

Apache *airavata*



Science Gateway Architectures: Open Source, Distributed Computing

Marlon Pierce, Suresh Marru

Science Gateway Group, Indiana University

marpierc@iu.edu, smarru@iu.edu



I590 Special Topics Course: Spring 2016

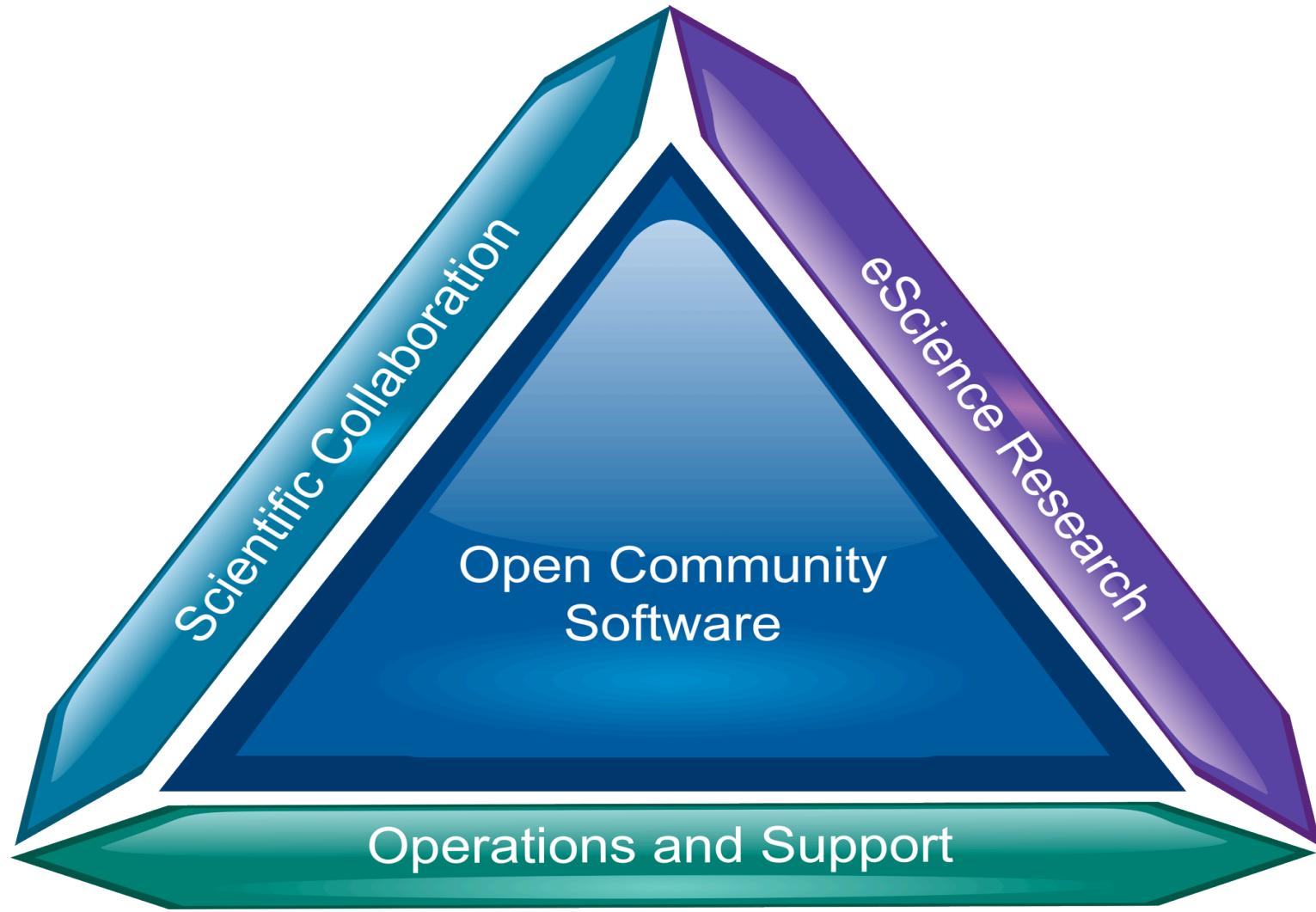
<http://homes.soic.indiana.edu/classes/spring2016/info/i590-marpierc>

Or

<http://s.apache.org/i590-spring2016>

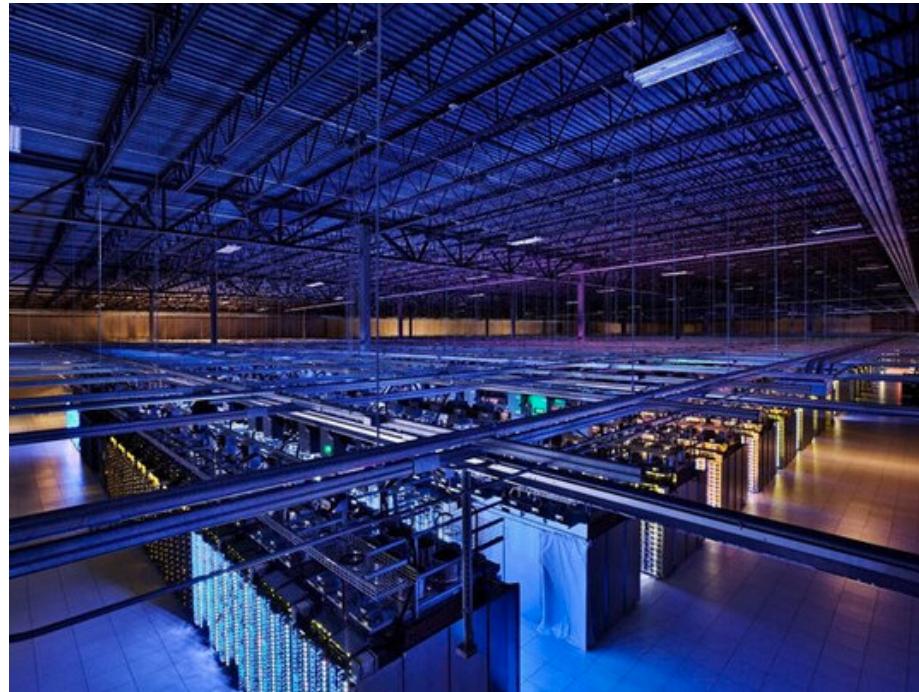


A Balancing Act



Science Gateways Group

SciGaP



Goal 1: teach basic distributed computing concepts by applying them to science gateways.



