LUDDY
SCHOOL OF INFORMATICS,
COMPUTING, AND ENGINEERING

CSCI-B 649 Topics in Systems:
Applied Distributed Systems

Spring 2022 Course Introduction

January 11th 2021

Suresh Marru, Marlon Pierce

Cyberinfrastructure

- "Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked by high speed networks to make possible scholarly innovation and discoveries not otherwise possible."
  - In short, it typically refers to information technology research systems that democratizes particularly powerful and advanced capabilities such as supercomputers.

Cyberinfrastructure Integration Research Center (CIRC)

CIRC's core mission is to accelerate research, discovery and collaboration through the creation, integration and operation of user-centric cyberinfrastructure that benefits scientific communities.

Course Instructors

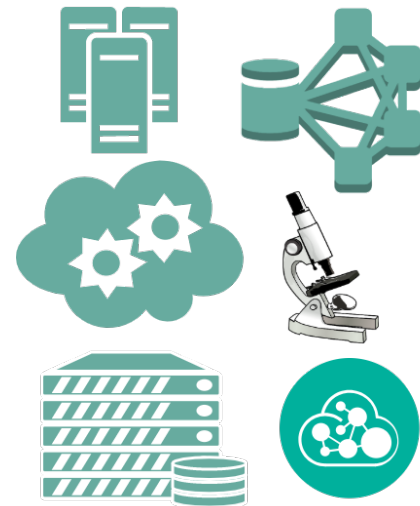Marlon Pierce   Suresh Marru

Class Introductions

What are we expecting you to get out of this class?

- A fusion of conceptual skills and "scientific way" of making choices.

- The course is tailored to use tools and technologies relevant in 2022 but our expectation is you will learn how to make choices not necessarily be a tutorial on a buzzy technology.

