



Apache Airavata: Building Gateways to Innovation

Suresh Marru
Indiana University

Thanks to the Airavata PMC

- **Aleksander Slominski (Incubation Mentor)**
- Amila Jayasekara
- **Ate Douma (Incubation Mentor)**
- Chathura Herath
- Chathuri Wimalasena
- **Chris A. Mattmann (Incubation Mentor)**
- Eran Chinthaka
- Hechan Surivayarachchi
- Lahiru Gunathilake
- Marlon Pierce
- Patanachai Tangchaisin
- Raminder Singh
- Saminda Wijeratne
- Shahani Markus Weerawarana
- Srinath Perera
- Suresh Marru (Chair)
- Thilina Gunarathn

Apache Airavata became an Apache TLP in September 2012. Thanks also to our incubator champion, **Ross Gardler** and to **Paul Freemantle** and **Sanjiva Weerawarna** for serving as mentors.

What's the Point of This Talk?

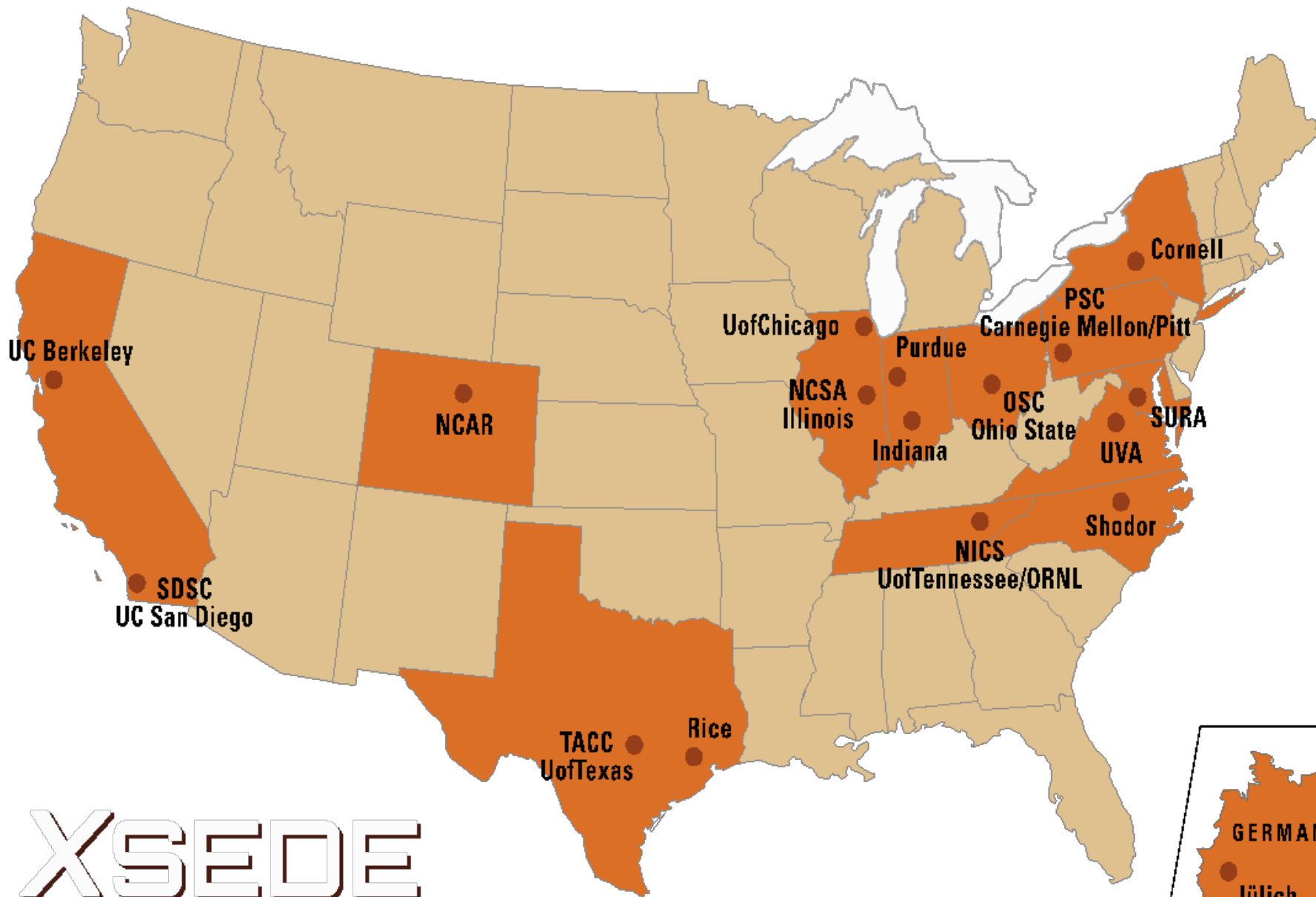
- Don't let history overly constrain the future.
- Broaden awareness of Airavata within the Apache community.
- Look for new collaborations outside the groups that we normally work with.

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

See talk by the NSF’s Dr. Dan Katz
2:30 pm during Thursday’s session.



XSEDE

Extreme Science and Engineering
Discovery Environment

Science Gateways: Enabling & Democratizing Scientific Research



Advanced Science Tools

Computational
Resources

Scientific
Instruments

Algorithms and
Models

Archived Data
and Metadata

Knowledge and Expertise

<http://sciencegateways.org/>

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 largely derived from NSF-funded academic research.

Apache *airavata*



Why Do We Care about Apache?



Two...No, Three Reasons

- Open Governance
 - Software should belong to those interested in contributing to it, regardless of funding.
- Broadening our developer community
- Making better connections with Apache.
 - We couldn't build Airavata with out the rest of Apache.



Cyberinfrastructure: How Open is Open Source Software?

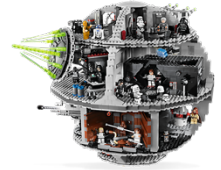
- What's missing?
 - Open source licensing
 - Open standards
 - Open codes (GitHub, SourceForge, Google Code, etc)



We also need open governance

Open Community Software and Governance

- Open source projects need diversity, governance.
 - Reproducibility
 - 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



Airavata's Apache Dependencies

Apache Axis2	Workflow Interpreter & WS-messenger services
Apache CXF	Registry API Front-end implementation
Apache OpenJPA, Derby	Registry API Back-end implementation
Apache Whirr, Hadoop	Enabling cloud bursting
Apache Shiro, Commons	Base for the security framework in Airavata
Apache Xmlbeans, Xmlschema, Axiom	Defining serializable descriptors
Apache Tomcat	Hosting the service frameworks