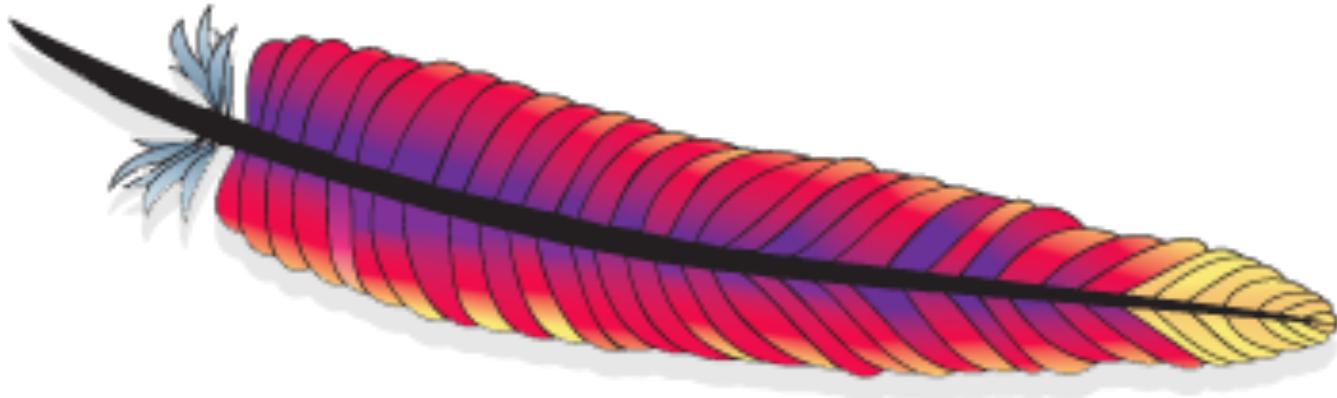
You Don't Choose Chaos Monkey...
Chaos Monkey Chooses You



@RealGeneKim, genek@realgenekim.me

Goal 2: explore new architectures,
methodologies, and technologies:
Microservices, DevOps





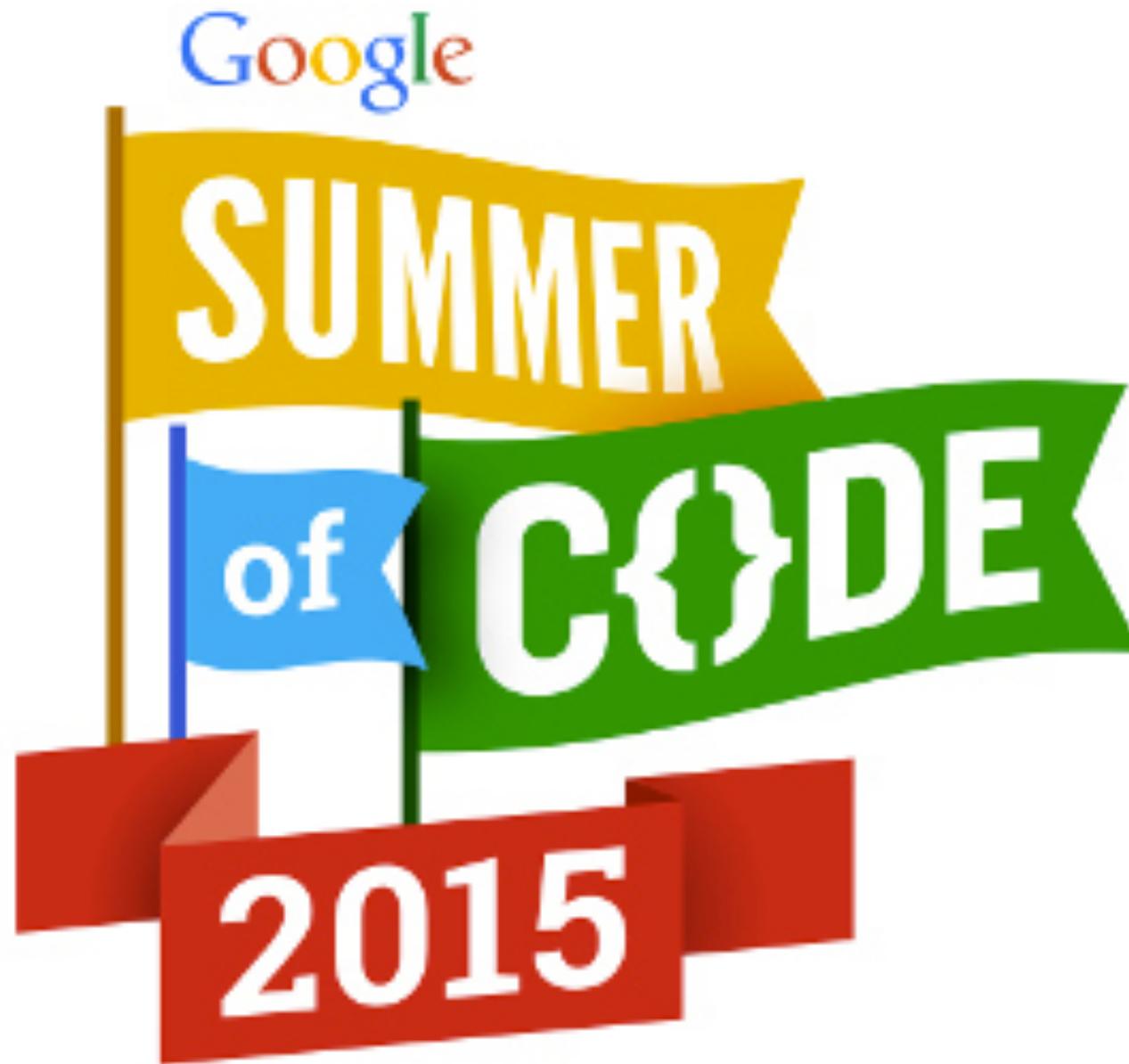
Goal 3: teach open source
software practices



