the globus alliance
www.globus.org

What Happens When Cloud Computing Meets HPC

Dr. Dan Fraser

Director, CDIGS

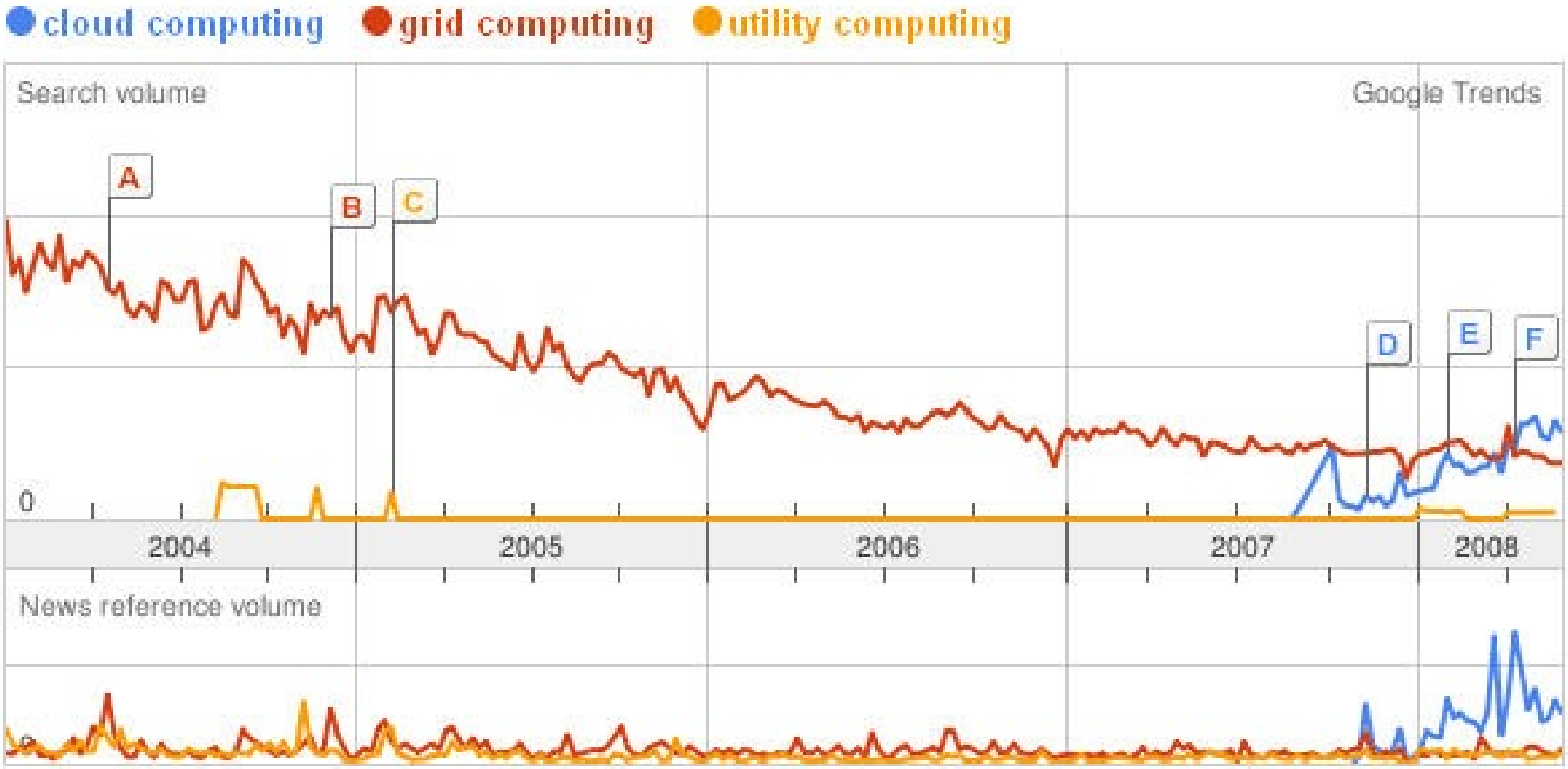
(Community Driven Improvement of Globus Software)