  - Our definition of a good student is someone who understand the difference between the two.
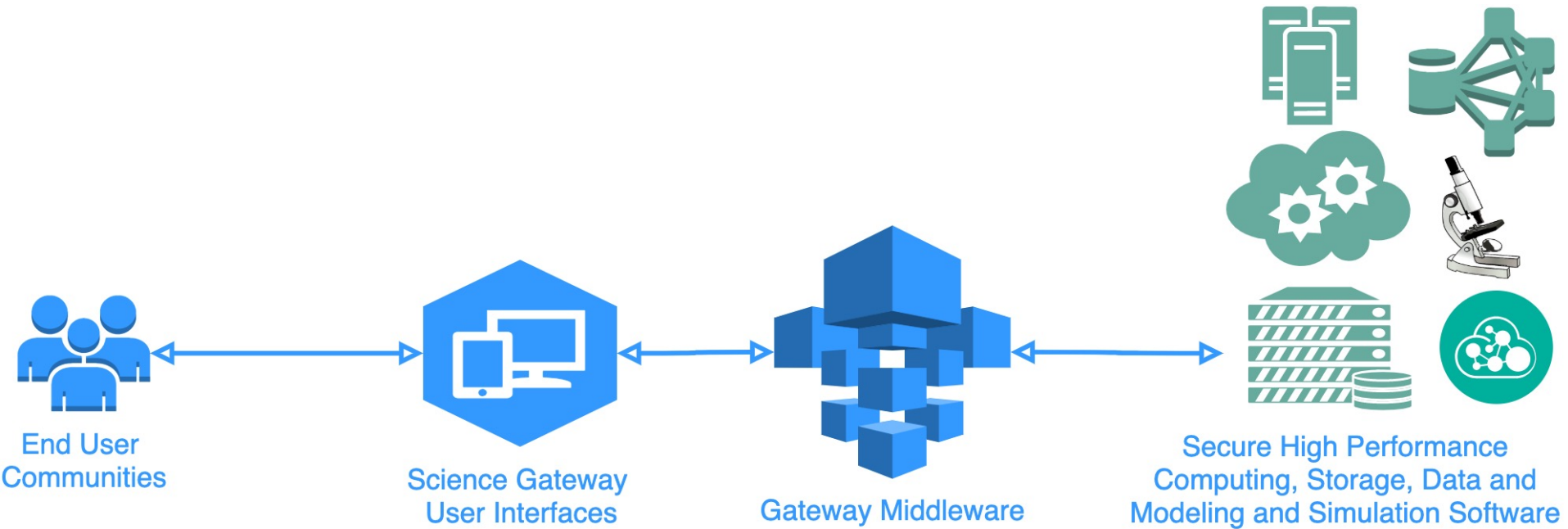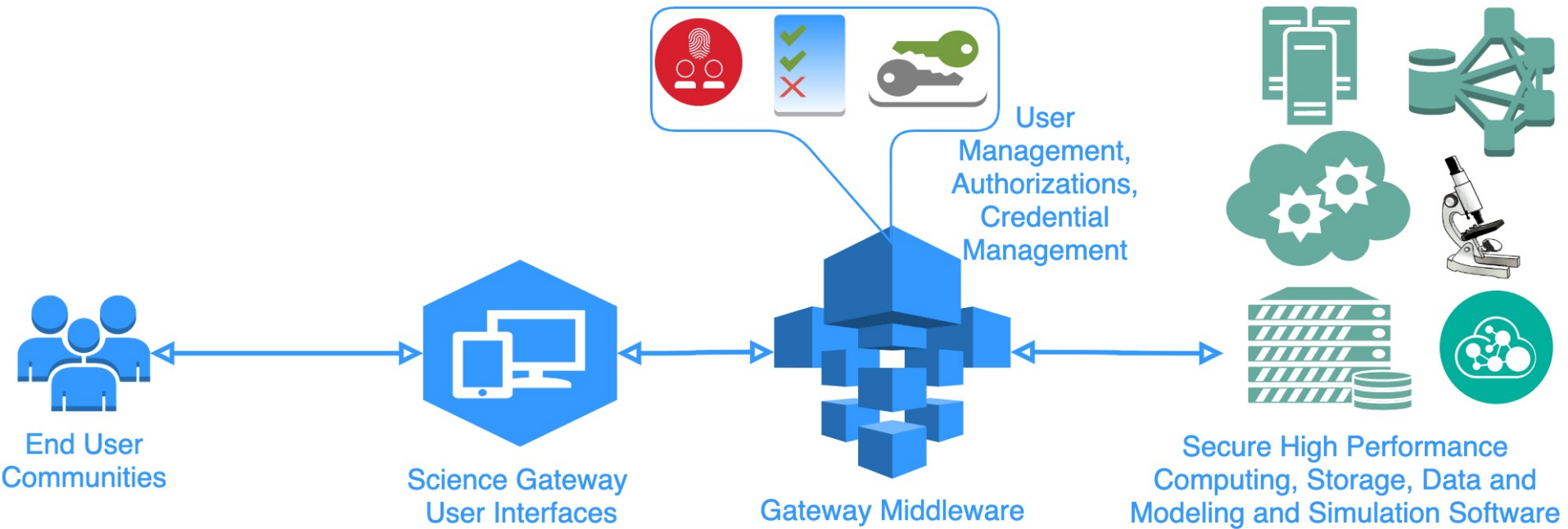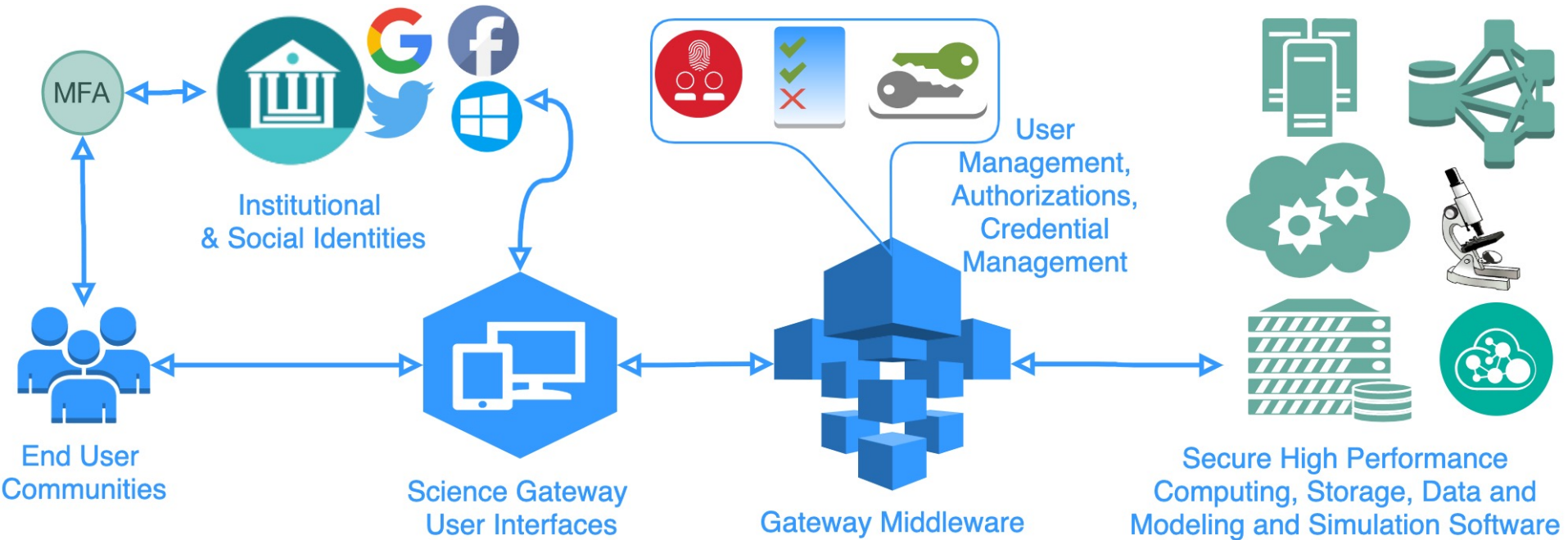
Scientists