Structure of the Course

- Students will be divided into development teams
- Each team will develop an open source science gateway system from scratch.
 - Biweekly assignments
 - Grades entirely from projects
- Each assignment will illustrate one or more distributed computing topics.
- Students will learn how to operate services as well as write good software.
- Students will learn open source methodologies.





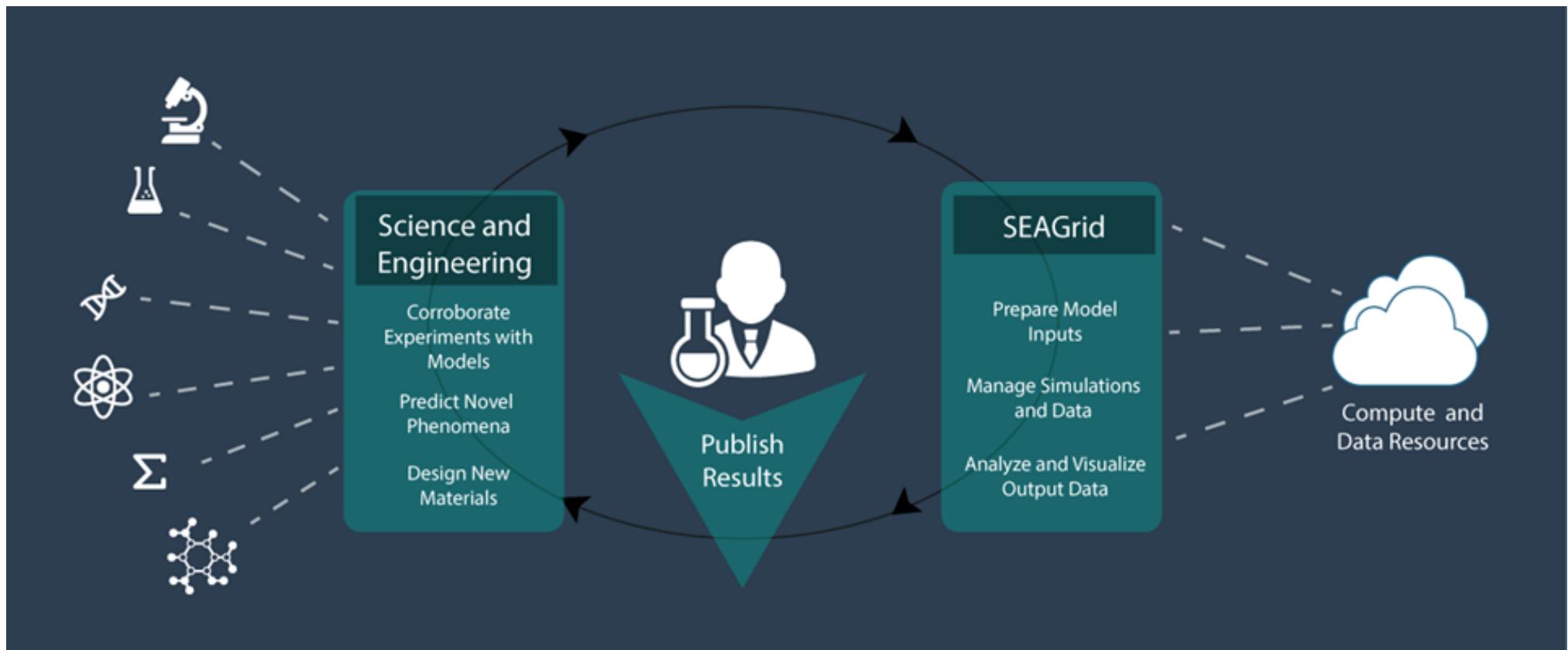
Student Application Deadline: March 25, 2016





What Is a Science Gateway?