http://www.cdigs.org

Outline

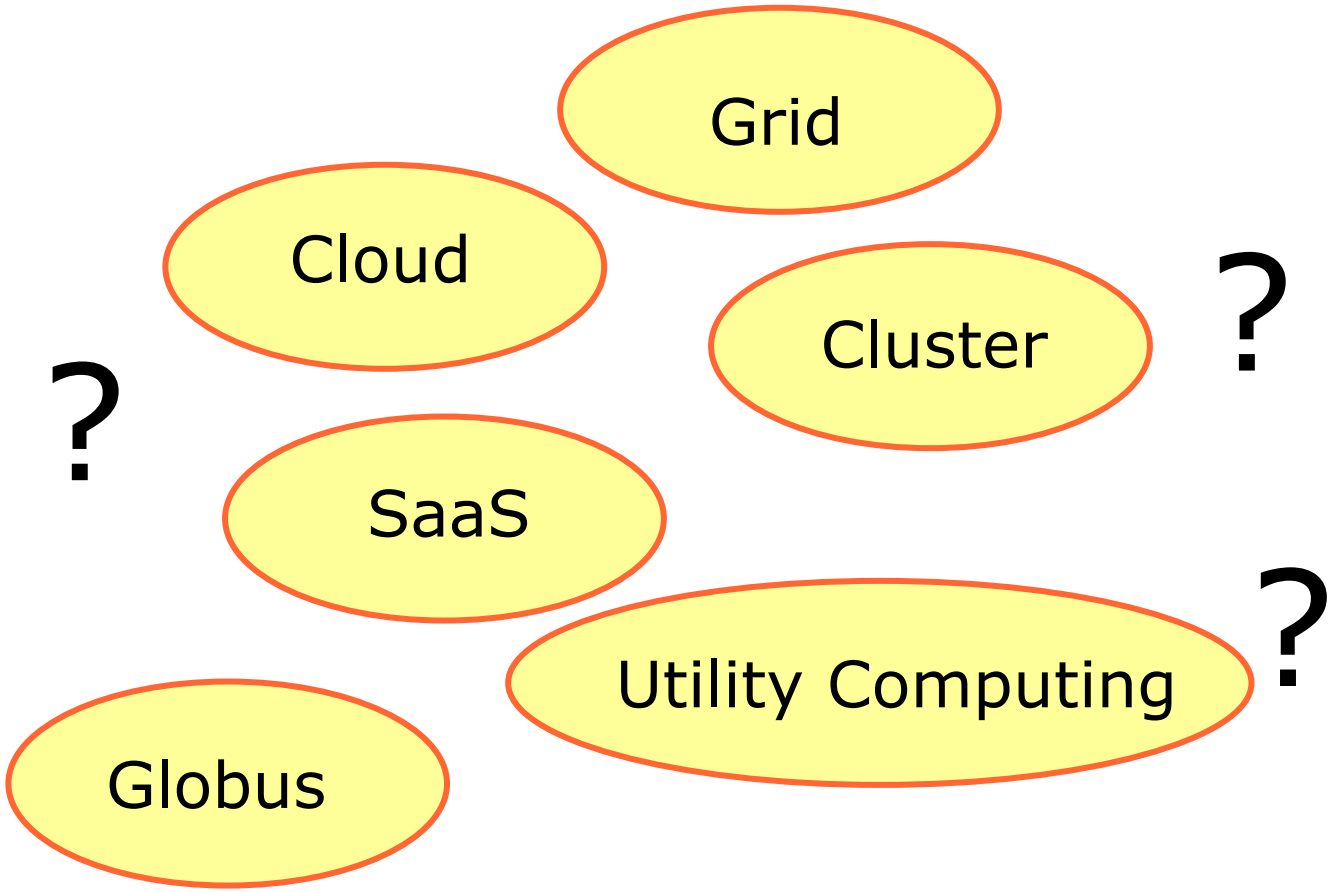
- Intro to Cloud Computing and Concepts
- Cloud Computing's Impact on HPC
- A Brief Look at Grid, Globus, and Clouds
  - Globus Incubator Program
  - Open Source EC2-like Capability
- Impact and Opportunity for Supercomputing Centers
- Dan's Head in the Clouds

"Cloud" Computing is ~1 yr old



Michael Sheehan's GoGrid Blog, July 25, 2008
http://linux.sys-con.com/node/587717

