



INDIANA UNIVERSITY
**SCHOOL OF INFORMATICS,
COMPUTING, AND ENGINEERING**

Applied Distributed Systems

January 14th, 2020

Suresh Marru, Marlon Pierce

smarru@iu.edu, marpierc@iu.edu

Todays Outline

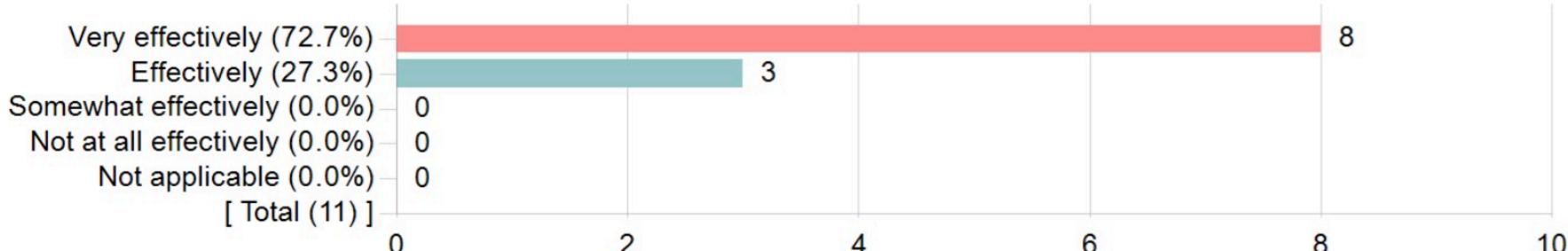
- What To Expect
- Course Logistics
- Course Topic Overview
- Open Discussion



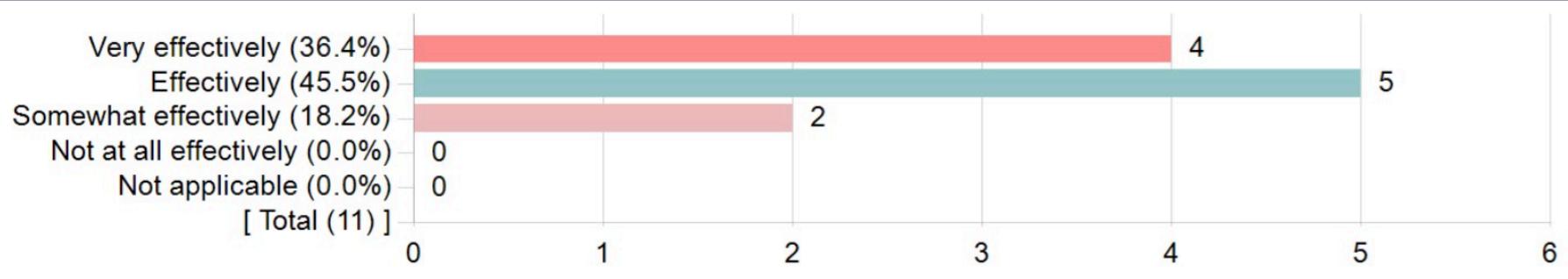
INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

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



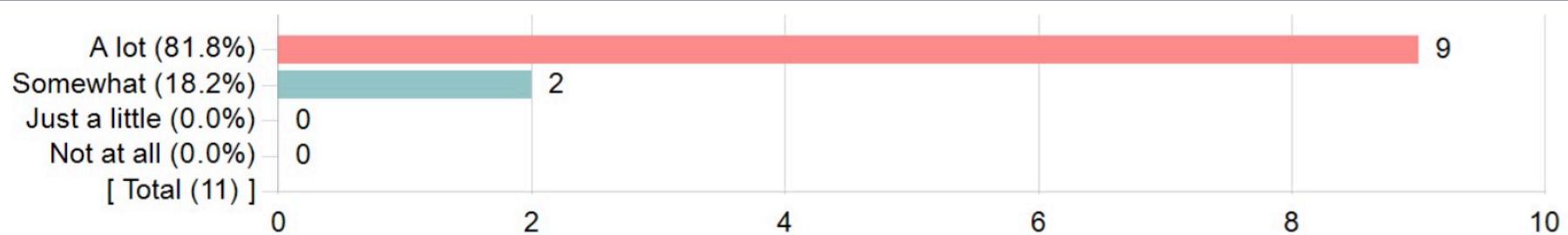
How effectively was class time used to help you learn?



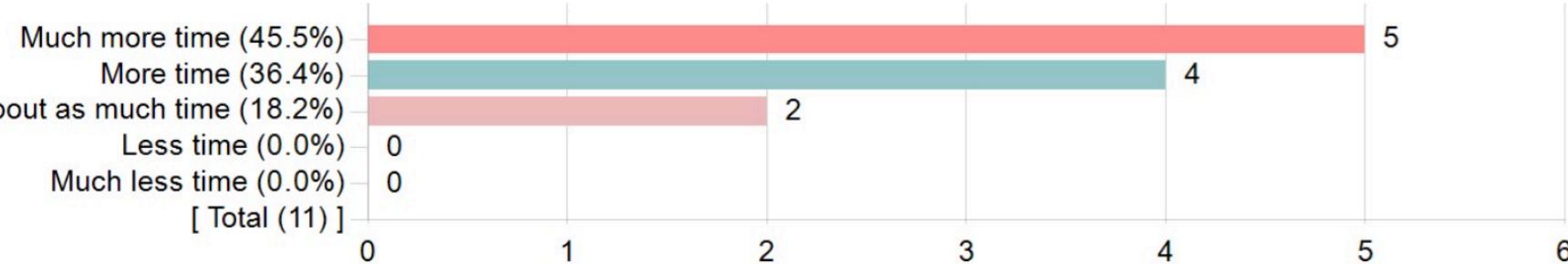
How much did the instructor motivate you to do your best work?