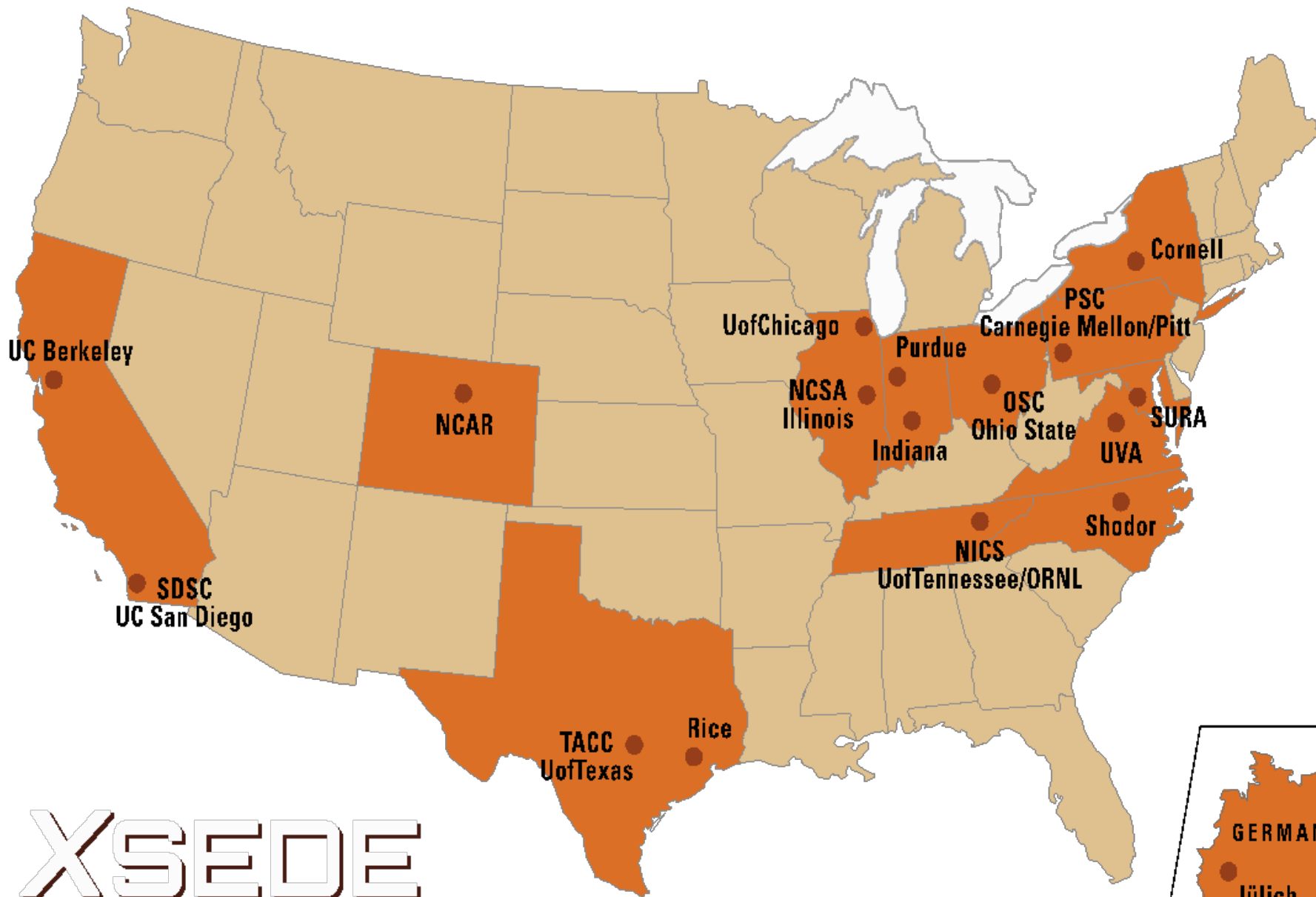
Some Collaboration Opportunities

Apache OODT	Workflow Interpreter & WS-messenger services
Apache Casandra	Increase reliability & availability through data replication
Apache Hadoop	By introducing capabilities of Hadoop we enable the use of data visualization tools available for hadoop
Apache Click, Flex, Rave, Shindig	Web base XBayya client, Airavata gadgets, Airavata dashboard

Apache *airavata*



Science Gateways, Scientific Workflows, and Cyberinfrastructure



XSEDE

Extreme Science and Engineering
Discovery Environment

Realizing the Universe for the Dark Energy Survey (DES) Using XSEDE Support

(Pis: A. Evrard (UM) and A. Kravtsov (UC))

Lb2600 full-sky light cone projected density 9: 234.0 to 260.0 $h^{-1}\text{Mpc}$

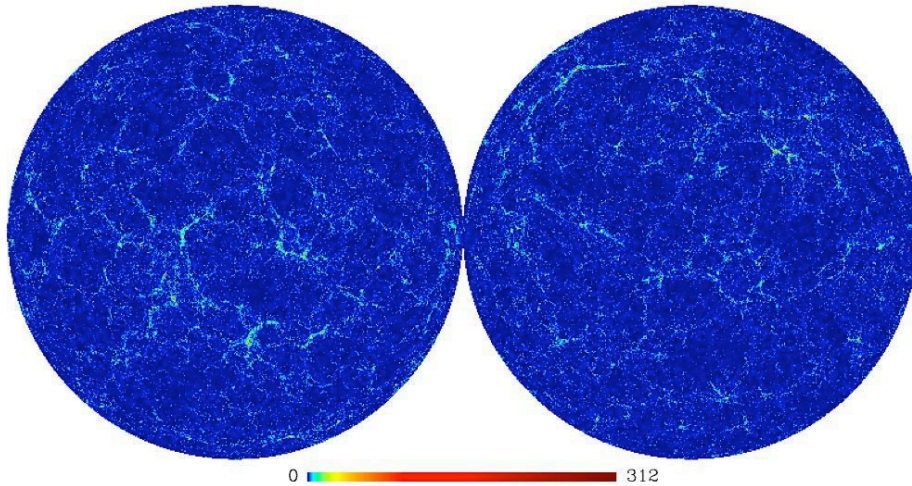


Fig. 1 The density of dark matter in a thin radial slice as seen by a synthetic observer located in the 8 billion light-year computational volume. Image courtesy Matthew Becker, University of Chicago.

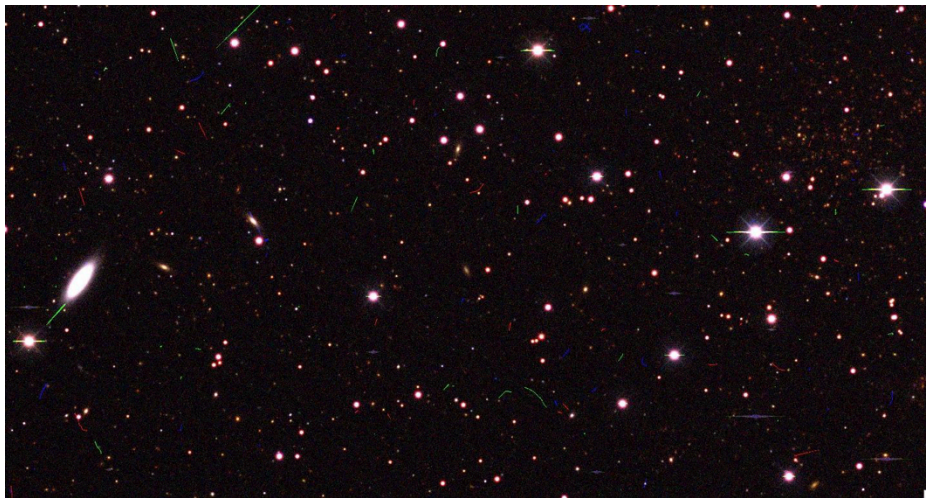
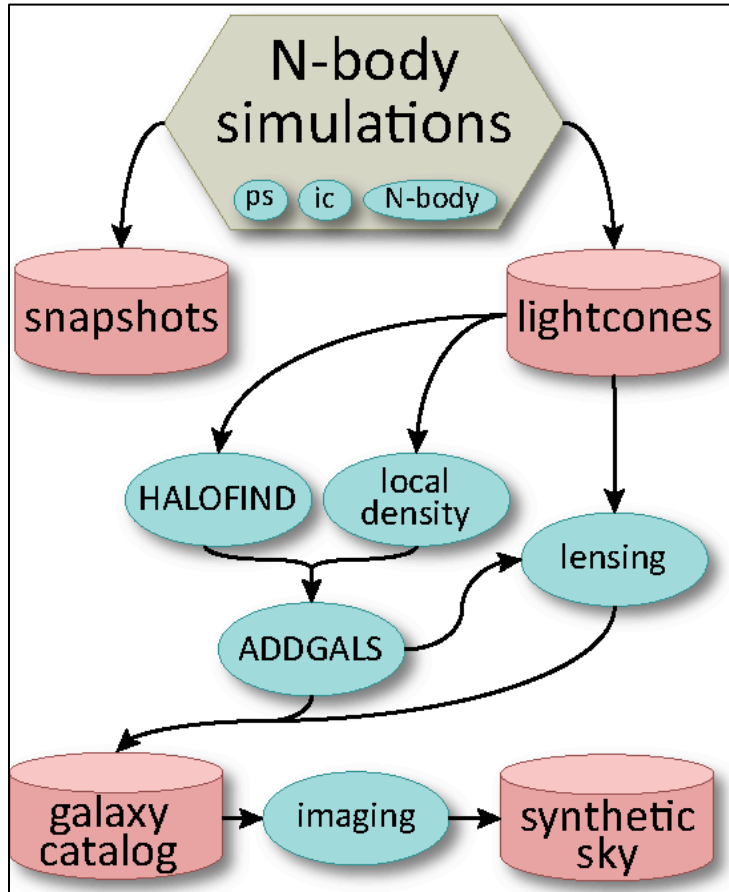


Fig. 2: A synthetic 2x3 arcmin DES sky image showing galaxies, stars, and observational artifacts. Courtesy Huan Lin, FNAL.