Sorting out the Pieces



SaaS = Software as a Service

One can categorize each component

Utility Computing

SaaS = Software as a Service

SaaS

Usage Model

BUT…

Cloud

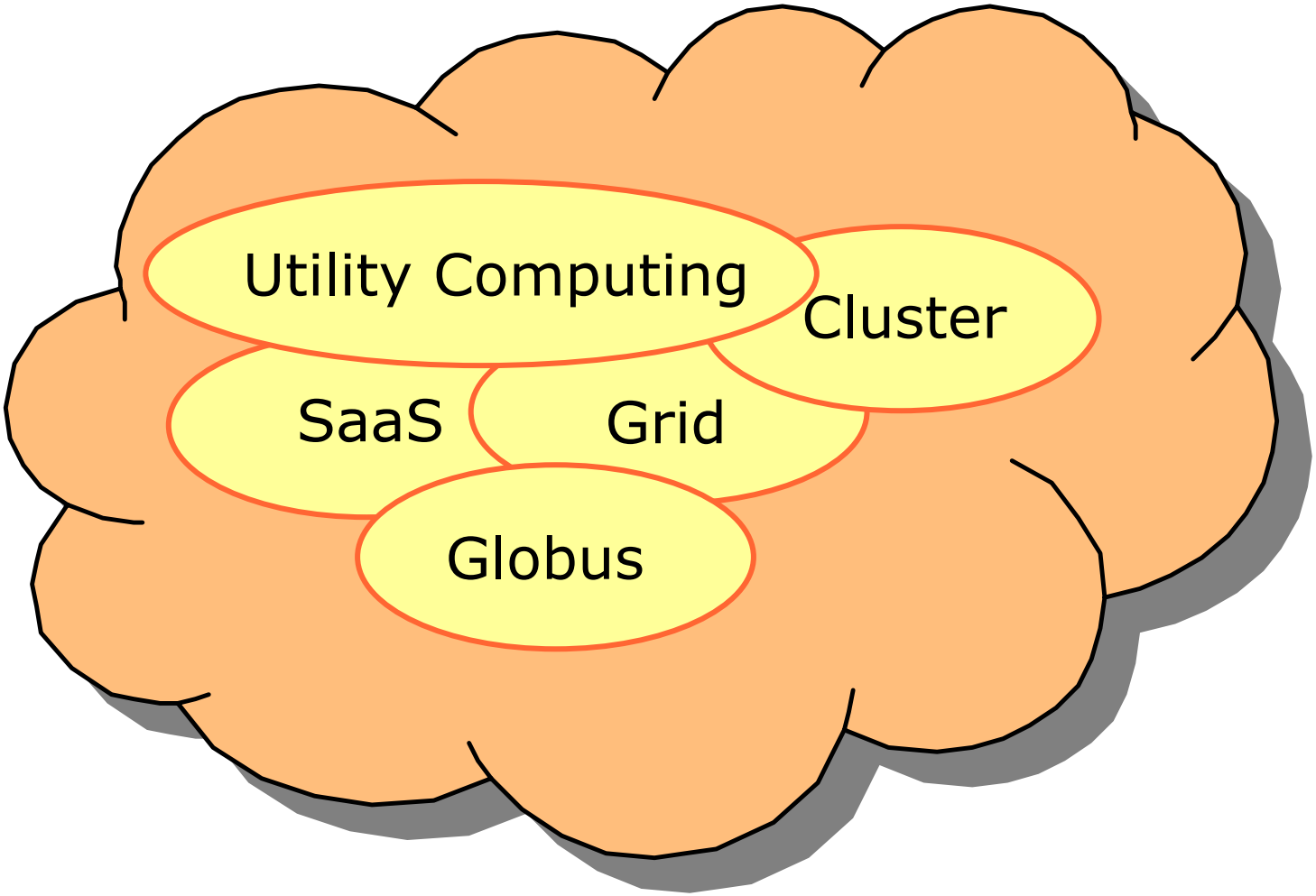
Cluster

Grid

Globus

Infrastructure

Clouds can have any/all of these



And the descriptions often overlap !

What makes a Cloud?

- Virtual Machines
- VM Manager (Amazon EC2, …)
  - Scalability
- File system Infrastructure
- Remote access (portal)
- Cost?
  - One reason the EC2 is successful is because of the low cost for cpu/data movement.
- Security?

Key Parts of Cloud Definition

Where is the value?

- Much of the value is in the Virtual Machines
- What are VMs used for?
  - Server Consolidation (Fermilab)
  - Disaster recovery (commercial)
  - Component Isolation (sandboxing)
  - Hardware Independence (any OS on any Box)
  - Cluster Computing
    - E.g. Deploy a classroom environment
    - E.g. Deploy a multi-use cluster with ROCKS
- Adding VM Management takes this to the "clouds"
  - Access resources on-demand
  - Isolate Users from each other
  - Schedule VM usage

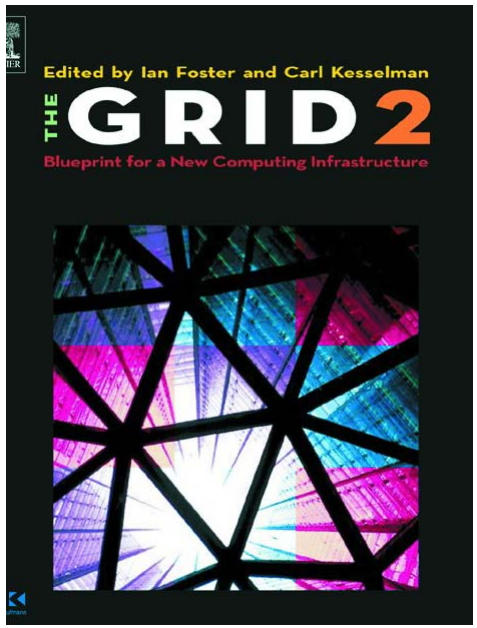
Where is the HPC value?

- Much of the value is in the Virtual Machines
- What are VMs used for?
  - X Server Consolidation (Fermilab)
  - X Disaster recovery (commercial)
  - X Component Isolation (sandboxing)
  - √ Hardware Independence (any OS on any Box)
  - √ Cluster Computing
    - E.g. Deploy a classroom environment
    - E.g. Deploy a multi-use cluster with ROCKS
- Adding VM Management takes this to the "clouds"
  - √ Access resources on-demand
  - √ Isolate Users from each other
  - √ Schedule VM usage

What is a Grid?

Enable *"coordinated resource sharing & problem solving in dynamic,* ***multi-institutional*** *virtual organizations."*

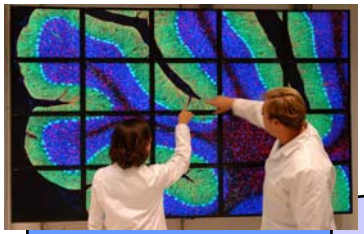
(Source: **"The Anatomy of the Grid"**)



What does Globus do?

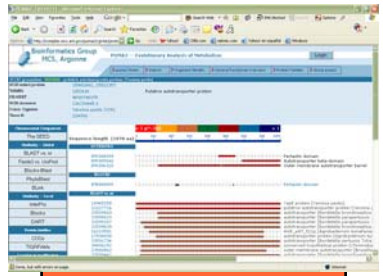
- Globus provides a
  - Secure…
  - Uniform Remote Job Submission Interface…
  - Plus numerous capabilities that make the environment "useful."
    - Data movement, Job monitoring, Service discovery, Security credential mgmt, Uniform data interfaces, …
- Many Globus components can be used as stand-alone software products
    - GridFTP, RLS, Index service, MyProxy

Creating a Useful Environment



Tool

User Application



Tool

Uniform interfaces, security mechanisms, Web service transport, monitoring

Workflow

User Svc

Host Env

Registry

Credent.