Supercomputers, Storage, Data and Modeling and Simulation Software

CYBERINFRASTRUCTURE INTEGRATION RESEARCH CENTER
PERVASIVE TECHNOLOGY INSTITUTE

End User Communities ↔ Science Gateway User Interfaces ↔ Gateway Middleware ↔ Secure High Performance Computing, Storage, Data and Modeling and Simulation Software

End User Communities ↔ Science Gateway User Interfaces ↔ Gateway Middleware ↔ Secure High Performance Computing, Storage, Data and Modeling and Simulation Software

User Management, Authorizations, Credential Management

MFA

Institutional & Social Identities

End User Communities

Science Gateway User Interfaces

User Management, Authorizations, Credential Management

Gateway Middleware

Secure High Performance Computing, Storage, Data and Modeling and Simulation Software

CYBERINFRASTRUCTURE INTEGRATION RESEARCH CENTER
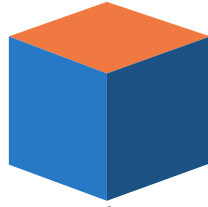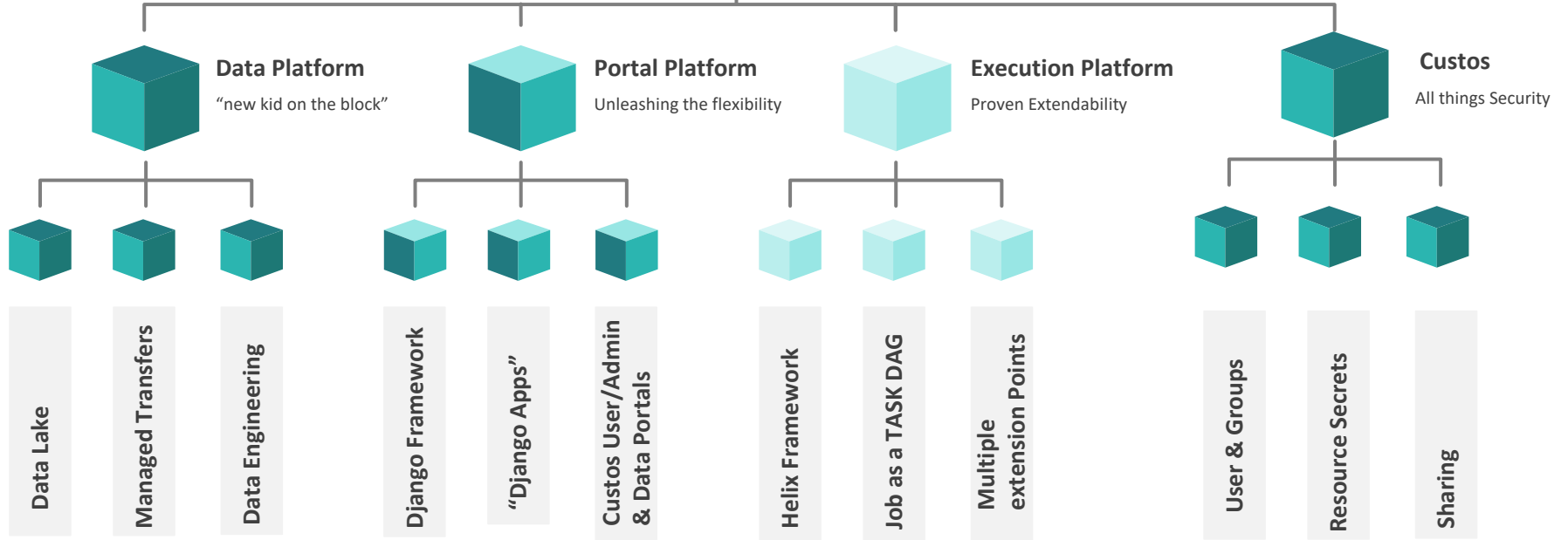PERVASIVE TECHNOLOGY INSTITUTE