SEAGrid.org is an Apache Airavata-powered gateway



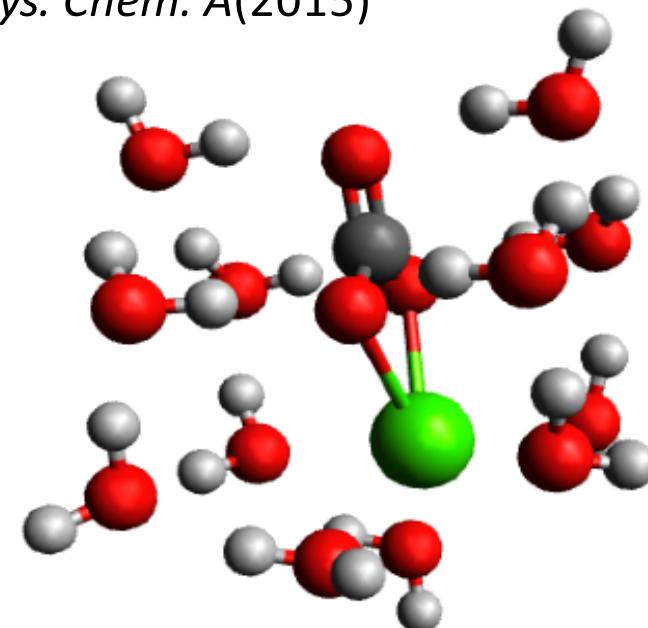
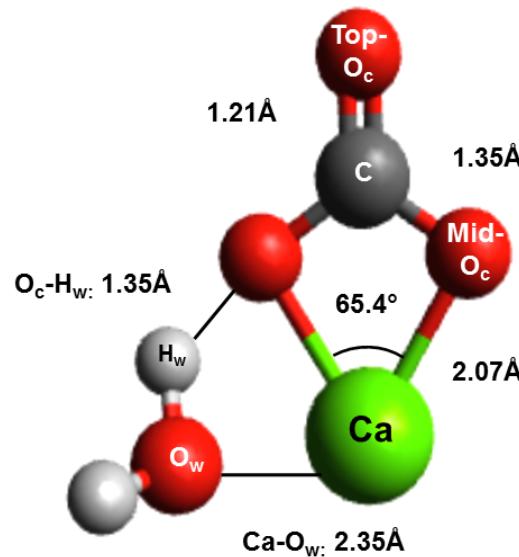
sgg@iu.edu



What is the chemistry of hydrated calcium carbonate?

- Bio-mineralization of skeletons and shells
- Geological CO₂ sequestration
- Cleanup of contaminated environments

Lopez-Berganza, et al. *J Phys. Chem. A*(2015)



sgg@iu.edu

SciGaP

$\text{CaCO}_3 \cdot x\text{H}_2\text{O}$ Initial
guess

SEAGrid.org enabled workflow



Stampede Supercomputer
TINKER
Monte Carlo Molecular
Mechanics
(Minimize Torsional Energy in
<20,000 steps)

$x=x+1$

Stampede Supercomputer
DFTB+
Approximate DFT-Based



-2-3 CaCO_3 Equilibrium
Structures
-Thermochemistry (E,H,G,
etc.)
-Vibrational Frequencies

Comet Supercomputer

Gaussian09
Ab initio Quantum Chemistry

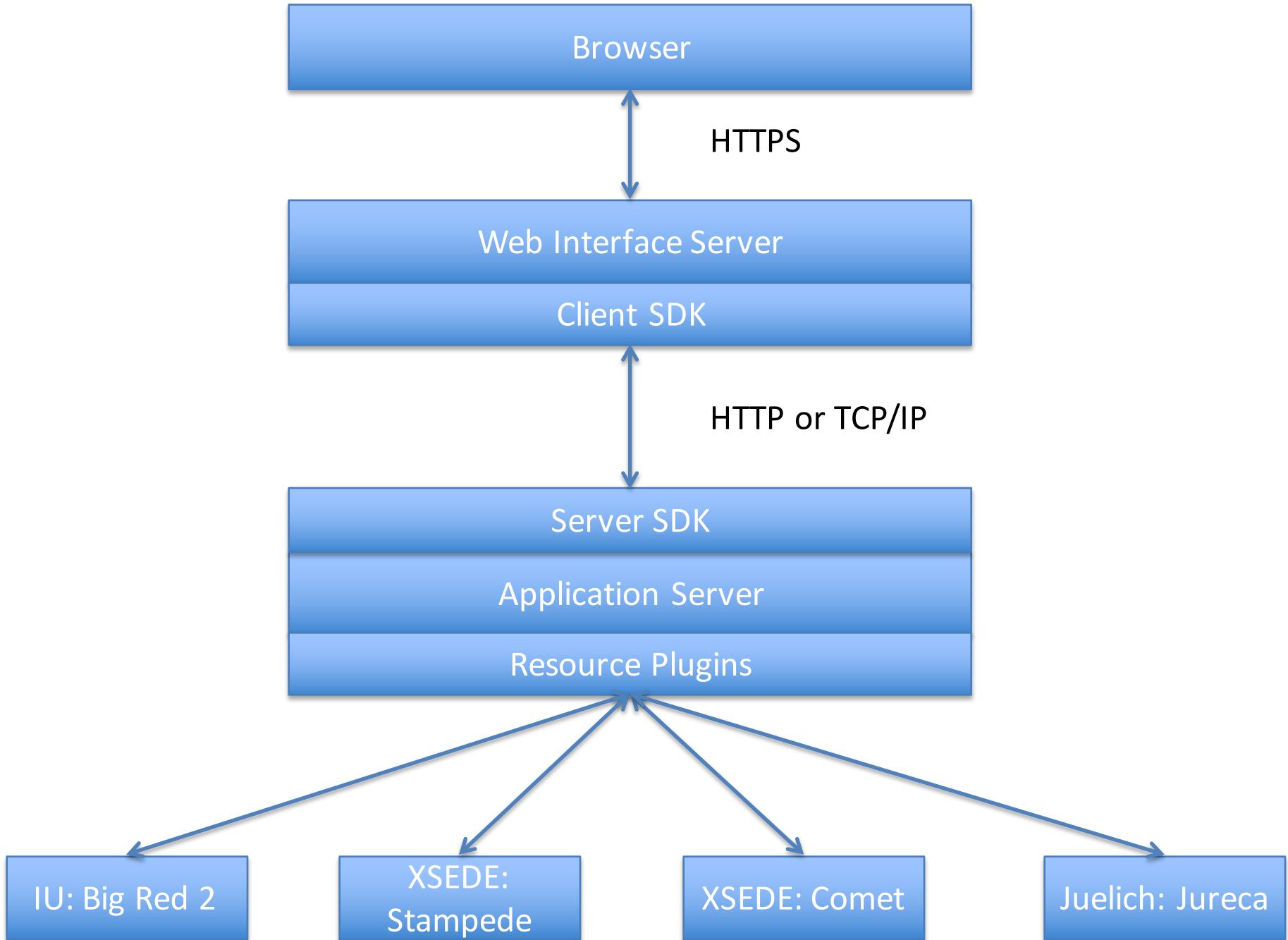


Lopez-Berganza, et al. *J Phys. Chem. A*(2015)



sgg@iu.edu





Challenges for Gateways

- Providing a rich user experience
- Defining an API for the application server
- Defining the right sub-components for the application server.
- Implementing the components, wiring them together correctly.
- Supporting multiple gateway tenants
- Fault tolerance for components
- State management
- Continuous delivery
- Security management
- Supporting full scientific exploratory cycle



What is Cyberinfrastructure?