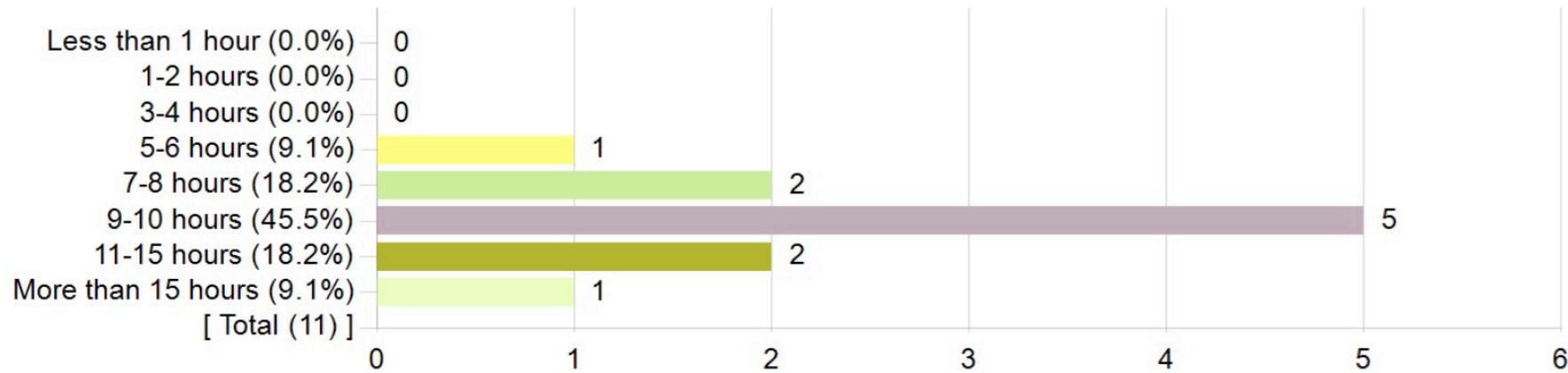
How much did the instructor emphasize student learning and development?



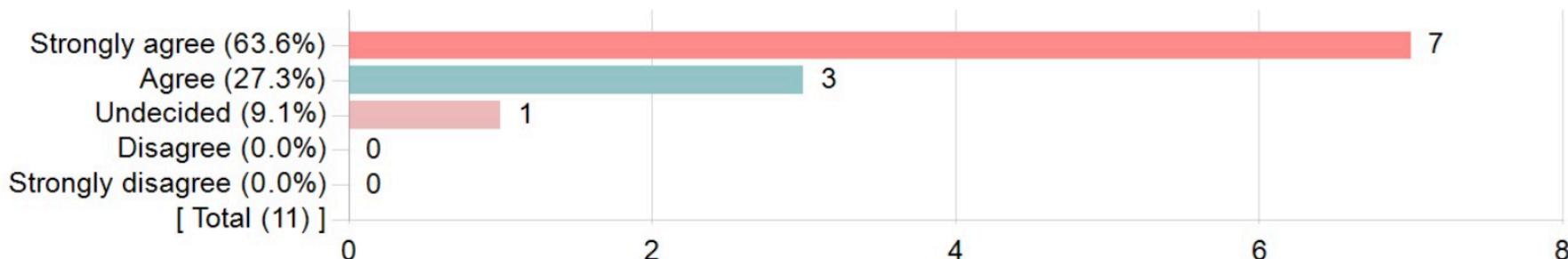
Compared to other courses you've taken, how much time did this course require?



In a typical week, about how much time did you devote to this course? (Do not count scheduled class time, labs, etc.)



I developed skill in critical thinking in this course.



What did you like most about this course and instructor?

Comments

Learning so many tools.

Best thing about projects is that I got to play around with instances on jetstream and there is no spoon feeding on how to go about setting it up. I learnt a lot in the process.

Explains in comprehensive manner

Topic Explanations were clear

Was able to learn new tools and technologies. The course was flexible with the project ideas being chosen.

Ample time was given to work on assignments

Structure of the Class

- We will have 3 project-based assignments
 - 90% of your grade
 - 25 points/project as a team of 3-4
 - 5 points/project for peer review (individual)
- The first two assignments will be due before semester break.
 - Each team will get the same assignment to build a science gateway using distributed systems concepts
- The third assignment will be for each team to apply your understanding to open problems in Apache Airavata.
- 10% of your grade will be attendance and classroom interactions.



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Class Format

- We will do a mixture of traditional lectures, interactive lectures, and flipped classrooms.
- Lectures will alternate between technology overviews and core concepts
 - “What is Kubernetes and how do you use it?”
 - “What are the architectural choices for building distributed systems?”
- We’ll also set aside “hackathon” time occasionally as we get near assignment deadlines.



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Sources of Truth

- Refer to the course's Canvas site for the authoritative information on deadlines, assignment details, assignment points, and grades.
- You will submit all assignments through Canvas.
- You can get lecture slides from <https://courses.airavata.org>
- All your work will go into GitHub.
 - Your code, your issues, your documentation, your peer reviews



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Should You Take This Class?