Apache Airavata

- **Gateway Developers:** Open source software for building science gateways
- **Users:** Use it to transfer data and execute remote applications and pipelines on distributed resources
- **Teams:** Create, organize, clone, and share computational experiments
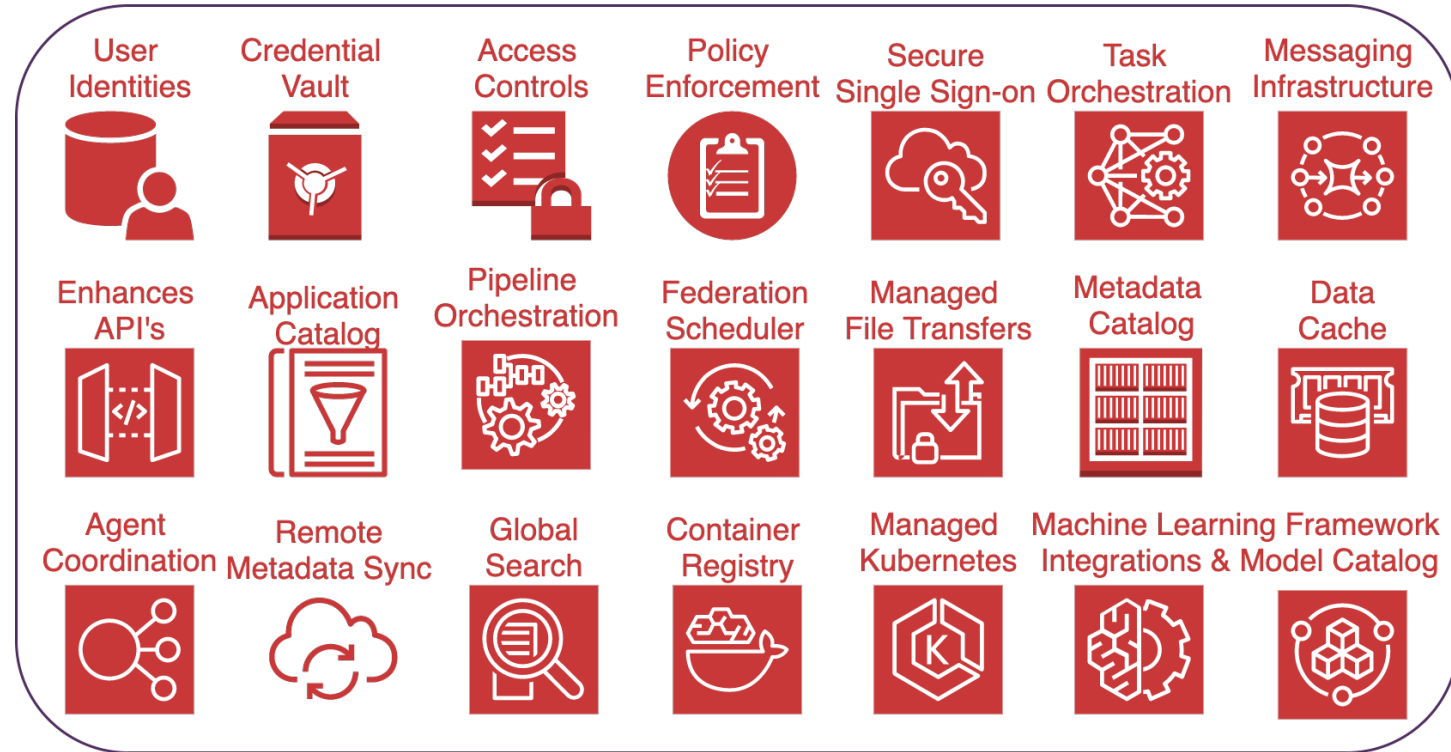- **Software Providers:** Make scientific software available as a service

Middleware (Distributed Systems Components)

User Identities — Credential Vault — Access Controls — Policy Enforcement — Secure Single Sign-on — Task Orchestration — Messaging Infrastructure

Enhances API's — Application Catalog — Pipeline Orchestration — Federation Scheduler — Managed File Transfers — Metadata Catalog — Data Cache

Agent Coordination — Remote Metadata Sync — Global Search — Container Registry — Managed Kubernetes — Machine Learning Framework Integrations & Model Catalog

Distributed Systems Everywhere

Develop Data Services and visualizations for
https://worldview.earthdata.nasa.gov/