Scientific Distributed Computing



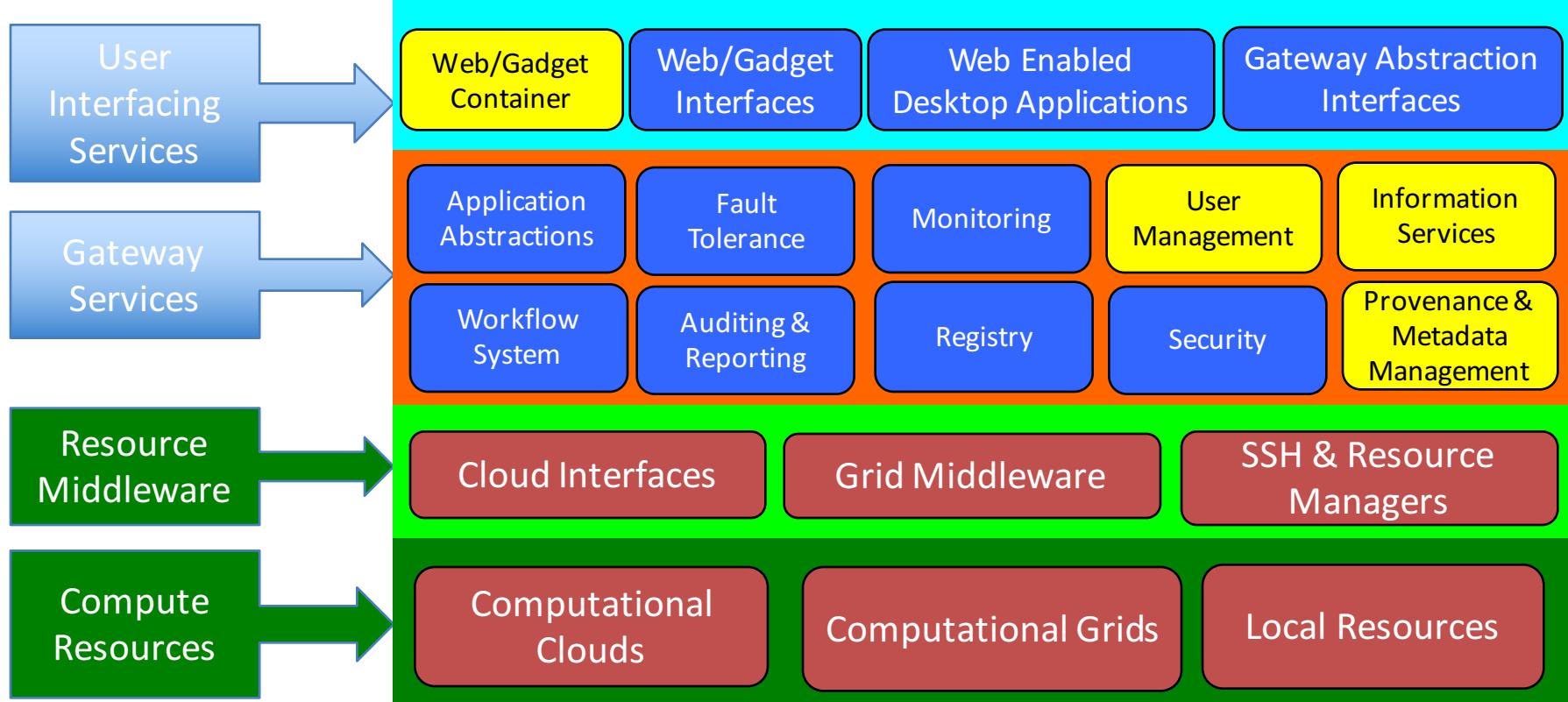
What Is Cyberinfrastructure?

“Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible.”