- We expect you to do a lot of work for the class
- We only require you to be able to write code and have a basic understanding of network protocols like HTTP and TCP/IP.
- We expect you will find the class challenging, rewarding, and enjoyable
- Make your semester plans accordingly
 - We'll offer the class again in Spring 2021



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

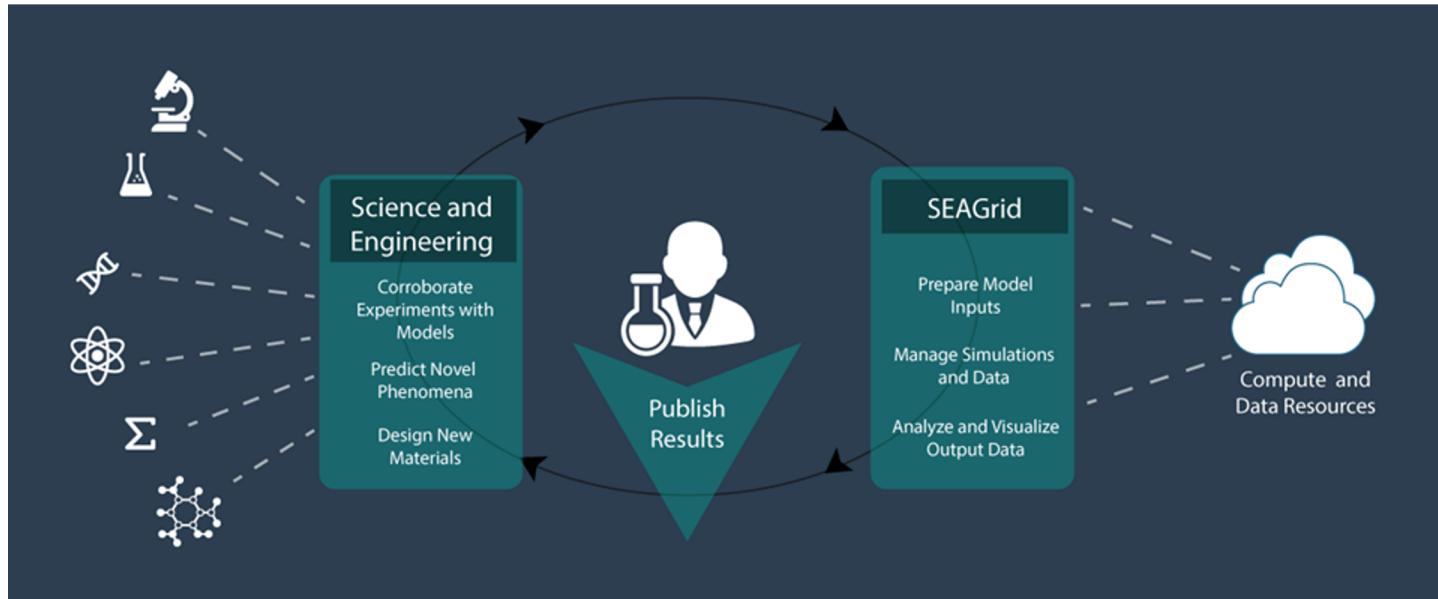
Applied Distributed Systems

- We will build user-centric distributed systems that support scientific research.
 - Science gateways
 - Cyberinfrastructure
- This course will be project-based.
- You will build distributed systems.



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING



SEAGRID.org is an Apache Airavata-powered gateway

Hydrated Calcium Carbonate in Action



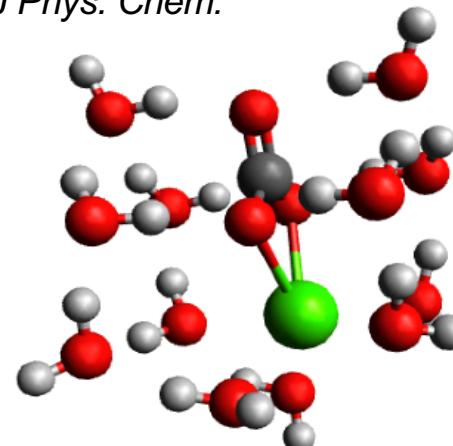
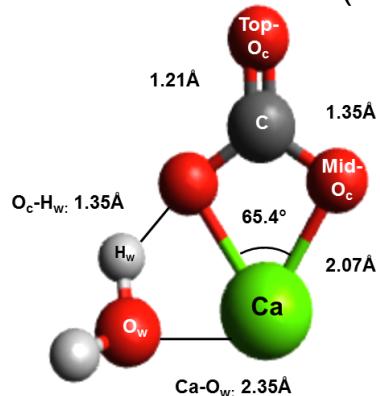
INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

What is the chemistry of hydrated calcium carbonate?