DAIS

GRAM

User Svc

Host Env

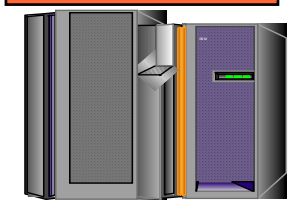
GridFTP

TeraGrid
EMPOWERING DISCOVERY

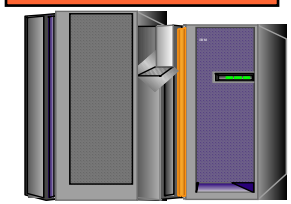
Open Science Grid

Computers

Specialized resource

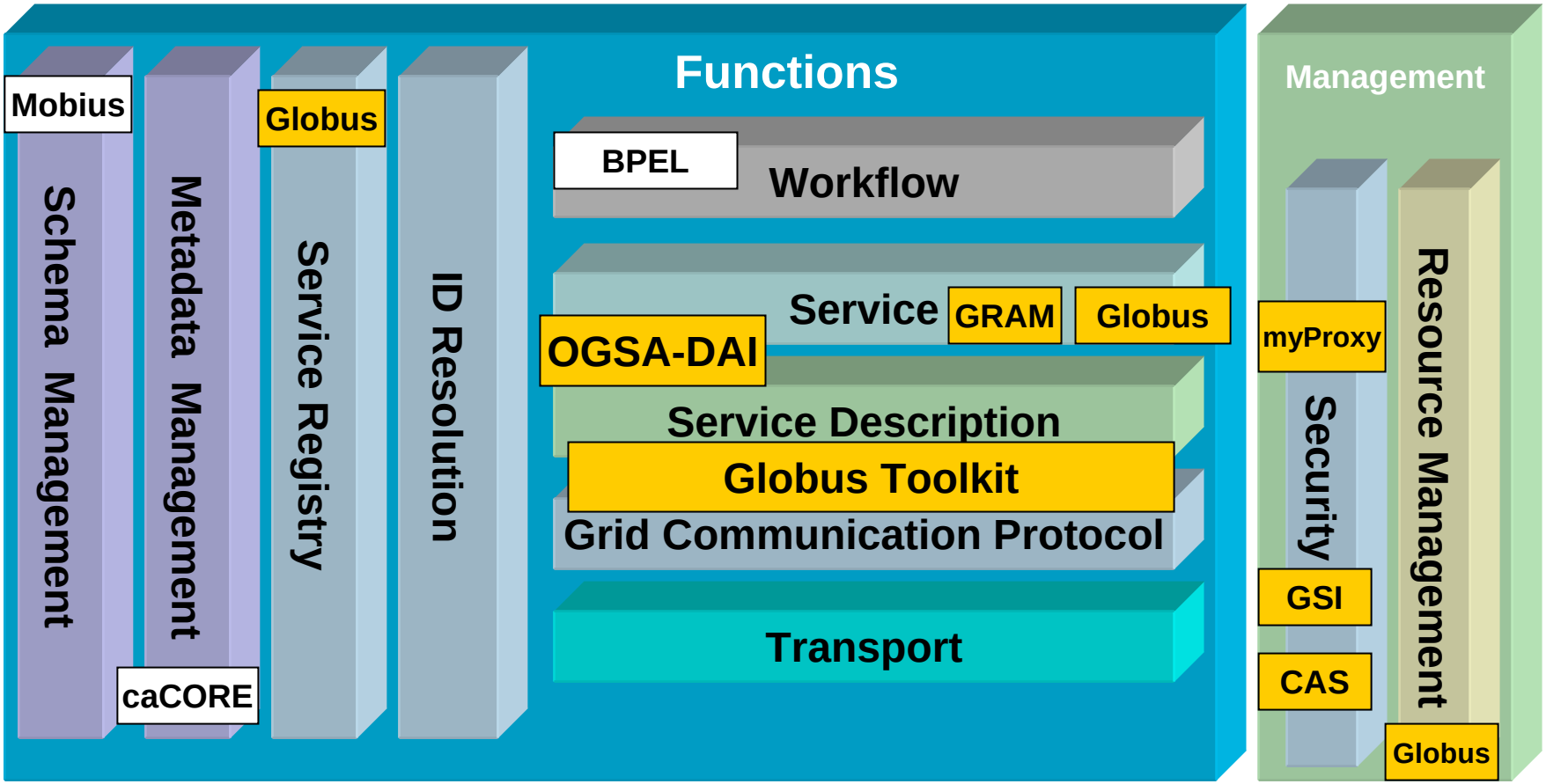


Storage



Database

Cancer Biomedical Informatics Grid

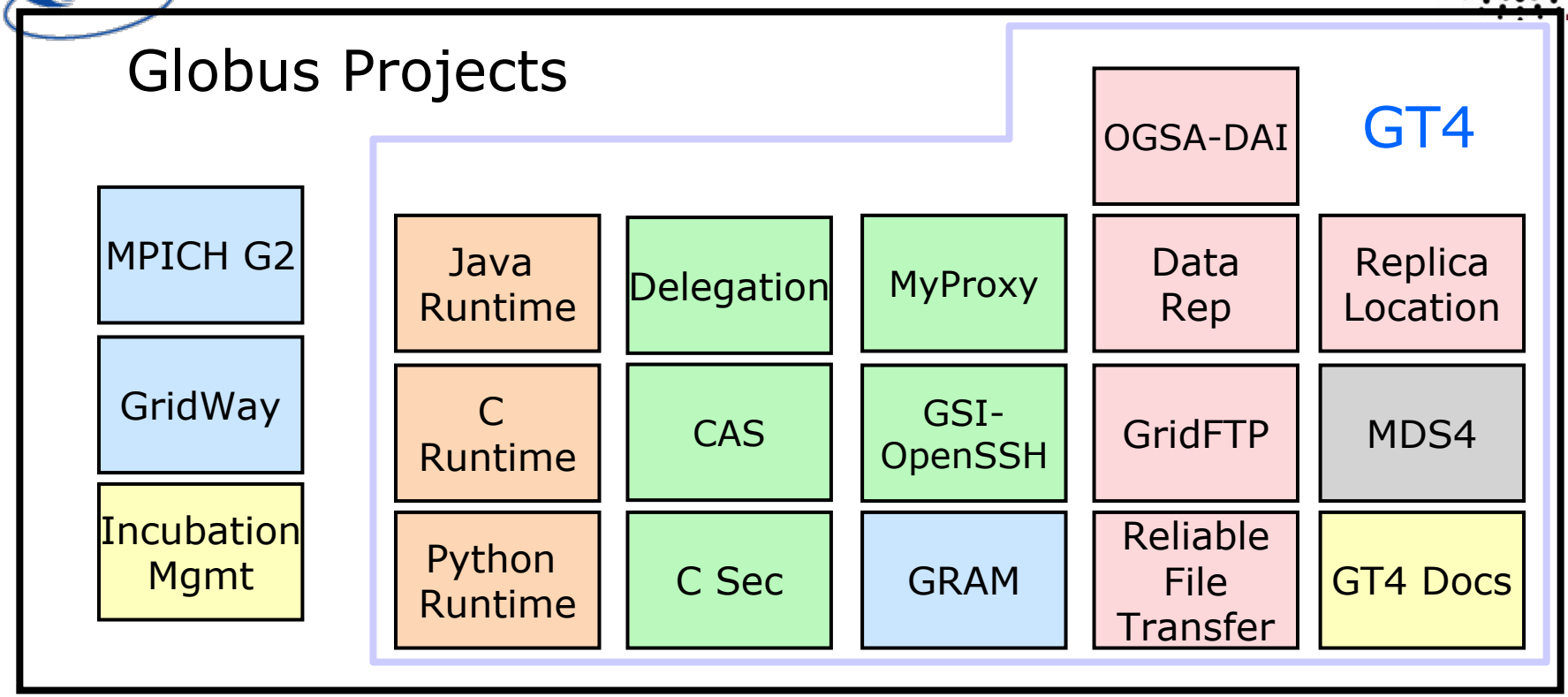


Spans 60 NIH cancer centers across the U.S.

caBIG™ cancer Biomedical Informatics Grid™ an initiative of the National Cancer Institute

