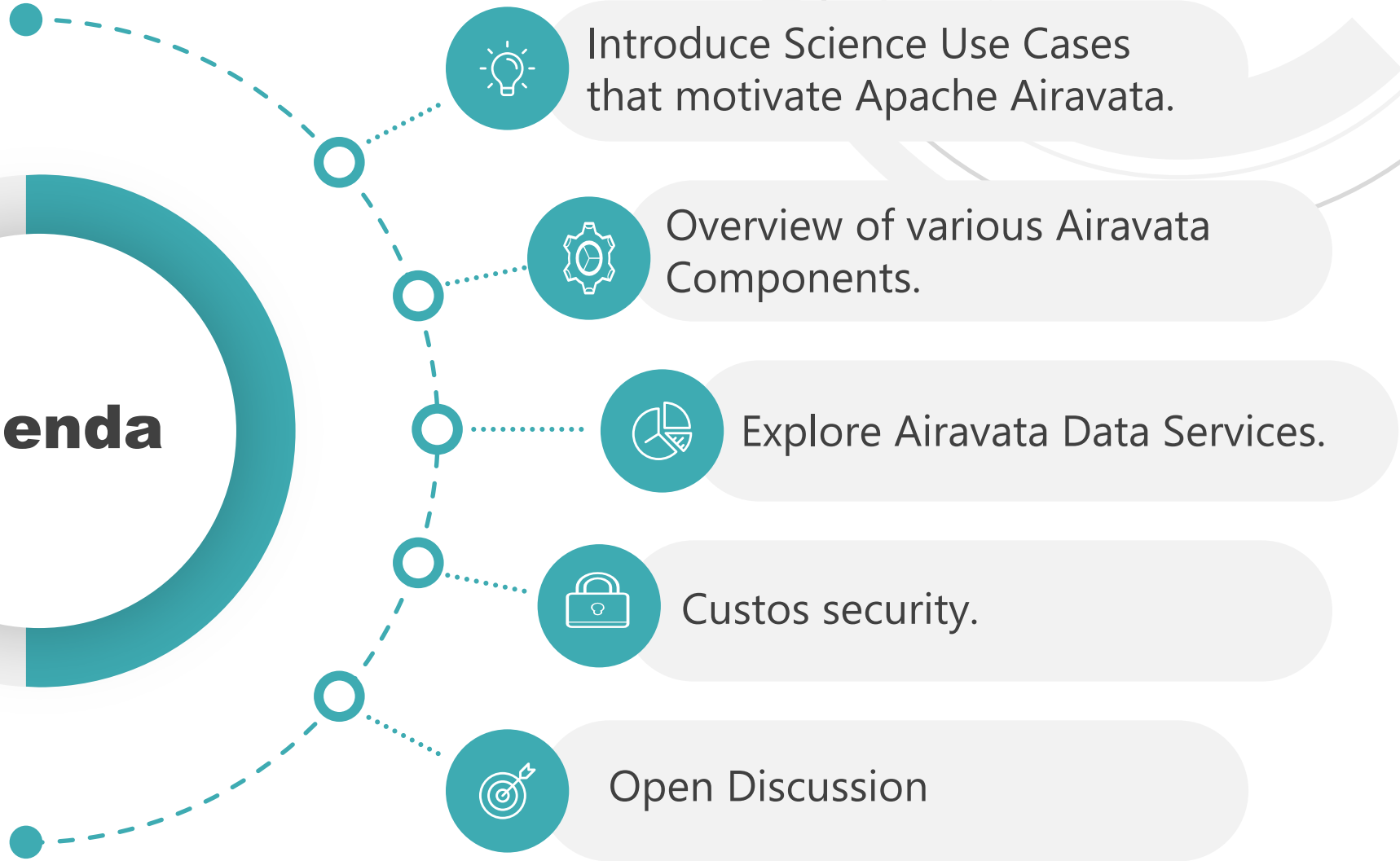
Apache Airavata

Suresh Marru
Indiana University
NASA Contractor (IMPACT)

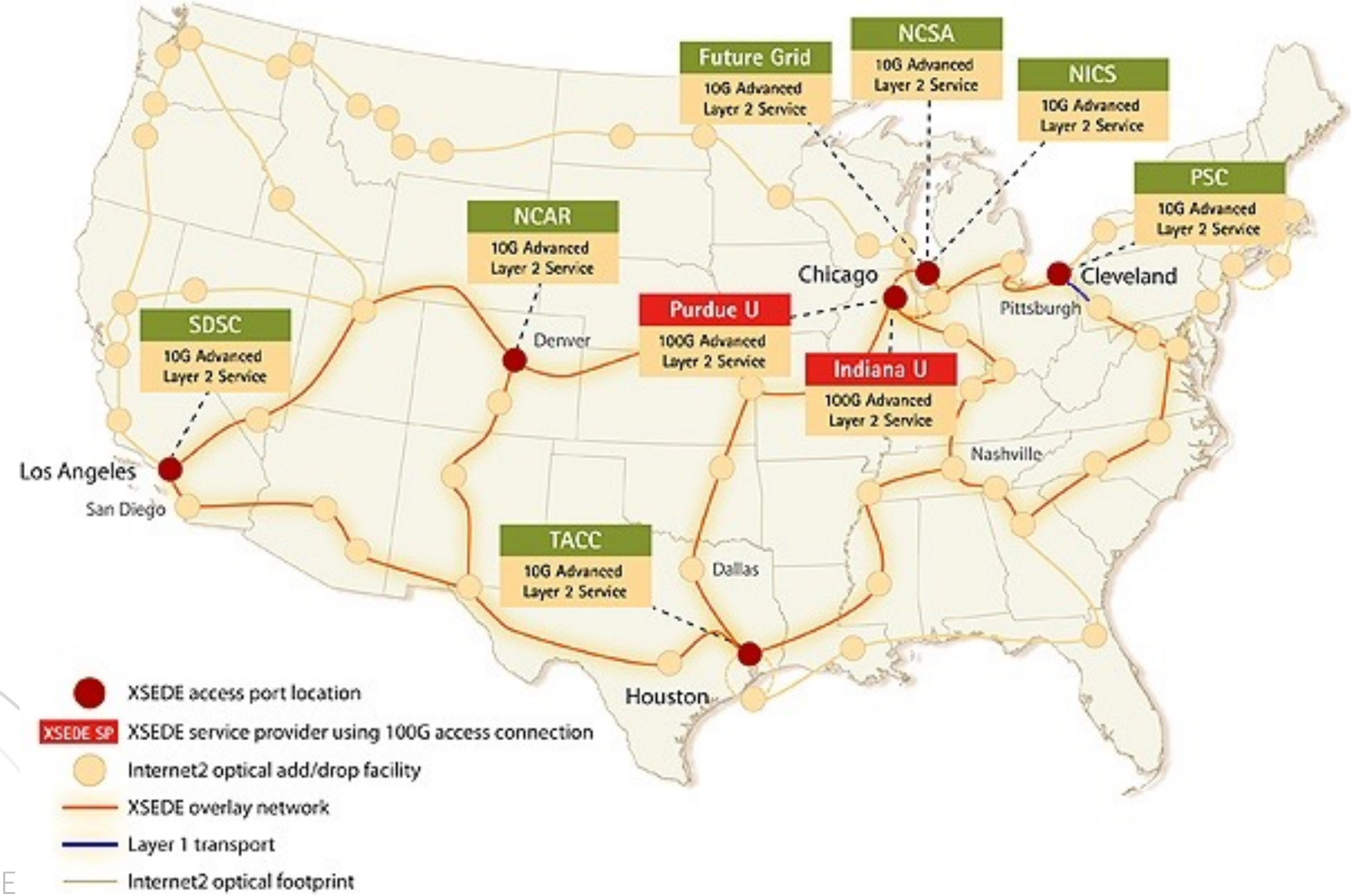
APACHE AIRAVATA

Agenda



- Introduce Science Use Cases that motivate Apache Airavata.
- Overview of various Airavata Components.
- Explore Airavata Data Services.
- Custos security.
- Open Discussion

APACHE AIRAVATA



APACHE AIRAVATA

Example: COVID-19 HPC Consortium

The COVID-19 High Performance Computing (HPC) Consortium members manage a range of computing capabilities that span from small clusters to some of the largest supercomputers in the world.