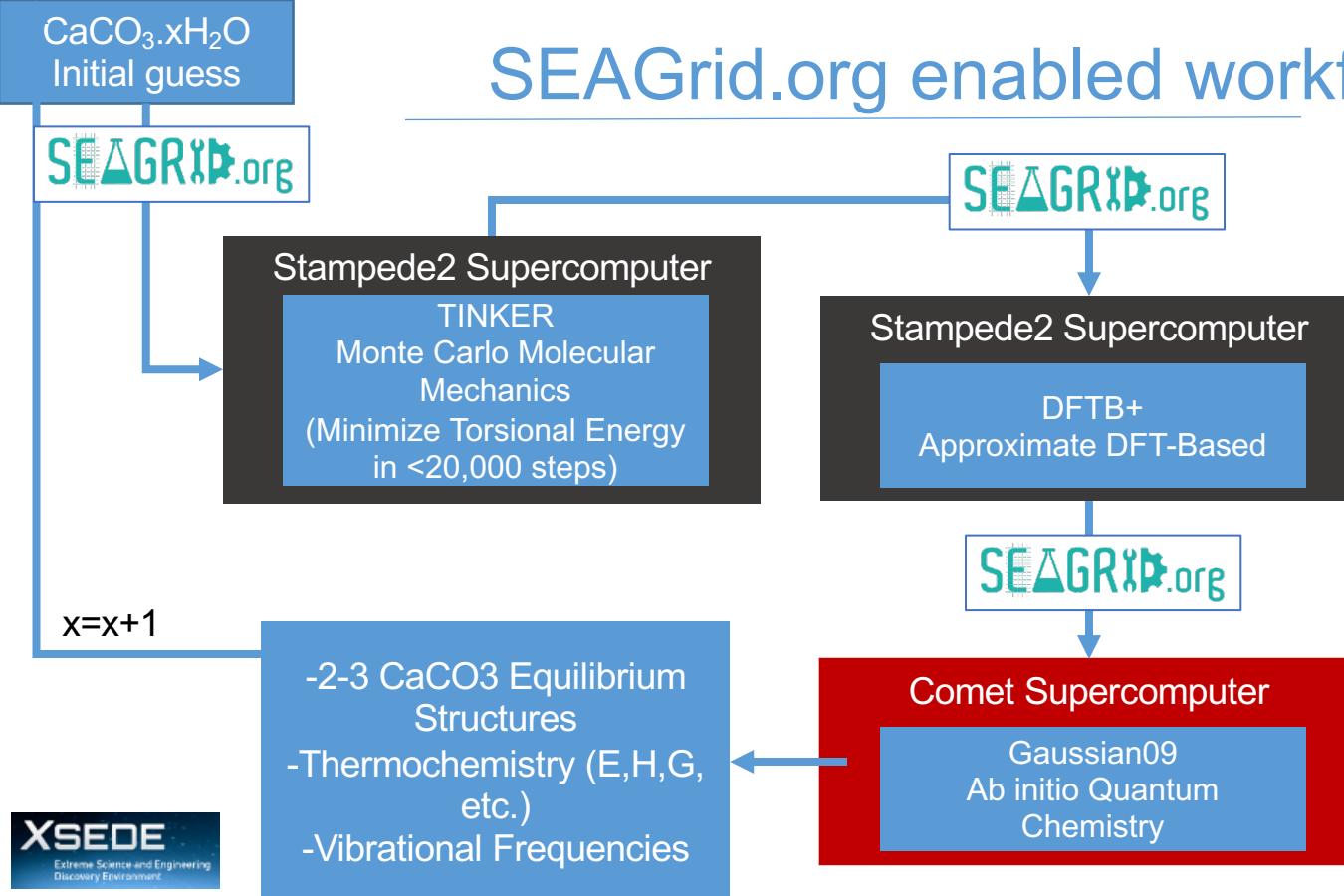
- Bio-mineralization of skeletons and shells
- Geological CO₂ sequestration
- Cleanup of contaminated environments

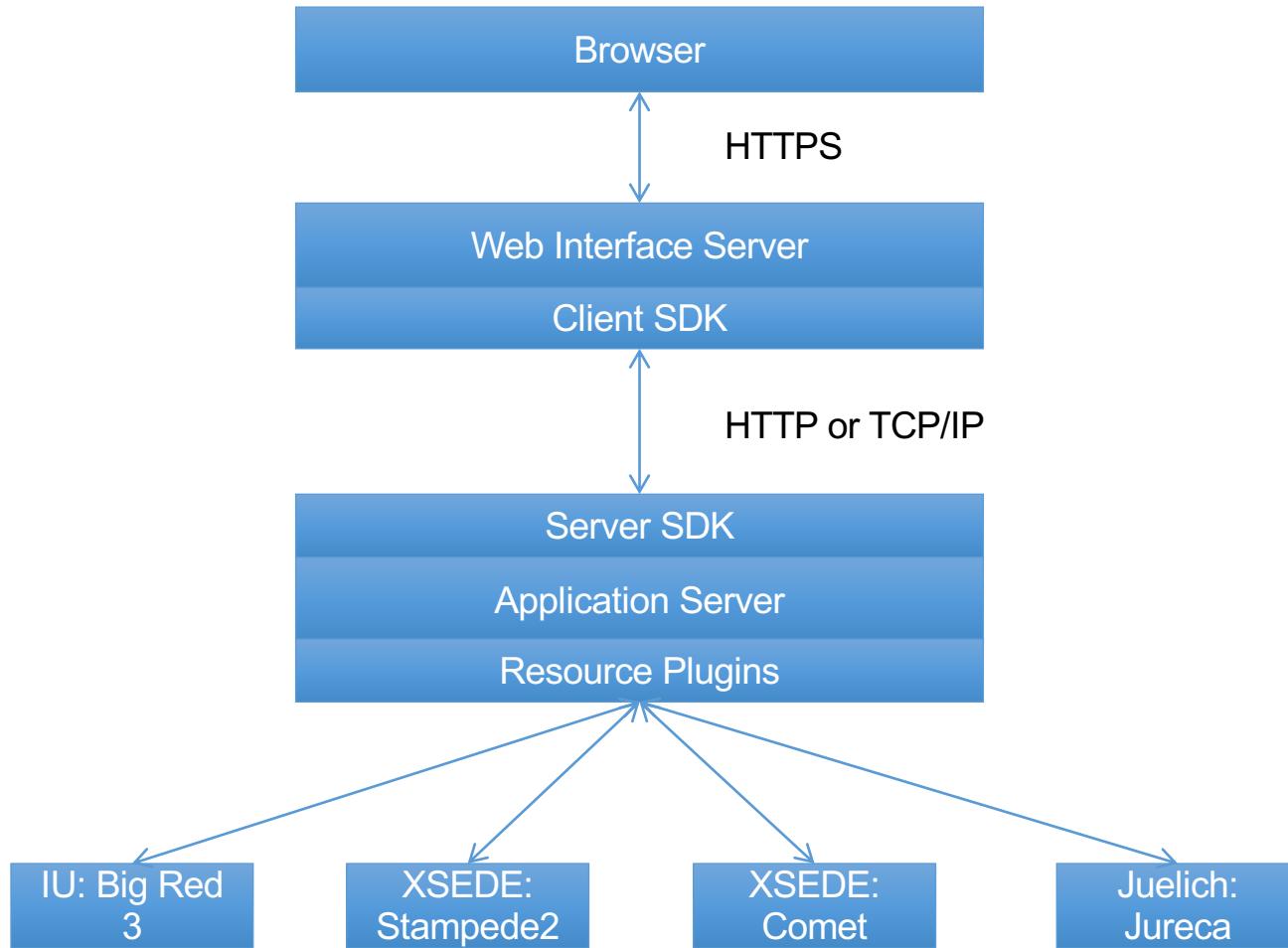
Lopez-Berganza, et al. *J Phys. Chem. A*(2015)