- The Dark Energy Survey (DES) is an upcoming international experiment that aims to constrain the properties of dark energy and dark matter in the universe using a deep, 5000-square degree survey of cosmic structure traced by galaxies.
- To support this science, the DES Simulation Working Group is generating expectations for galaxy yields in various cosmologies.
- Analysis of these simulated catalogs offers a quality assurance capability for cosmological and astrophysical analysis of upcoming DES telescope data.
- These large, multi-staged computations are a natural fit for workflow control atop XSEDE resources.

DES Application	Component Description
CAMB	<i>Code for Anisotropies in the Microwave Background</i> is a serial FORTRAN code that computes the power spectrum of dark matter, which is necessary for generating the simulation initial conditions. <i>Output is a small ASCII file describing the power spectrum.</i>
2LPTic	<i>Second-order Lagrangian Perturbation Theory initial conditions</i> code is an MPI based C code that computes the initial conditions for the simulation from parameters and an input power spectrum generated by CAMB. <i>Output is a set of binary files that vary in size from ~80-250 GB depending on the simulation resolution.</i>
LGadget	<i>LGadget</i> is an MPI based C code that evolves a gravitational N-body system. The outputs of this step are system state snapshot files, as well as lightcone files, and some properties of the matter distribution, including the power spectrum at various timesteps. <i>The total output from LGadget depends on resolution and the number of system snapshots stored, and approaches ~10 TB for large DES simulation boxes.</i>

DES as a Workflow



Processing steps to build a synthetic galaxy catalog.

There are plenty of issues:

- **Long running code:** Based on simulation box size L-gadget can run for 3 to 5 days using more than 1024 cores.
- **Local HPC provider policies:** XSEDE resource provider's job scheduling policy does not allow jobs to run for more than 24 hours in normal queue
- **Do-While Construct:** Restart service support is needed in workflow. Do-while construct was developed to address the need.
- **Data size and File transfer challenges:** L-gadget produces 10~TB for large DES simulation boxes in system scratch so data need to moved to persistent storage ASAP
- **File system issues:** More than 10,000 lightcone files are doing continues file I/O. This can cause problems with the HPC resource's file system (usually Lustre-based in XSEDE).

Break for the DES Movie

Apache Airavata in Action

Domain	Description
Astronomy	Image processing pipeline for One Degree Imager instrument on XSEDE
Astrophysics	Supporting workflow of Dark Energy Survey simulations working group on XSEDE
Bioinformatics	Supported workflow executions on Amazon EC2 for BioVLAB project
Biophysics	Manage large scale data analysis of analytical ultracentrifugation experiments on XSEDE and campus resources
Computational Chemistry	Manage workflows to support computational chemistry parameter studies for ParamChem.org on XSEDE
Nuclear Physics	Workflows for nuclear structure calculations using Leadership Class Configuration Interaction (LCCI) computations on DOE resources

Airavata Culture

- Java code base
- Airavata 0.6 is out, working on 0.7
 - What is in a release?
 - Sprint/scrum + Apache =?
- Work through dev mailing list and Jira.
- Actively engage students
 - GSOC
 - Thanks to Shahani W.
- Engage through XSEDE advanced support
 - Find new

