



APACHE
AIRAVATA

Science Gateway Architectures: Applied Open Source, Distributed Computing

Marlon Pierce, Suresh Marru

Science Gateways Research Center, Indiana University

marpierc@iu.edu, smarru@iu.edu

B649 Special Topics Course: Fall 2017

- For syllabus information and presentation material, see the course website
 - <http://courses.airavata.org>
- For project assignments and other communications, use Canvas



What Is a Science Gateway?

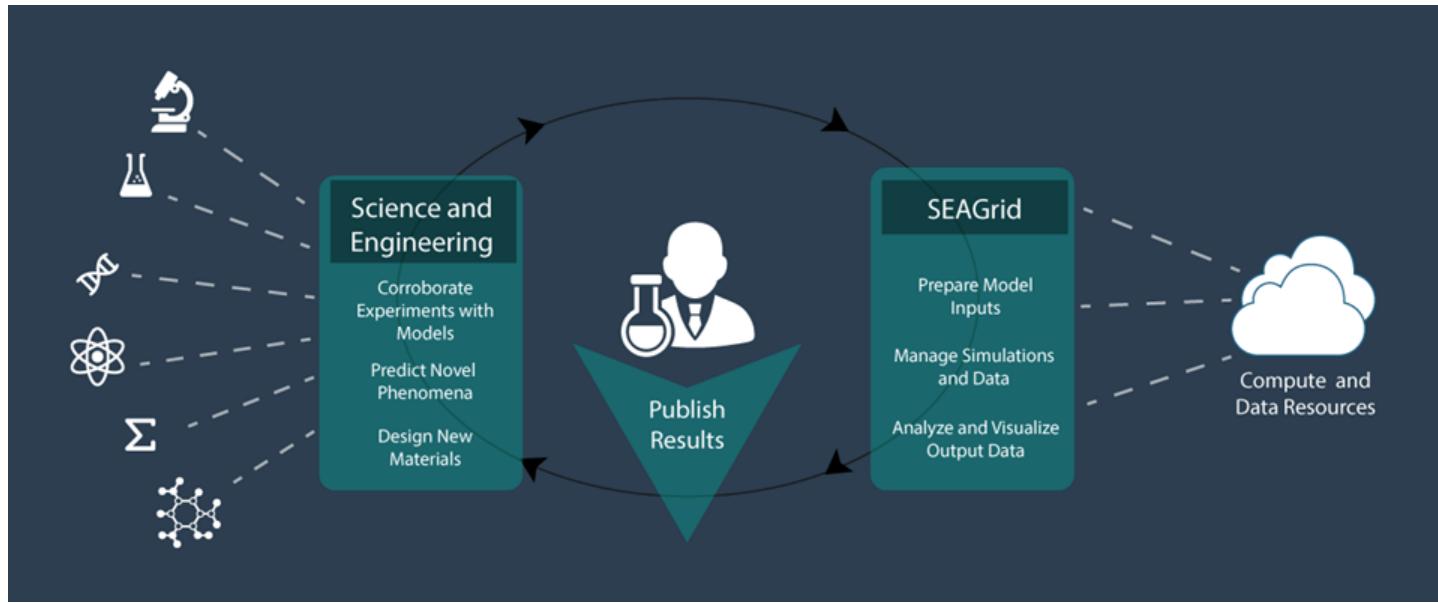
A Science Gateway is a user-centric environment for conducting online computational scientific research

What if Mark Zuckerberg Read More Chemistry and Biology Books?





What Is a Science Gateway?



SEAGRID.org is an Apache Airavata-powered gateway

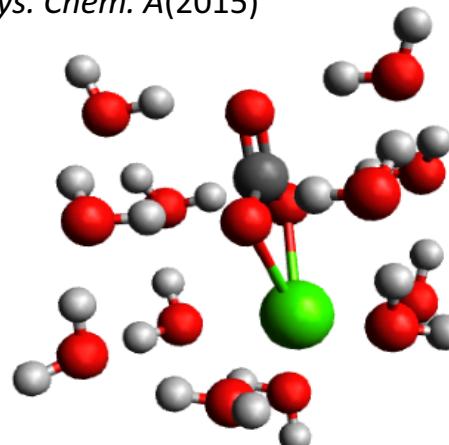
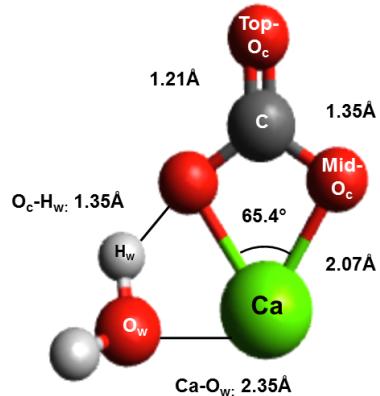
Hydrated Calcium Carbonate in Action



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)