Members

Industry

- IBM
- Amazon Web Services
- AMD
- D. E. Shaw Research
- Dell Technologies
- Google Cloud
- Hewlett Packard Enterprise
- Microsoft
- NVIDIA
- Intel

- Massachusetts Institute of Technology
- Rensselaer Polytechnic Institute
- University of Illinois
- University of Texas at Austin
- University of California - San Diego
- Carnegie Mellon University
- University of Pittsburgh
- Indiana University
- Massachusetts Green High Performance Computing Center (MGHPCC)
- University of Wisconsin-Madison
- Ohio Supercomputer Center
- UK Digital Research Infrastructure
- CSCS – Swiss National Supercomputing Centre
- SNIC PDC – Swedish National Infrastructure for Computing, Center for High Performance Computing
- Arizona State University
- University of Alabama-Birmingham
- Purdue University

Department of Energy National Laboratories

- Argonne National Laboratory
- Lawrence Livermore National Laboratory
- Los Alamos National Laboratory
- Oak Ridge National Laboratory
- Lawrence Berkeley National Laboratory
- Sandia National Laboratories
- Idaho National Laboratory

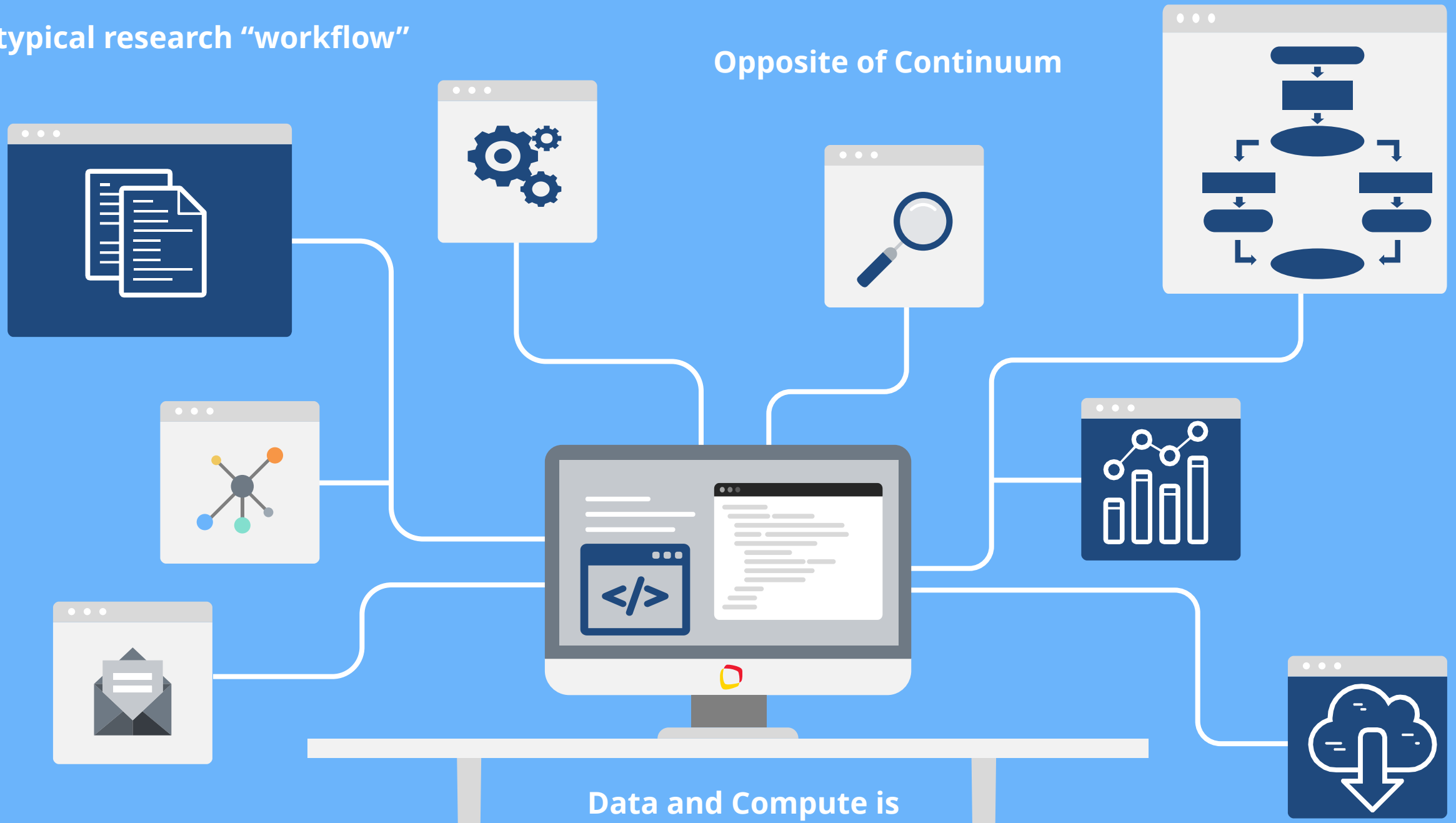
Federal Agencies

- National Science Foundation
  - XSEDE
  - Pittsburgh Supercomputing Center (PSC)
  - Texas Advanced Computing Center (TACC)
  - San Diego Supercomputer Center (SDSC)
  - National Center for Supercomputing Applications (NCSA)
  - Indiana University Pervasive Technology Institute (IUPTI)
  - Open Science Grid (OSG)
  - Purdue University Research Computing (RCAC)
- NASA

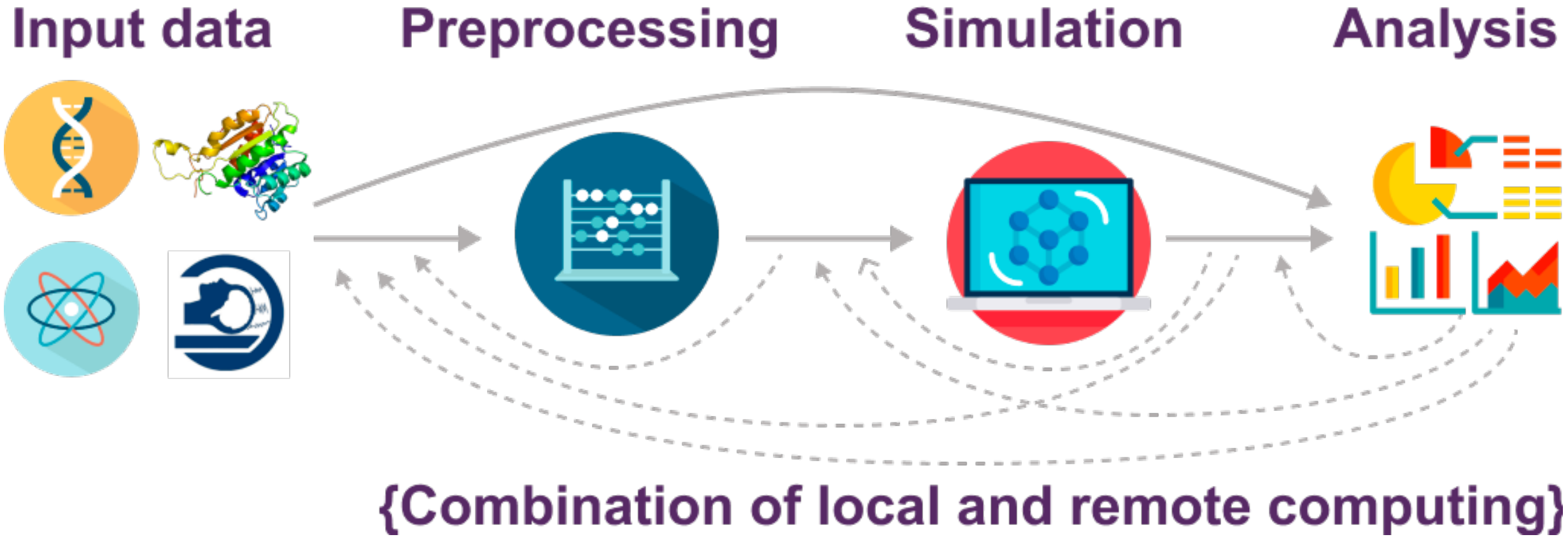
AIRAVATA

A typical research “workflow”