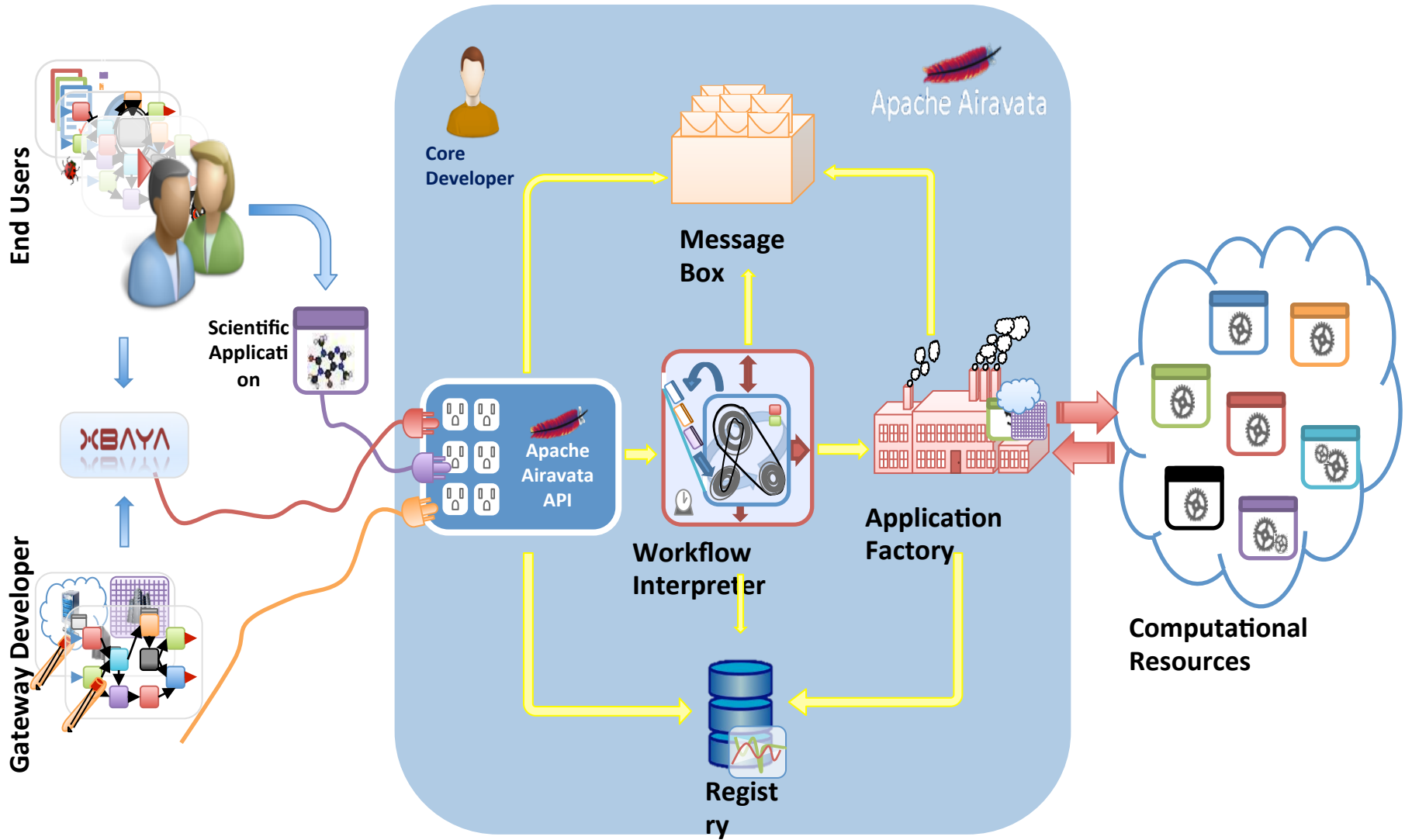


Apache *airavata*



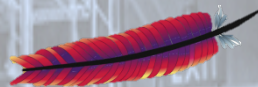
Apache Airavata Overview

Apache Airavata



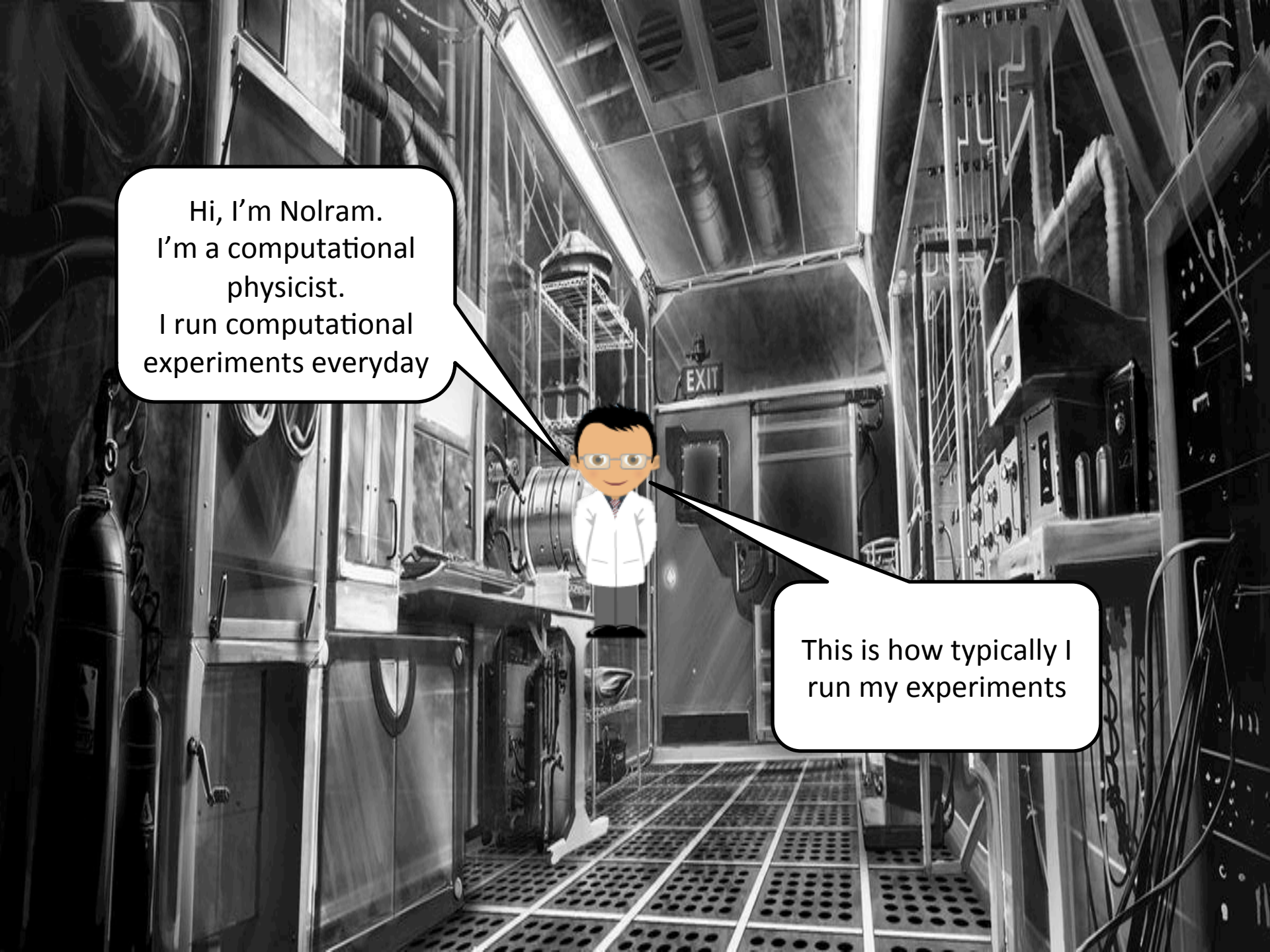
Apache Airavata Components

Component	Description
XBaya	Workflow graphical composition tool.
Registry Service	Insert and access application, host machine, workflow, and provenance data.
Workflow Interpreter Service	Execute the workflow on one or more resources.
Application Factory Service (GFAC)	Manages the execution and management of an application in a workflow
Messaging System	WS-Notification and WS-Eventing compliant publish/subscribe messaging system for workflow events
Airavata API	Single wrapping client to provide higher level programming interfaces.



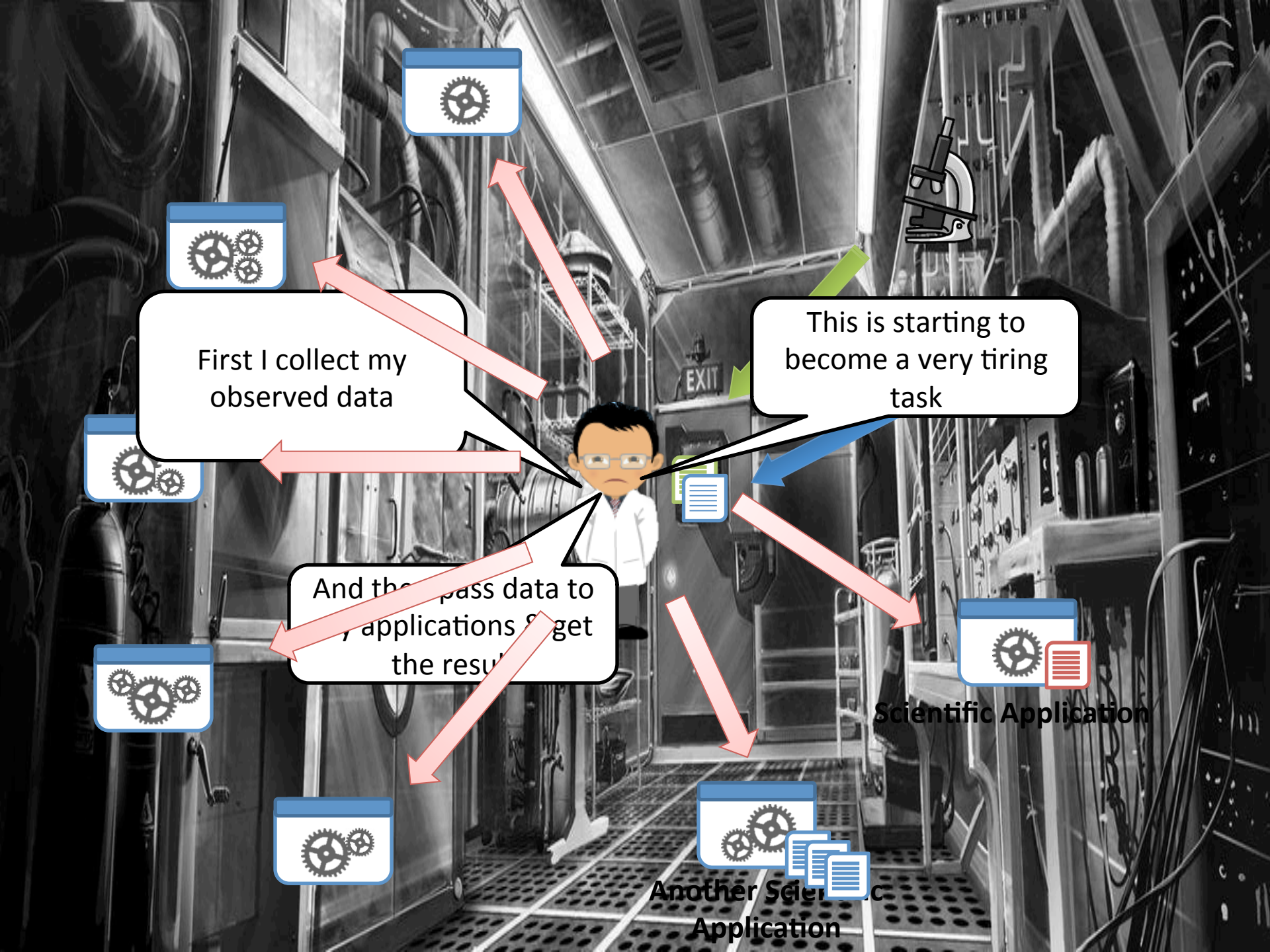
Apache Airavata

An Architectural introduction



Hi, I'm Nolram.
I'm a computational
physicist.
I run computational
experiments everyday