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

SEAGrid.org enabled workflow





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 (“transactions”)
- Continuous integration and deployment
- Security management



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

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

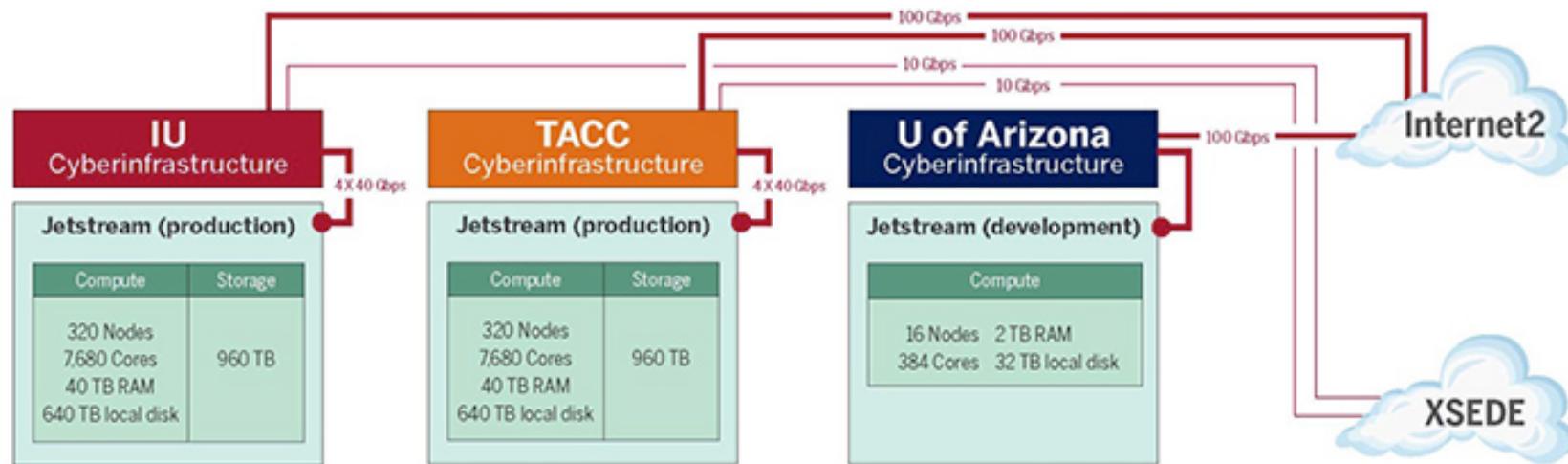


INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING



Science Engineering Cloud based on OpenStack



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

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



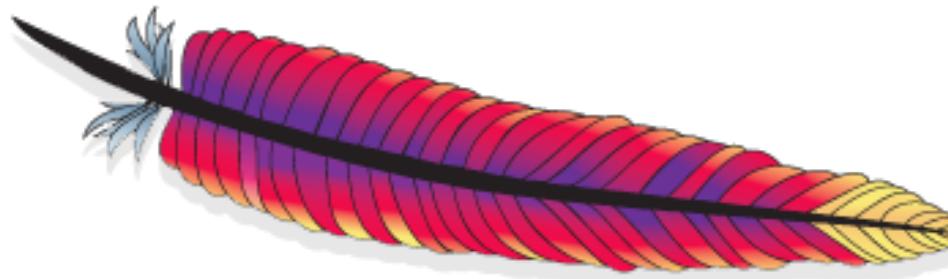
@RealGeneKim, genek@realgenekim.me



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Goal 3: Teach open source software practices



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Why Do We Teach This Class?