End product will need to look like this -
https://earthdata.nasa.gov/covid19/

IMPACT
Interagency Implementation
and Advanced Concepts Team

Spring 2022 Project

Microservices

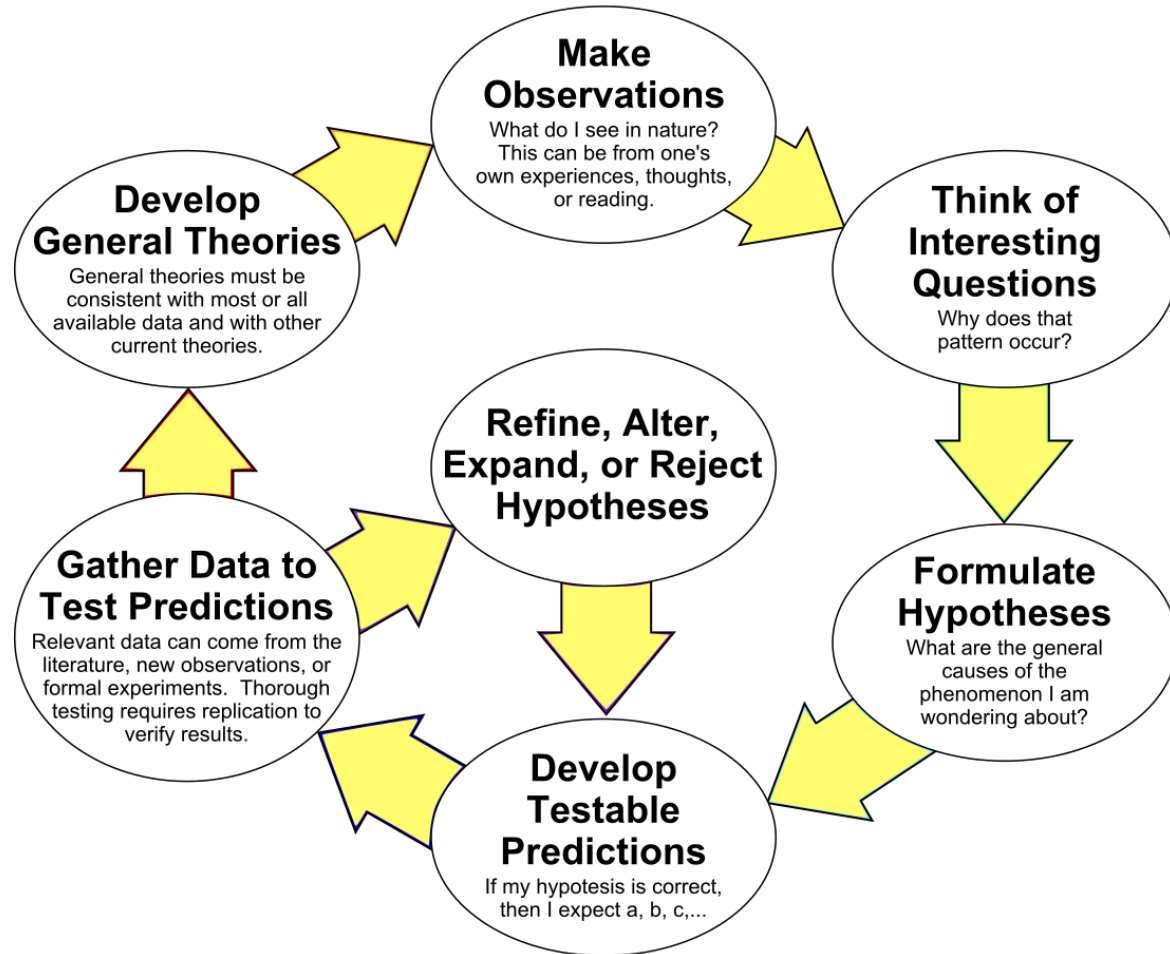Platform

HTTP

Cloud Storages
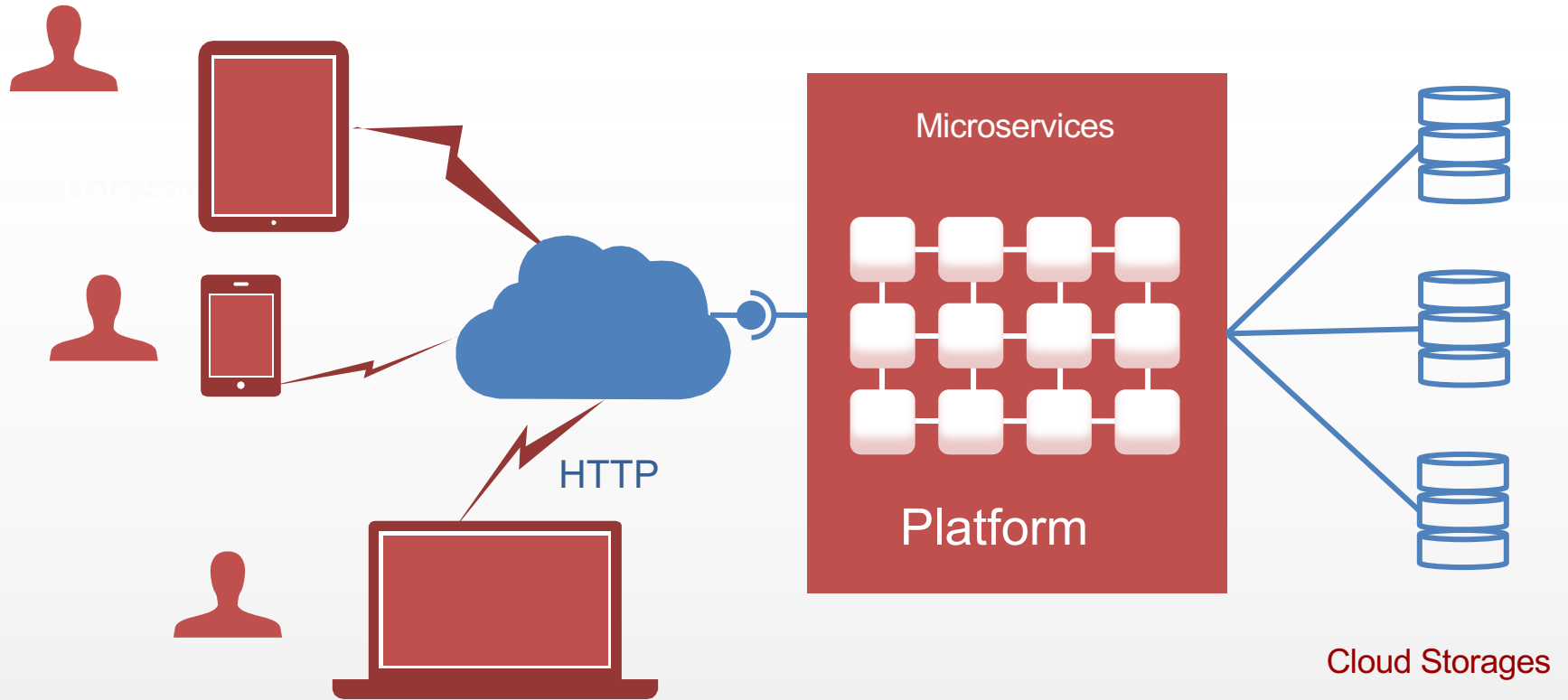
Applied Distributed Systems

- We will build user-centric distributed systems that mimic scientific research.

- This course will be project-based.

- You will build distributed systems.

The Scientific Method as an Ongoing Process

Make Observations
What do I see in nature? This can be from one's own experiences, thoughts, or reading.

Develop General Theories
General theories must be consistent with most or all available data and with other current theories.

Think of Interesting Questions
Why does that pattern occur?

Refine, Alter, Expand, or Reject Hypotheses

Gather Data to Test Predictions
Relevant data can come from the literature, new observations, or formal experiments. Thorough testing requires replication to verify results.

Formulate Hypotheses
What are the general causes of the phenomenon I am wondering about?

Develop Testable Predictions
If my hypotesis is correct, then I expect a, b, c,...

Structure of the Class

- We will have 4 project-based assignments
  - 90% of your grade
  - Projects will done by a team of 3 but graded individually.
  - 5 Bonus points/project for peer review (individual)
- Recorded Mid-term and final presentations
  - Opportunities to showcase selected presentations on
- 10% (bonus) will be attendance, classroom interactions and peer project interactions (auditable github communications, whatsapp msgs do not count).

Characteristics of a Good Technology Base

✓You are continually improving your code base

✓You are strategically adding major new capabilities

✓You get improvements expeditiously into production

✓You can replace key personnel

✓You get meaningful contributions

✓You have boring operations: the system as a whole doesn't break, security upgrades aren't a major hassle, etc.