This is how typically I
run my experiments



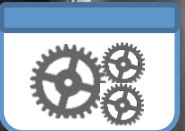
First I collect my observed data

This is starting to become a very tiring task

And then I pass data to my applications to get the results

Scientific Application

Another Scientific Application



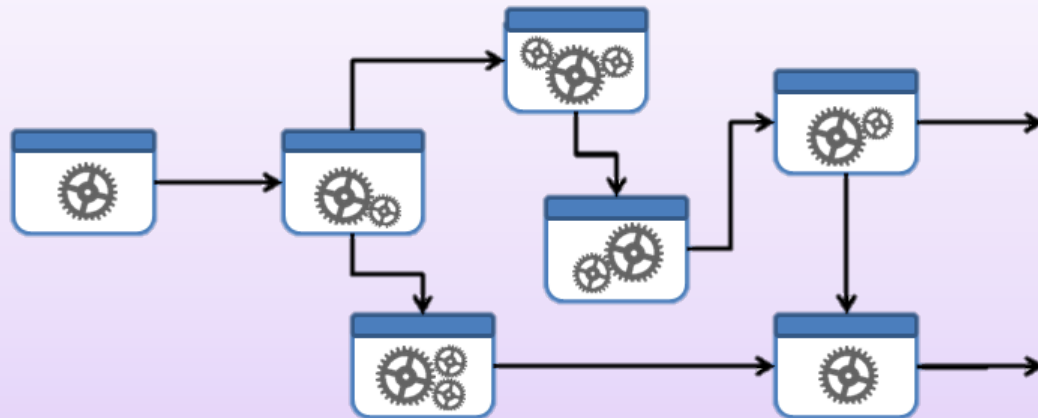
Scientists from many different fields face this problem everyday.

What is a workflow you ask?



The solution is to use a workflow-powered science gateway to manage the experiment online.

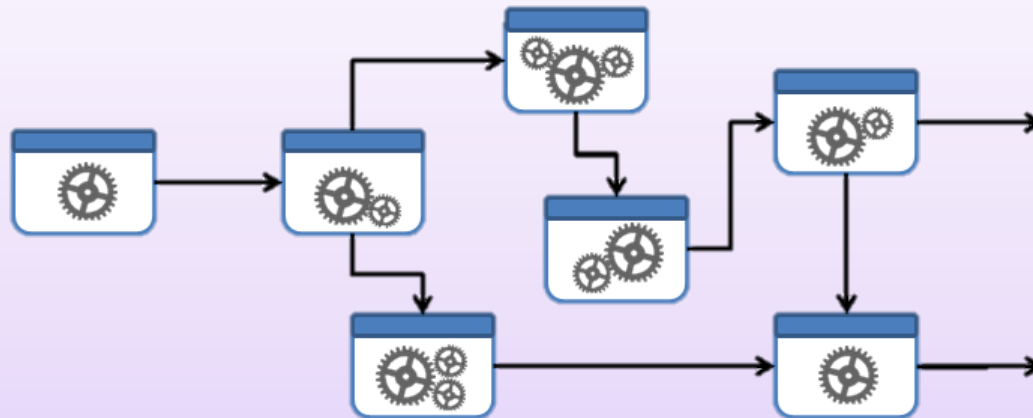
Well, you just saw one in our previous animation...





We introduce **Apache Airavata**, a system capable of composing, managing, executing, and monitoring small to large scale applications and workflows

Want to see how it works?

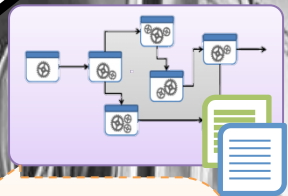


A Typical Workflow

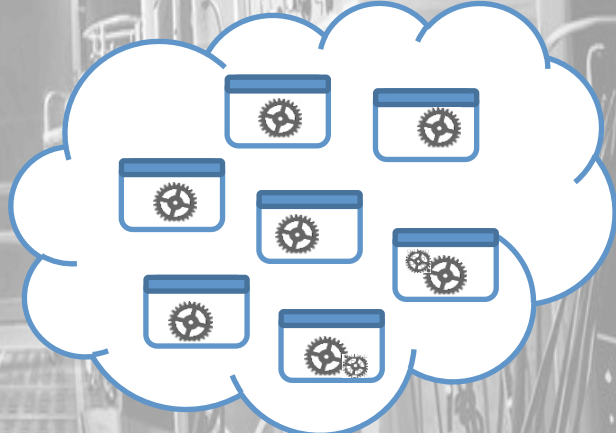
... and while I wait for results,
Airavata will notify me with
progress updates of my experiment



Results



Progress of the experiment

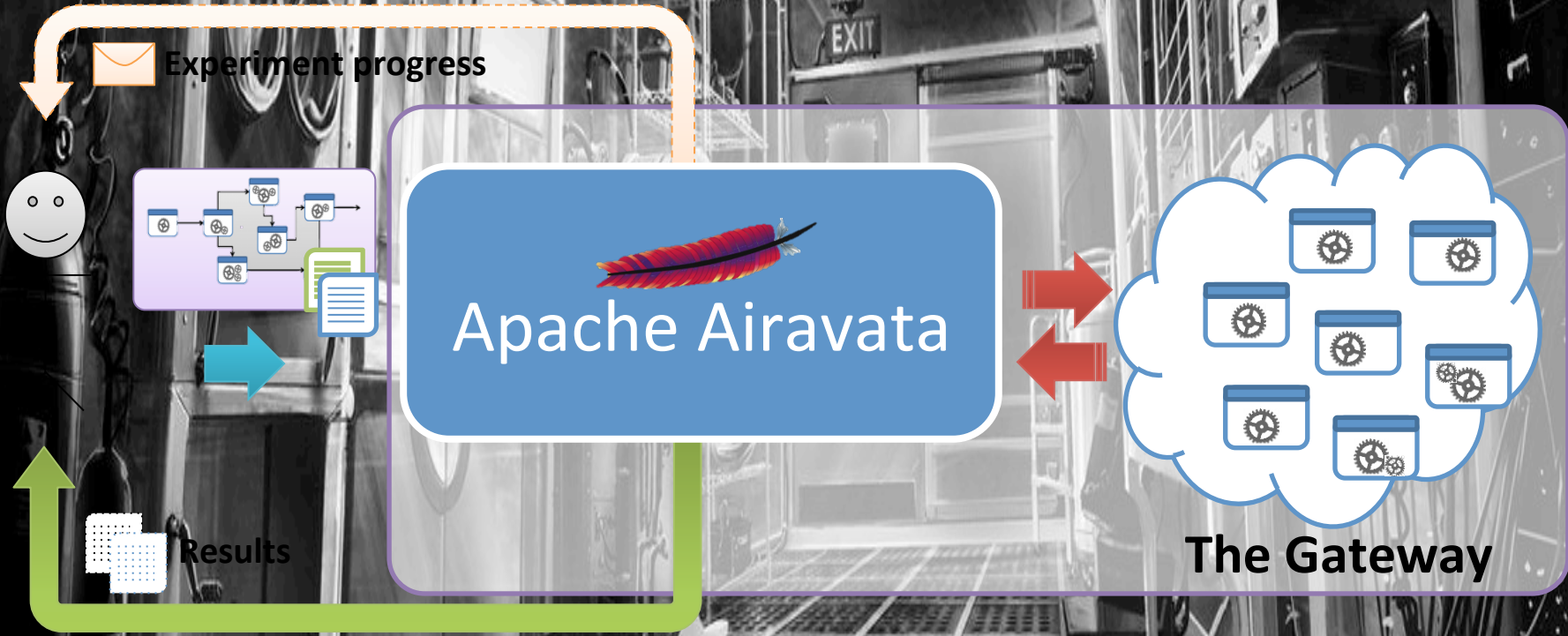


The Gateway





Let's look closely how Airavata manages workflows.