—Craig Stewart, Indiana University



Cyberinfrastructure Layers



Color Coding



Airavata components



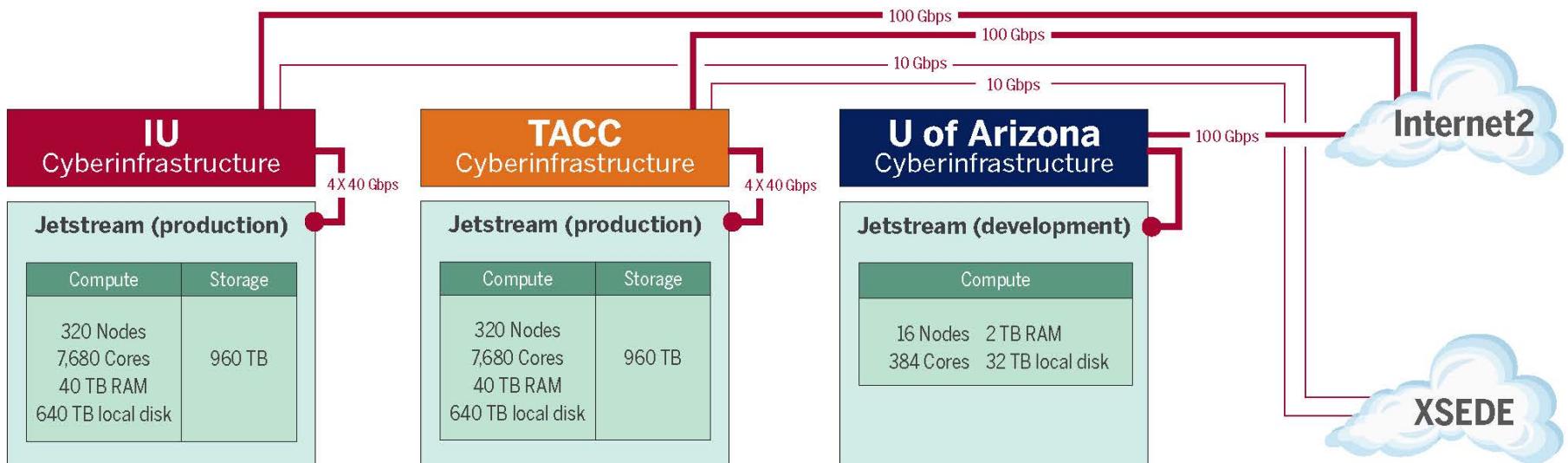
Complimentary gateway components



Dependent resource provider components



jetstream



- Geographically Distributed Cloud, 0.5 PetaFLOPS
- High-speed connections to Internet2 and local connections to Wrangler disk storage at IU and TACC
- Globus-based large scale file movement



Apache *airavata*



Developing Computational Science Gateways using Apache Airavata

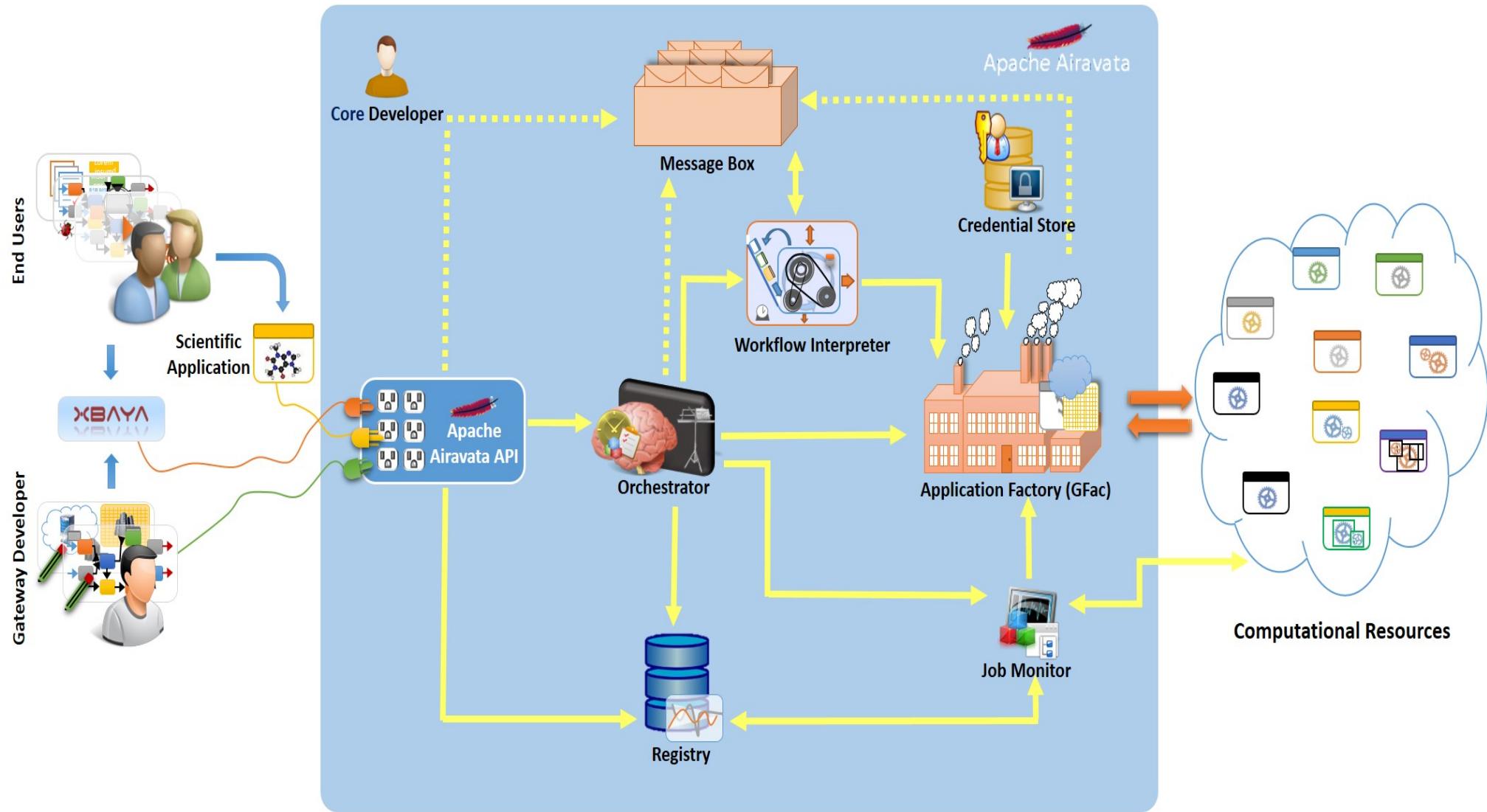


What Is Apache Airavata?

- Science Gateway software system to
 - Compose, manage, execute, and monitor distributed, computational workflows
 - Wrap legacy command line scientific applications with Web services.
 - Run jobs on computational resources ranging from local resources to computational grids and clouds
- Airavata software is derived from NSF-funded academic research.