Opposite of Continuum

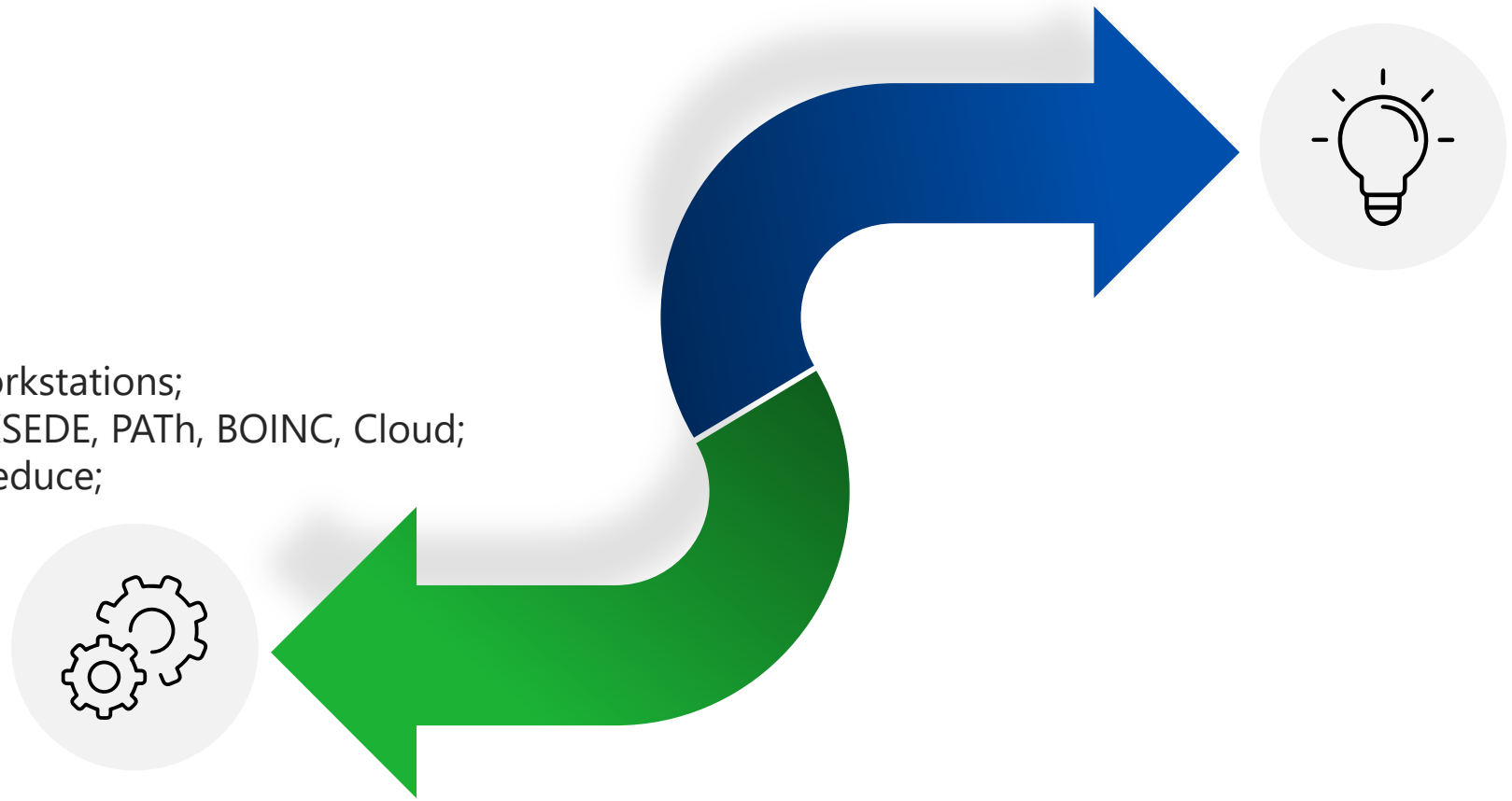


Science is not linear but very iterative at all steps



Apache Airavata: Bridging Science and Cyberinfrastructure

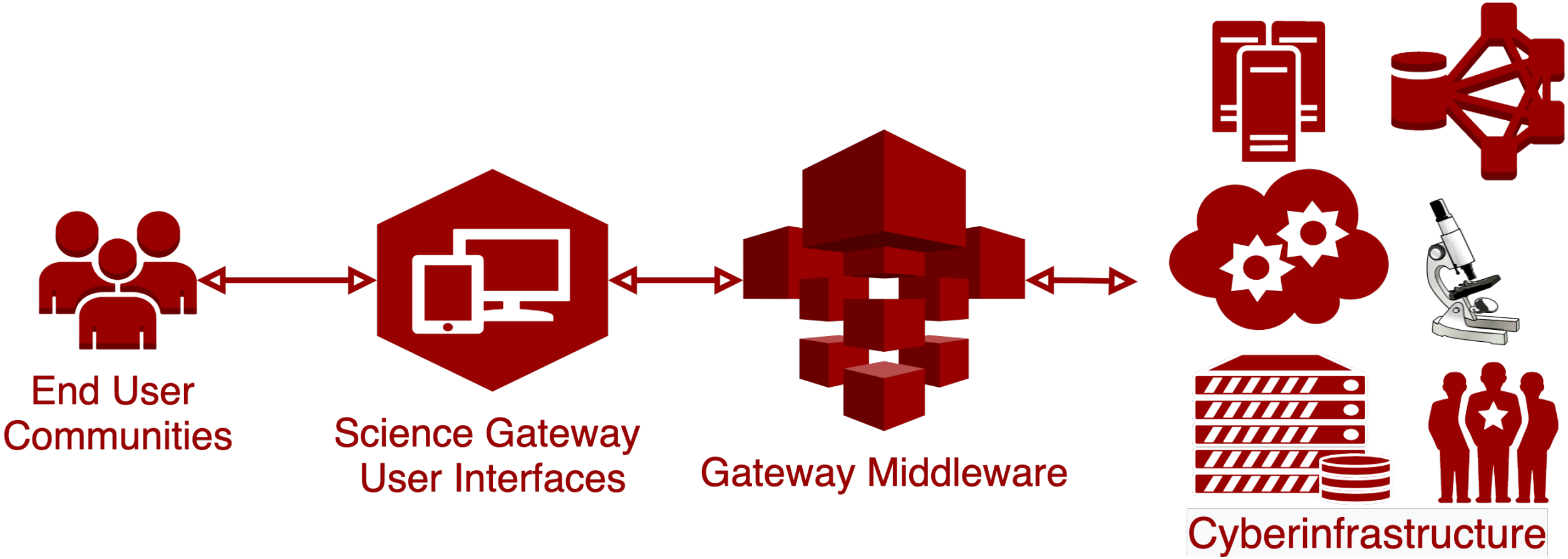
Diverse Science Use cases

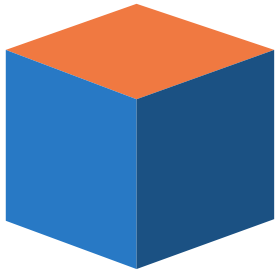


Powerful Laptops, workstations;
Campus Resources, XSEDE, PATh, BOINC, Cloud;
MPI, OpenMP, MapReduce;
AI, ML, DL …
…….

Diverse Cyberinfrastructure

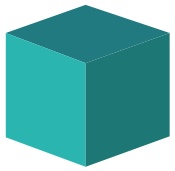
Science Gateways:
Science-centric views to Cyberinfrastructure





Apache Airavata

PMC & Integrated Platform



Data Platform

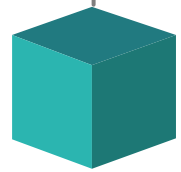
"new kid on the block"

Portal Platform

Unleashing the flexibility

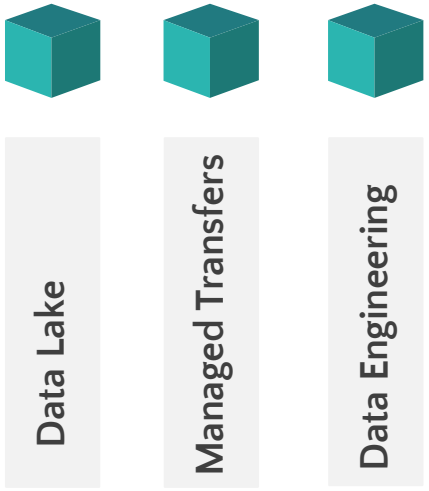
Execution Platform

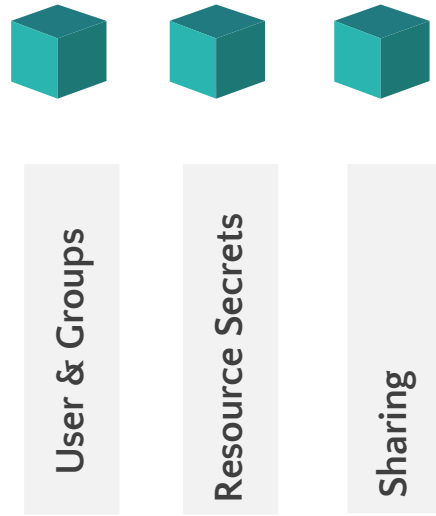
Proven Extendability



Custos

All things Security





What Is Apache Airavata?

Middleware for managing users, workflow executions, data transfers, and metadata for science gateways

Airavata can integrate university, XSEDE, and other resources that your gateway needs.

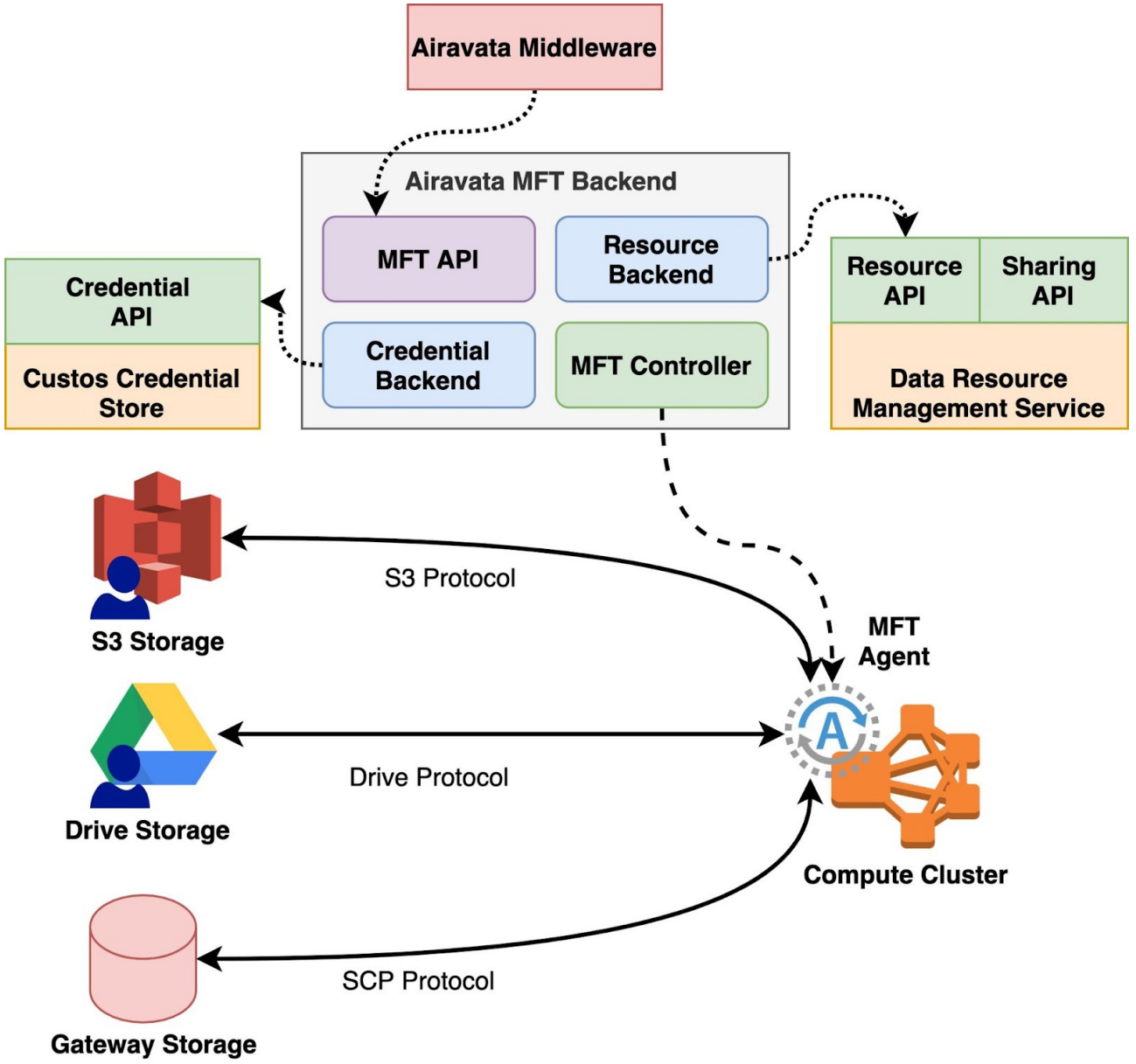
https://github.com/apache/airavata-custos