Slide credit: Peter Covitz, National Institutes of Health

Globus Software: dev.globus.org



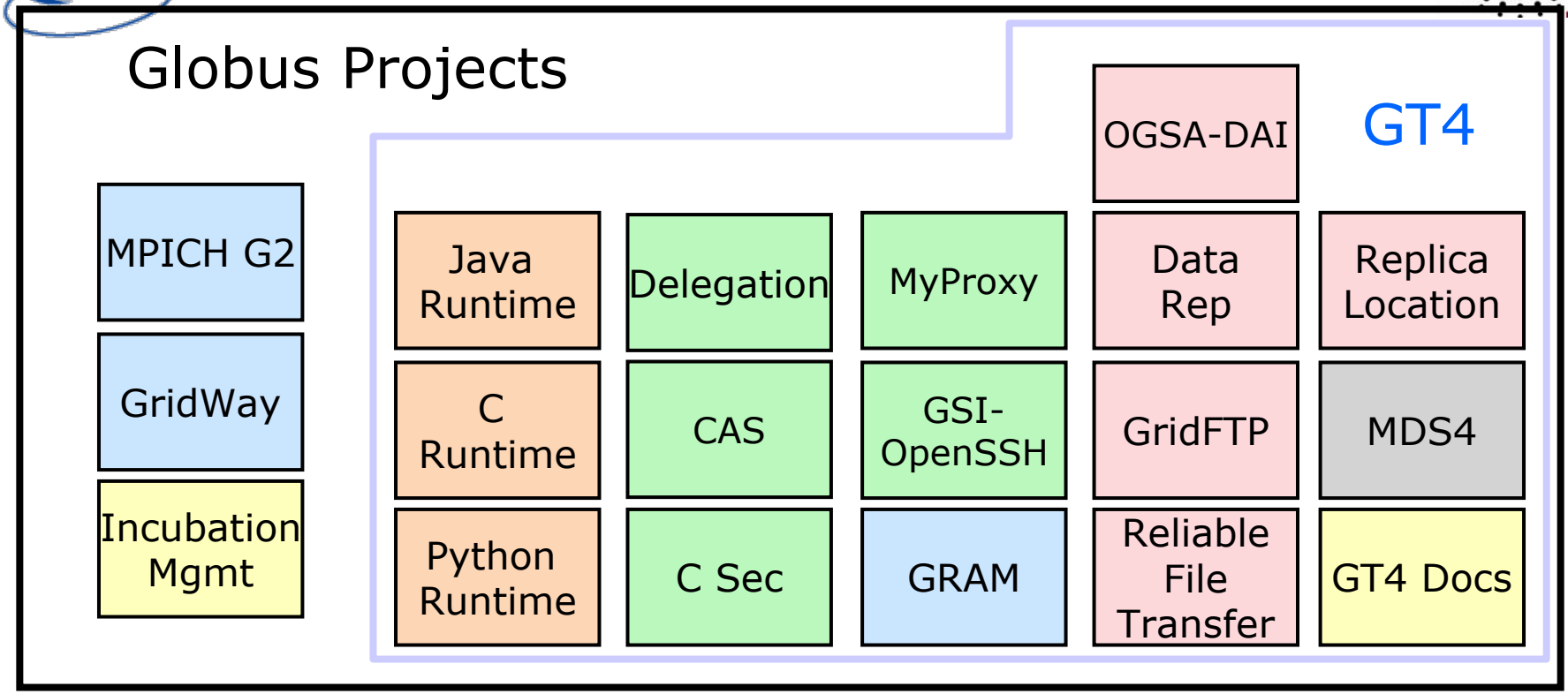
Incubator Projects

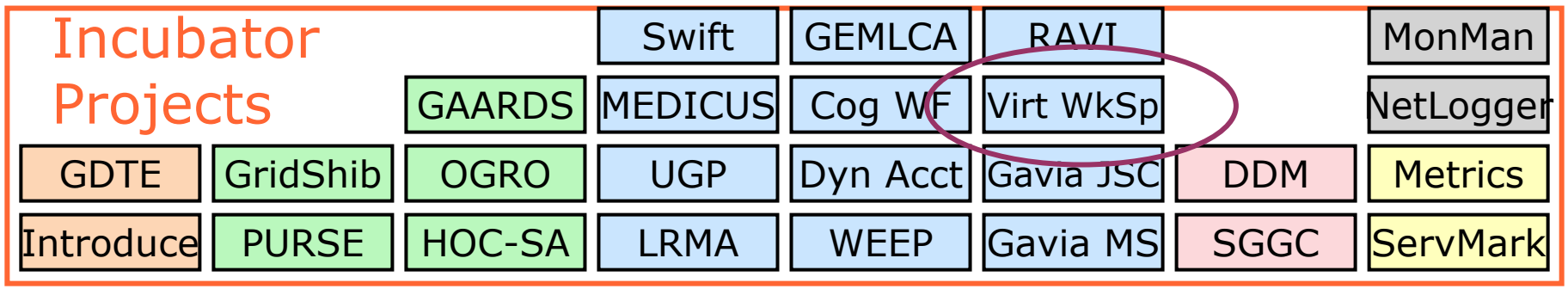
Common Runtime	Security	Execution Mgmt	Data Mgmt	Info Services	Other

Incubator Projects

- Contributed from teams around the world
  - Must utilize a "Globus" open source License
  - Code can be sold, used by others, adapted…