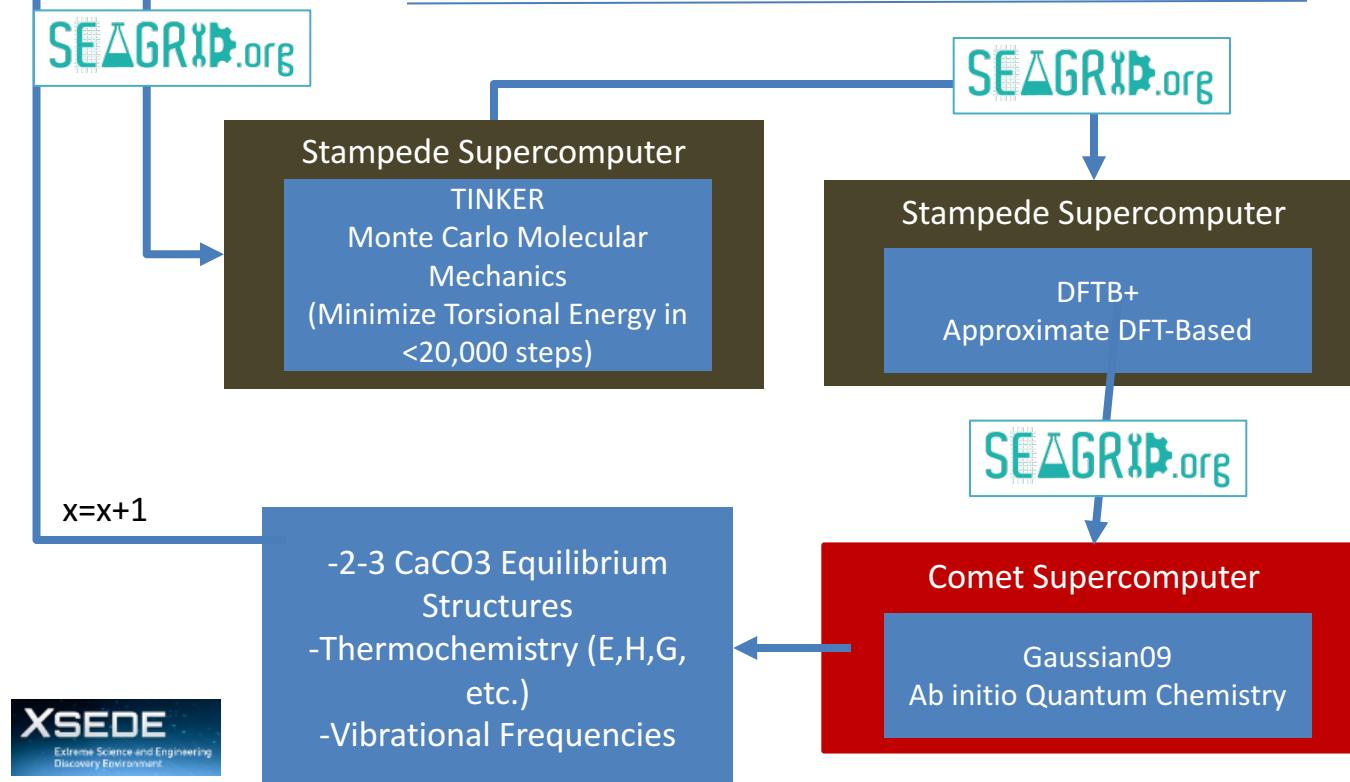


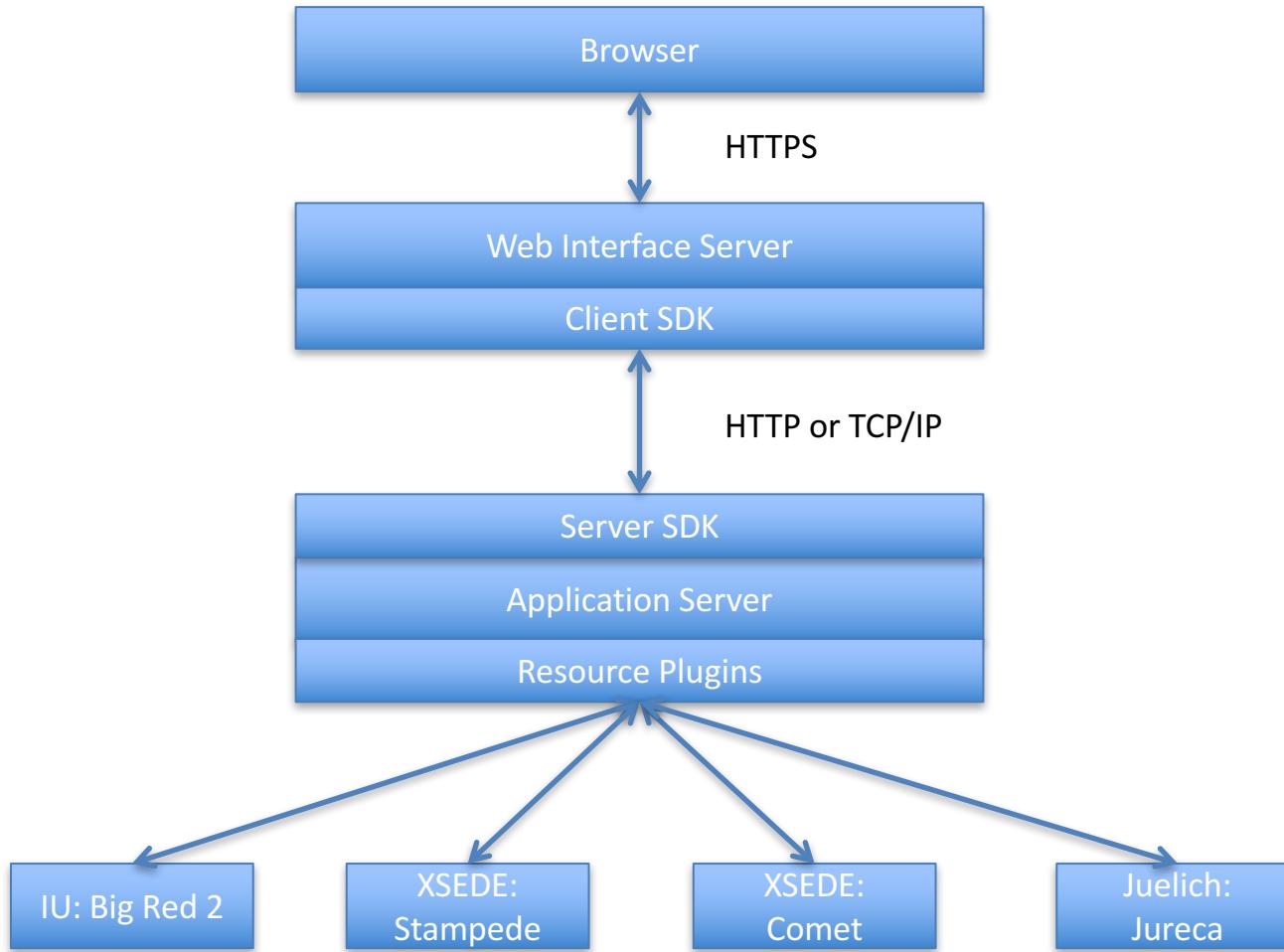
CaCO₃.1H₂O

CaCO₃.12H₂O