Custos Security

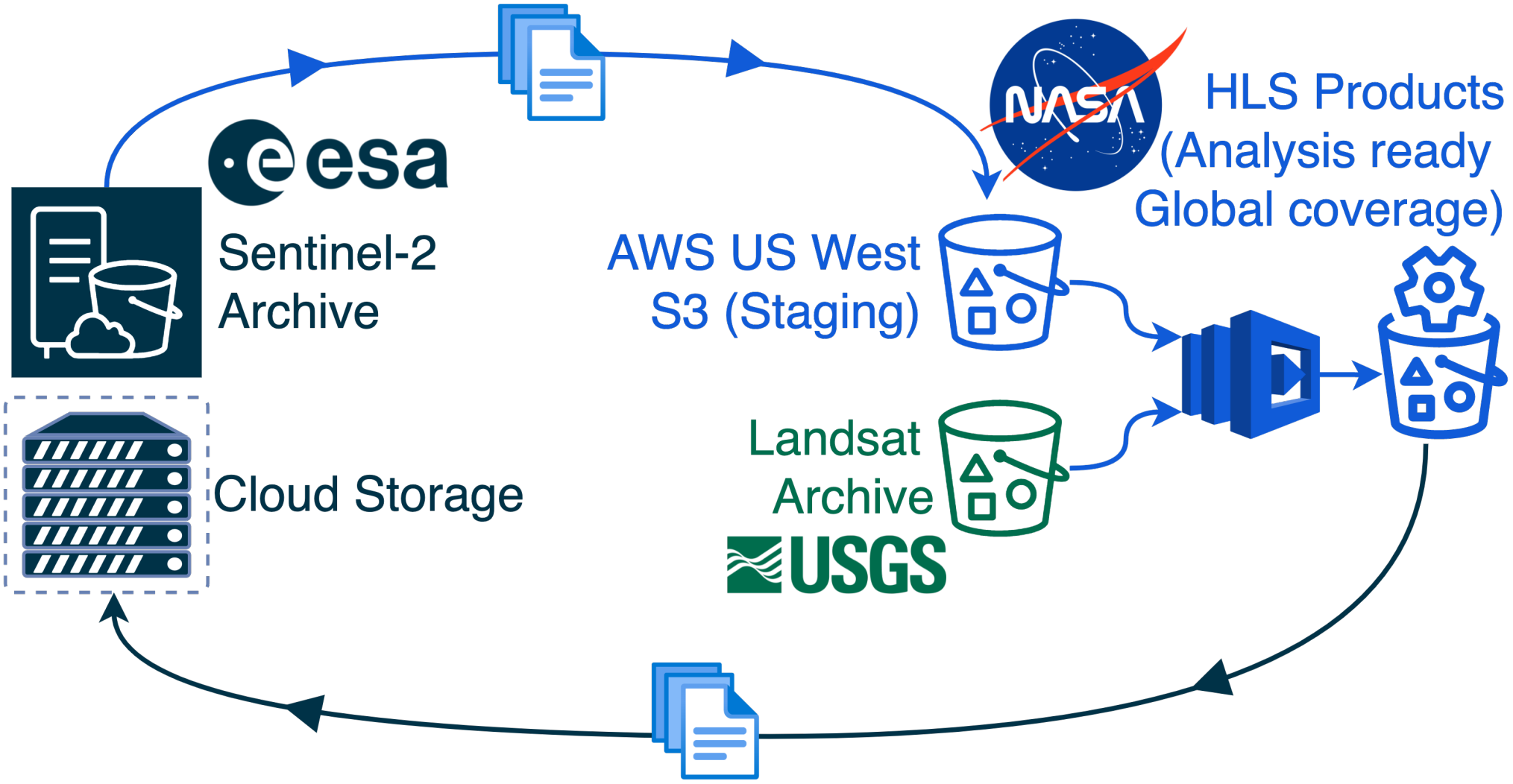
All inclusive open-source software and services

Federated Authentication;
Fine Grained Authorizations;
User, group and secret management;
controlled data sharing.