Airavata Components



Airavata and SciGaP

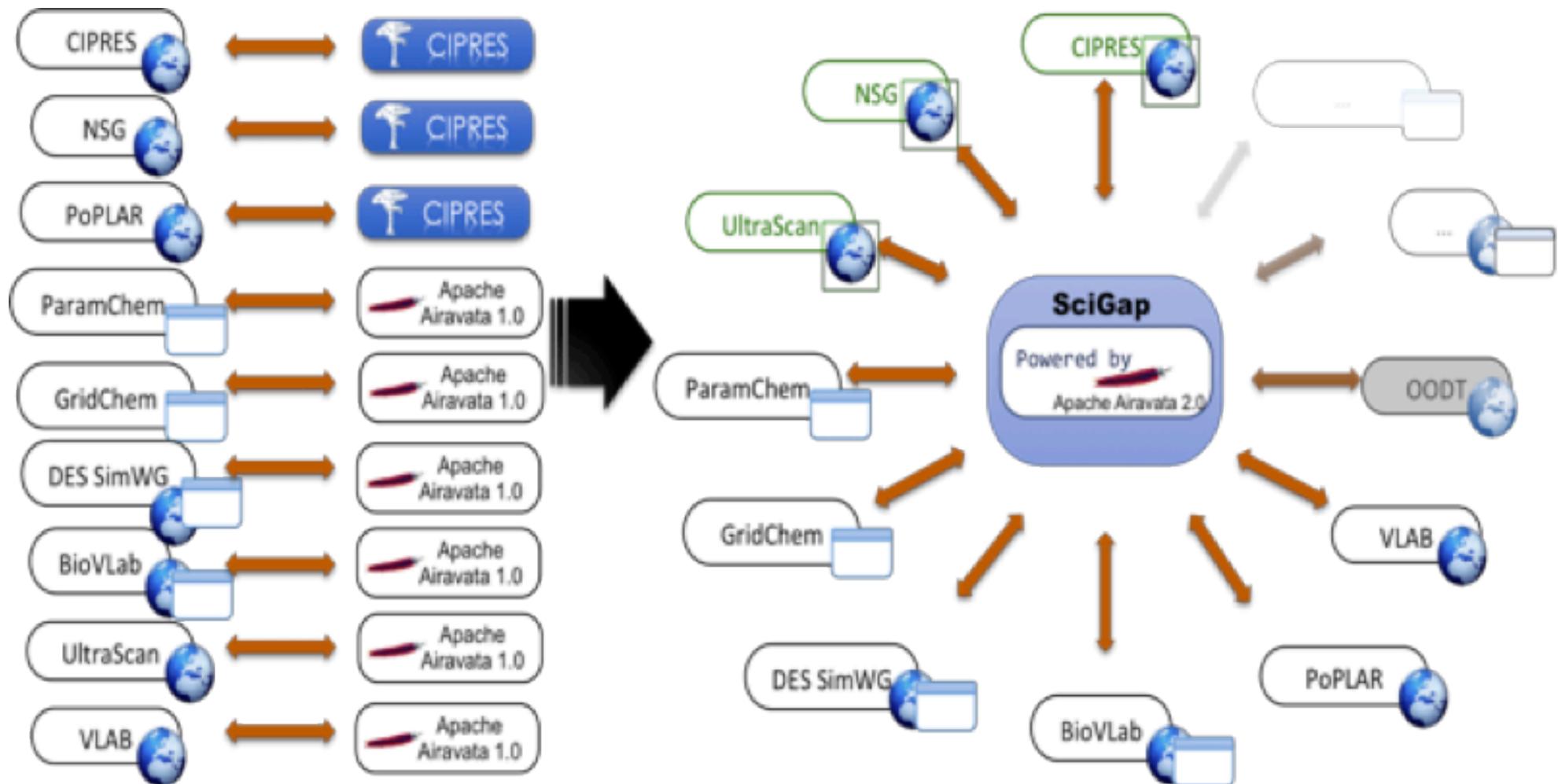
- SciGaP: Airavata as a multi-tenanted Gateway Platform as a Service
- Goal: We run Airavata so you don't have to.
 - Scalable support
- Challenges:
 - Centralize system state
 - Make Airavata more cloud friendly, elastic