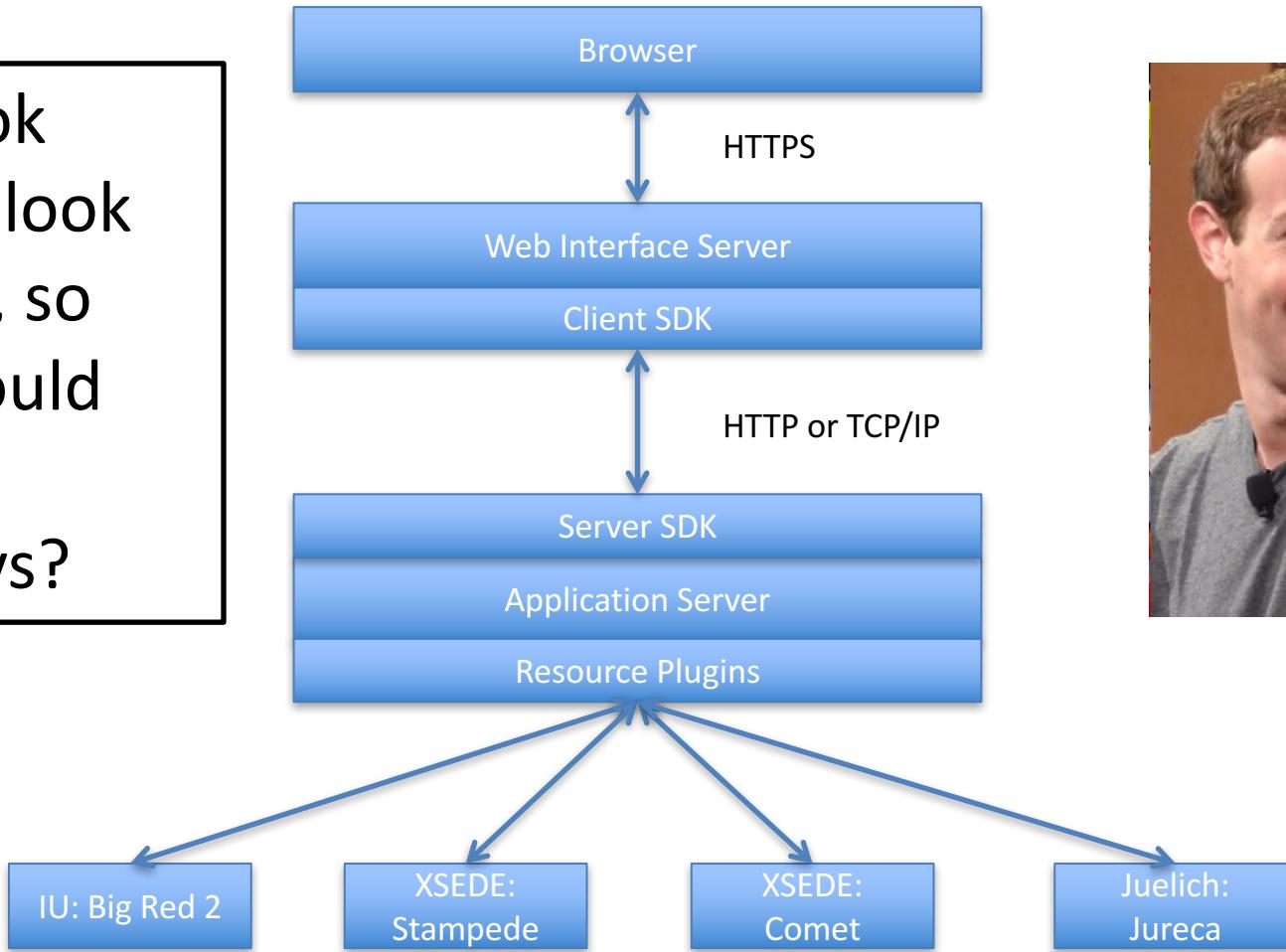
$\text{CaCO}_3 \cdot x\text{H}_2\text{O}$
Initial guess

SEAGrid.org enabled workflow





Facebook
doesn't look
like this, so
why should
science
gateways?



Challenges for Science 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



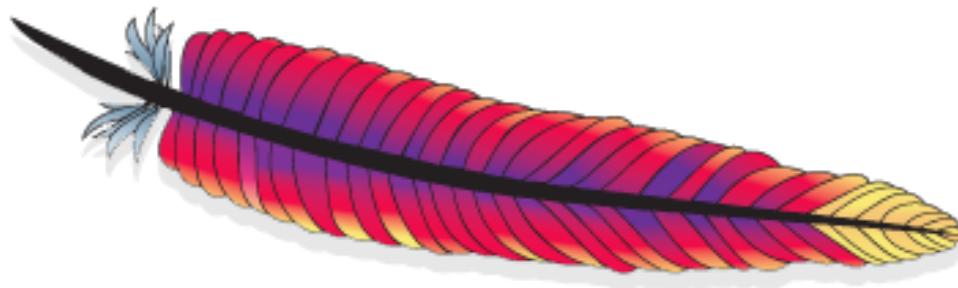
Goal 1: Apply basic distributed computing concepts to Science Gateways.

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



@RealGeneKim, genek@realgenekim.me

Goal 2: Apply new architectures,
methodologies, and technologies to
Science Gateways: Microservices,
DevOps



Goal 3: Teach open source
software practices



APACHE
AIRAVATA

Developing Computational Science
Gateways using Apache Airavata

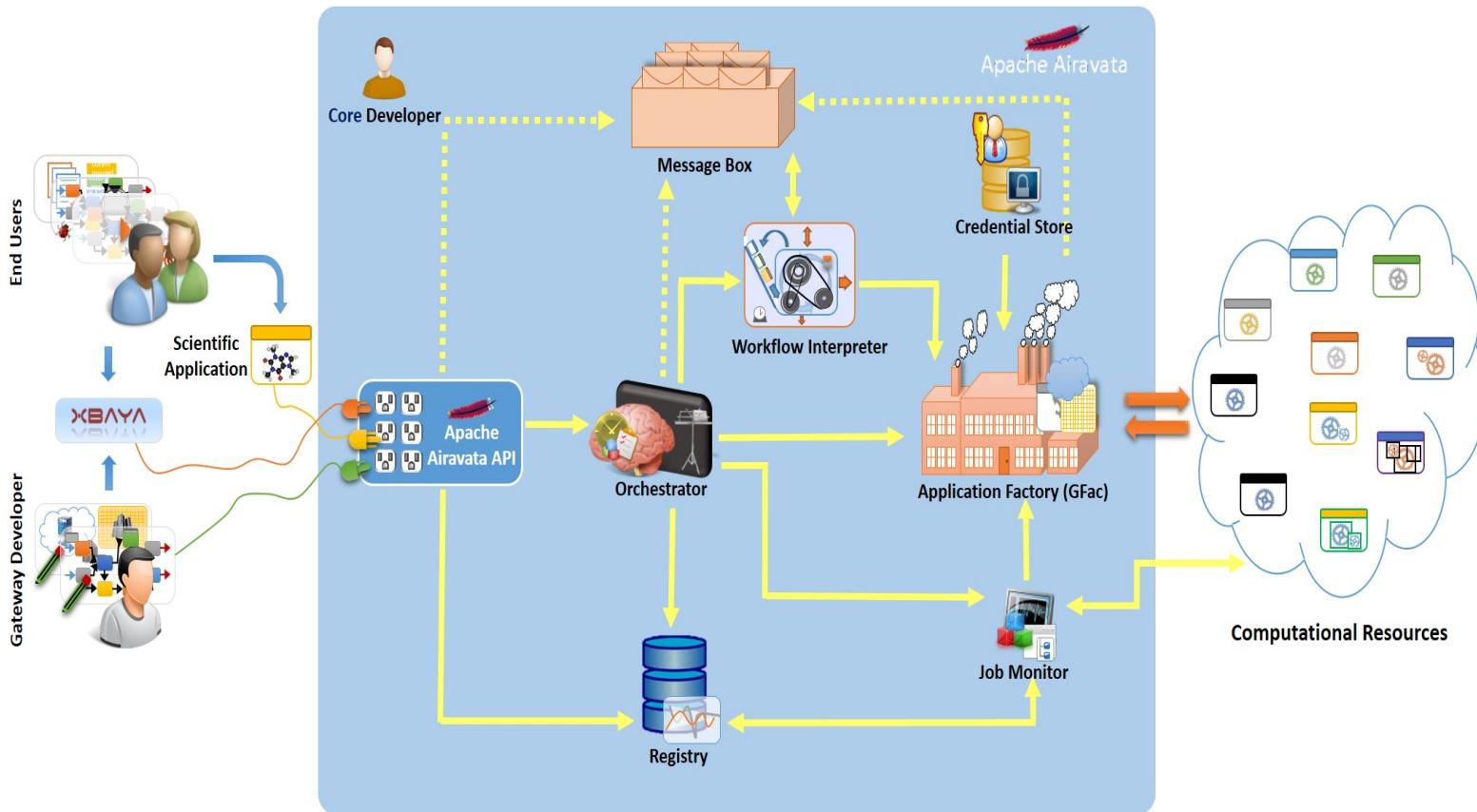
What Is Apache Airavata?

- Open source middleware to support Science Gateways
 - 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
 - Record, preserve, search, and share metadata about computational experiments
- Hosted version of Apache Airavata provides multi-tenanted Platform as a Service.
 - SciGaP

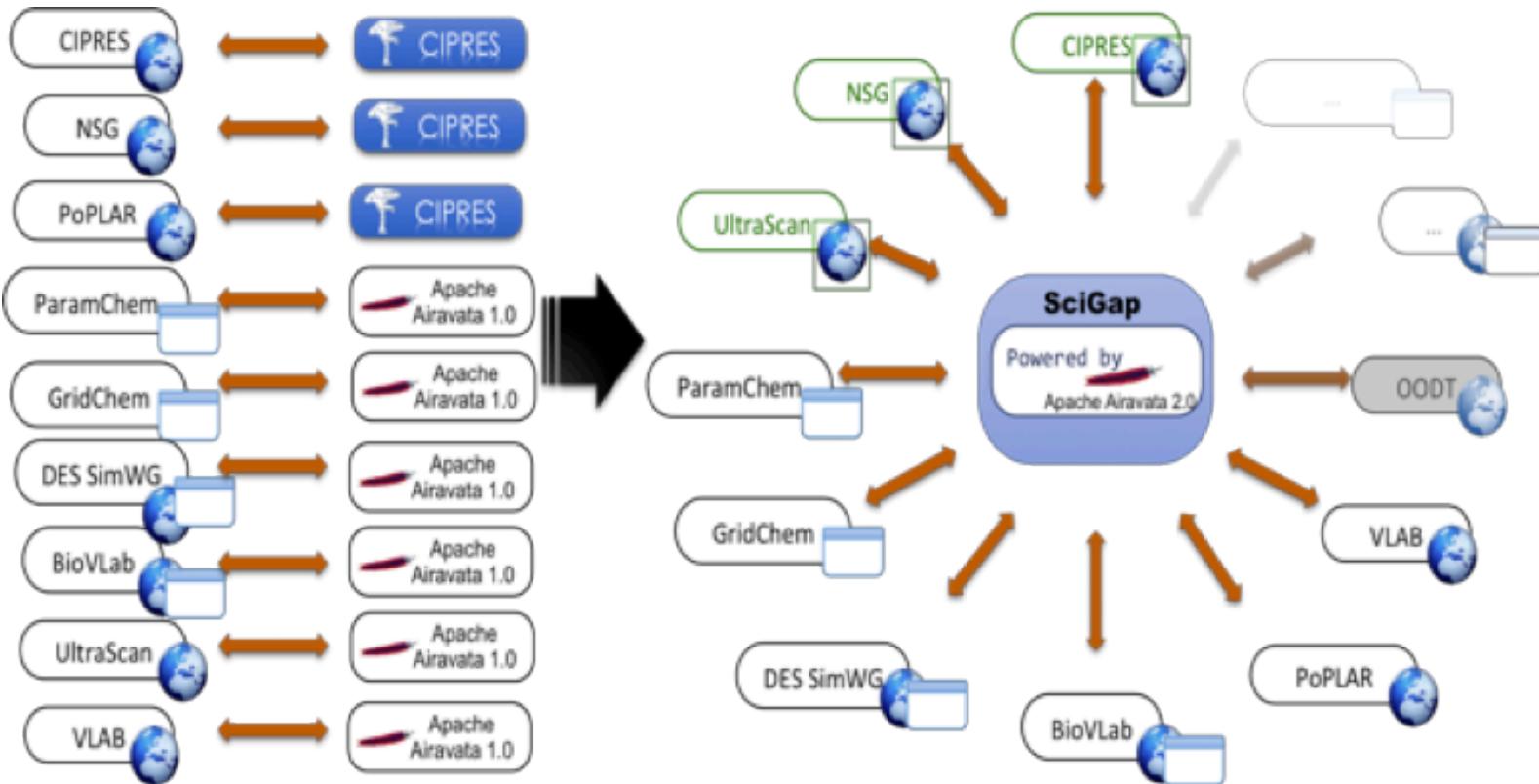


APACHE
AIRAVATA

Airavata Components



Airavata as a Service: SciGaP



Structure of the Course

- All students work individually
- 80% of your grade is from homework submissions
- 20% comes from mid-term and final presentations and reports
- Everyone will have a github repository within the
<https://github.com/airavata-courses> organization
- See the syllabus (courses.airavata.org) for more details
- We will provide you with Jetstream accounts that can be used to create VMs for running your services.

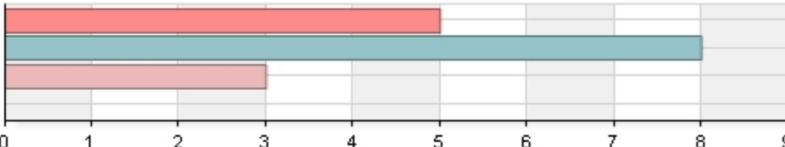
<http://courses.airavata.org/>

Project Grading

- Each project is assigned for 2 weeks and worth 10 points
 - 8 assignments
- You get 8 points for your submission and 2 points for peer reviewing 2 other submissions
- For each assignment
 - Week 1: You must submit your project by Monday midnight of the first week
 - Week 2: Peer reviewers are assigned on Tuesday and must complete by Monday midnight of second week
 - Decision: each peer either approves or disapproves the submission
 - At least one approval is needed before the GAs will grade
 - Use GitHub issues for peer reviews and approvals
- Tuesday of Week 2: course GAs start grading all approved assignments
- If the grader executes the assignment, the submitter gets 8 points and the peer reviewers each get 1 point
 - If not, the submitter loses 8 points and the approving reviewers lose 1 point
- Peer reviewers who disapprove an assignment get their points

How effectively was class time used to help you learn?

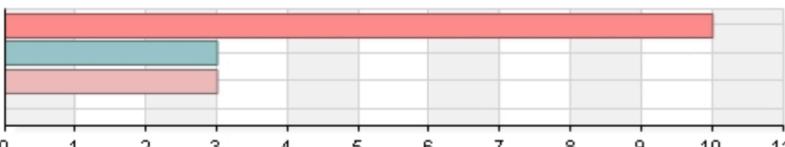
Very Effectively	5	31.3%
Effectively	8	50.0%
Somewhat Effectively	3	18.8%
Not at all Effectively	0	0.0%
Total	16	



Question	Course			Department (INFO)			Institution		
	Mean	Response Count	Standard Deviation	Mean	Response Count	Standard Deviation	Mean	Response Count	Standard Deviation
How effectively was class time used to help you learn?	3.1	16	0.7	2.9	3416	1.0	3.2	89829	0.9

How effectively did out-of-class work (assignments, readings, practice, etc.) help you learn?

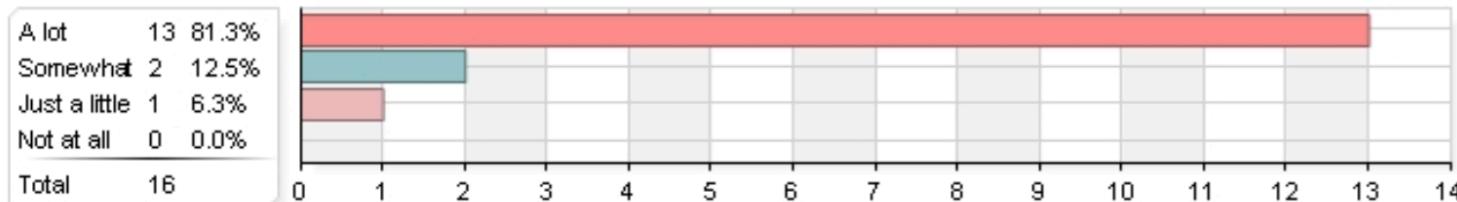
Very Effectively	10	62.5%
Effectively	3	18.8%
Somewhat Effectively	3	18.8%
Not at all Effectively	0	0.0%
Total	16	



Question	Course			Department (INFO)			Institution		
	Mean	Response Count	Standard Deviation	Mean	Response Count	Standard Deviation	Mean	Response Count	Standard Deviation
How effectively did out-of-class work (assignments, readings, practice, etc.) help you learn?	3.4	16	0.8	3.0	3437	1.0	3.1	88762	0.9

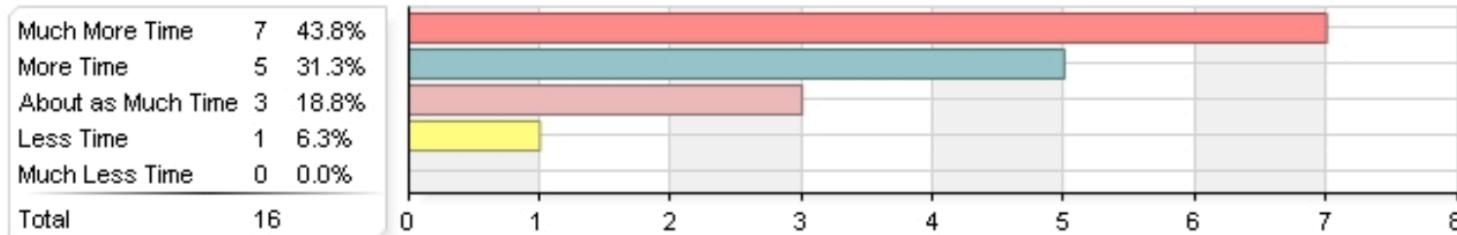
Learning occurs outside the classroom.

How much did the instructor emphasize student learning and development?



Question	Course			Department (INFO)			Institution		
	Mean	Response Count	Standard Deviation	Mean	Response Count	Standard Deviation	Mean	Response Count	Standard Deviation
How much did the instructor emphasize student learning and development?	3.8	16	0.6	3.5	3686	0.8	3.6	94081	0.7

Compared to other courses you've taken, how much time did this course require?



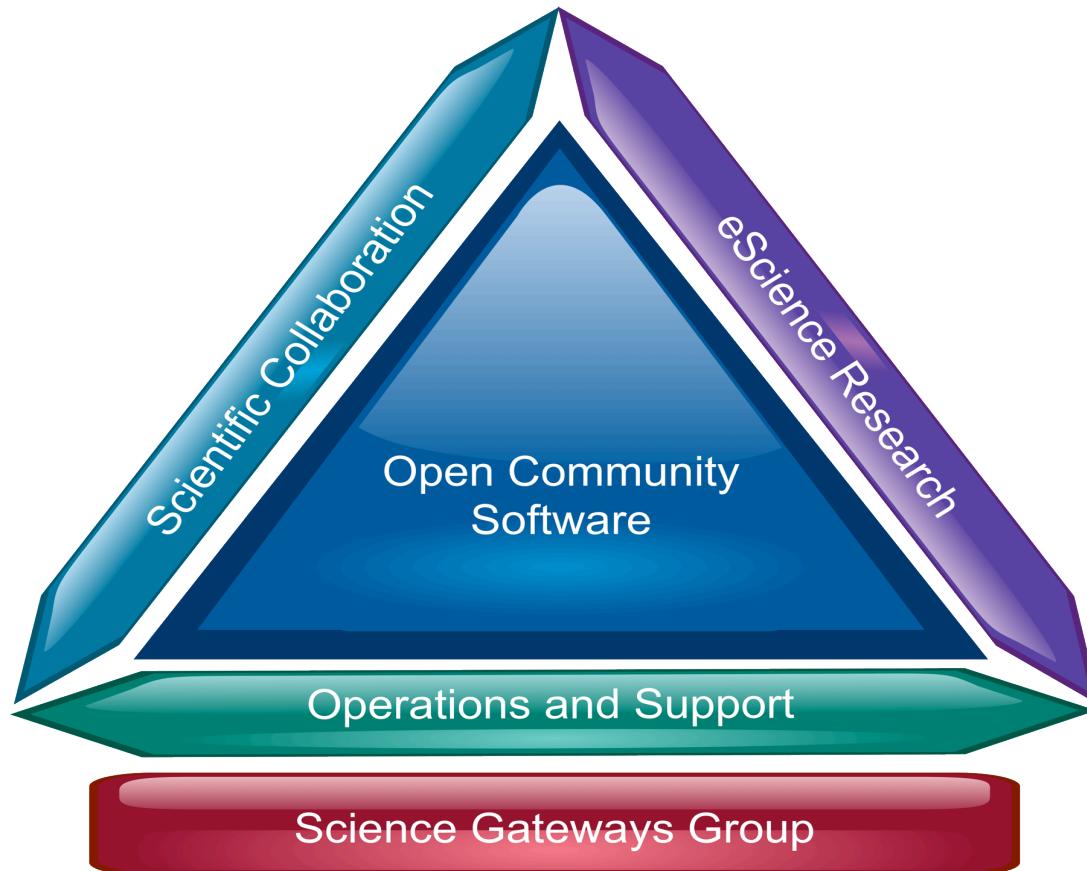
What You Need to Do Before Thursday

- Create a GitHub account if you don't have one
- Send us your GitHub user name so that we can add you to the Airavata Courses GitHub organization
- Go to <https://portal.xsede.org/> and create an account
 - Send us your user name
- If possible, please make these user names consistent with your IU network ID.
- Submit this through Canvas (0-point assignment)

More Information

- Contact:
 - marpierc@iu.edu, smarru@iu.edu
 - Join dev@airavata.apache.org, users@airavata.apache.org,
architecture@airavata.apache.org
- Websites:
 - Apache Airavata: <http://airavata.apache.org>

A Balancing Act



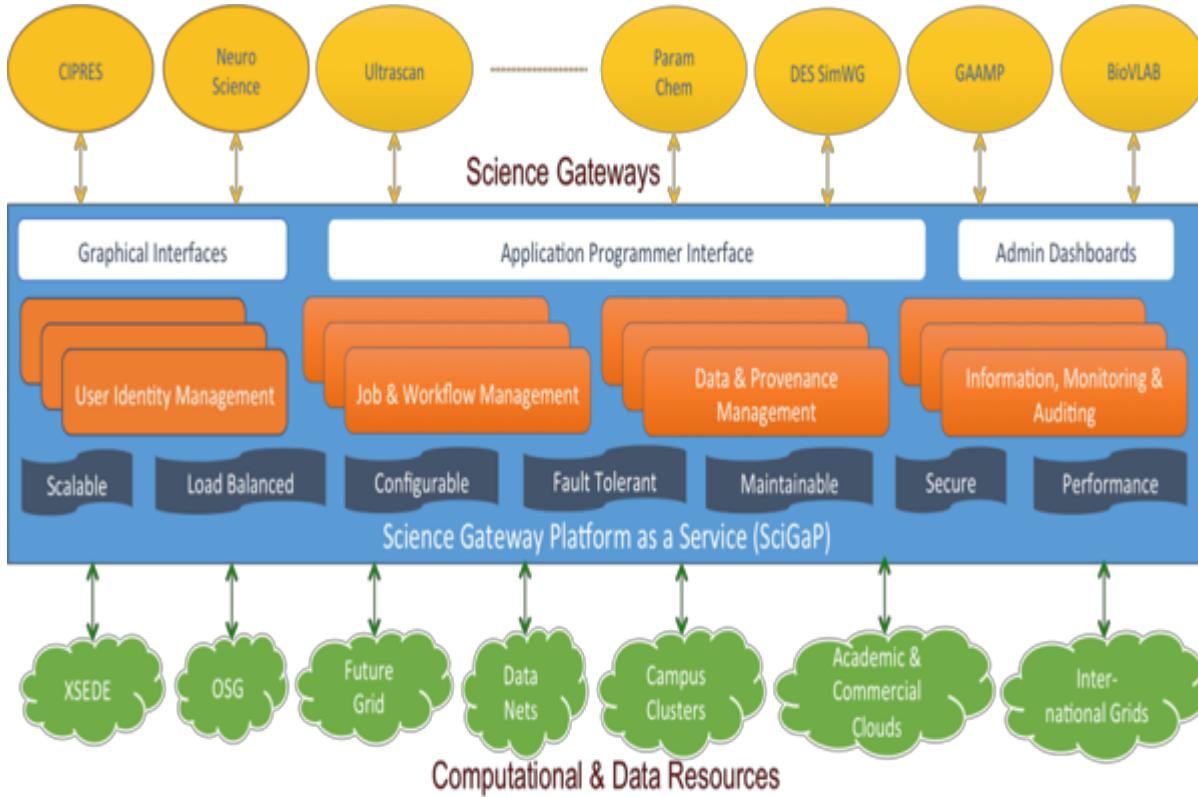
What is Cyberinfrastructure?

Scientific Distributed Computing

What Is Cyberinfrastructure?

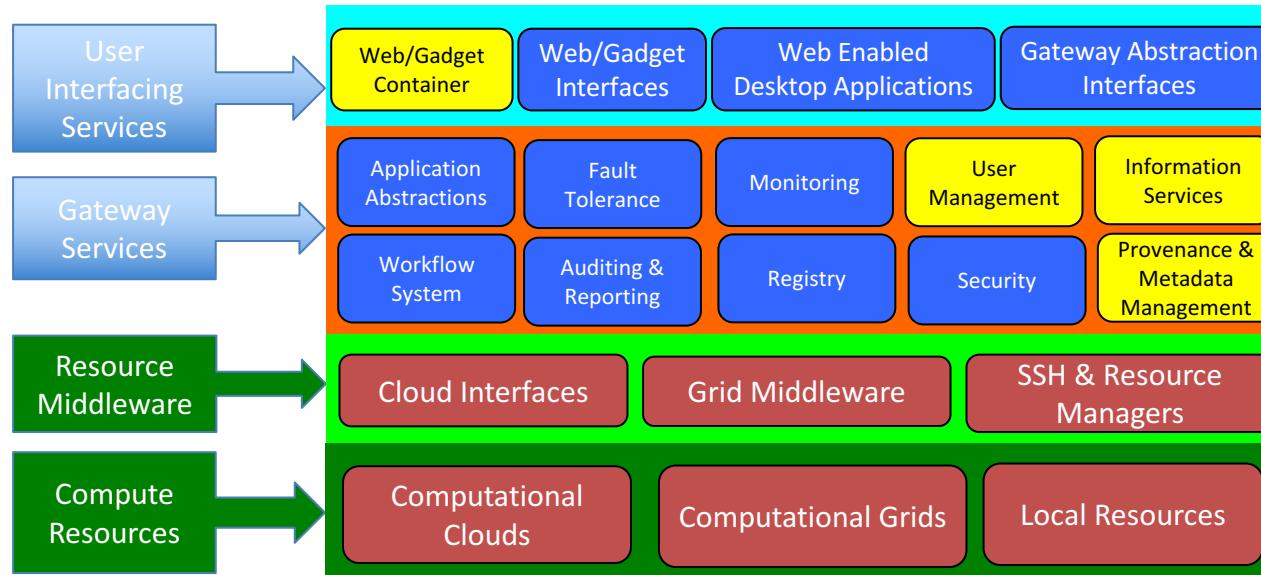
“Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible.”

—Craig Stewart, Indiana University



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

Cyberinfrastructure Layers



Color Coding



Airavata components



Complimentary gateway components



Dependent resource provider components