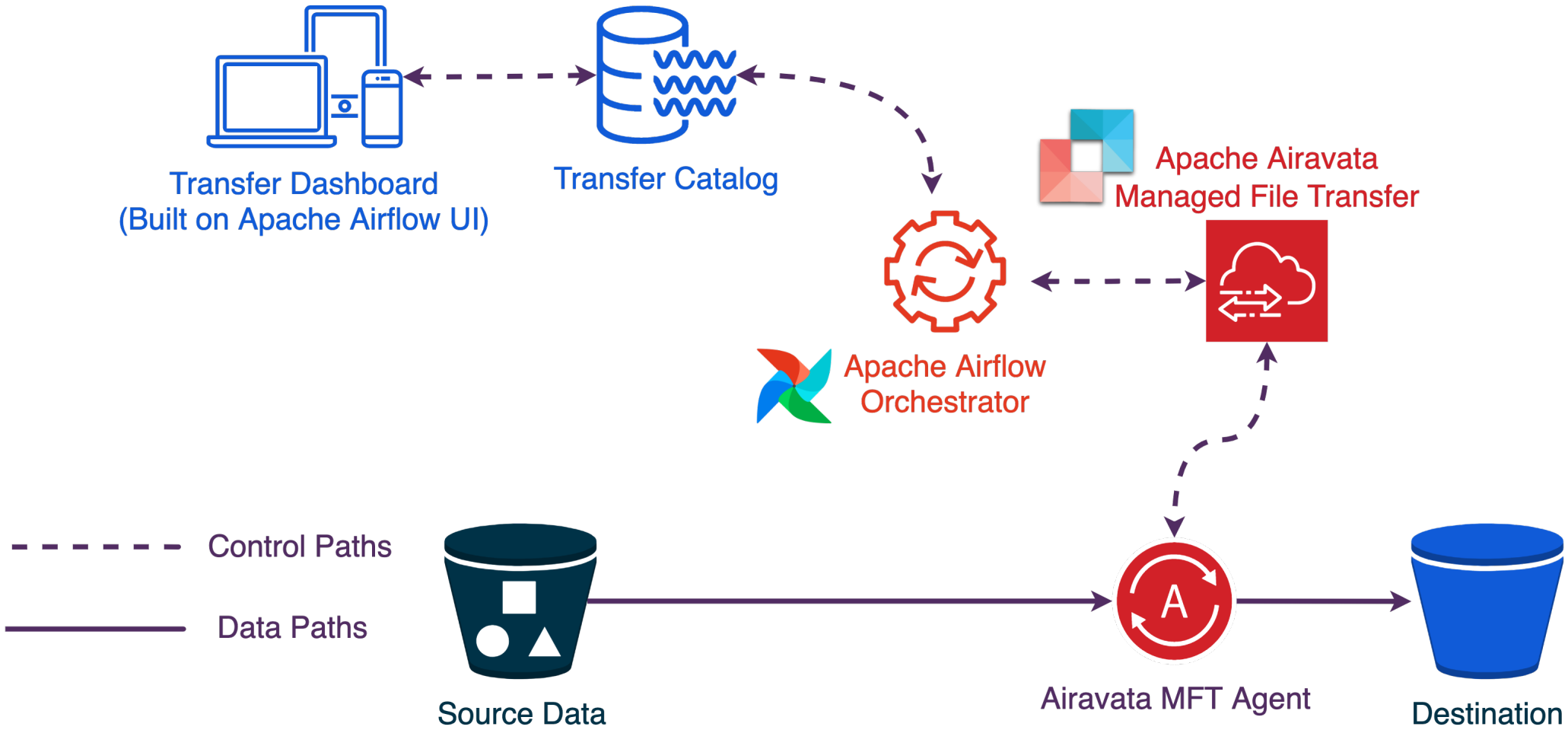
https://github.com/apache/airavata-mft



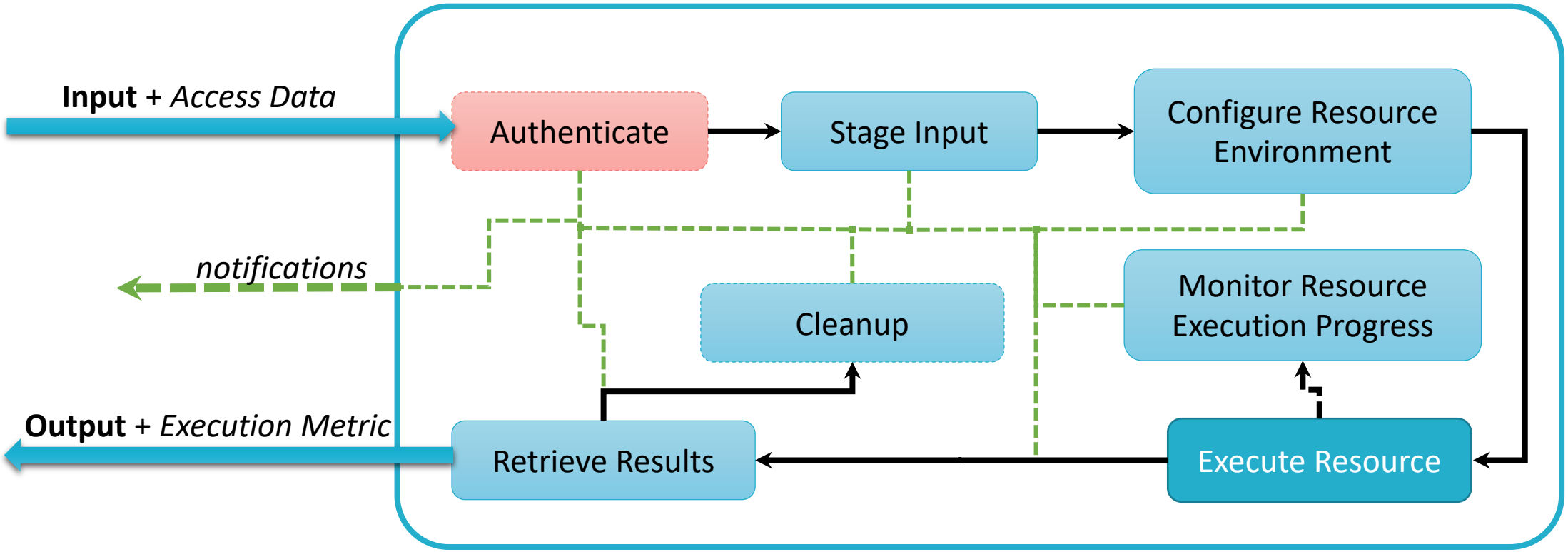
Data Transfer Use Cases



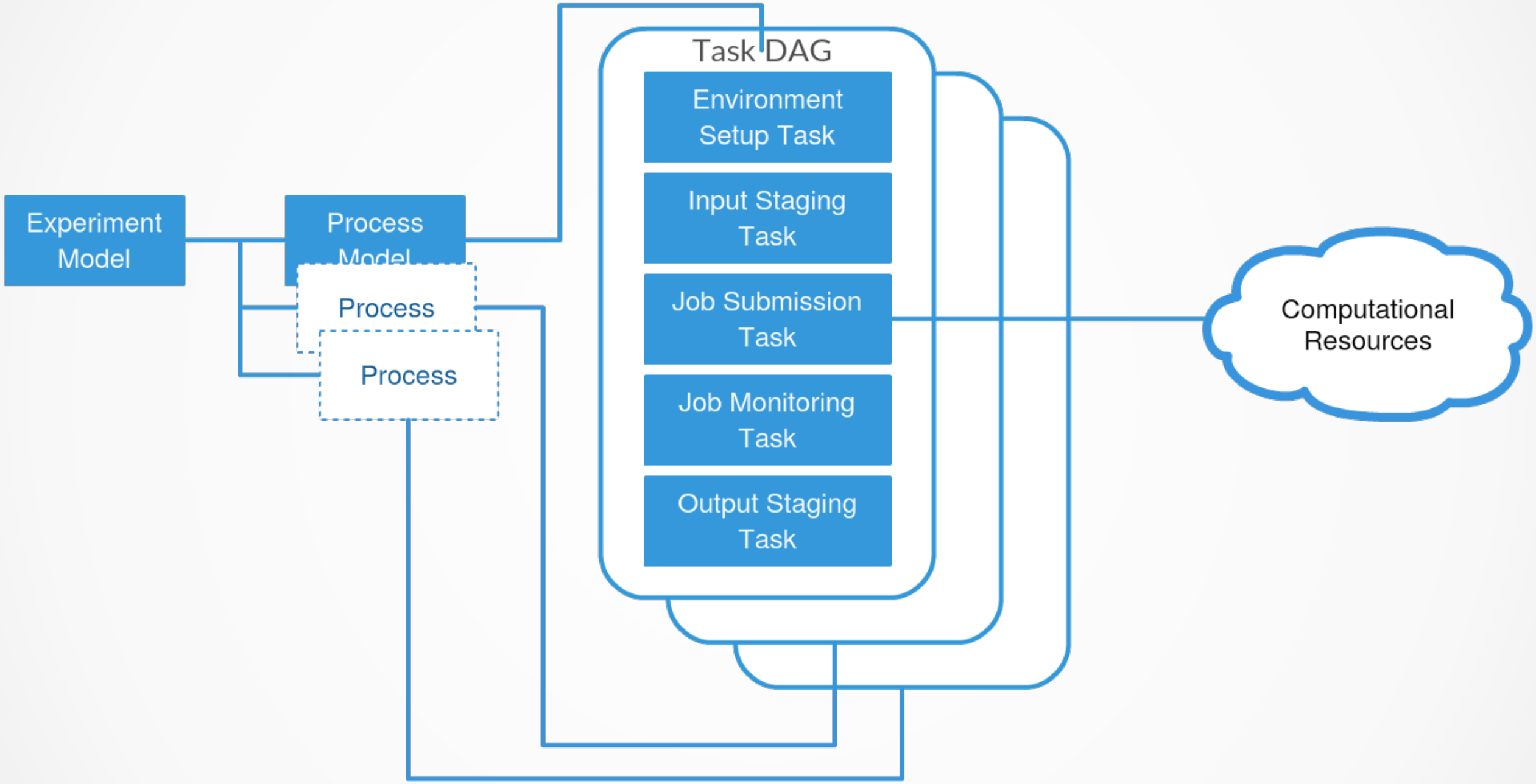
MFT + Airflow: Cloud to Cloud transfers



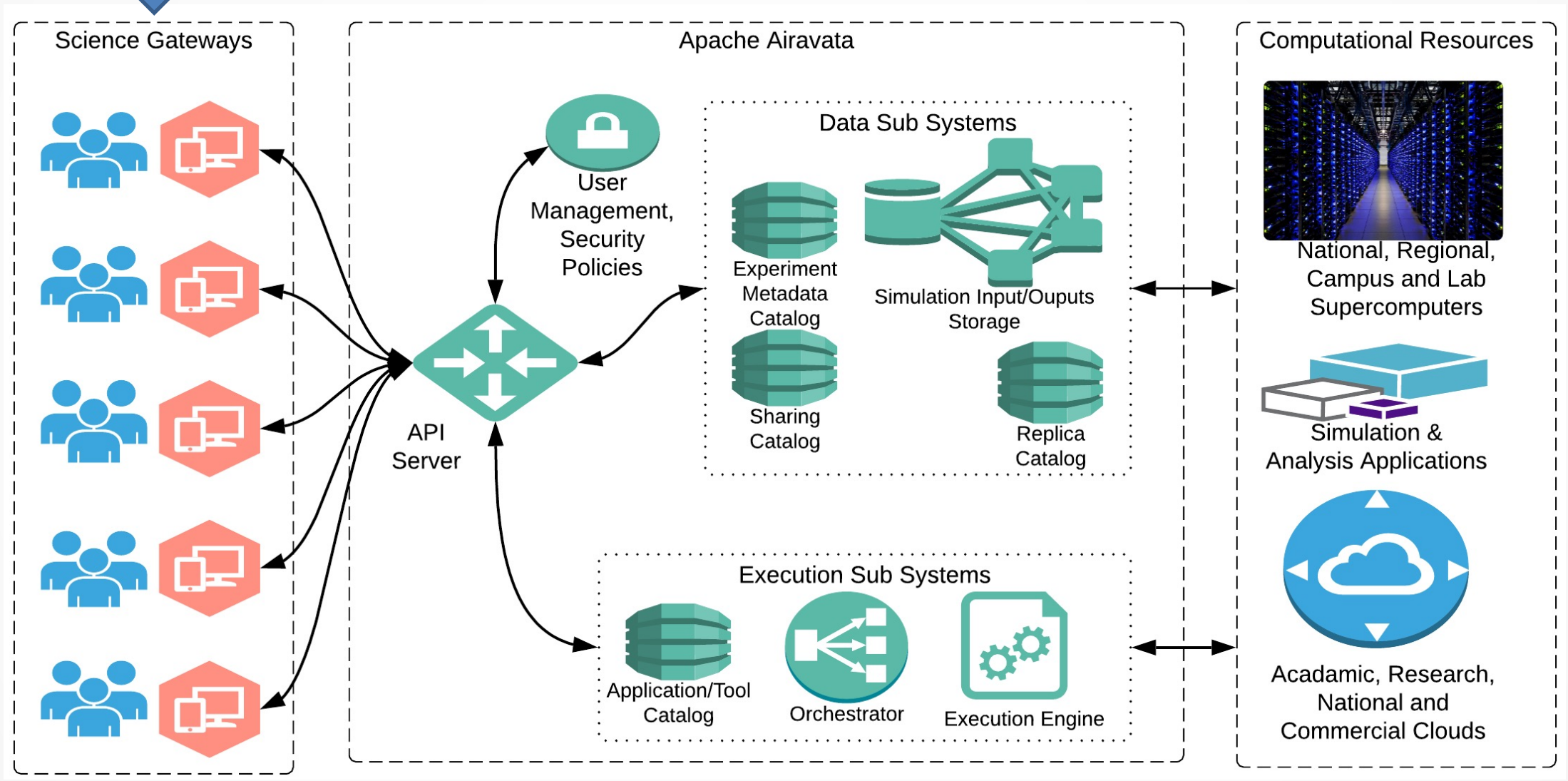
Apache Airavata Abstract Job Management



APACHE AIRAVATA


Experiment Model
Process Model
Process
Process
Task DAG
Environment Setup Task
Input Staging Task
Job Submission Task
Job Monitoring Task
Output Staging Task
Computational Resources

Django Portal for Airavata



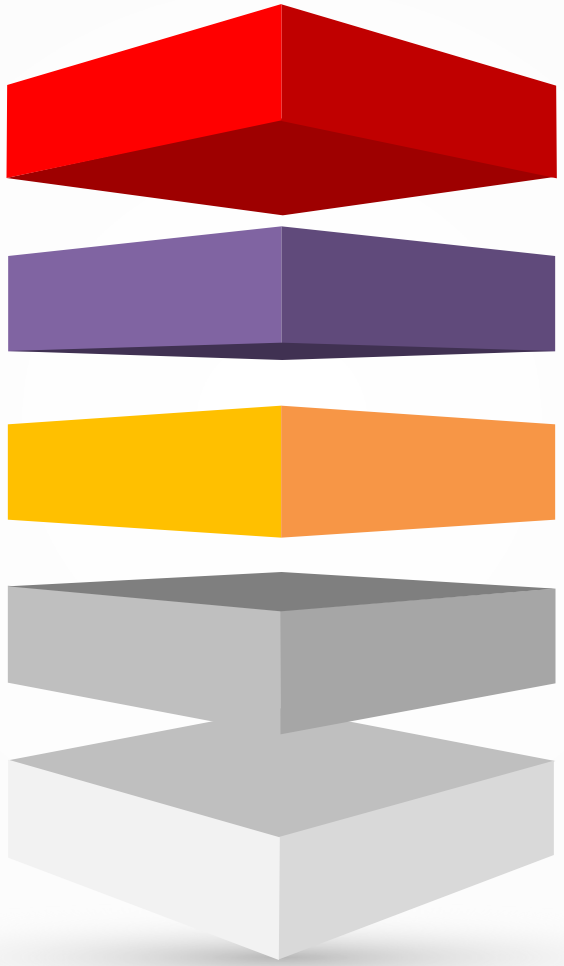
What's in Out of the box Apache Airavata?

A turnkey Django gateway to work with Airavata middleware

Customizable the look and feel

Extension points

Full stack Cyberinfrastructure



"Science-Centric" User Interfaces (based on Django Framework).

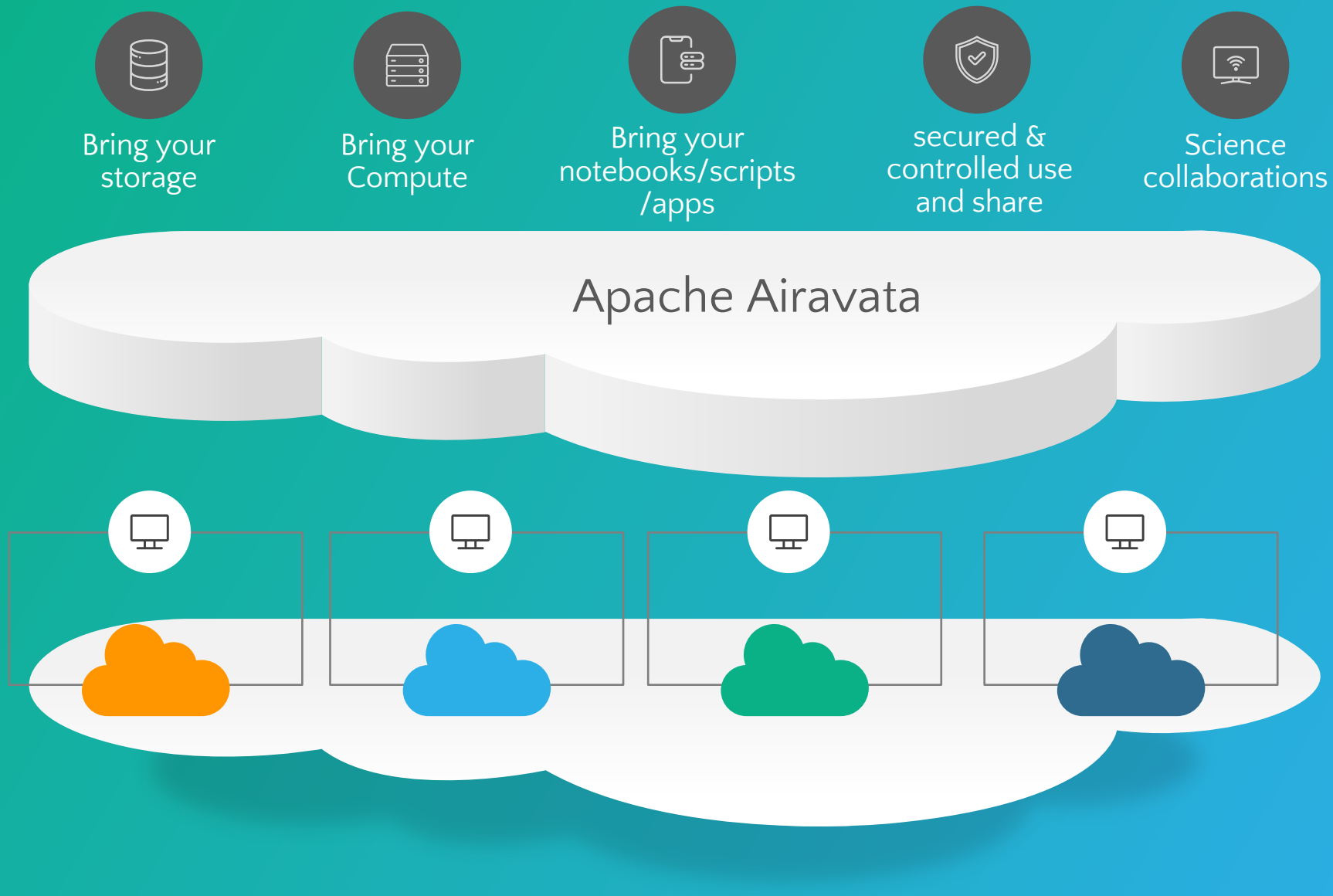
Custos Security Services (based on CILogon, Keycloak, Vault).

Apache Airavata Data Services (based on MFT, Neo4j).