- Each project has its own "Committers"
  - Committers govern the project
- Globus Provides Infrastructure & Oversight
  - Project site, e-mail lists, some publicity
  - Overall project approval, & follow-up
- You can add your Incubator:
  - http://dev.globus.org/

Globus Software: dev.globus.org



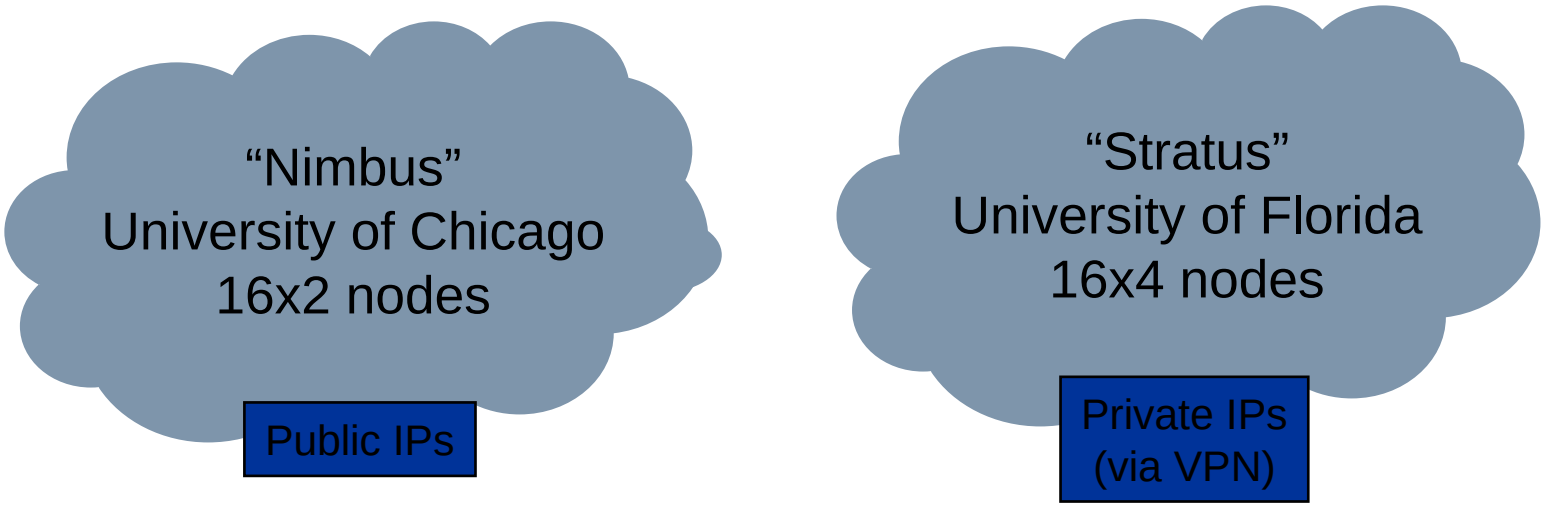


Common Runtime | **Security** | **Execution Mgmt** | **Data Mgmt** | **Info Services** | **Other**