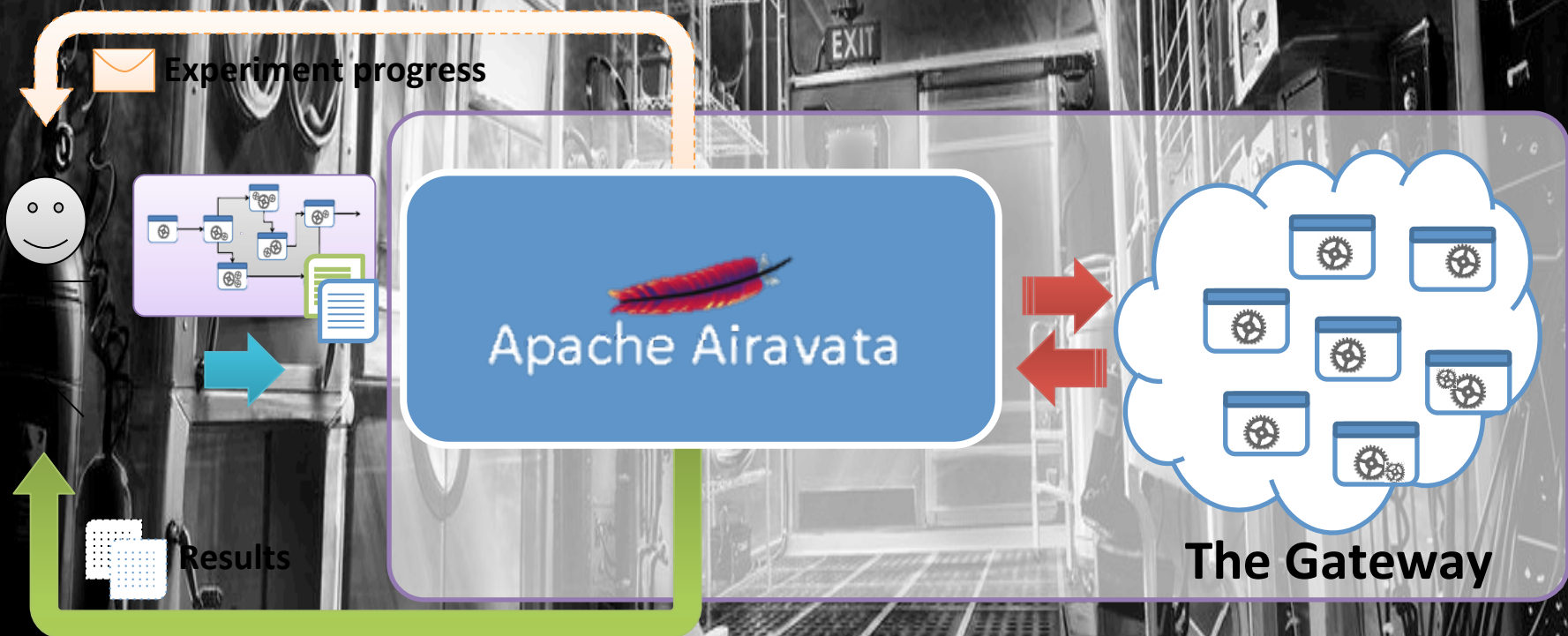
Experiment progress

Apache Airavata

The Gateway

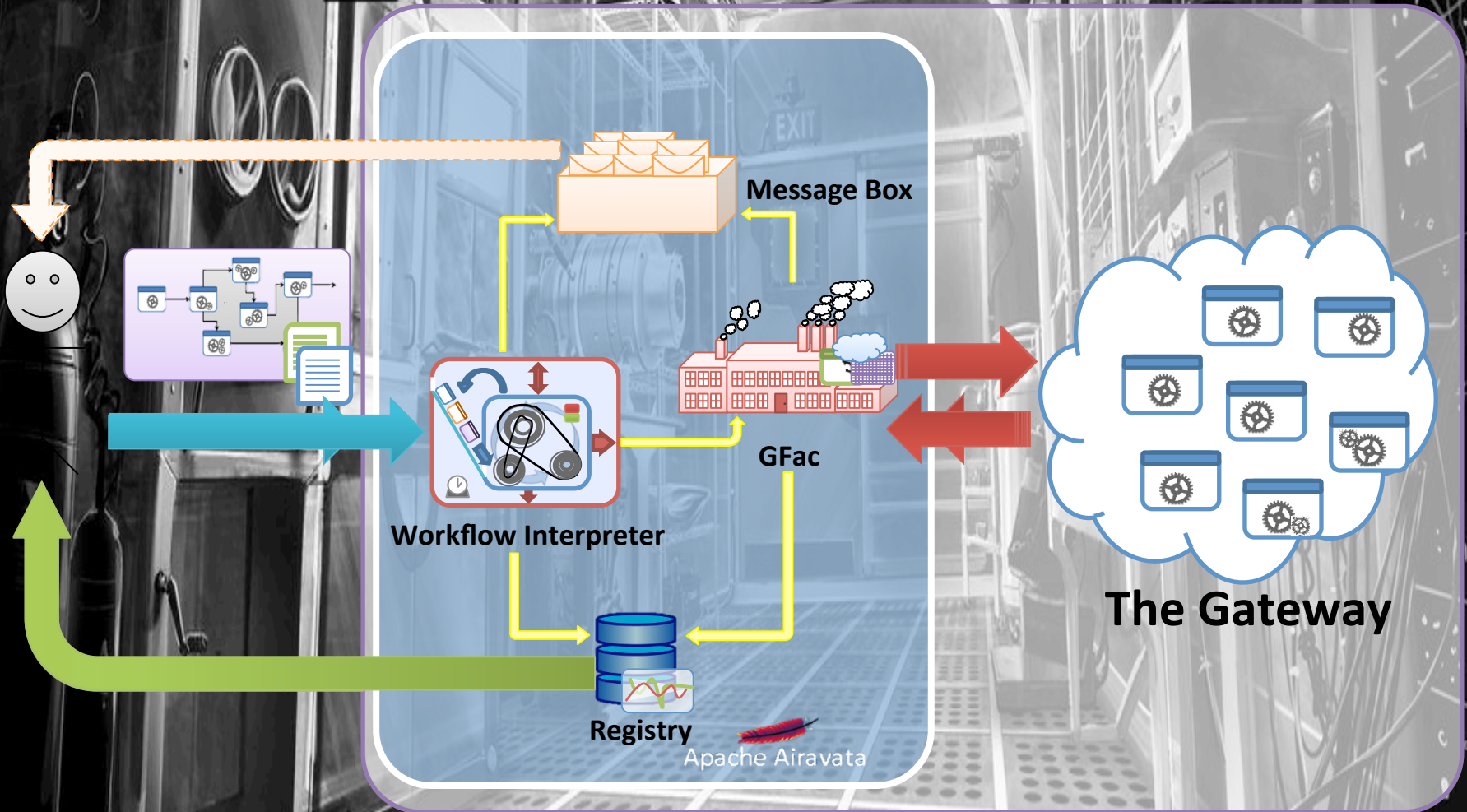
Results

Let's look closely how Airavata manages workflows.



4. The Message Box

Records the progress of the workflow execution



Now you have a basic understanding of what Airavata is, why it is useful & how it works.