Apache Airavata Execution Services (based on Apache Helix framework).

Elastic Virtual Clusters, Container Infrastructure.

Apache Airavata Ecosystem



Integrated local, campus, national resources; research or community data and codes.

Subscribe to Airavata's hosted service platform: SciGaP

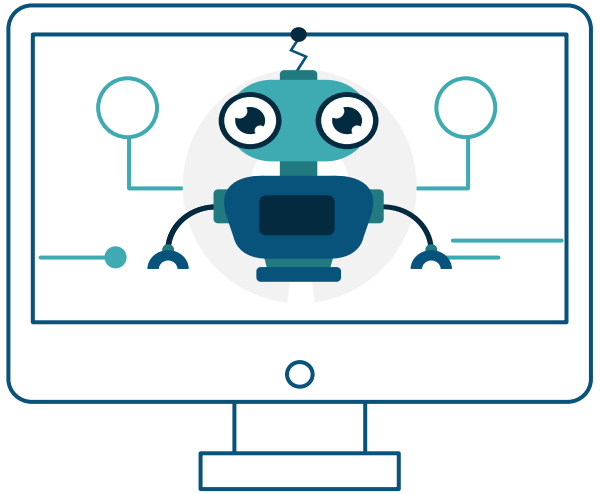
Science Gateways

KEY TAKEAWAY:

What can you do with ApacheAiravata?

Bring your science

- Science goals
- Data
- Applications



Bring your resources

- Distrusted Storage
- Compute Resources
- Security & Access Control Policies

"Convince" Eroma to give you a production ready Science Gateway in 2 hours of her "available" time

APACHE AIRAVATA

Airavata Data Stack

APACHE AIRAVATA

InterACTWEL

Enabling Long-term AI-driven Decision Support for Adaptive Management of Water, Energy, and Land Resources in Watershed Communities

Oregon State University

INDIANA UNIVERSITY

POWERED BY APACHE AIRAVATA

SGCI Science Gateways Community Institute

USDA NIFA

AWARD NUMBER: 2017-67003-26057

NSF

AWARD NUMBERS: #1547611, #1339774.

