large scale distributed computing into high throughput capabilities accessible by a diverse group of users ranging from international collaborations to single-PI research teams. The five teams of the InDHTC will address these challenges by developing a framework that is based on four underlying principles:

Subject: [Chtc-users] Daily CHTC OSG glidein usage 2011-05-04
From: condor@cm.chtc.wisc.edu
Date: Wed, 4 May 2011 00:15:02 -0500
To: chtc-users@cs.wisc.edu

Total Usage between 2011-05-03 and 2011-05-04

Group Usage Summary

User	Hours	Pct
1 Spalding	23943.7	78.5%
2 Chemistry	6190.7	20.3%
3 zhijun@submit.chtc.wisc.edu	374.5	1.2%
TOTAL	30509.0	100.0%

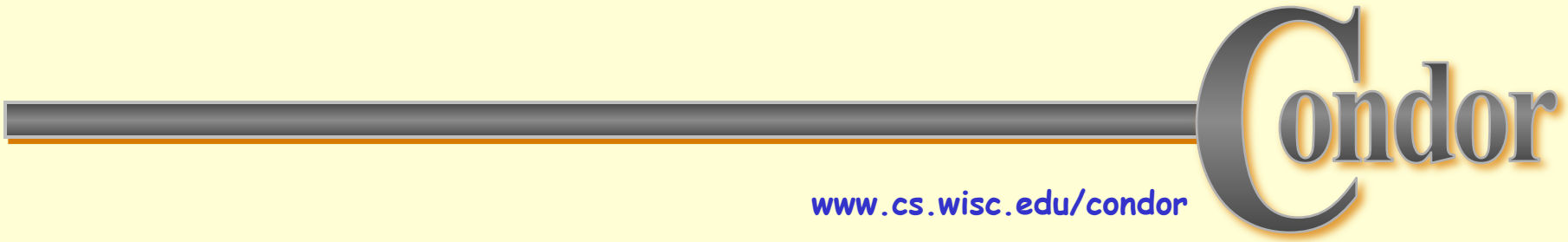
GRID WORKSHOP
and
GRID TUTORIAL

CHEP 2000

International Conference on
COMPUTING IN HIGH ENERGY
AND NUCLEAR PHYSICS
February 7 - February 11, 2000 - Padova, Italy

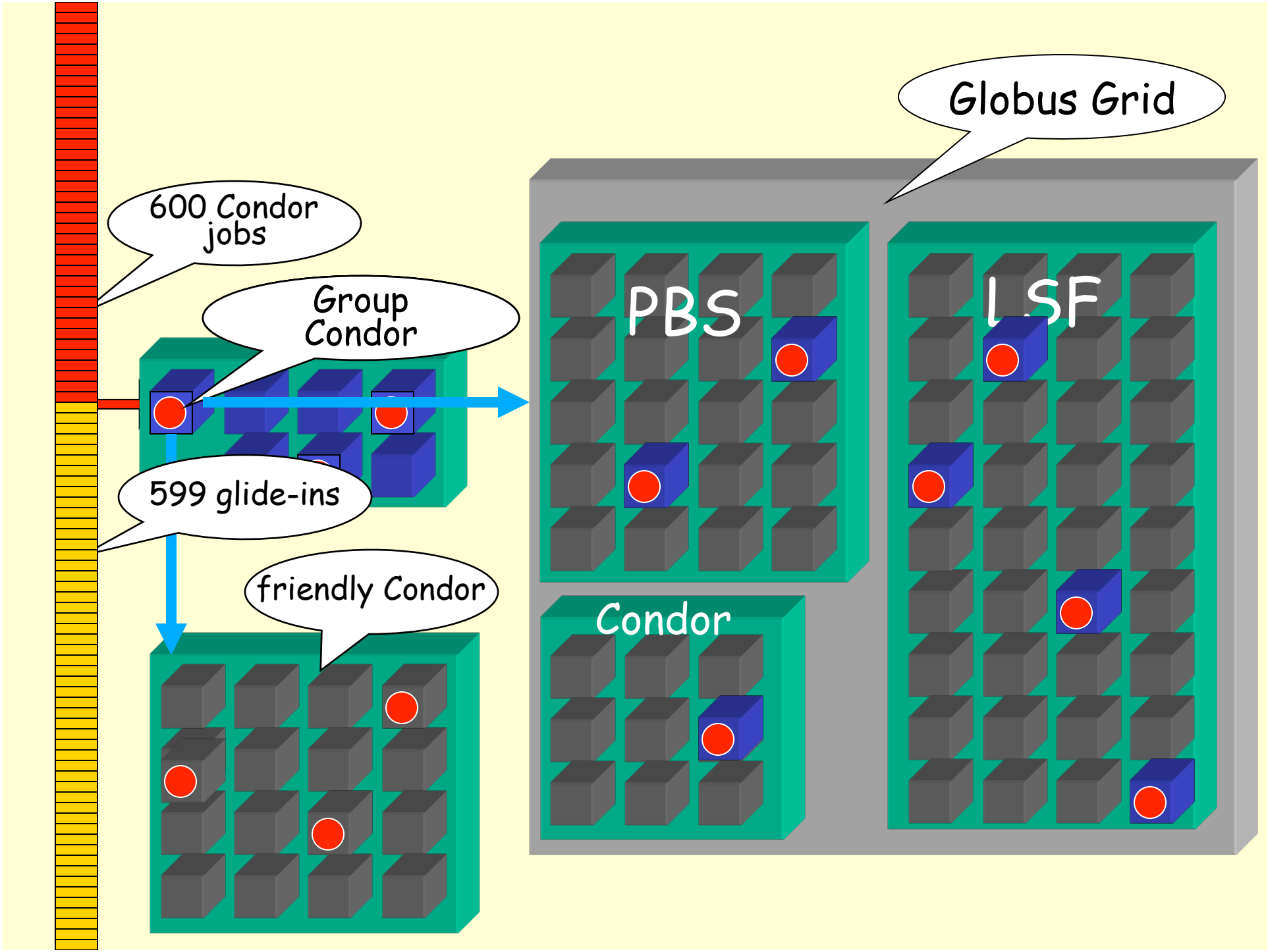
Step IV - Think big!

- Get access (account(s) + certificate(s)) to Globus managed Grid resources
- Submit 599 "To Globus" Condor glide-in jobs to your personal Condor
- When all your jobs are done, remove any pending glide-in jobs
- Take the rest of the afternoon off …



A "To-Globus" glide-in job will ...

- ... transform itself into a Globus job,
- submit itself to Globus managed Grid resource,
- be monitored by your personal Condor,
- once the Globus job is allocated a resource, it will use a GSIFTP server to fetch Condor agents, start them, and add the resource to your personal Condor,
- vacate the resource before it is revoked by the remote scheduler


Globus Grid
600 Condor jobs
Group Condor
599 glide-ins
friendly Condor
PBS
LSF
Condor

Thank you for building such

a wonderful HTC community