Being a Part of
Airavata Community



Being a Part of Airavata Community

Play with different popular Apache technologies & tools

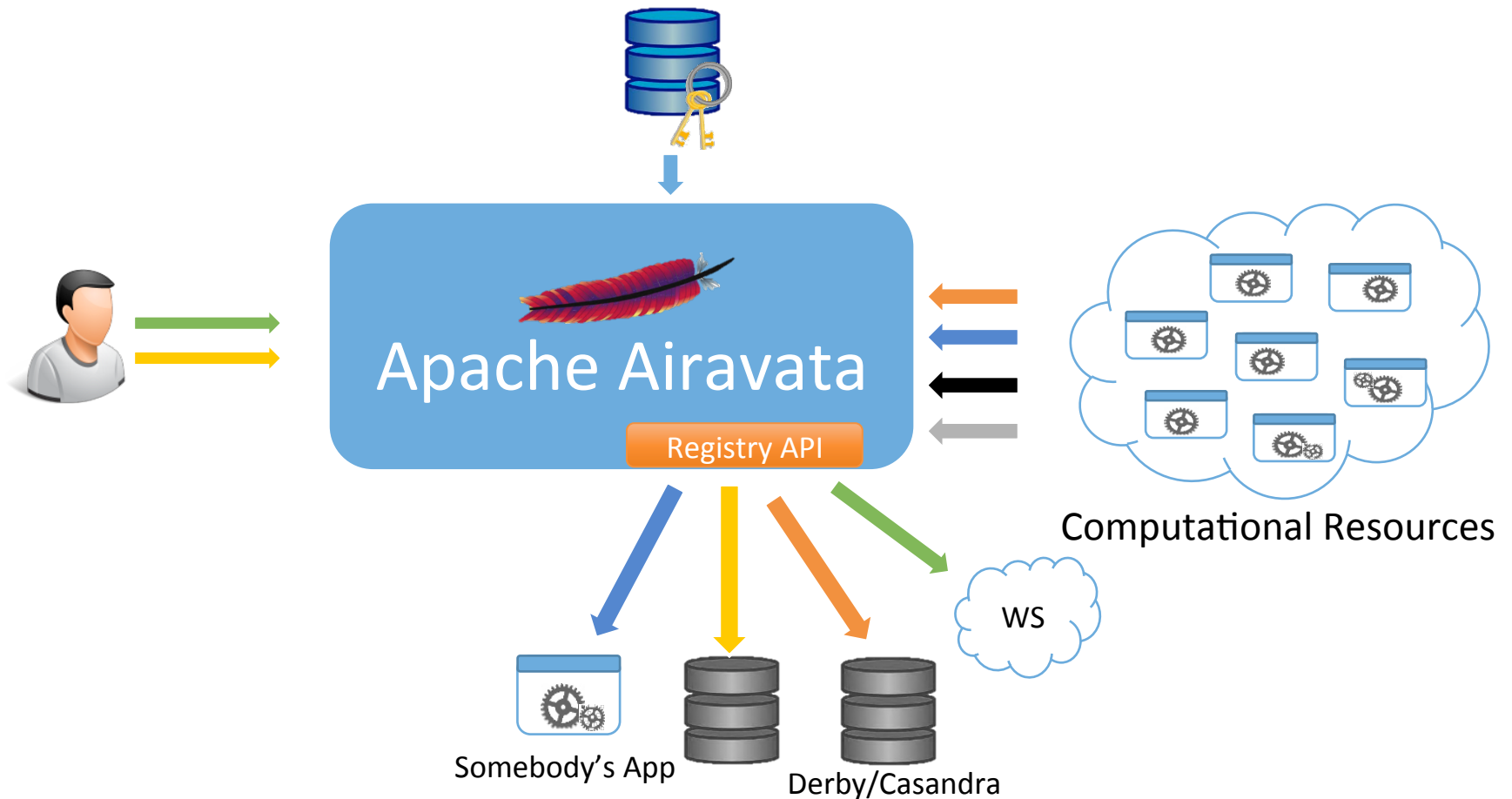
Experiment with the Cloud, the Grid... it's all **here**...

Learn & Engage with a multidisciplinary community



The recent impact from
the community...

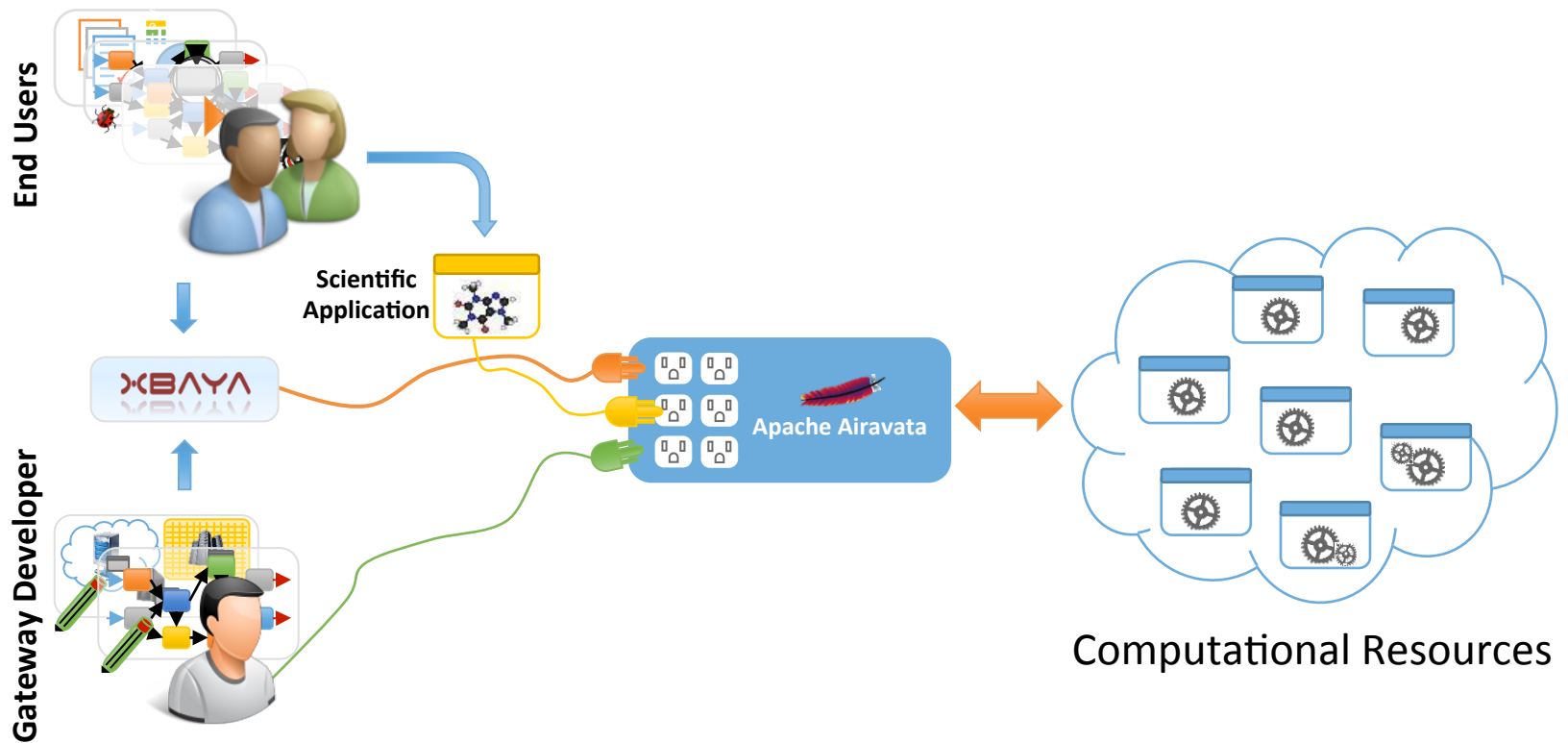
A Pluggable & Customizable Framework for Registries