1. We are looking for students who like what we do and want to contribute to Apache Airavata.
2. Technologies change, and we need to keep up ourselves.



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

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



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

The Changing Way for Developing and Delivering Software

Microservices vs Monolithic Applications



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Monolithic Applications: Traditional Software Releases

- Software runs on clients' systems
- Software releases may be frequent, but they are still distinct
 - Firefox
 - OS system upgrades
- Traditional release cycles
 - Extensive testing
 - Alpha, beta, release candidates, and full releases
- Extensive recompiling and testing required after code changes
- Code changes require the entire release cycle to be repeated



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Microservices: Software as a Service

- Does your software run as an online service?
- Traditional release cycles don't work well
 - May make releases many times per day
 - Test-release-deploy takes too long
- You can be a little more tolerant of bugs discovered after release if you can fix quickly or roll back quickly.
- Get new features and improvements into production quickly.



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

What Is a Microservice?

- Develop a single application as a suite of small services
- Each service runs in its own process
- Services communicate with lightweight mechanisms
 - “Often an HTTP resource API”
 - But that has some problems
 - Messaging and hybrid approaches
- Independently deployable by fully automated deployment machinery.
- Minimum of centralized management of these services,
 - May be written in different programming languages
 - May use different data storage technologies.



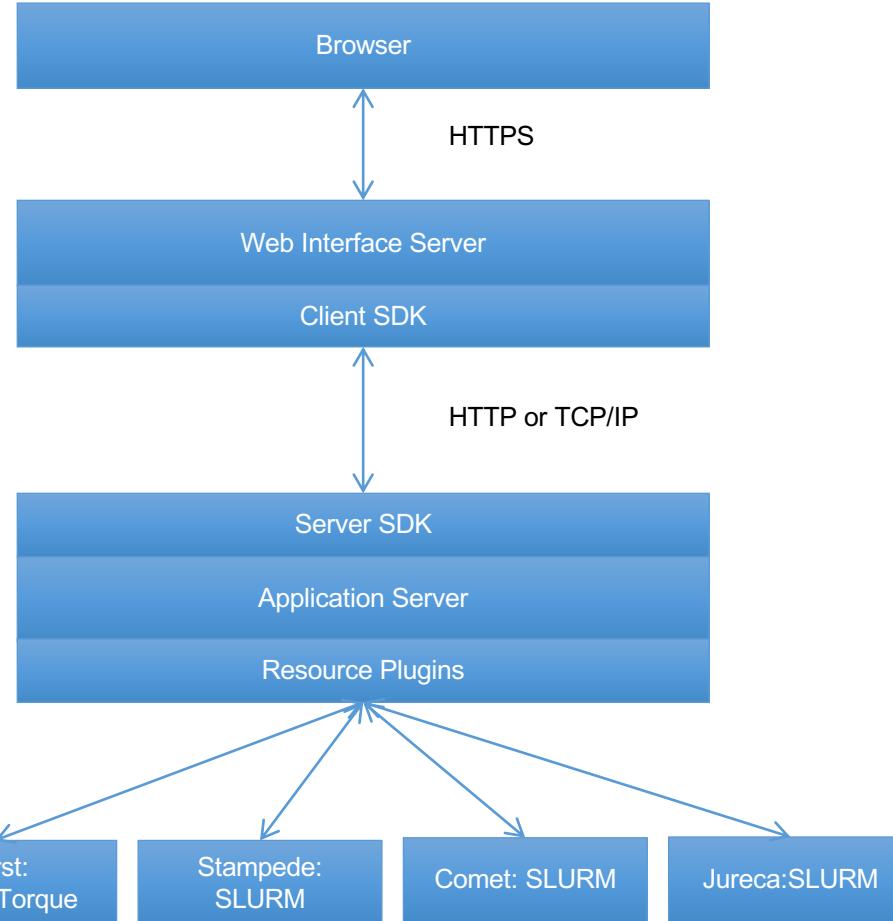
INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

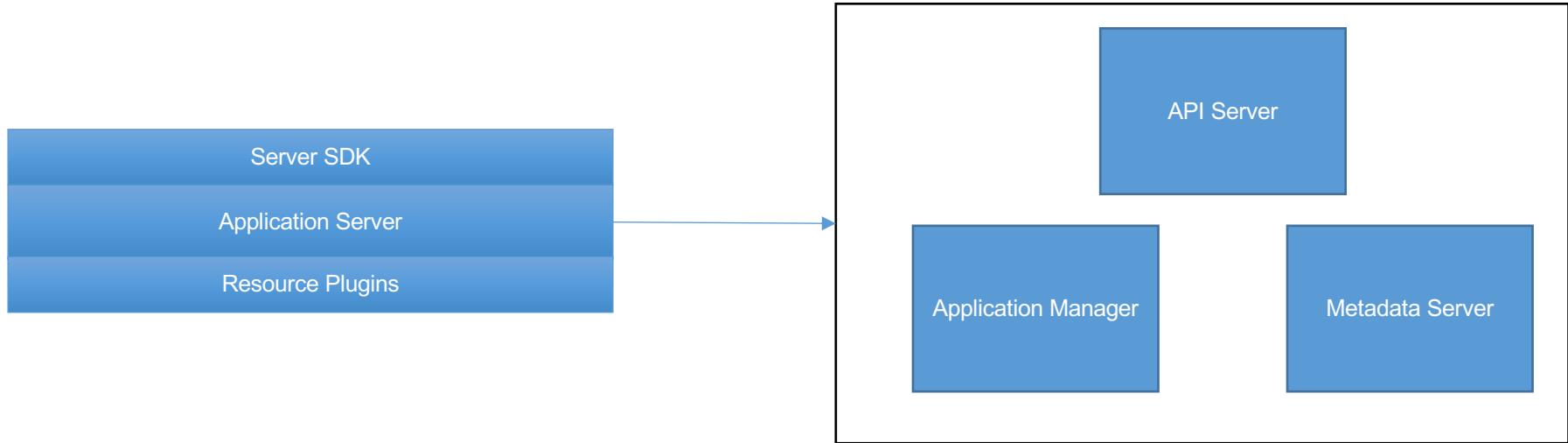
<http://martinfowler.com/articles/microservices.html>