✓Parts of your base get reused in other projects.
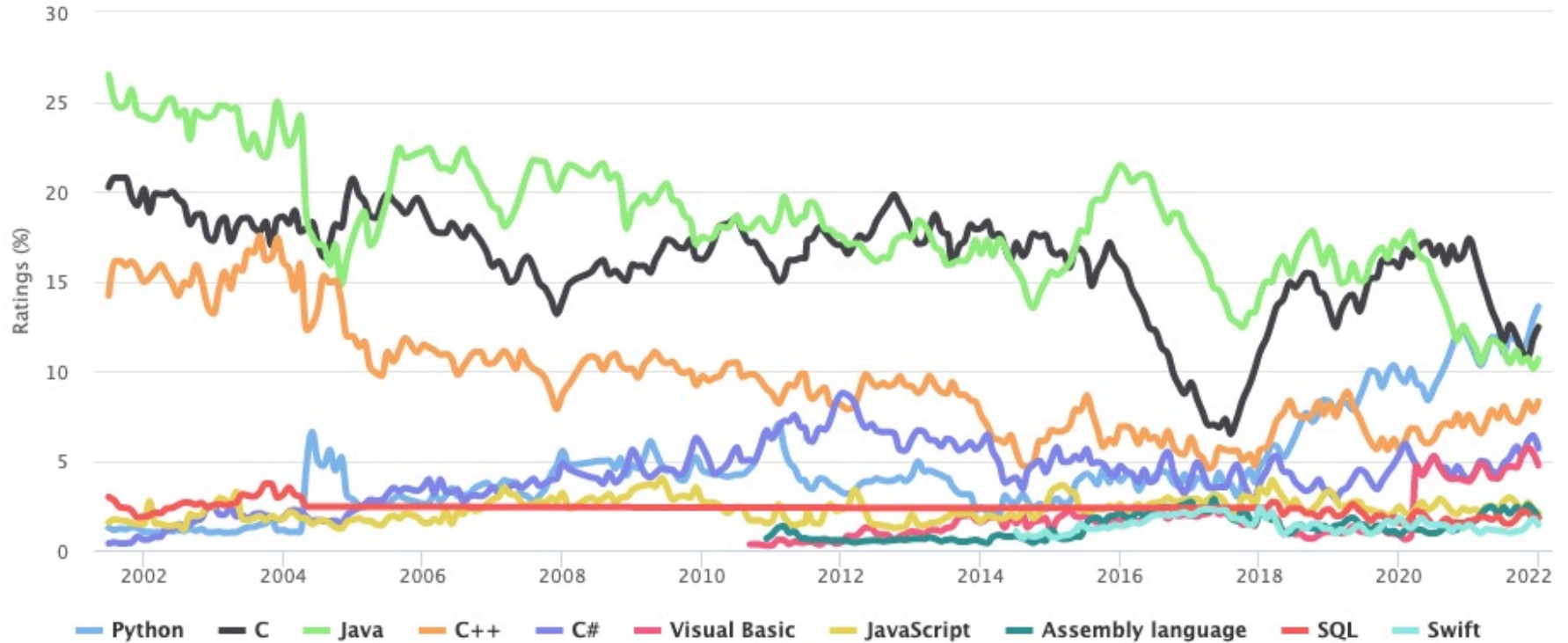
Cloud-Native Architecture Principles

Each service is broken by a functional capability

Services should be able to evolve independently, scale independently.
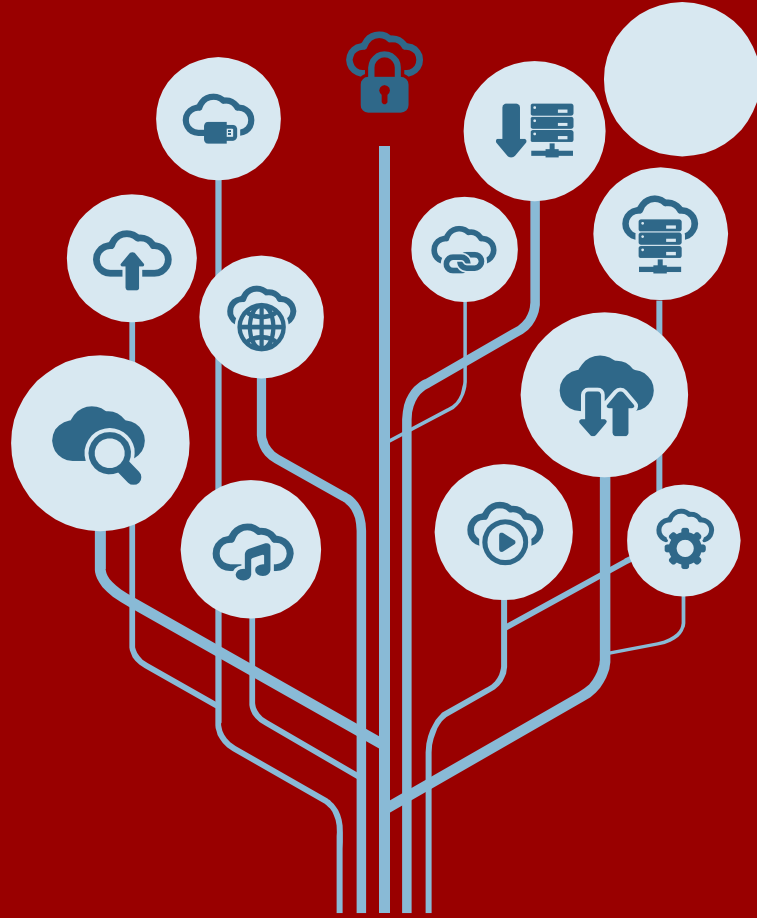
TIOBE Programming Community Index
Source: www.tiobe.com