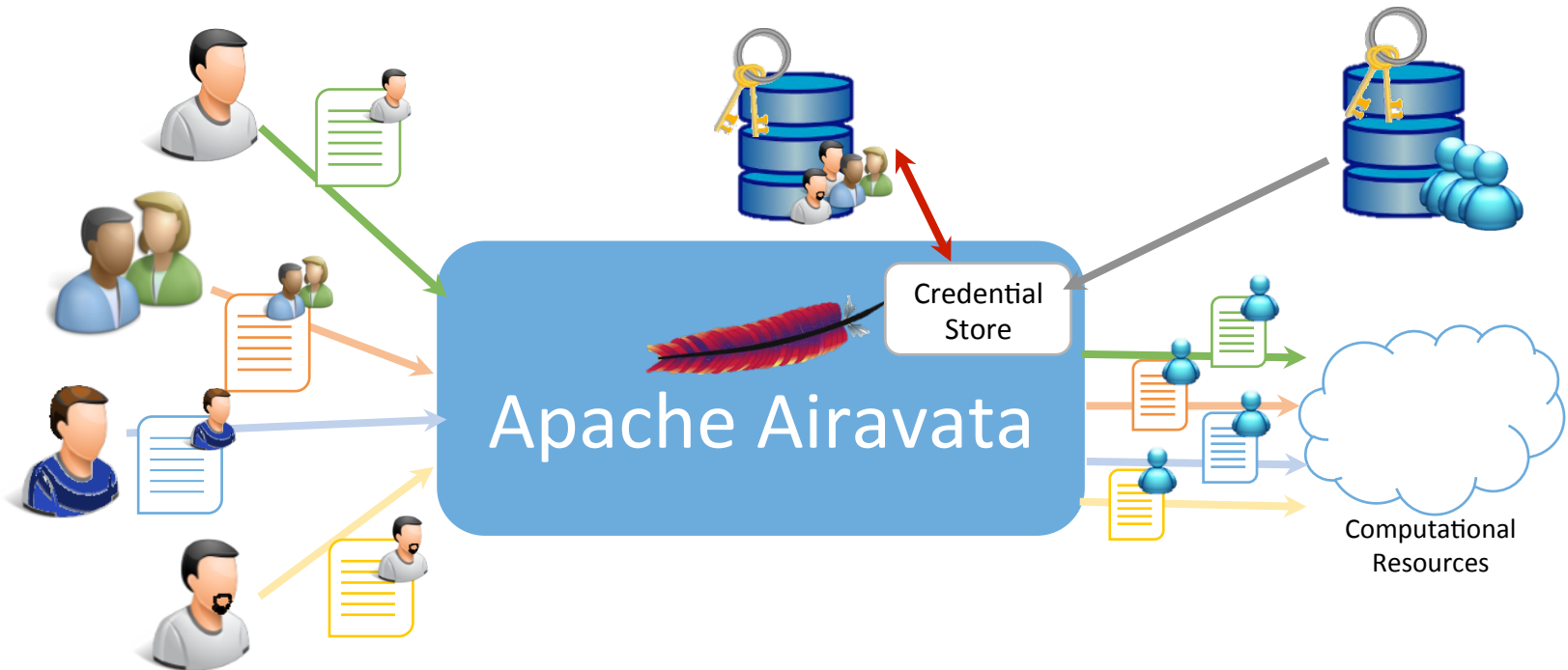
Support for Cloud-Bursting Applications



A Stable API for Airavata



Solutions for Unique Security Requirements



UNICORE Support

Airavata as a Service

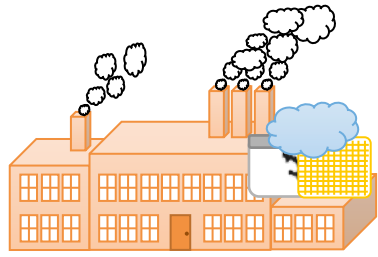
Real-time Debugging
Workflows

An Extendable Application
Factory

The Concept of steering Apps &
Workflows

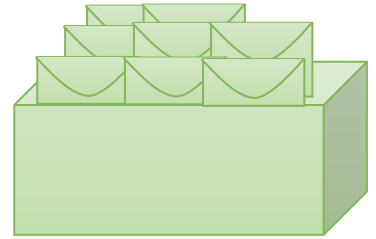


Impact from Airavata to
the community...



A Generic Application Factory

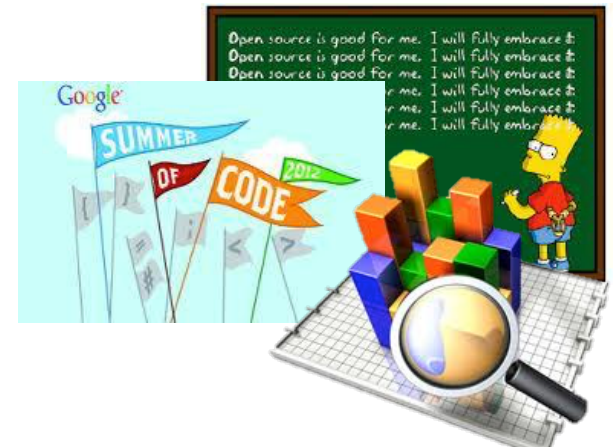
A Pub-Sub Messaging Framework



A Credential Store



A Student Introduction



Creating New Ties...





Apache CXF



Apache Commons



Apache Tomcat





**Extend Airavata from your project or
extend your project from Airavata**



Or just come up with your own idea
to make Airavata better

Useful Workflow Components

Enhanced Data Layer (eg: NoSQL)

CLI/Graphical Tools

(Plugins,Gadgets,Mobile Apps etc.)

Multitenant Support

Data Visualization

Providers for Computing
Resources

Throttling Support

Airavata Easy Deployment

- Airavata Deployment Studio (ADS)
- FutureGrid
- One button configurable deployment
 - OpenStack, EC2, Eucalyptus
 - Ubuntu, CentOS, Redhat
 - X86, 64-bit
 - Airavata 0.6

ADS Sneak Peak

Bootstrap, from Twitter - Chromium

localhost:5000/#

Apache Airavata Home About Contact

Logged in as heshan

CONFIGURATION

- Dashboard
- Build instance
- Configure services

MONITORING

- Dashboard
- Monitor instances
- Monitor services

RESOURCES

- Apache Airavata
- FutureGrid
- Science Gateway Group
- Indiana University

Airavata Deployment Studio!

Welcome to Airavata Deployment Studio (ADS) - one stop shop for Airavata configuration and deployment. This web application let's you build your own Airavata server from scratch. That is from the underlying hardware architecture, Operating System to specific Airavata release version.

[Learn more »](#)

Configuring

This application let's you configure cloud computing environments like Eucalyptus, EC2 and Openstack with Apache Airavata.

[View details »](#)

Monitoring

Monitoring dashboards let's a user to monitor his up and running server instances and their services.

[View details »](#)

Why ADS?

Using ADS will make the user's/scientist's life easy during the installation process and help him/her to evaluate Airavata quickly without going through the hassle of running shell commands.

[View details »](#)

localhost:5000/#

ADS Sneak Peak ...

Bootstrap, from Twitter - Chromium

localhost:5000/create

Apache Airavata Home About Contact

Logged in as heshan

CONFIGURATION

- Dashboard
- Build instance
- Configure services

MONITORING

- Dashboard
- Monitor instances
- Monitor services

RESOURCES

- Apache Airavata
- FutureGrid
- Science Gateway Group
- Indiana University

Create Airavata image

Select values to configure the image

Infrastructure	OpenStack
Architecture	64 bit
Memory	512MB
Operating System	Ubuntu
Java version	Java 1.7
Tomcat version	Tomcat 6
Airavata version	Airavata 0.6

Generate Image

Further Information

- Contact: marpierc@iu.edu, smarru@iu.edu
- Apache Airavata: <http://airavata.apache.org>
- You can contribute to Apache Airavata!
 - Join the mailing list: dev@airavata.apache.org
- YouTube presentation on Apache and NSF Cyberinfrastructure:
<http://www.youtube.com/watch?v=AN7LoQct17U>



References

- Images from
 - <https://encrypted-tbn2.gstatic.com>
 - <http://xmlbeans.apache.org>
- <http://airavata.apache.org/>
- <https://cwiki.apache.org/confluence/display/AIRAVATA/index>