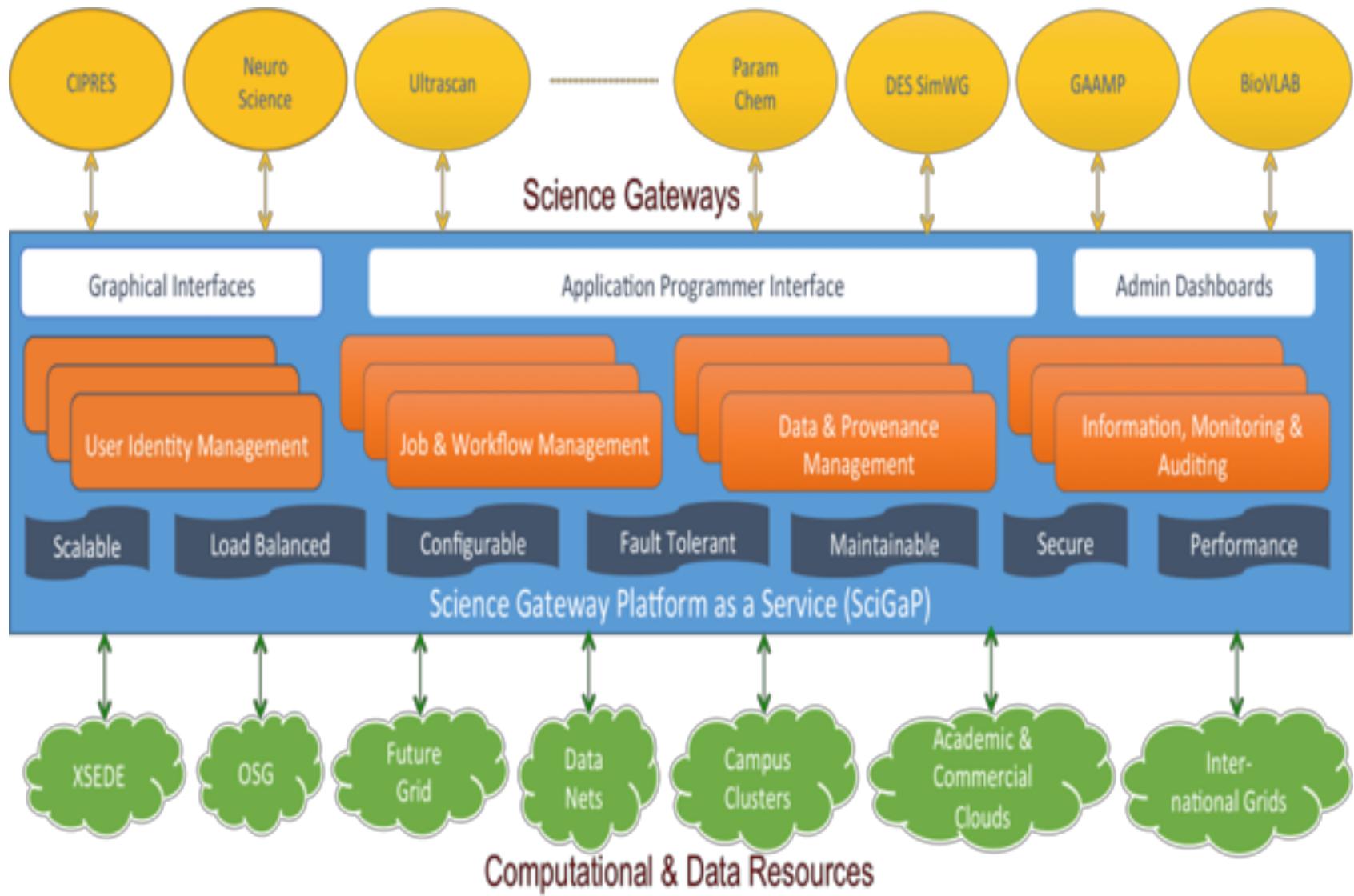


<http://scigap.org>



SciGaP Vision





SciGaP Role and Goals: Improve sustainability by converging on a single set of hosted infrastructure services

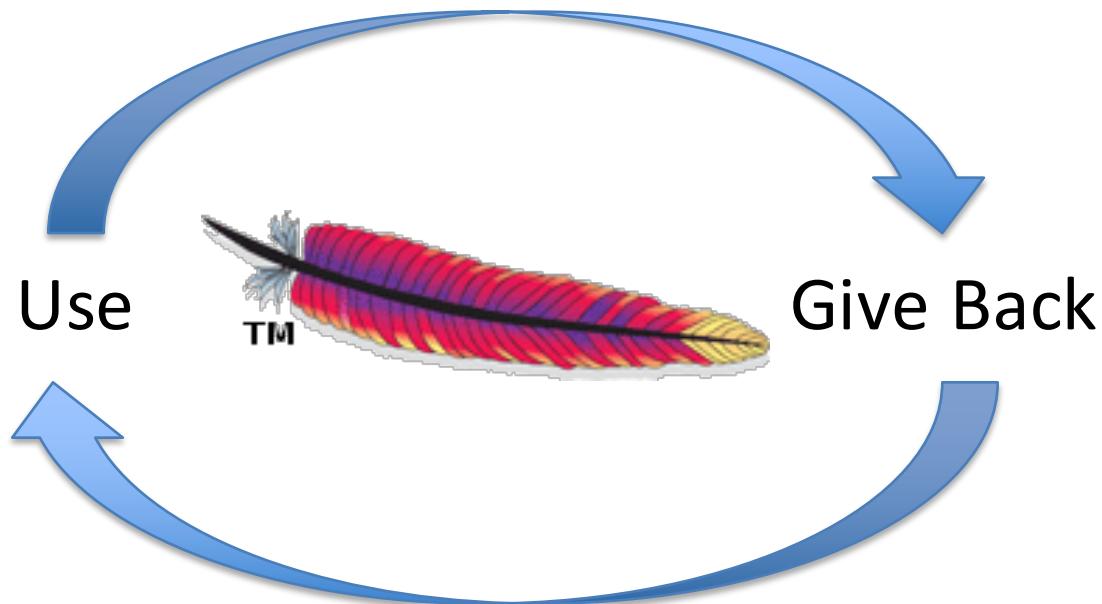


Contributing to the Airavata Framework

Component	Research Opportunities
Registry	Better support for Thrift-generated objects; NoSQL and other backend data stores; fault tolerance
Orchestrator	Pluggable scheduling: work stealing internal scheduling implementations, super-scheduling over multiple resources
GFAC	Apache Storm workflow management; elastic packaging.
Messenger	Investigate AMQP, Kafka, and other newer messaging systems
Workflow Interpreter	Alternative workflow processing engines.
Overall	Rewire component micro-services.

We took our software to the Apache Software Foundation to encourage more scientific collaborations. Community-owned software. **GET INVOLVED THROUGH GSOC.**

“Apache” Means “Open”



<http://theapacheway.com/>



How open is open source?

- What's missing?
 - Open source licensing
 - Open Standards
 - Open Code (GitHub)



We also need open governance

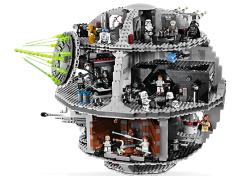


Open Source Software and Governance

- Open source projects need diversity, governance.
 - Sustainability
- Incentives for projects to diversify their developer base.
- Govern
 - Software releases
 - Contributions
 - Credit sharing.
 - Members are added
 - Project direction decisions.
 - IP, legal issues
- Our approach: Apache Software Foundation



Compete



Collaborate