Globus & Cloud Computing

- Virtual Workspaces is a Globus Incubator
- An Open Source EC2-like Management System
  - You can run on the cloud
  - You can even build your own cloud

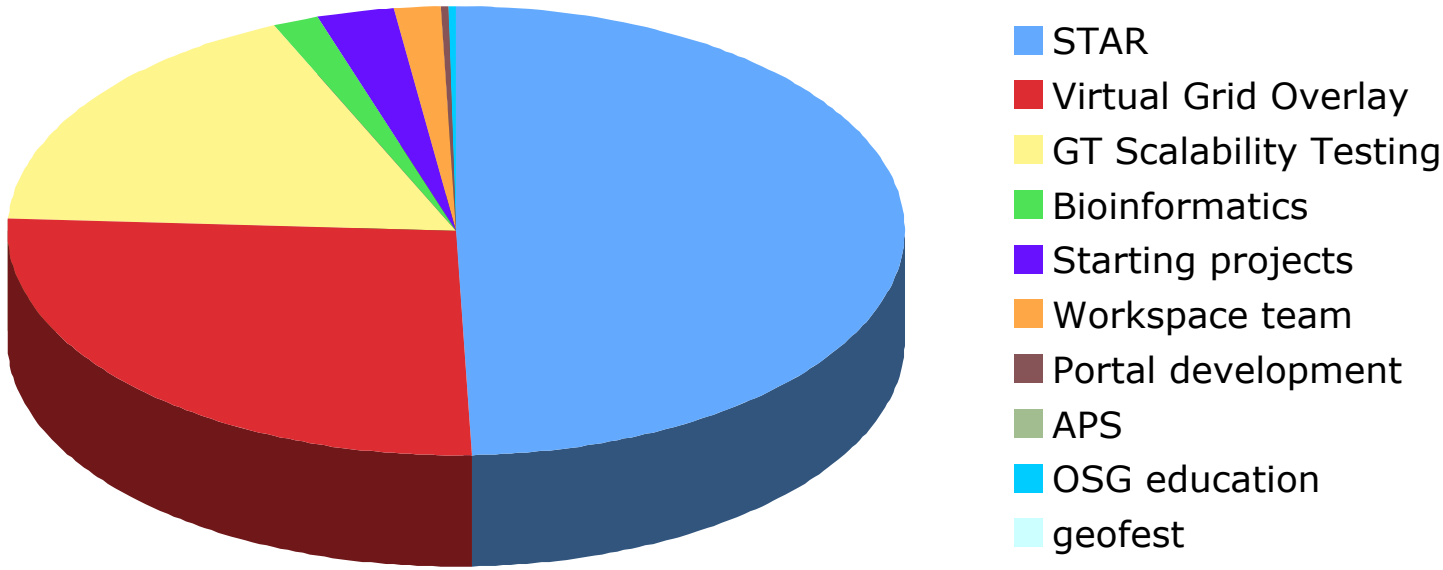
Science Clouds



- Powered by workspace tools
- EC2-like interfaces (PKI credential vs credit card)
- More clouds on the way
- http://workspace.globus.org/clouds

Who Runs on the Science Clouds?