Recall the Gateway Octopus Diagram

We will focus
on this piece



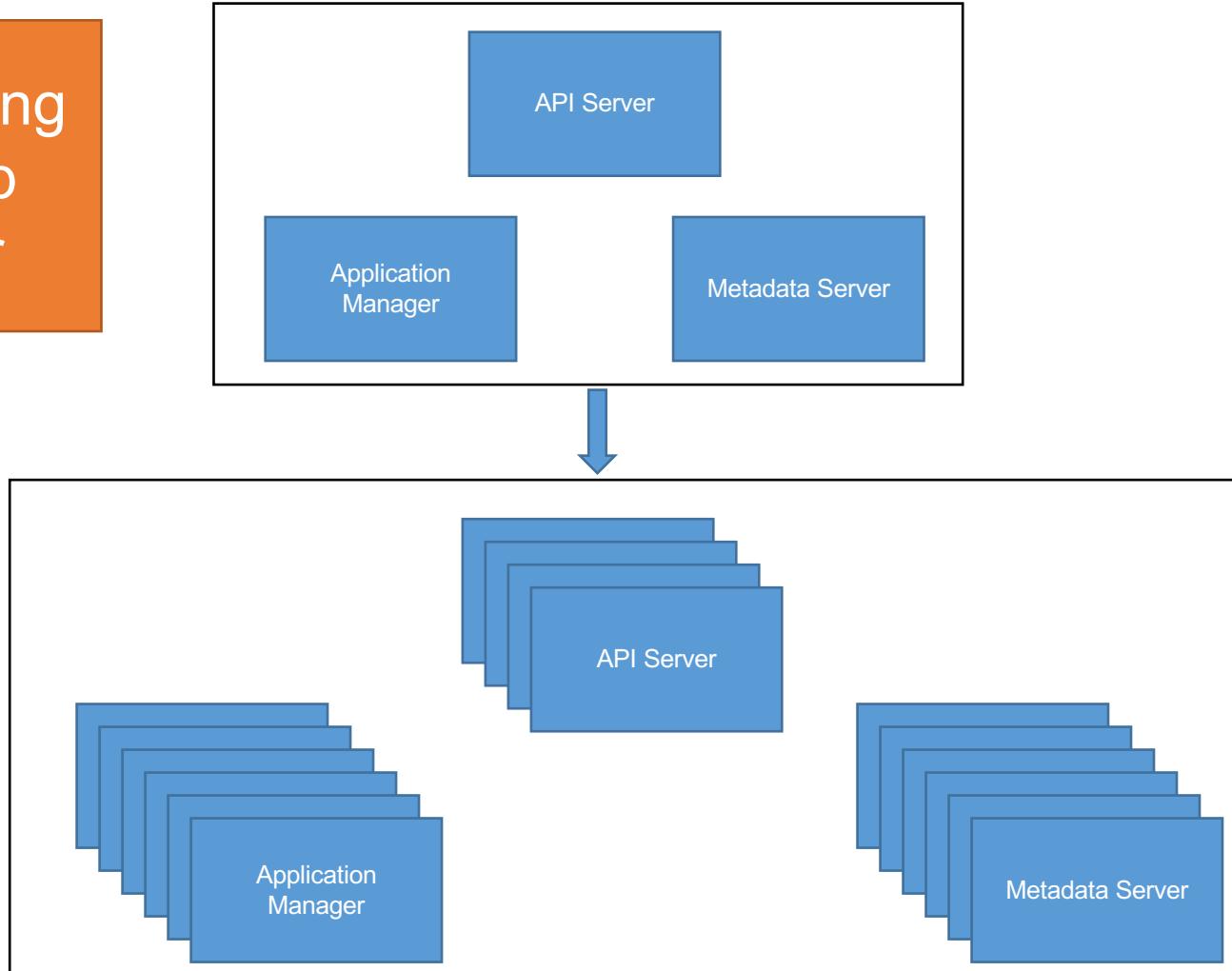
Basic Components of the Gateway App Server