InterACTWEL: Web-based Decision Support for Adaptation Scenario Planning in Local Communities

Who?

- InterACTWEL aims to empower communities in watersheds

Researchers

Managers & Agencies

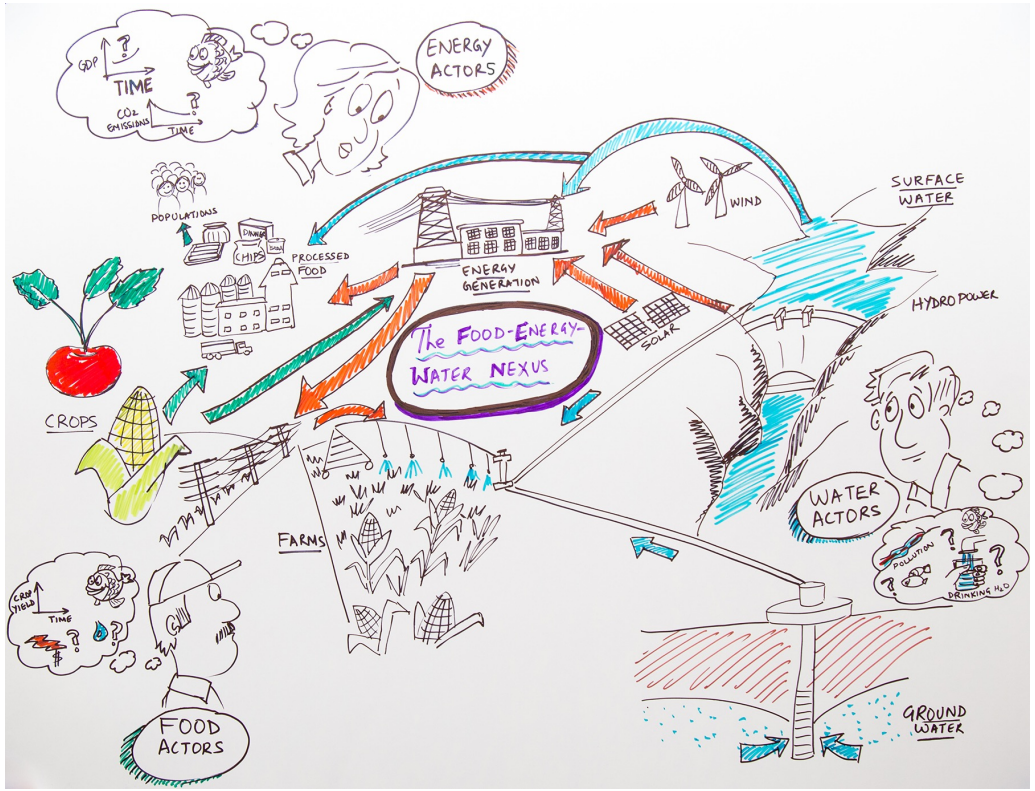
Watersheds

Do What?

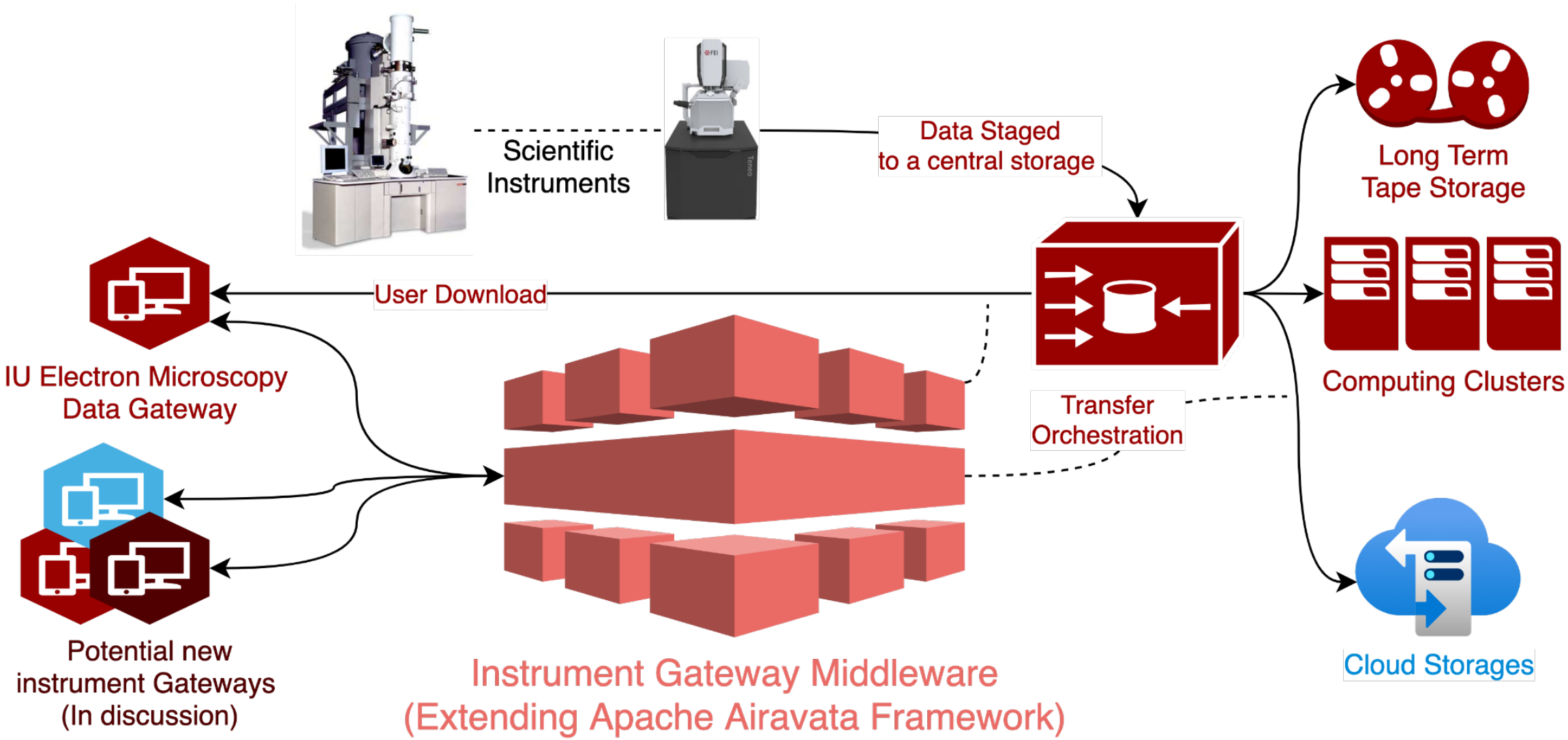
- To coordinate and plan adaptation strategies for responding to stresses and changes to water, energy, and land resources.

How?

- By providing data-driven services for planners and managers to simulate, visualize, evaluate, and share scenarios of multi-sectoral adaptation actions that reduce their community's vulnerability to stresses

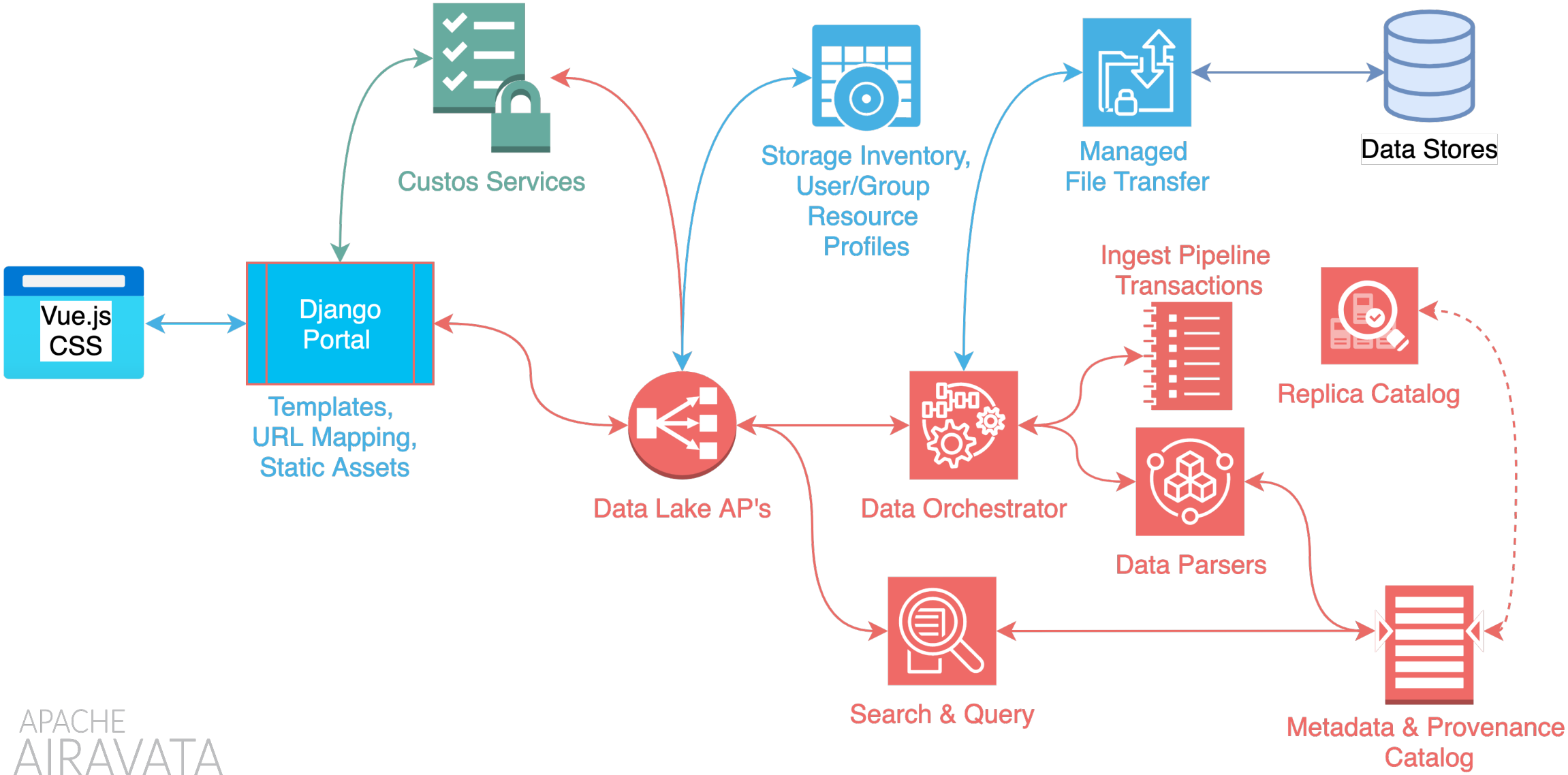


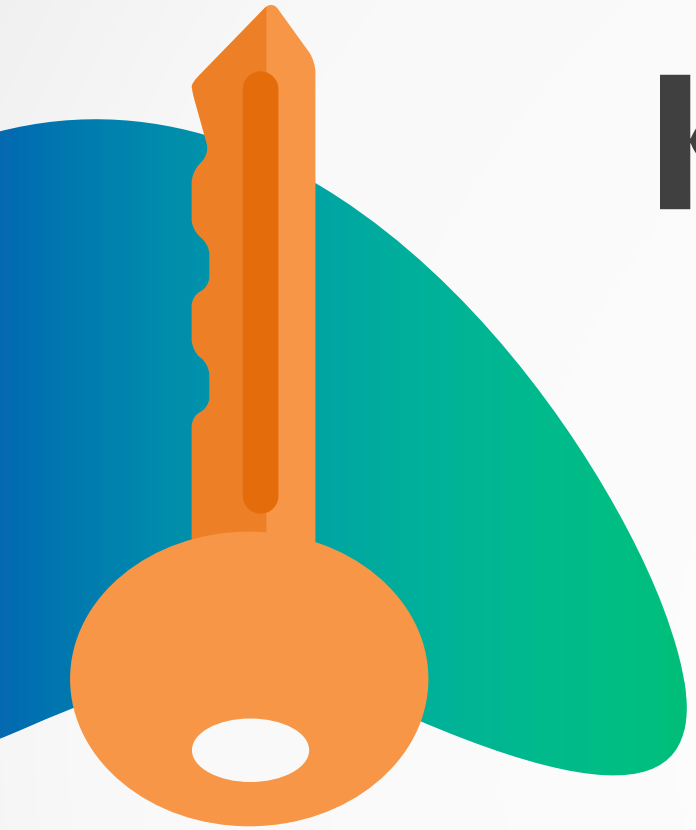
Instrument Gateways



APACHE AIRAVATA

Data Lake Abstract Architecture





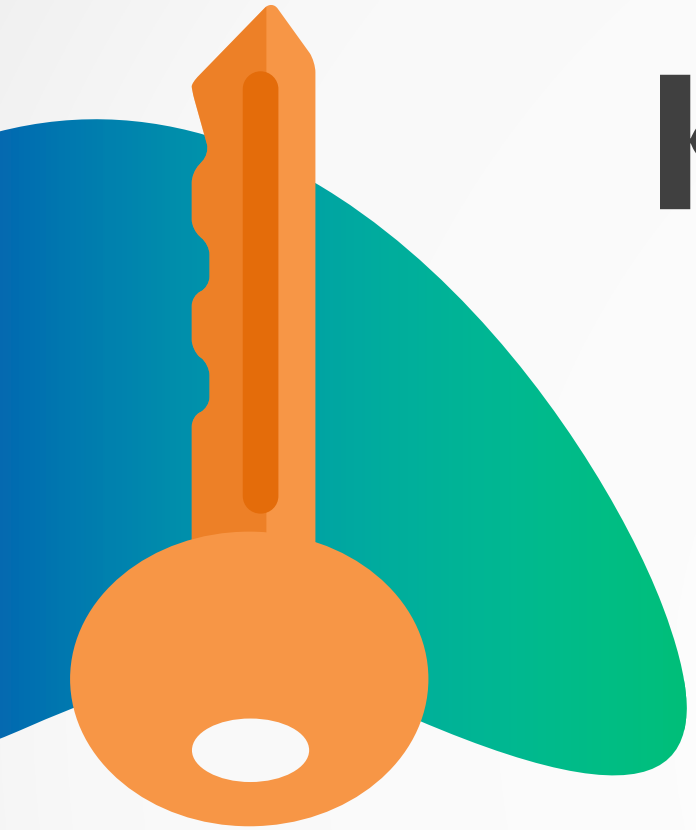
KEY TAKEAWAYS

- Airavata Data Stack can be used independently or in conjunction with execution services.
- Data services are integrated with Custos security services for fine-grained authorization.
- Data Stack integrates Airavata Managed File Transfer Services.

Custos Features

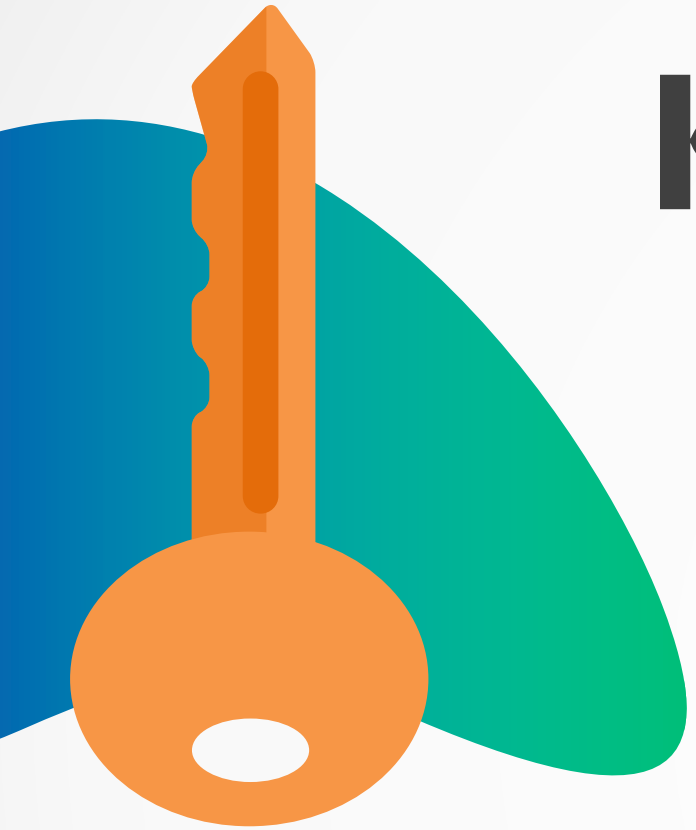
- Local and federated authentication
- Authorization
  - Role based authorization
  - Group based authorization
  - Attribute based authorization
- User management
- Sharing and permissions management
- Secrets (resource credential) management

We target end-to-end science gateway usage scenarios that integrate these features.



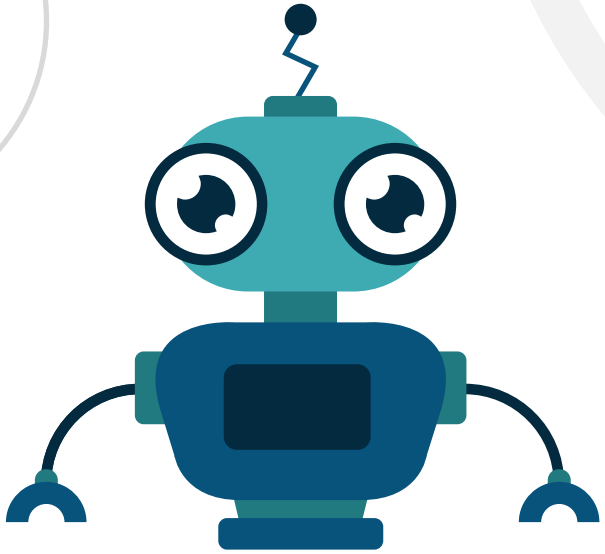
KEY TAKEAWAYS

- Custos services can be used independently similar to Galaxy or Hathi Trust Research Centre.
- Custos security services provide foundational capabilities across of the spectrum.
- Custos integrates and simplifies use of CILogon, keycloack, Vault



KEY TAKEAWAYS

- Use a turnkey Django gateway to work with Airavata middleware.
- Customize the look and feel .
- Create custom applications in a local Django installation



THANK YOU

Suresh Marru - smarru@apache.org

APACHE AIRAVATA