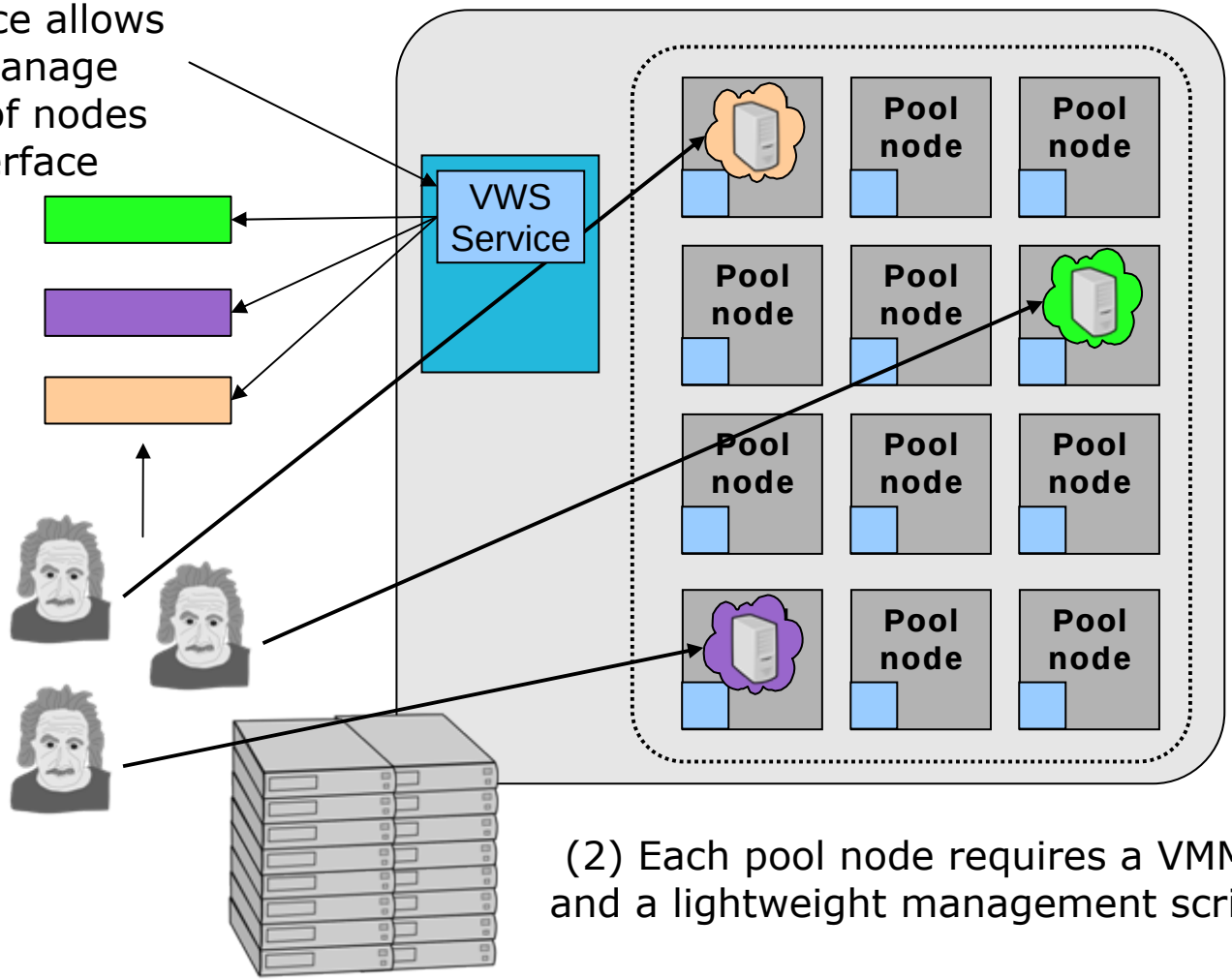
- Nimbus utilization breakdown since March 4th
- ~30 Communities



Interacting With Workspaces

(1) The workspace service allows users to deploy and manage workspaces on a pool of nodes through a WSRF interface

(3) Information on each workspace is published as WSRF Resource Properties ao that users can find out information about their workspace (e.g. what IP the workspace was bound to) or subscribe to notifications on changes



(2) Each pool node requires a VMM and a lightweight management script

STAR

- Motivation for STAR
  - Resources **with the right configuration** are hard to find
    - Complex environments: correct versions of operating systems, libraries, tools, etc all have to be installed.
    - Require validation
- Virtual Workspace: an OSG STAR cluster
  - OSG cluster
    - OSG CE (headnode), gridmapfiles, host certificates, NSF, PBS
  - STAR worker nodes: SL4 + STAR conf
- Requirements
  - One-click virtual clusters
  - Migration: nimbus/scientific resources -> EC2

STAR (cont.)

- From proof-of-concept to production runs
  - ~2 years ago: proof-of-concept
  - Last September: EC2 runs of up to 100 nodes (production scale)
  - Testing for full production deployment
- Performance
  - Within 10% of expected performance for applications
- Work by Jerome Lauret, Doug Olson, Leve Hajdu, Lidia Didenko
- Long-lived community of many
- Similar work for other HEP communities (Alice and Atlas), bioinformatics, geofest, and others

The Supercomputing Center "Threat"

- Grid computing provides uniform access to computational resources
  - Computational resources become commodities
  - Supercomputing Centers offer a variety of applications, libraries, and support
- Cloud Computing Makes Use of Virtual Machines where applications, libraries and dependencies can be hidden
  - Supercomputing Centers can become commodities in themselves
- Ok so "threat" may be a bit overstated
  - Problems don't go away quite so easily (shell game)
  - But shake-outs can/do happen along the way…

The Opportunity

- Be the Supercomputing Center that enables cloud computing!
  - (Gradually) turn the center into a big cloud
    - Today's clouds have only ~16 VMs
  - Conduct Research in VMs, VM Management, and VM Maintenance
  - Develop Tools to make Cloud Computing accessible to the scientists
  - Become the center of HPC Cloud expertise

So what happens when HPC meets Cloud computing?

We don't really know

because

the possibilities are just now emerging!

Dan's Head in the Clouds

- What if… scientists could:
  - Download and use a VM that would make it easy to parallelize their application;
  - And test it in parallel right on their laptop.
- What if… scientists could:
  - Run a converter to change one VM type to another;
  - Or enable a VM created at one center to automatically run other places even though the infrastructure may be different (VMWare, Xen, RPATH, …)
- What if… scientists could:
  - Select applications and components from a list;
  - Select some of their own applications;
  - Push a button to create a cluster-ready VM image;
  - Then push another button to automatically deploy them.
- And the list goes on …

Conclusion

- HPC cloud computing is an emerging technology
- There are big opportunities for leadership to develop in this space.
- Using VMs is only the beginning. There must also be collections of tools for managing and maintaining VMs …