Legend: Python, C, Java, C++, C#, Visual Basic, JavaScript, Assembly language, SQL, Swift

Programming Language "polyglotism"

Polygot – use at least 3 programming languages
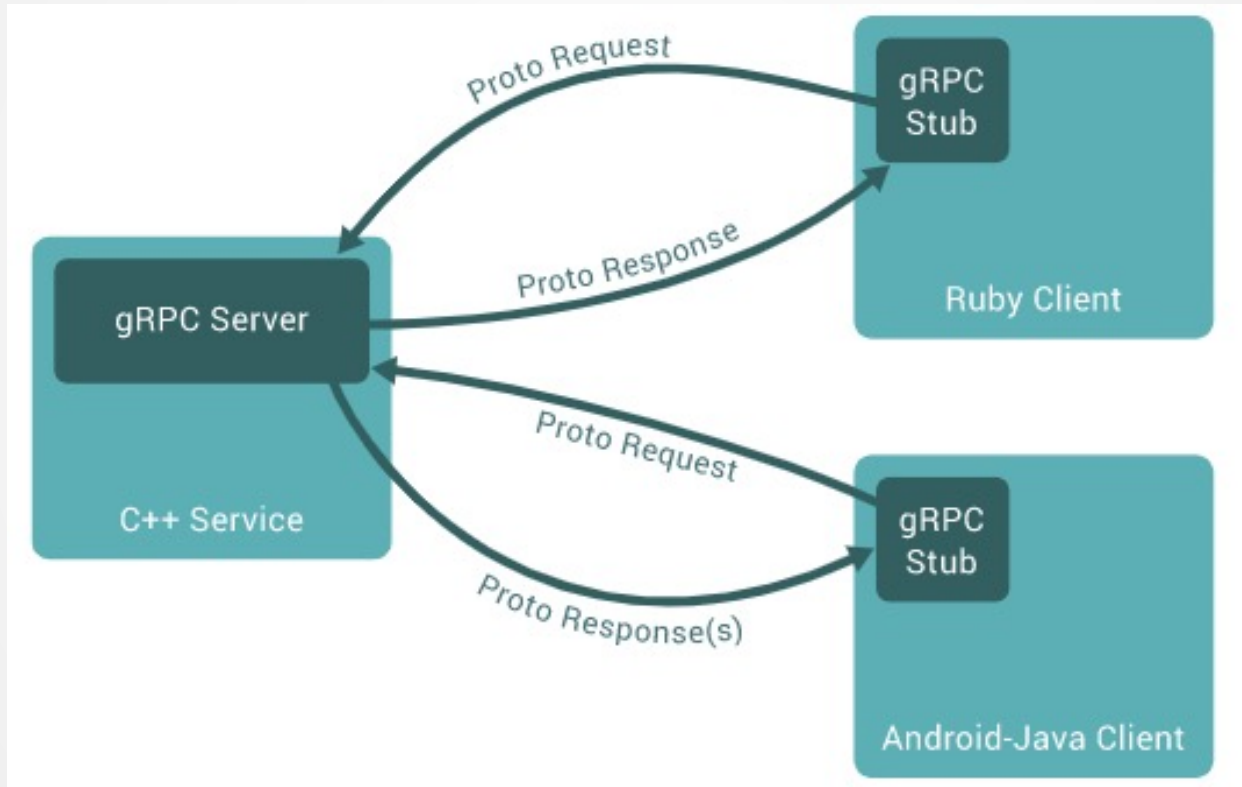
JAVA

C++

PYTHON

HTML/CSS/JS

CODING

All components (including UI) need to use a build framework: Make, Maven, Bower…

gRPC, Thrift, Protocol Buffers

Cyber Security at all layers:

Go beyond
Authentication and
Authorization

Securing all
communications

Project Mechanics

- Create your project team.

- We will populate your team repo

- Use all GitHub software engineering tools to start working on your project.

- Make your repos and wiki's ready for peer-review.

- Peer-reviews will be your open source user community, your project team is the PMC - https://www.apache.org/foundation/governance/pmcs.

- You submit the project for grading.

- TA's will grade the work of the team and peer reviewers and the team's response to peer reviews.