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Decoupling the App Server



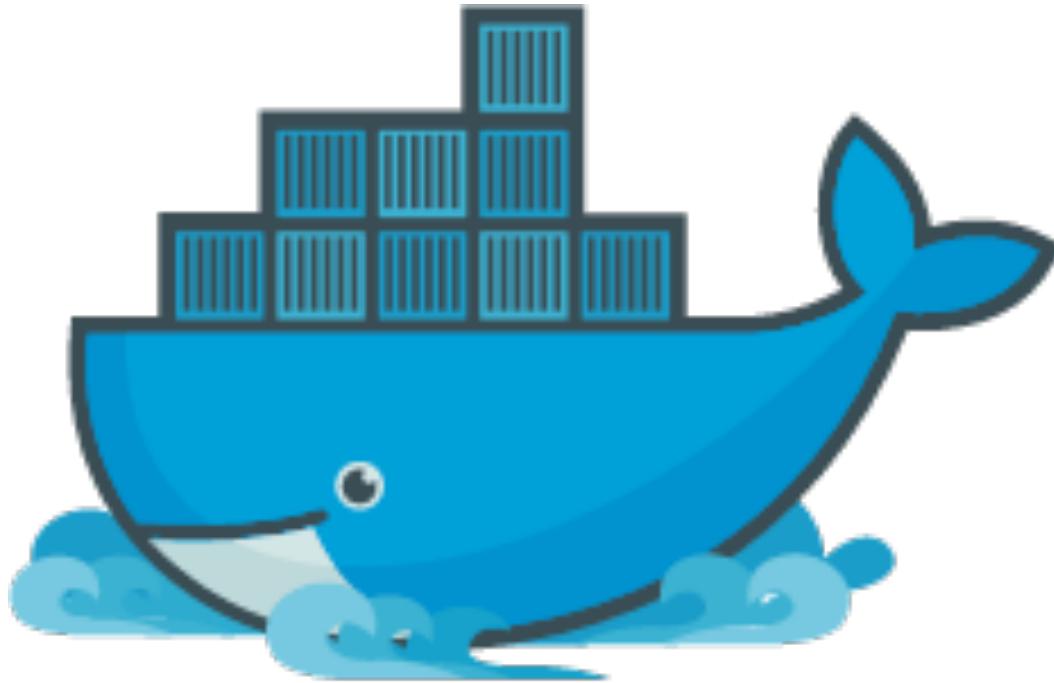
How Do We Package and Where Do We Run All Those MicroServices?

On the Cloud? In the Matrix?



INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Virtualization, Containers, Docker



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

How Do Microservices Communicate?

Push, Pull e.t.c



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Messaging Systems: RabbitMQ, Apache Kafka



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

How Can Components Expose their APIs and Data Models to Other Components?

And can we make this programming language
independent?



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

API and Metadata Model Design

Apache Thrift™



envoy



RESTful API
GET PUT POST DELETE



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

How Can I Discover, Monitor, and Manage Services?

Can we learn some lessons from distributed systems research?



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Distributed State Management: Consul, ETCD, Zookeeper



INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

How Do I Manage Logs from Microservices

And detect if there are problems



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING



Prometheus



How Can I Secure Microservices?

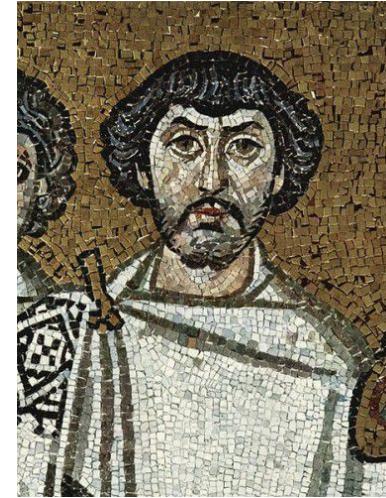
How do I manage user identities, authentication and authorization?



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Security: OAuth2 and OpenIDConnect



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

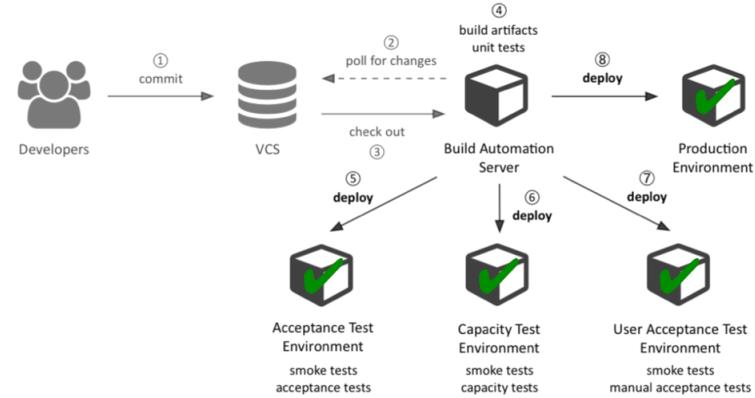
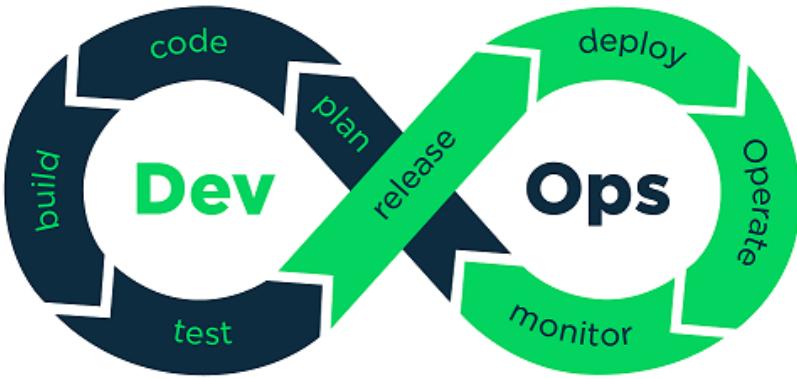
How Can We Automate All of This?

How can we make our infrastructure reproducible?



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING



Next Lecture

- More details about the first two project assignments
- Recap for any new students
- Bring your questions



INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING