

Cyber Infrastructure for Molecular Science Communities

Sudhakar Pamidighantam

Department of Chemistry and
Cyberinfrastructure Integration Research Center

Indiana University

pamidigs@iu.edu

Cybertraining Workshop 14 June 2021

Outline

- What is cyberinfrastructure and Science gateway
- Apache Airavata Collaboration Platform
- SEAGrid Science Gateway
- CIRC Collaborations

Acknowledgements



**CYBERINFRASTRUCTURE
INTEGRATION RESEARCH CENTER**
PERVASIVE TECHNOLOGY INSTITUTE



POWERED BY



APACHE
AIRAVATA

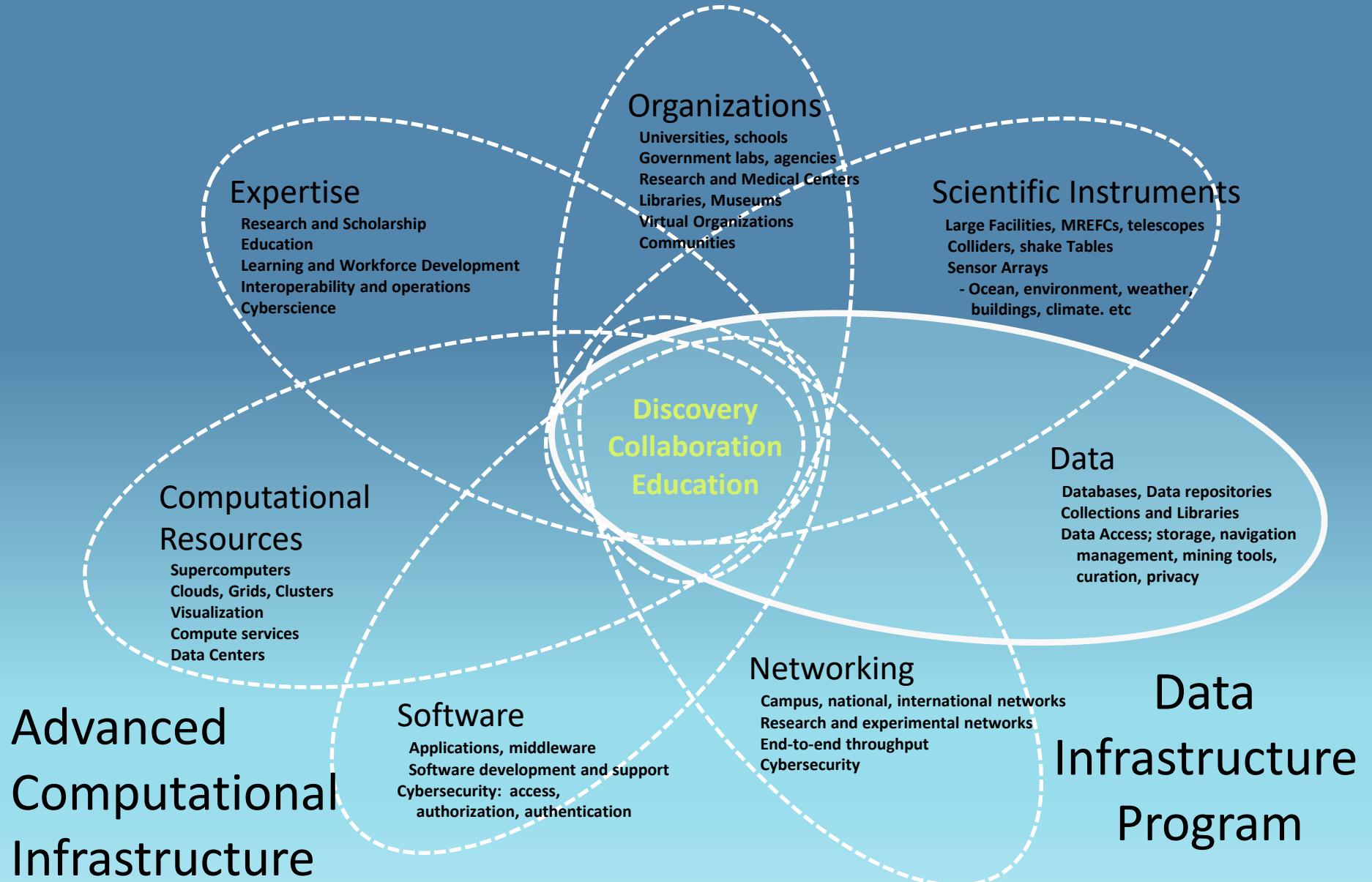


Cyberinfrastructure

- Laboratory, Local to Campus, State, National and International Resources of Data, Computing, Networks and Software frameworks integrated together to provide seamless access to the scientist for research and teaching.

NSF Supercomputer Centers XSEDE, Frontera, Regional and Local systems

NSF CIF21 Major Areas



Current XSEDE Resources

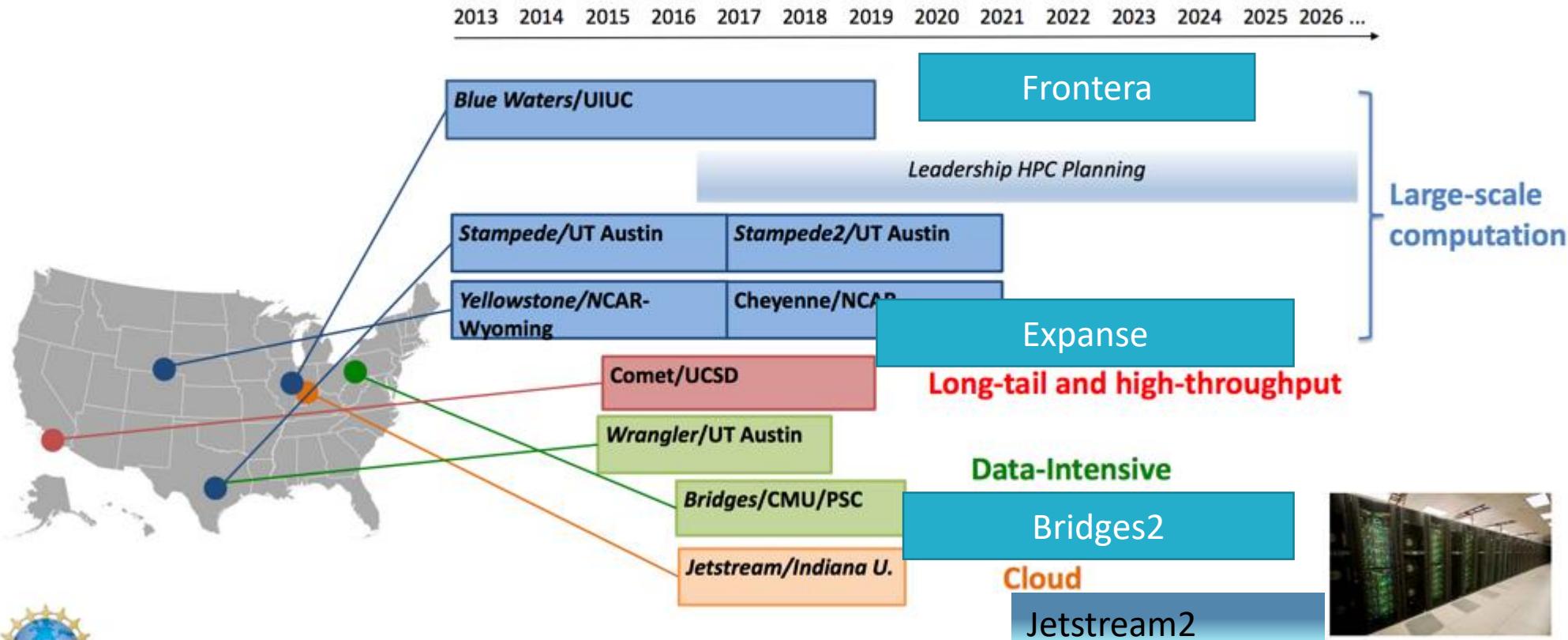
Resource	Organization	Type
HP/NVIDIA Interactive Visualization and Data Analytics System (Maverick)	TACC	vis
IU/TACC (Jetstream)	Indiana U	compute
IU/TACC Storage (Jetstream Storage)	UT Austin	storage
LSU Cluster (superMIC)	LSU CCT	compute
Open Science Grid (OSG)	OSG	compute
PSC Bridges GPU (Bridges2 GPU)	PSC	compute
PSC Large Memory Nodes (Bridges2 Large)	PSC	compute
PSC Regular Memory (Bridges2)	PSC	compute
PSC Storage (Bridges Pylon)	PSC	storage
SDSC Comet GPU Nodes (Expanse GPU)	SDSC	compute
SDSC Dell Cluster with Intel Haswell Processors (Expanse)	SDSC	compute
SDSC Medium-term disk storage (Data Oasis)	SDSC	storage
Stanford University GPU Cluster (XStream)	Stanford U	compute
TACC Data Analytics System (Jetstream)	TACC	compute
TACC Dell/Intel Knights Landing, Skylake System (Stampede2)	UT Austin	compute
TACC Long-term Storage (Jetstream Storage)	TACC	storage
TACC Long-term tape Archival Storage (Corral)	TACC	storage

NSF HPC Resources

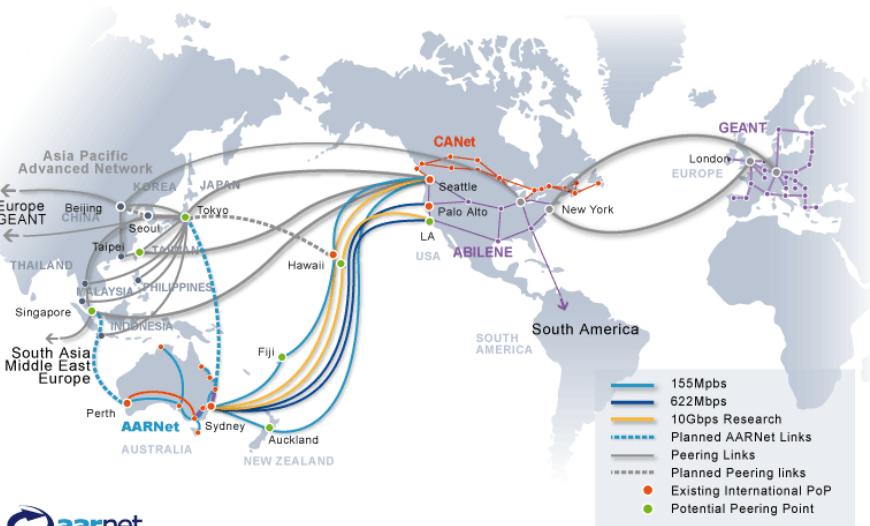
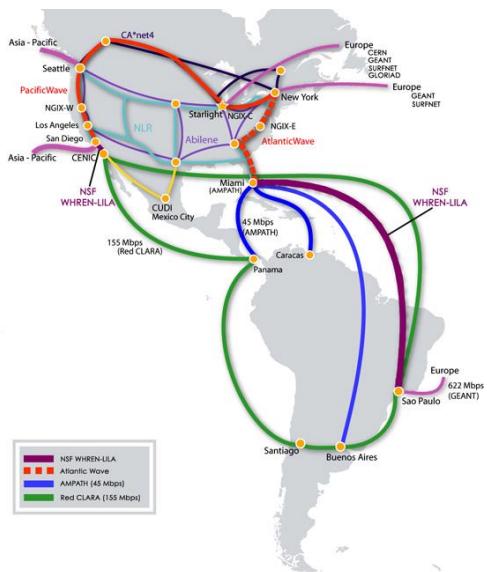
High Performance Computing

NSF-supported National Computing Resources

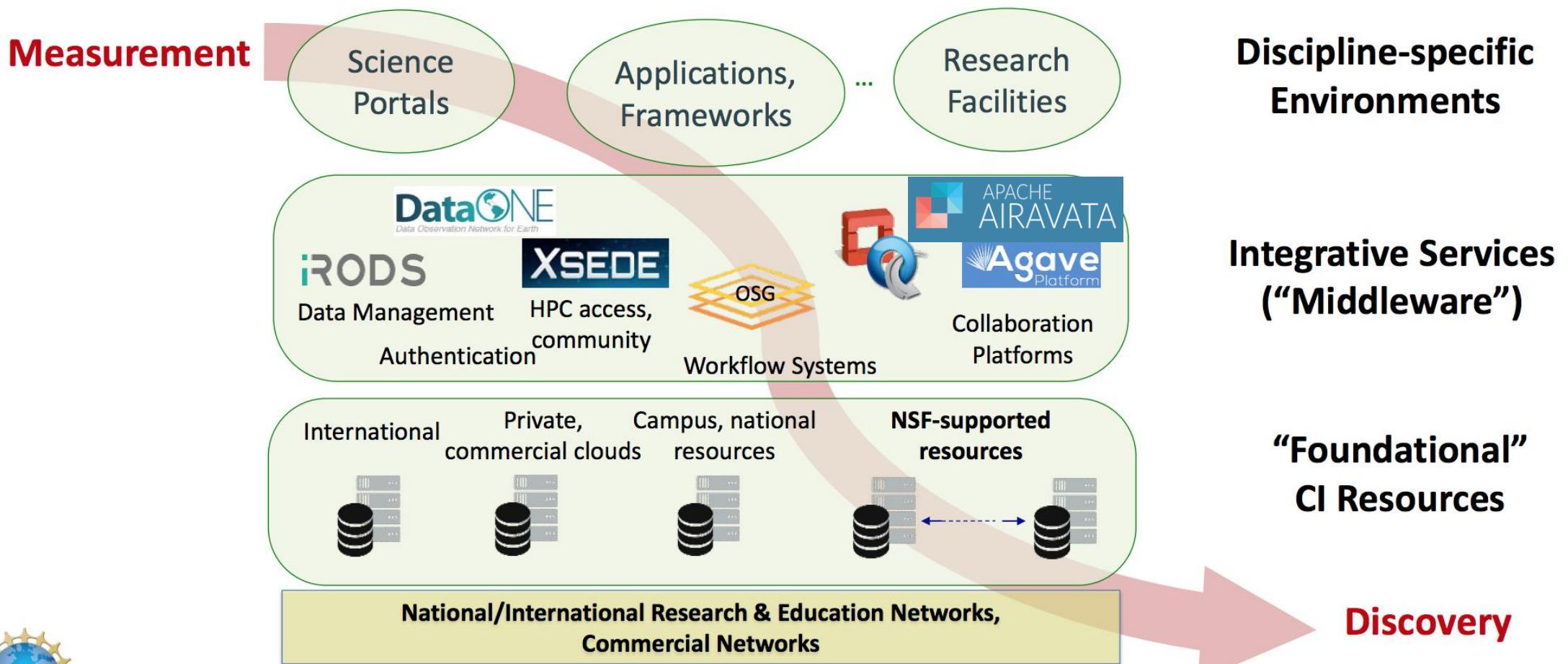
Complements Larger Aggregate Investments from Universities and other Agencies



VO-Global: International R&E Networking



Emerging discovery pathways at scale: Architecture view



Science Gateways

Web interfaces and middleware for integrating distributed computing, automating expertise, controlling access, managing results, and speeding up your critical computational workflows

Science gateways are Web and desktop interfaces to high performance computing clusters, computing clouds.

Science gateways encode expertise

- Running specific scientific application
- Running jobs on diverse, nonlocal machines
- Moving data to and from world-wide resources

Science gateways enable sharing of results

Science gateways make results recoverable and reproducible

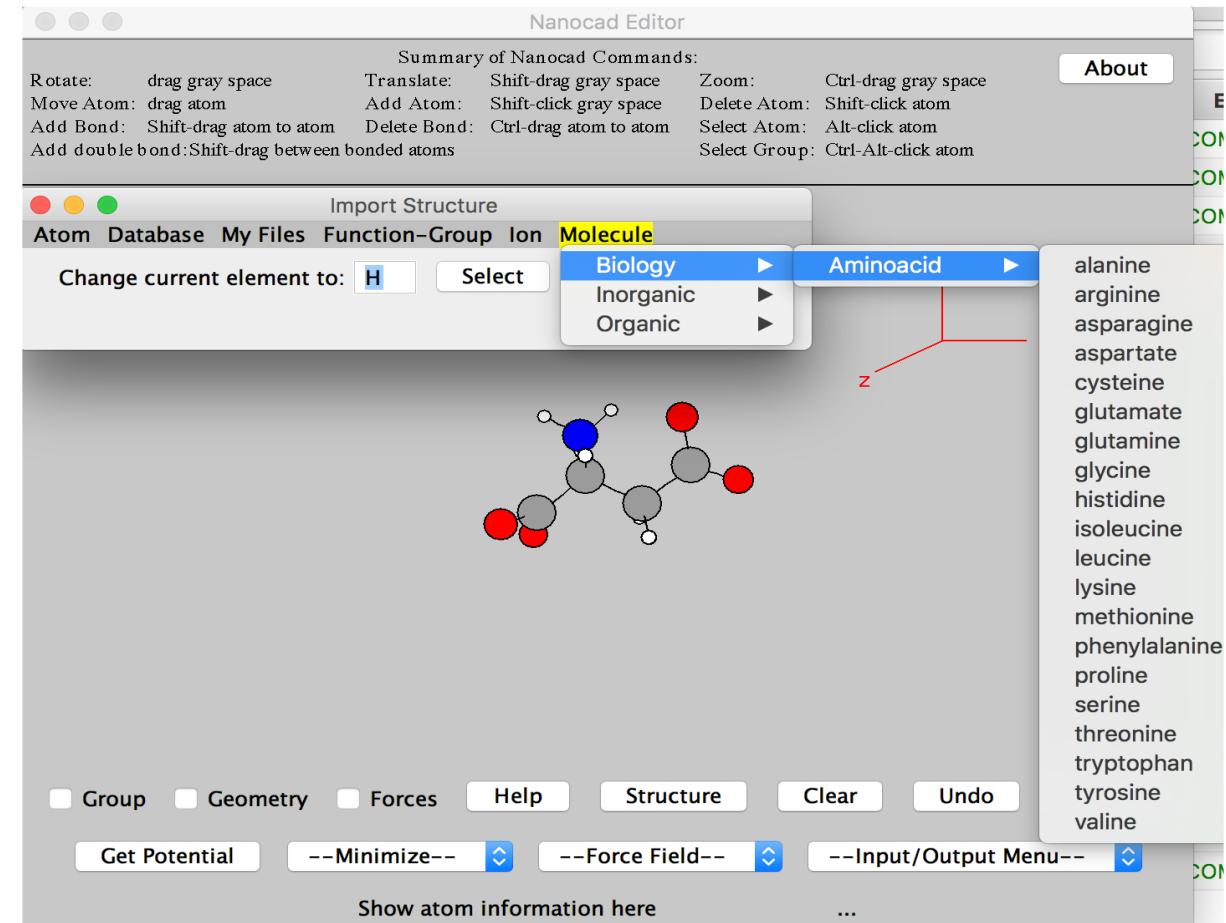
Technology Adoption Choices

```
import sys
conf = json.load(open('tosconfig.json', 'r'))
tos_session = requests.Session()
tos_project_list = [x['chargeCode'] for x in tos_session.get(conf['api_url'] + '/projects/resource/Jetstream', auth=(('tos-jetstream', conf['tos_pass']))).json()['result']]
auth = v3.Password(auth_url=conf['os_auth_url'], user_id=conf['os_user_id'], password=conf['os_password'], project_id=conf['os_project_id'])
sess = session.Session(auth=auth)
keystone = client.Client(session=sess)

connection = pymysql.connect(host=conf['mysql_host'], user=conf['mysql_user'], passwd=conf['mysql_pass'], db='ceilometer')
cursor = connection.cursor()

cursor.execute('select id,generated from event where event_type_id=28 and generated >?F order by generated % conf["last_generated"]')

su_table = {'m1.tiny': 1, 'm1.small': 2, 'm1.medium': 6, 'm1.large': 10, 'm1.xlarge': 24, 'm1.xxlarge': 44}
project_cache = {}
user_cache = {}
for item,generated in cursor.fetchall():
    try:
        cursor.execute(**"select trait_text.value from trait_text where event_id=%d and trait_text.key='state'"% item)
        if cursor.fetchone()[0]=='active':
            cursor.execute(**"select trait_text.value from trait_text where event_id=%d and trait_text.key='project_id'"% item)
            project_id = cursor.fetchone()[0]
            cursor.execute(**"select trait_text.value from trait_text where event_id=%d and trait_text.key='user_id'"% item)
            user_id = cursor.fetchone()[0]
            cursor.execute(**"select trait_text.value from trait_text where event_id=%d and trait_text.key='host'"% item)
            host = cursor.fetchone()[0]
            cursor.execute(**"select trait_text.value from trait_text where event_id=%d and trait_text.key='instance_id'"% item)
            instance_id = cursor.fetchone()[0]
            cursor.execute(**"select trait_text.value from trait_text where event_id=%d and trait_text.key='instance_type'"% item)
            instance_type = cursor.fetchone()[0]
            cursor.execute(**"select trait_datetime.value from trait_datetime where event_id=%d and trait_datetime.key='audit_period_beginning'"% item)
            audit_period_beginning = datetime.utcnow().timestamp(cursor.fetchone()[0])
            cursor.execute(**"select trait_datetime.value from trait_datetime where event_id=%d and trait_datetime.key='audit_period_end'"% item)
            audit_period_end = datetime.utcnow().timestamp(cursor.fetchone()[0])
            cursor.execute(**"select trait_datetime.value from trait_datetime where event_id=%d and trait_datetime.key='launched_at'"% item)
            launched_at = datetime.utcnow().timestamp(cursor.fetchone()[0])
            su = su_table[instance_type] * (audit_period_end - audit_period_beginning).total_seconds() / 3600
            if project_id not in project_cache:
                try:
                    project_cache[project_id] = keystone.projects.get(project_id).name
                except:
                    project_cache[project_id] = project_id
            if project_id in tos_project_list:
                if user_id not in user_cache:
                    user_cache[user_id] = keystone.users.get(user_id).name
                d = {'endUTC': audit_period_end.strftime('%Y-%m-%dT00:00:00'), 'project': project_cache[project_id], 'queueName': host, 'queueUTC': launched_at.strftime('%Y-%m-%dT00:00:00'), 'resource': 'Jetstream', 'schedulerId': instance_id + '-' + item, 'startUTC': audit_period_beginning.strftime('%Y-%m-%dT00:00:00'), 'sus': su, 'username': user_cache[user_id], 'cpus': su_table[instance_type]}
                print(d)
    except:
        pass
```

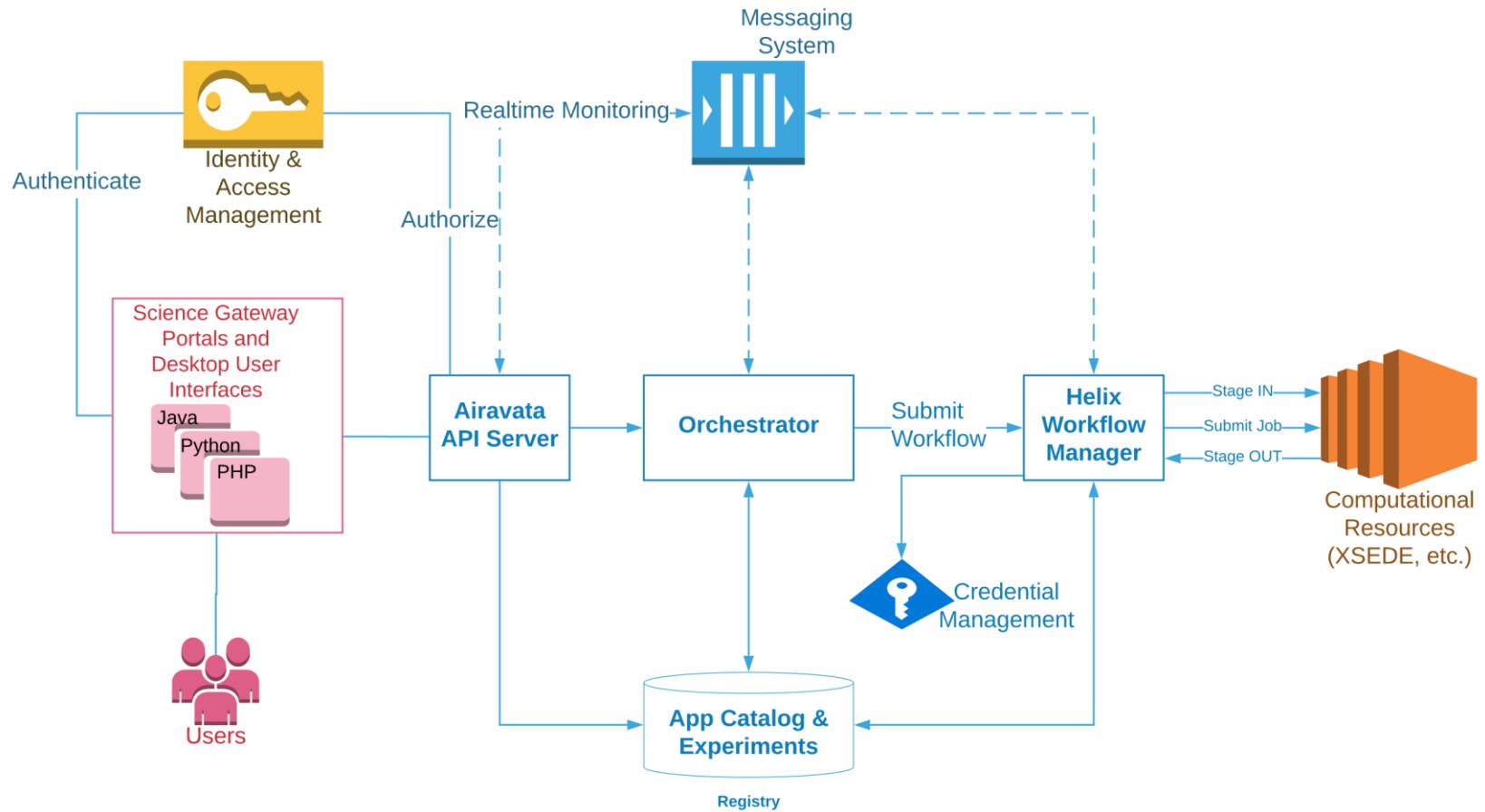


Apache Airavata



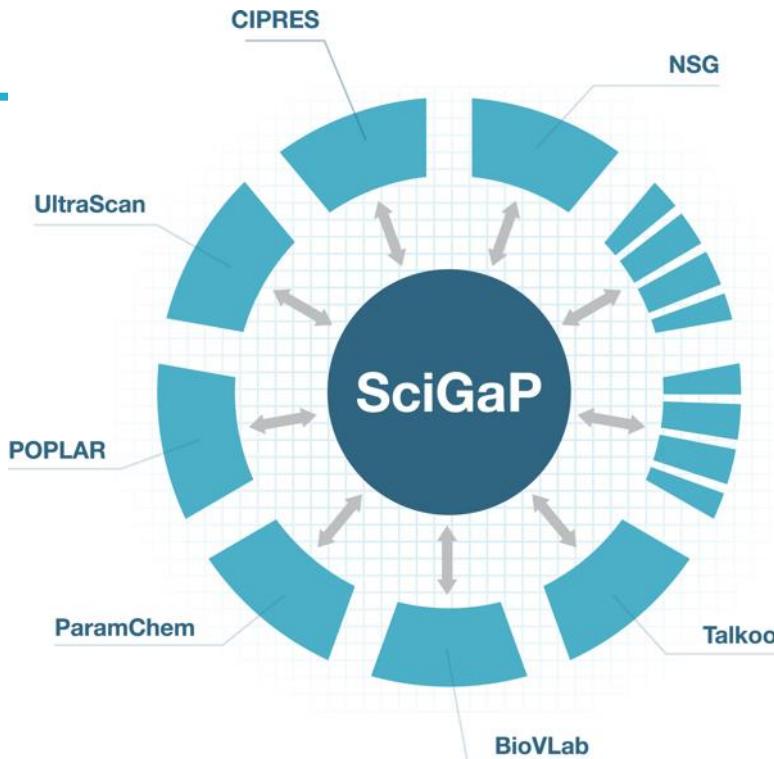
APACHE AIRAVATA

- Open source science gateway framework (airavata.apache.org)
- Enable sharing of
 - Applications
 - Compute/storage resources
 - Simulation results
- API and full-featured UI clients



SciGaP Hosting Services

- Airavata is multi-tenanted
- The SciGaP project runs a hosted instance of Airavata, supports 30+ science gateways
- Web Portal hosting
- Common use case: Software-as-a-service science gateway
- Consulting help also available
- Request a gateway at <https://scigap.org>
- Sustainably operated by SGRC after NSF funding through NSF SI2 program



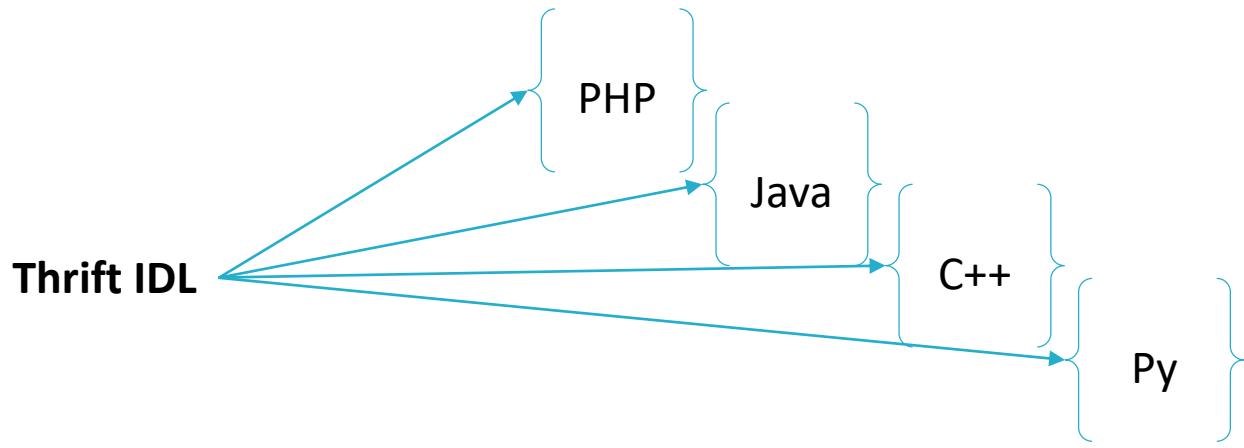
**SCIENCE GATEWAYS
RESEARCH CENTER**

INDIANA UNIVERSITY
Pervasive Technology Institute

Building an Airavata client

Option 1: Build your own

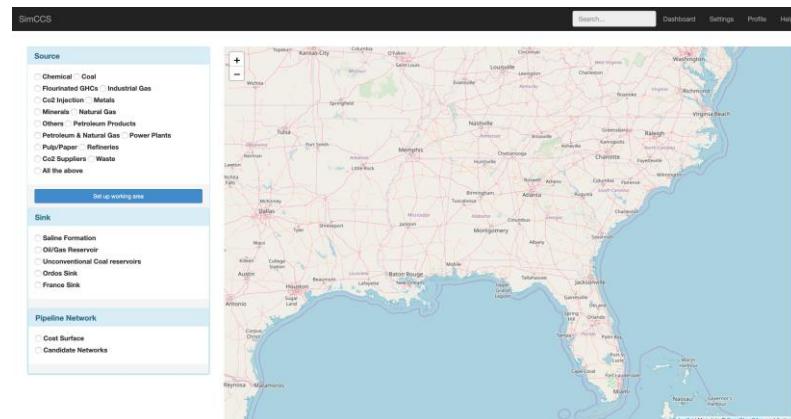
- Thrift based API
- Clients exist in Java, Python, PHP, but many more target languages possible
- Example code is available



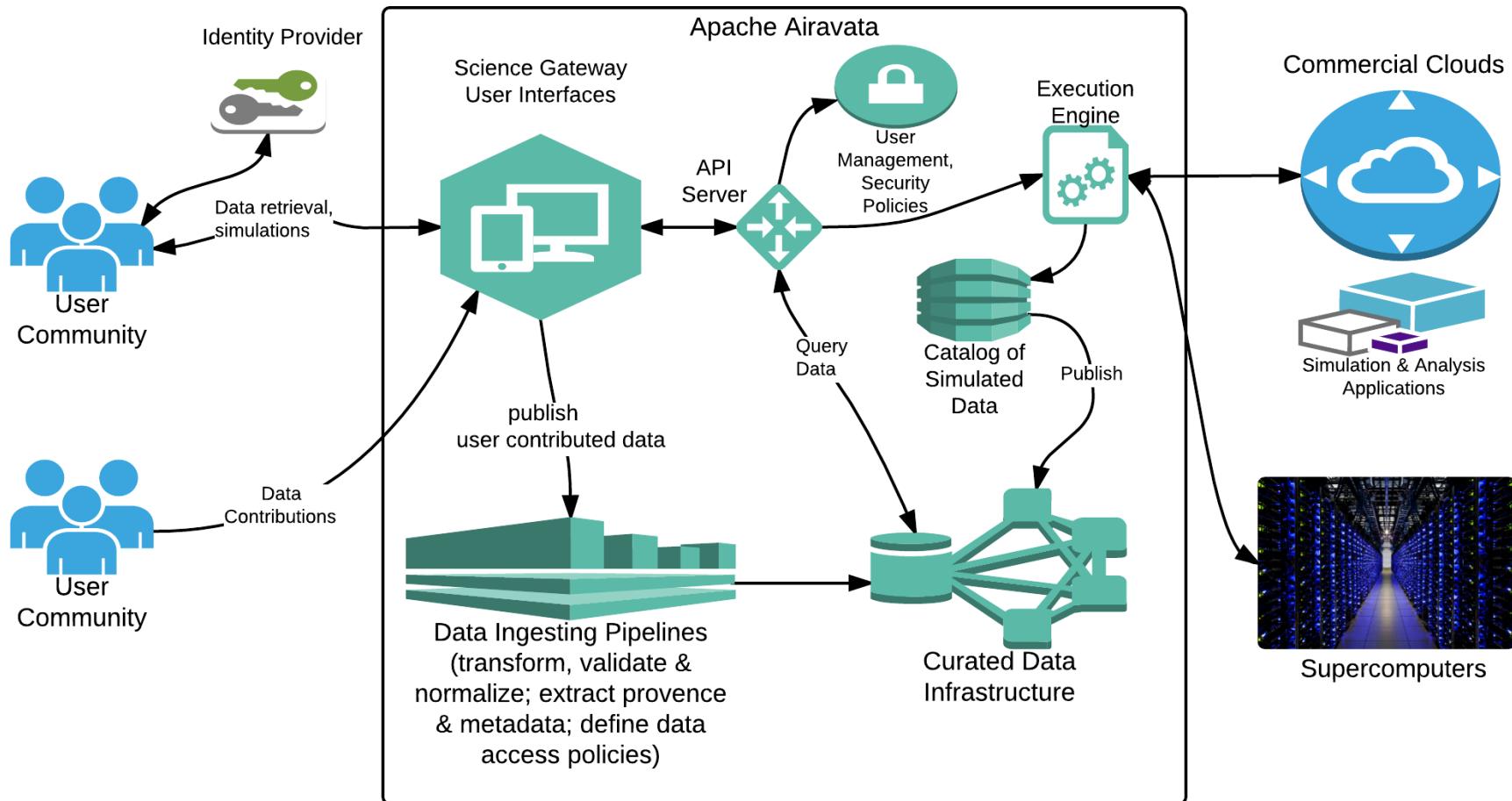
Option 2: Build extensions to Django Portal

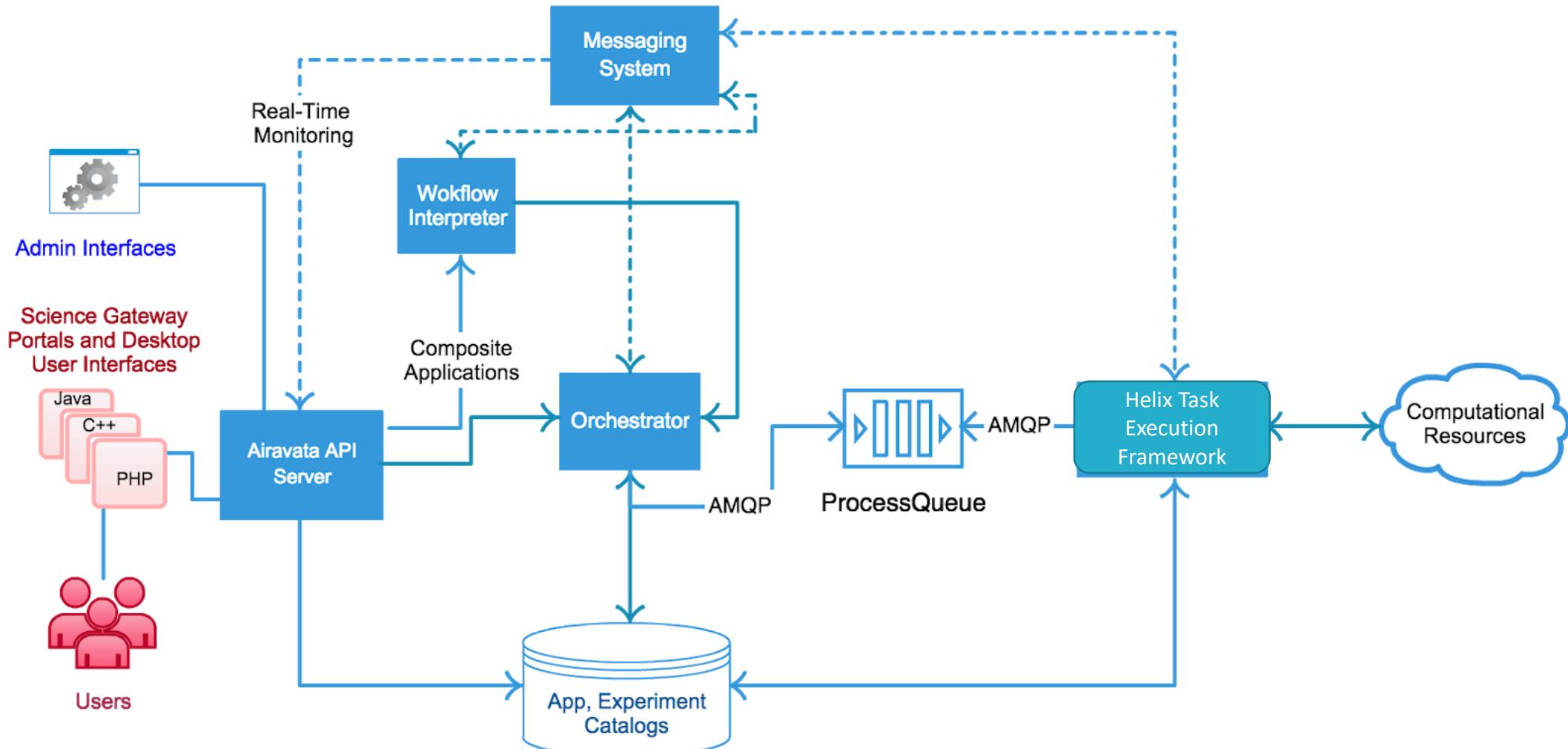
- New Django Portal (beta)
 - production availability this summer
- Extensibility
 - Add additional Django apps
 - Custom experiment input editors
 - Custom experiment output viewers
 - Leverage REST API and JS libraries

Example custom
Django app:
SimCCS Maptool



Science Gateway Architecture





What Is Apache Airavata?

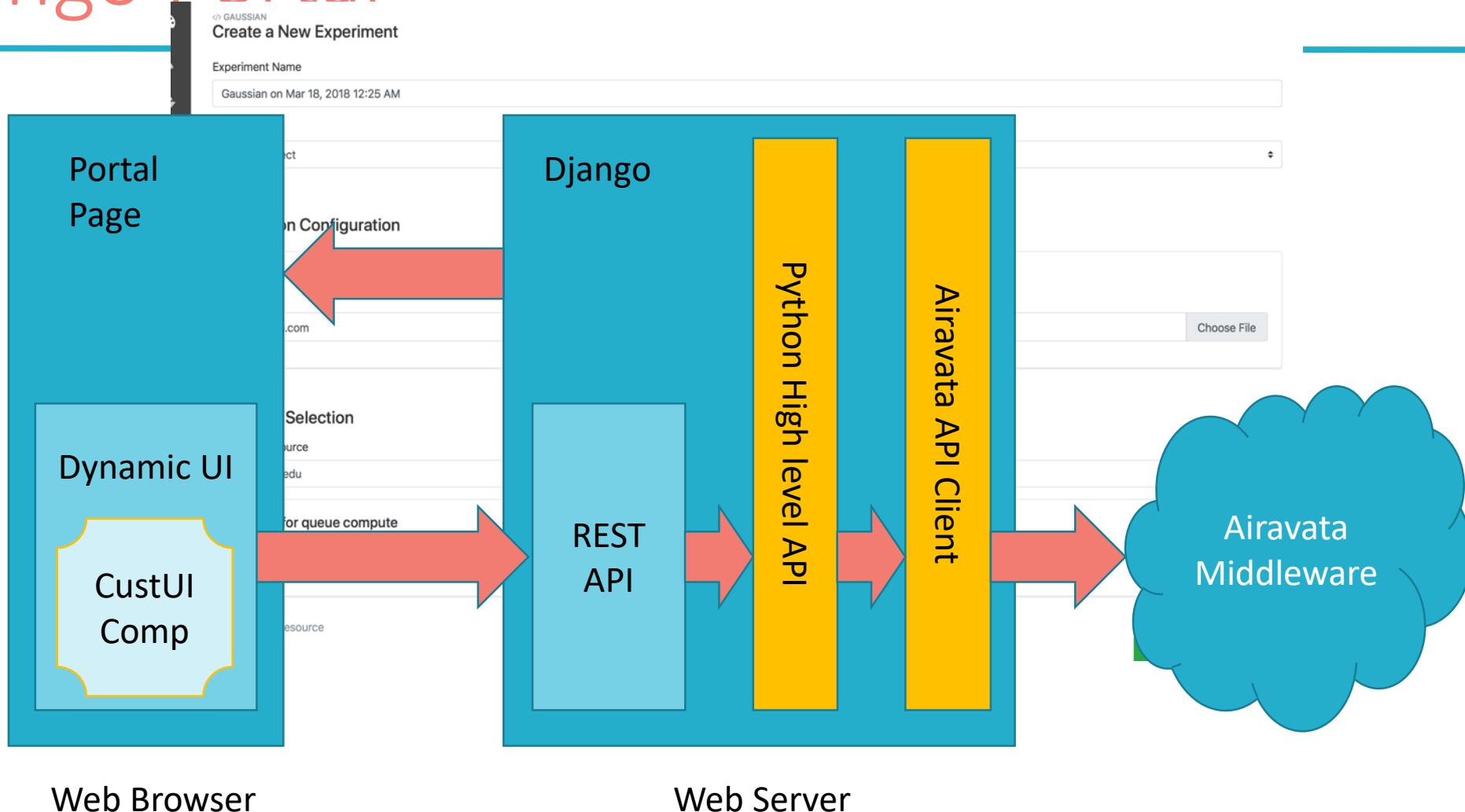
- Apache Airavata is software for building science gateways.
 - Don't start from scratch
- Airavata-based gateways integrate clusters and supercomputers from all over the world.
 - We can make your resources available to your team.
 - We can help you access supercomputers, clusters, and computing clouds from outside your institution or enterprise.

Some Gateways Built with Apache Airavata

Gateway	Description
UltraScan	Support for data analysis of analytical ultracentrifugation experiments
GeoGateway	Earthquake modeling and data access to support NASA and other researchers
IU Cybergateway	Campus gateway for accessing campus resources (in revision)
University of South Dakota Gateway	Campus gateway specializing in chemistry and bio applications. Other campus gateways: Oklahoma University , University of Utah, Georgia State University
dREG Science Gateway	Gateway for locating and understanding Transcriptional Regulatory Elements (TREs) that encode the temporal and spatial patterns of gene expression.
Oklahoma Innovation Institute Gateway	Gateway for computational chemistry and engineering applications
PHASTA Gateway	Gateway for computational engineering, finite element simulation

For a complete list, please see <https://circ.iu.edu/collaborations.html>

Django Portal



Custom UI Components

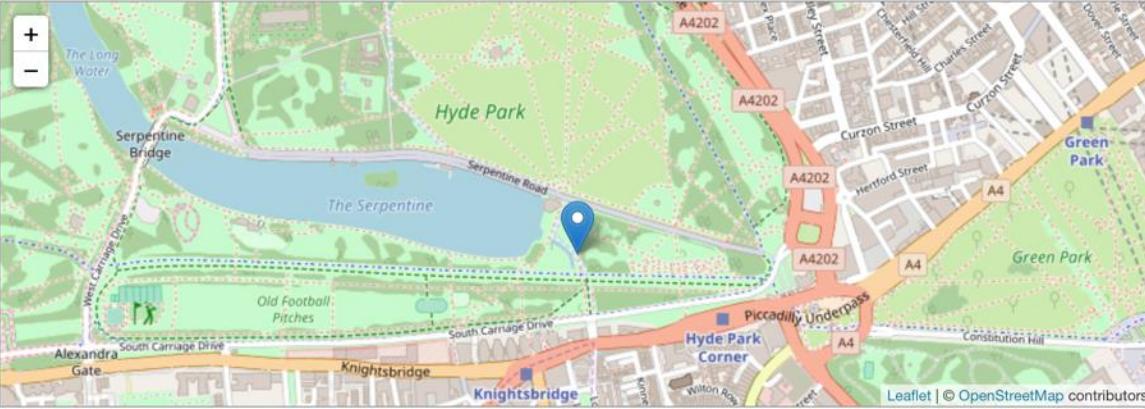


Image credit: leafletjs.com website

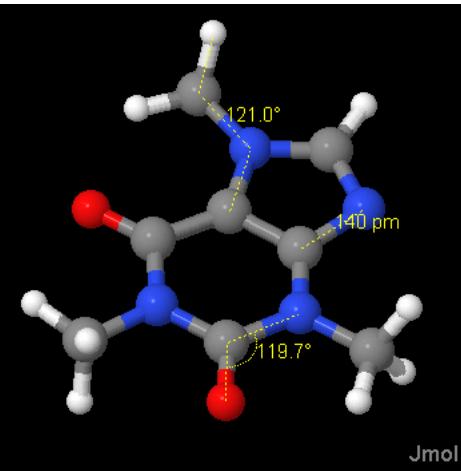
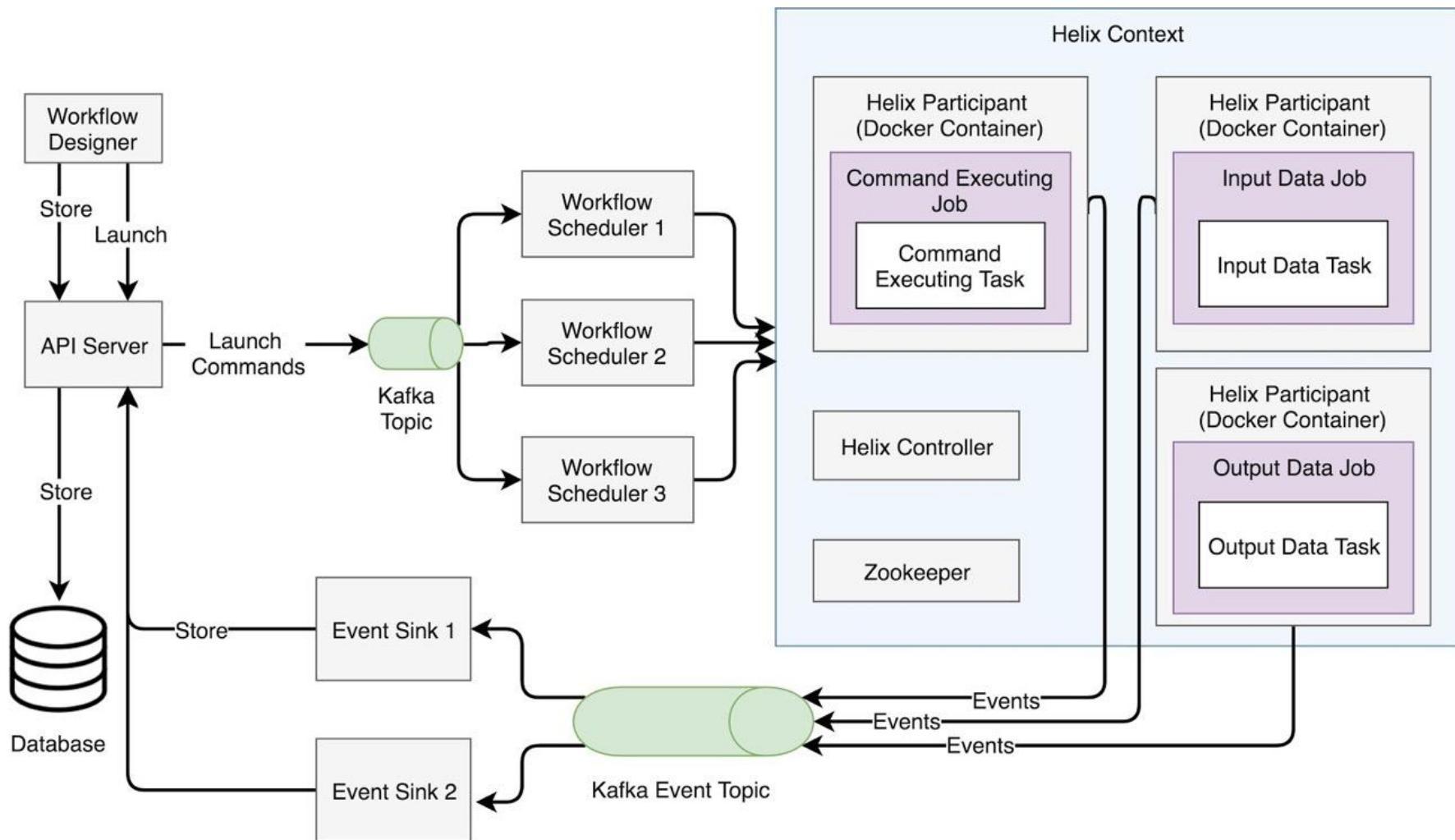


Image credit: jmol.sourceforge.net website

Helix-based Task Execution

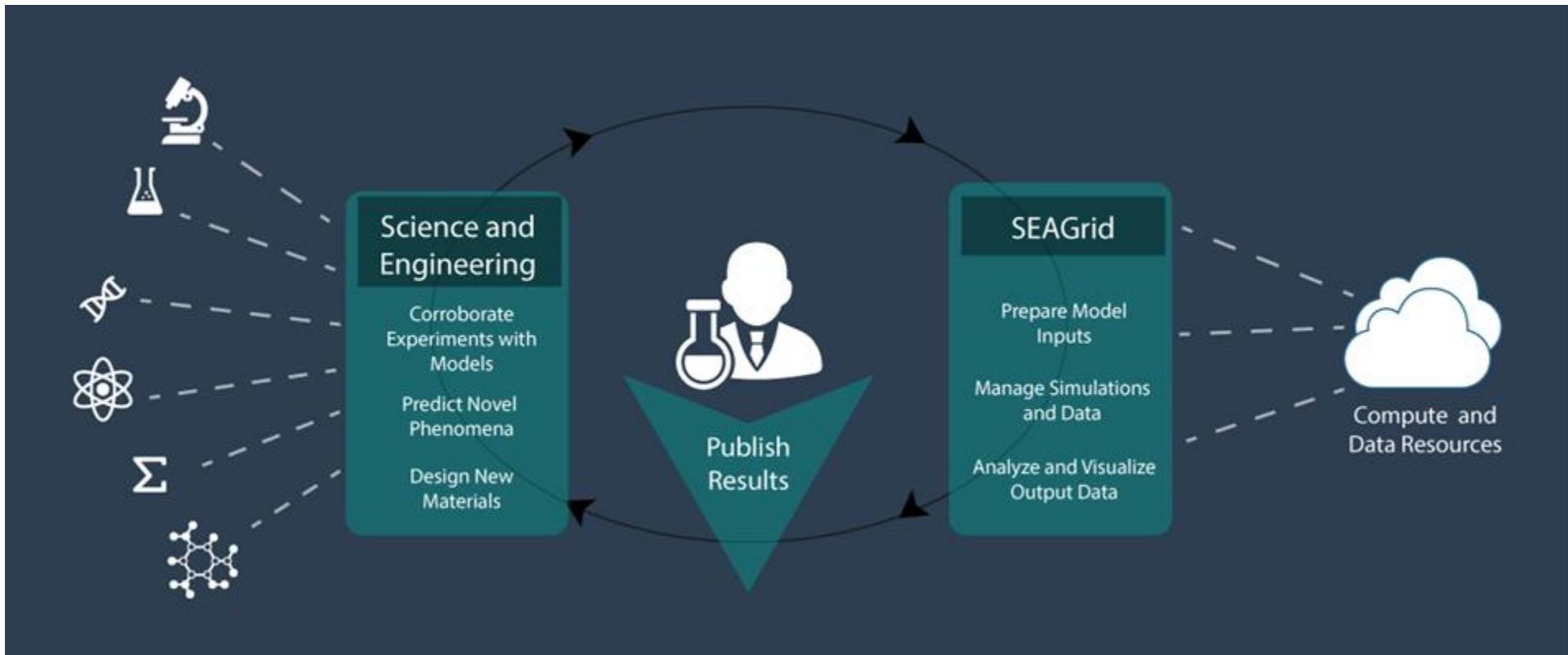


Group-based Authorization

- Currently, authorization based on statically defined roles
- Group-based authorization gives gateway admins more fine-grained control
- But also opens the door to allowing users to share resources and applications with other users

Apache Airavata Summary

- Apache Airavata has been used as the basis for several kinds of gateways, Single Application, Campus, Domain Specific, Multi-disciplinary etc..
- We can integrate Workflows, Data and Metadata, Data reuse mechanisms, sharing and Group wise access control
- Apache Airavata is Open Source Open Community software framework and can accommodate contributors from many disciplines
- Education, Training, Work-force development opportunities to get involved are available



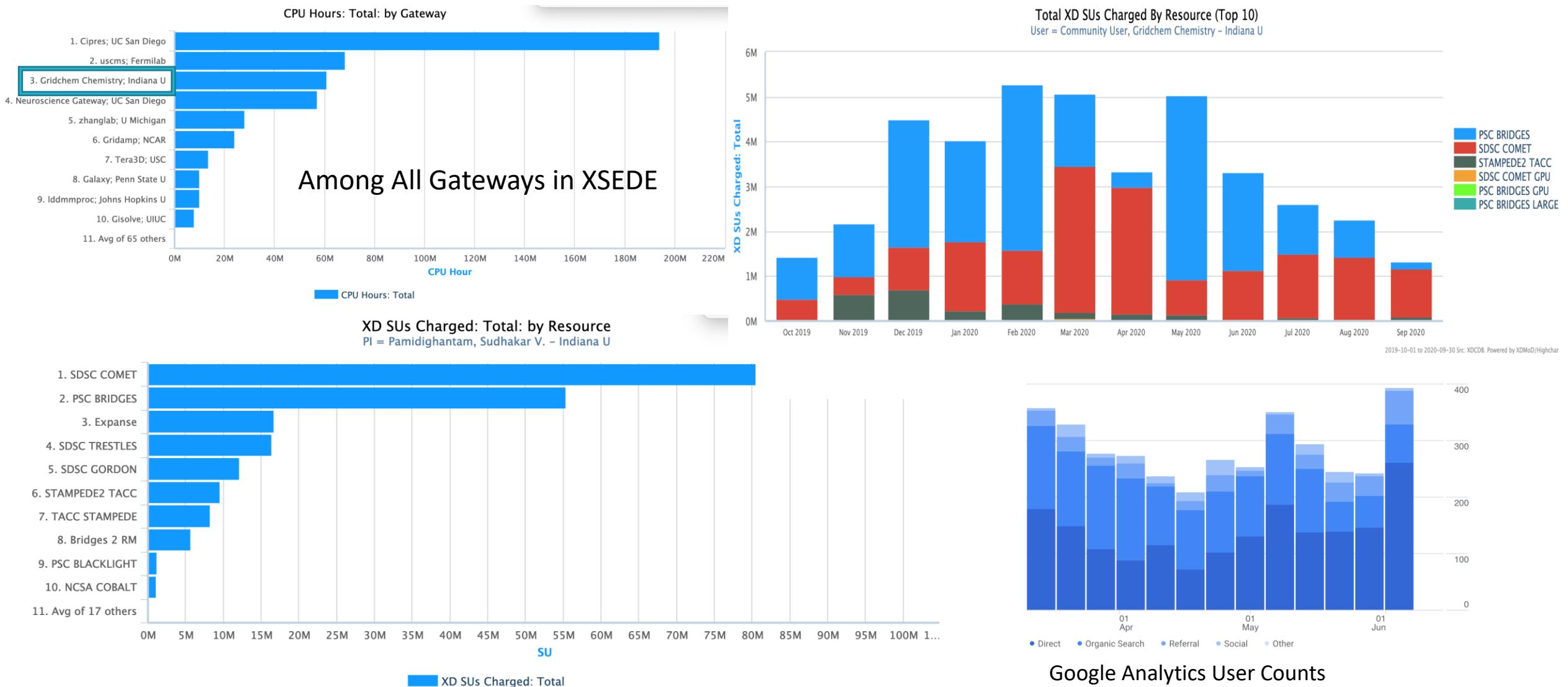
SEAGRID.org is an Apache Airavata-powered gateway

SEAGrid 16 Years in Service

- NSF XSEDE Gateway 2021 Allocation valued at \$2,974,326.00
- NSF XSEDE Resources NCSA, SDSC, TACC, PSC, IU, LSU
- Bigred3 and Carbonate (IU), CCR U. Buffalo
- 1662 Registered Users, ~ 400 Active
- 222M XD SUs for 222K Jobs Since 2005
- More than 80 Publications since 2015
- Mainly Chemistry Applications
- Desktop Client for Pre and Post processing
- Dynamic Information Services (RSS, HPC Load data, Queue Prediction) Inherent workflow capability – Checkpoint Reuse, High throughput and Coupled Applications
- Allocations, PI specific user and resource management and Job level usage monitoring
- Consulting – Adaptive Services
- Data Archive

Application	Site	Host
Abaqus 6.14-1/6.13-3	SDSC	Comet
Abinit	Multiple	TACC/SDSC
AceMD	SDSC	Comet-GPU
Amber	Multiple	Systems in XSEDE
AutoDock-Vina	TACC,SDSC	Stampede2,Comet
BoltzTrap	SDSC	Comet
Charmm	SDSC	Comet
CP2K	SDSC	Comet
CPMD	SDSC	Comet
Dalton	PSC	Bridges
DDScat	SDSC	Comet
DFTB+	TACC	Stampede2
Gamd-Namd	SDSC	Comet
Gamess	SDSC	Comet, Bigred3, Karst
Gaussian 09/16	SDSC	Comet, Bridges, Bigred3, CO3
GNAT	TACC	Stampede2
Graph-MBT-ONIOM'	PSC	Bridges
Gromacs	Multiple	Systems in XSEDE
Lammps	Multiple	Systems in XSEDE
Molcas	TACC	Stampede
Nek5000	SDSC	Comet, Bigred3
NWChem	Multiple	Systems in XSEDE
Octopus	TACC	Stampede2
OpenMM	SDSC	Comet
Orca	TACC,PSC	Stampede2, Bridges
PolyUMod	SDSC	Comet
PSI4	SDSC	Comet, Stampede2
Quantum Espresso/6.1	Multiple	Systems in XSEDE

SEAGrid Usage



SEAGrid Gateway Admin View

The screenshot shows the SEAGrid Gateway Admin View homepage. At the top, there's a navigation bar with links for Download, Documentation, Publications, SEAGrid Data, About, Contact, and Admin Dashboard. Below the navigation bar, the page title is "Gateway: seagrid" and the subtext is "Let's get started!". There are several sections: "See what's happening in your projects" with links to Browse Projects, Browse Experiments, and Experiment Statistics; "Manage Users Access" with a link to Browse Users; "Manage Computing and Storage Resources and Preferences for your Gateway" with links to Compute Resources, Gateway Management, Storage Resources, and Credential Store; and "Manage Application Modules, Interfaces and Deployments" with links to Browse Application Modules, Browse Application Interfaces, and Browse Application Deployments.

The screenshot shows the SEAGrid Gateway Admin View user management page. The left sidebar includes links for Experiment Statistics, Users, Compute Resources, Storage Resources, App Catalog, Module, Interface, Deployment, Gateway Management, Credential Store, and Notices. The main content area is titled "Users:" and contains a table with columns: First Name, Last Name, Username, Email, User Enabled, and Role. A search bar at the top right allows users to search by username. The table lists 16 users, each with a checked "User Enabled" checkbox and a "Check All Roles" button. The users listed are: qwer, asdf, 123456, asdas@dasd.com; Shreeram, Sridharan, 2skera, skera2@uky.edu; ALDO, GUZMAN DUXTAN, 697004ag, aldo.guzman@unmmsm.edu.pe; Adetunji, Adeleke, aadeleke, aaadelek@iupui.edu; Aashish, Ahuja, aahuja, aahuja@calmi2.org; Alexei, Bykhovski, ab4k, ab4k@virginia.edu; a, c, abcde, 1594541251@qq.com; Abhijit, Mitra, abhijitm, abhijit.mitra@manhattan.edu; shentan, chen, achenie, chen.1038@osu.edu; D., Clabo, aclabo, dclabo@fm Marion.edu; Aaron, Culich, aculich@berkeley.edu, aculich@berkeley.edu; Srirangam, Addepalli, addepall, sri rangam.v.addepalli@ttu.edu.

First Name	Last Name	Username	Email	User Enabled	Role :
qwer	asdf	123456	asdas@dasd.com	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
Shreeram	Sridharan	2skera	skera2@uky.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
ALDO	GUZMAN DUXTAN	697004ag	aldo.guzman@unmmsm.edu.pe	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
Adetunji	Adeleke	aadeleke	aaadelek@iupui.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
Aashish	Ahuja	aahuja	aahuja@calmi2.org	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
Alexei	Bykhovski	ab4k	ab4k@virginia.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
a	c	abcde	1594541251@qq.com	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
Abhijit	Mitra	abhijitm	abhijit.mitra@manhattan.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
shentan	chen	achenie	chen.1038@osu.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
D.	Clabo	aclabo	dclabo@fm Marion.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
Aaron	Culich	aculich@berkeley.edu	aculich@berkeley.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>
Srirangam	Addepalli	addepall	sri rangam.v.addepalli@ttu.edu	<input checked="" type="checkbox"/>	<input type="button" value="Check All Roles"/>

SEAGrid Admin Dashboard Compute Resource Browser

Experiment Statistics

Users

Roles

Compute Resources

Browse

Storage Resources

Browse

App Catalog

Module

Interface

Deployment

Gateway Preferences

Credential Store

Search Compute Resources

Compute Resource Name

Name	Id	Enabled	View
Test.org	Test.org_9a4b5180-839e-499c-bca0-da957d115f90	<input checked="" type="checkbox"/>	
bigred2.uits.iu.edu	bigred2.uits.iu.edu_ac140dca-3c88-46d8-b9ed-875d96ea6908	<input checked="" type="checkbox"/>	
karst.uits.iu.edu	karst.uits.iu.edu_a9a65e7d-d104-4c11-829b-412168bed7a8	<input type="checkbox"/>	
lonestar.tacc.utexas.edu	lonestar.tacc.utexas.edu_0d2d81a2-af4f-48c3-8be9-2093ebe2b866	<input checked="" type="checkbox"/>	
gordon.sdsc.edu	gordon.sdsc.edu_bb11b481-fe7e-44d4-95ba-d3ffcd08bf08	<input checked="" type="checkbox"/>	
stampede.tacc.xsede.org	stampede.tacc.xsede.org_ea585ade-831f-4ad1-91c6-d897fb170e3b	<input checked="" type="checkbox"/>	
series.usda.gov	series.usda.gov_46bd9f3e-ab96-463b-ab5e-4e7cc527694f	<input checked="" type="checkbox"/>	
comet.sdsc.edu	comet.sdsc.edu_91b900df-0ee0-4909-89b3-98e8f64e1969	<input checked="" type="checkbox"/>	
hpc.usd.edu	hpc.usd.edu_a75b5f70-febf-4918-8448-858c158236a9	<input checked="" type="checkbox"/>	

SEAGrid Gateway Credential Store

The screenshot shows the SEAGrid Admin Dashboard at <https://seagrid.org/admin/dashboard/credential-store>. The left sidebar contains navigation links for Experiment Statistics, Users, Compute Resources, Storage Resources, App Catalog, Module Interface, Deployment, Gateway Management, and Credential Store. The main content area is titled "SSH Keys" and displays a table of tokens and their corresponding public keys. It also includes a "Generate a new token" button. Below this is a "Password Credentials" section with a table for registering new password credentials.

SSH Keys

Generate a new token

Token	Public Key	Delete
6a49b998-b4ba-467d-8d45-687fb9d09db5	ssh-rsa AAAAB3NzaC1yc2EAAAQABAAQCGiUpwGZ7hKT+6hNCjsJAZILoigqi/Y0/prDc9sLpsQ3A2kGdEz3ZUDbTi9Yn8ZID3cQD3l4Z2Hk5FkLHgMYnobfN2I0ysXwPQ9W0M6ox0BPMQdWN46nEURMXc7sZzQXZLJvGwHtUsgaKW1xKhevBr1/Z6iJM6eDyogwoHMOius/4FrzS+B9n97cw1BKXCClr05AXT6P0R1KW6AkoQLW8BiGprMozeN8FmPcoDaJMo+DOOnSSflCncGhbEAhAb0fJkwupaVwOqwTXk4R2BhHTwwMBYD+n/C07uovzb9g8eDhPKQZGxCrtORqSuu7EcDiMc+bgwm d68lfSizjN1nb	trash
4d4b8bb3-7209-4f18-a635-1539e5316cfca	ssh-rsa AAAAB3NzaC1yc2EAAAQABAAQDj0S+qTFJBBGiLHUavBjOMFwo3JpsDQYKDULJL7s9X47cc/JVJHYusxmS0FextX/Db2d5NXvLTcw/I90O7ldKRz6Czm+pc7apGCvo62N+g4ciW8AdGidvF5UL6USGJffh5wP7t3iSEKhC+EporSPkVY2vMdvb2GS3yVmEoy1xKd8gEk1vI91yZIDXTcAediQ+WhC/jKbsJlmNDqqRcJrK5FAaWtH2bL6iHnfPpDN/6W4u+ICQAxZ0Y58vvGBVClyDrV8uT/NrEbDat+bgknUgtvjsOKQuUi9ZLWkrPhz5bGtZjdBlzfFqzmd1ON2D8IEZUZjbDJu5lqq0Q0zmn	trash
3e7c6884-1e68-43d9-b3f9-fb6ea5b57ac4	ssh-rsa AAAAB3NzaC1yc2EAAAQABAAQCeFEyvZXkqeSD2iYr4DaJTwkGr/VHDwVpDDMhP3yy1JTfaSWEwlpzlhcoNdLPxcj3BqgFpJY1je51FnjuzQo6Cw4QFKrvEKx0PXzt7to2e+JiB9a9roaFRCjWC4UwsAn30hh4YM7zO5LbSC9uDeQGbnxD883fpSwgVC49BqLKKQHbAqERAHHGBh6hx/QzPuQrPfwJw8AA0EIREB1dP8C3uK/S1BcCmAvesTotzqO+P/mjRzdbYrLi9boXiJ1tu1QDC0fXJRqCor/TC36gqmMDh6T+B6BuSSk3pj03hcWFuxNr9GsTSssTL/Ts4RnliMPQvUUUHtY4IS3l25SaUJV	trash
3d65bf6d-2c9f-4166-a51b-e76e0022bd3b	ssh-rsa AAAAB3NzaC1yc2EAAAQABAAQCApfdu3cJ7rx44M41gEKVgRfefk/Vd/7dip9Xr9juX4jqmGuLoZR+hk287Hmbps6hZv23Av97XkCLIdH0pueDekthcBD04cRiQMDBp9NLz/9muqYVB1985psEbazYygagy/q/3ebi4CvEYd4chBee1V3GYyuJS2RklFo5/Otyf8x8auJSMniHgbTbVqtFEOYuPdXIRBfH37k55N7+2IXksIVgn1MDweHebQWJMx2LDckLfJitrTiwcuST55w6N/n8WYZM2t+ie3BjEr6tcGAVIuyxj7CbMYMziM/hM5J+oQGPY/IBtINZoxiZ0DeAC9SiEvE9fbTEW2iRWnQL	trash

Password Credentials

Register a new password credential

Token	Description	Delete
1ffb3180-f7fe-479b-ab60-2633cd369dd3		trash
c45d4eb1-c219-407a-b0dc-8c547673d4fc	Keycloak admin password for realm seagrid	trash

SEAGrid Application Catalog

Project: Experiment

Download Desktop App Documentation Publications About Contact

Admin Dashboard spamidig

Create a new Application Module

Existing Modules : Search by Module Name

- Echo Echo_19dc358d-d241-43d8-918c-f5a21a3b0845
- Amber_Sander Amber-Sander_e8ce4375-aa60-4dd9-aff5-61bbe1f275c0
- Gaussian Gaussian_57eb2905-1cd8-400e-ad40-cadfb8f434f
- AutoDock AutoDock_24d6d7da-1160-45b0-9958-63bcc9044804
- Gromacs Gromacs_17f97a2e-bdea-4075-9425-6bde4a8d1317
- Lammps_BR2 Lammps_1f150b05-6295-43c5-8040-27992693cef6
- NWChem NWChem_2f38a95e-c2aa-4db3-b1bf-2adb006a34bc
- Gamess_BR2 Gamess_BR2_5c979001-d994-47fb-a9af-5227ff5c5c55
- NEK5000 NEK5000_35922b60-f0ad-41dd-8b3c-7877617767d8
- Abinit Abinit_34e497ac-28cd-42a2-afe0-1ee99f173c6d
- Quantum_Espresso Quantum_Espresso_cf47dcc-c870-4e76-be68-784a1574426e
- Gromacs_CrayMPI Gromacs_CrayMPI_09bb24b1-2eae-4875-85f6-6352e7c0eb0f
- Lammps Lammps_4e799726-669e-4b87-9a7d-9d567d83c1c7

App Catalog

Module

Interface

Deployment

Gateway Preferences

Credential Store

0010-0020/e/00001

- Lammps Lammps_4e799726-669e-4b87-9a7d-9d567d83c1c7
- CP2K CP2K_3d1895c9-03e1-4ba7-96b3-94bf2c88c865
- Phasta_P Phasta_P_681394ed-212a-404e-b336-e5d41593016f
- Gamess Gamess_6ed12ada-a6a7-4df5-b044-b9a5e2b0c49b
- Tinker_Monte Tinker_Monte_b11b893a-626b-4778-997a-2da92537b4b7
- CPMD CPMD_6870dafa-e2e3-4fe6-a0e2-5b11c1436af4
- QChem QChem_9214bc96-2bf8-4ba2-b717-a1cd4dc98a4f
- DDSCat DDSCat_b01bf3eb-7e19-4ebe-bd18-32af7c81e5df
- DFTB+ DFTB+_9add4ce4-7f57-495a-bb1a-eff570b56441
- WRF WRF_17f339d1-92e3-4eb9-a72e-cb4961d685a9
- Gamess_Stampede Gamess_Stampede_1eb6c1f7-9292-456d-b1c3-1b88c3936b68
- Molcas Molcas_f7dd5a10-0f8d-43b0-9c88-ec1a648dd037
- Abaqus Abaqus_96a88449-b4cd-4a39-8ad2-26abfbdd4b3e

Application Interface Editor

Edit Application Interface

Application Name*
Gaussian

Application Description
Gaussian provides capabilities for electronic structure modeling.

Application Modules
Gaussian
[Add Application Module](#)

App Input Fields

Name*	Input-File
Value	
Type	URI
Application Argument	
Standard Input	False
User Friendly Description	Gaussian input file specifying desired calculation type, model chemistry,
Input Order	1
Data is Staged?	False
Is the Input required?	True
Required on command line?	True
Meta Data	

Continued...

[Add Application Input](#)

App Output Fields

Name*	Gaussian-Application-Output
Value	gaussian.log
Type	URI
Application Argument	
Data Movement	True
Is the Output required?	True
Required on command line?	True
Location	
Search Query	

App Output Fields

Name*	Gaussian-Standard-Out
Value	
Type	STDOUT
Application Argument	
Data Movement	False
Is the Output required?	True
Required on command line?	False
Location	
Search Query	

[Add Application Output](#)

[Update](#) [Cancel](#)

Application Deployment Editor

SEAGRID

Project Experience

Experiment Statistics

Users

Roles

Compute Resources

Browse

Storage Resources

Browse

App Catalog

Module

Interface

Deployment

Gateway Preferences

Credential Store

Edit Application Deployment

Application Module*
Gaussian

Application Compute Host*
comet.sdsc.edu

Application Executable Path*
/opt/gaussian/g09/g09

Application Parallelism Type*
SERIAL

Application Deployment Description
Gaussian provides capabilities for electronic structure modeling.

Module Load Commands
module load gaussian

Add Module Load Commands

Library Prepend Paths
Add a Library Prepend Path

Continued...

Deployment

Add a Library Prepend Path

Gateway Preferences

Credential Store

Library Append Paths

Add a Library Append Path

Environments

Add Environment

Pre Job Commands

Add Pre Job Command

Post Job Commands

Add Post Job Command

Update Cancel

Gateway Profile

Secure | <https://seagrid.org/admin/dashboard/gateway>

SEAGRID Admin Dashboard spamidig

Project Experiment Statistics
Experiment Experiment Statistics
Storage Storage

Experiment Statistics
Users
Compute Resources
Browse
Storage Resources
Browse
App Catalog
Module
Interface
Deployment
Gateway Management
Credential Store
Notices

Edit your Gateway Profile

Gateway - seagrid

SEAGRID Portal

seagrid Credential Store Token
3d65bf6d-2c9f-4166-a51b-e76e0022bd3b Set

+ Add a Compute Resource Preference

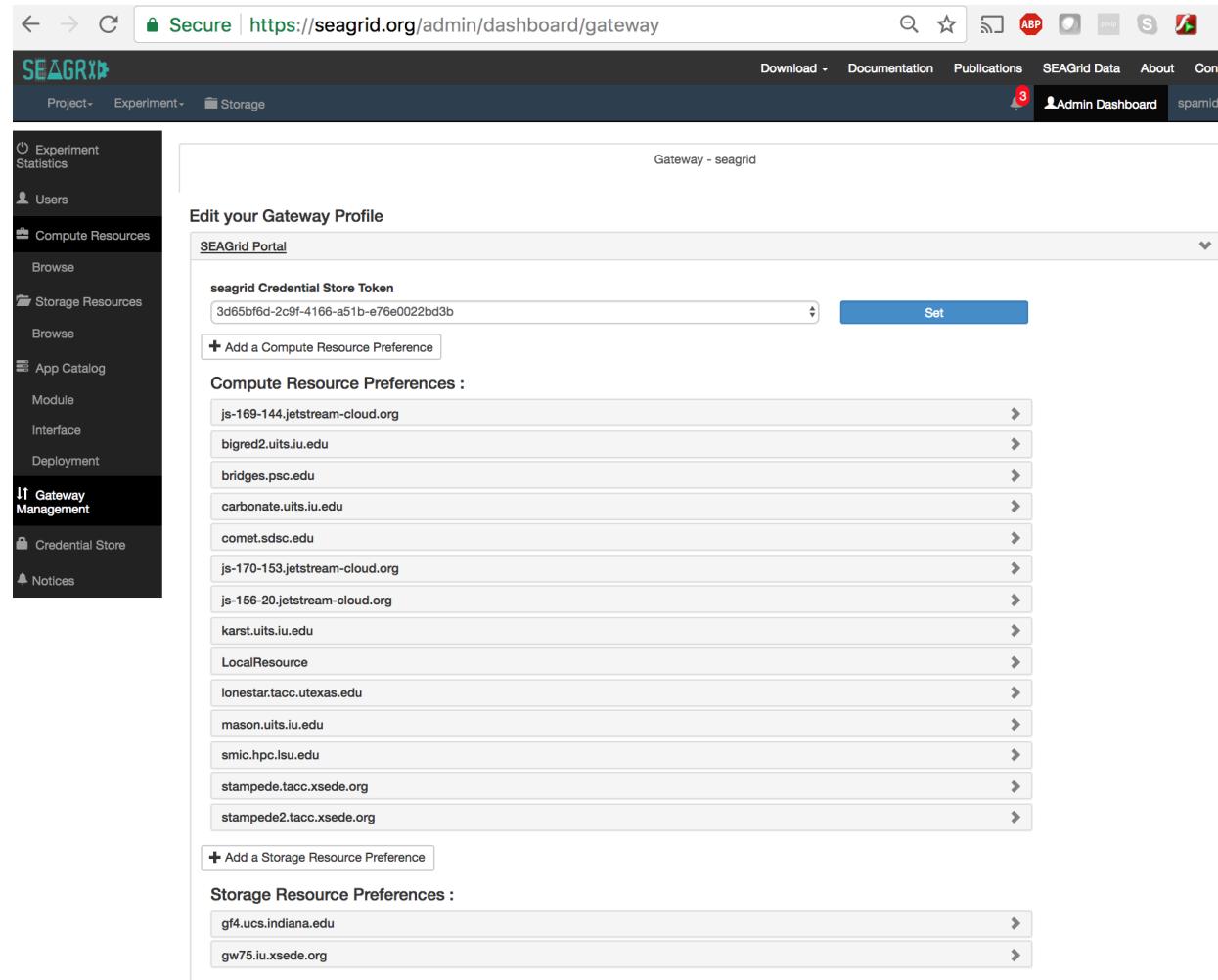
Compute Resource Preferences :

- js-169-144.jetstream-cloud.org
- bigred2.uits.iu.edu
- bridges.psc.edu
- carbonate.uits.iu.edu
- comet.sdsc.edu
- js-170-153.jetstream-cloud.org
- js-156-20.jetstream-cloud.org
- karst.uits.iu.edu
- LocalResource
- lonestar.tacc.utexas.edu
- mason.uits.iu.edu
- smic.hpc.lsu.edu
- stampede.tacc.xsede.org
- stampede2.tacc.xsede.org

+ Add a Storage Resource Preference

Storage Resource Preferences :

- gf4.ucs.indiana.edu
- gw75.iu.xsede.org



SEAGrid Experiment Creation

Create a new experiment

Experiment Name***Experiment Description****Project*****Application****Continue****Reset values****Application configuration****Application input****Input-File**[view file](#) alanine_b3lyp.inp

Gaussian input file specifying desired calculation type, model chemistry, molecular system and other parameters.

Enable Auto Scheduling **Compute Resource*****Select a Queue*****Node Count (Max Allowed Nodes - 72)****Total Core Count (Max Allowed Cores - 1728)****Wall Time Limit (Max Allowed Wall Time - 2880)**

minutes

Total Physical Memory MB**Notifications** Do you want to receive email notifications for status changes in the experiment?**Save****Save and launch****Start over**

SEAGrid Experiment Status

Experiment Summary



Enable Auto Refresh

ON

OFF

Experiment Id	Tests_3299e7ce-e87c-433c-96f7-f8514123ef07											
Name	Tests											
Description	Tests											
Project	Default Project											
Application	Gaussian											
Compute resource	comet.sdsc.edu											
Experiment Status	EXECUTING											
Job	<table border="1"><thead><tr><th>Name</th><th>ID</th><th>Status</th><th>Creation Time</th></tr></thead><tbody><tr><td>A920254545</td><td>1739851</td><td>QUEUED</td><td>2016-03-11, 9:45 AM - GMT-0600 (CST)</td></tr></tbody></table>				Name	ID	Status	Creation Time	A920254545	1739851	QUEUED	2016-03-11, 9:45 AM - GMT-0600 (CST)
Name	ID	Status	Creation Time									
A920254545	1739851	QUEUED	2016-03-11, 9:45 AM - GMT-0600 (CST)									
Creation time	2016-03-11, 9:45 AM - GMT-0600 (CST)											
Last Modified Time	2016-03-11, 9:45 AM - GMT-0600 (CST)											
Enable Auto Schedule	false											
Wall time	30											
CPU count	16											
Node count	1											
Queue	compute											
Inputs	alanine_b3lyp.inp											
Outputs	Experiment hasn't completed. Experiment Status is : EXECUTING											
Storage Directory	Open											
Errors												

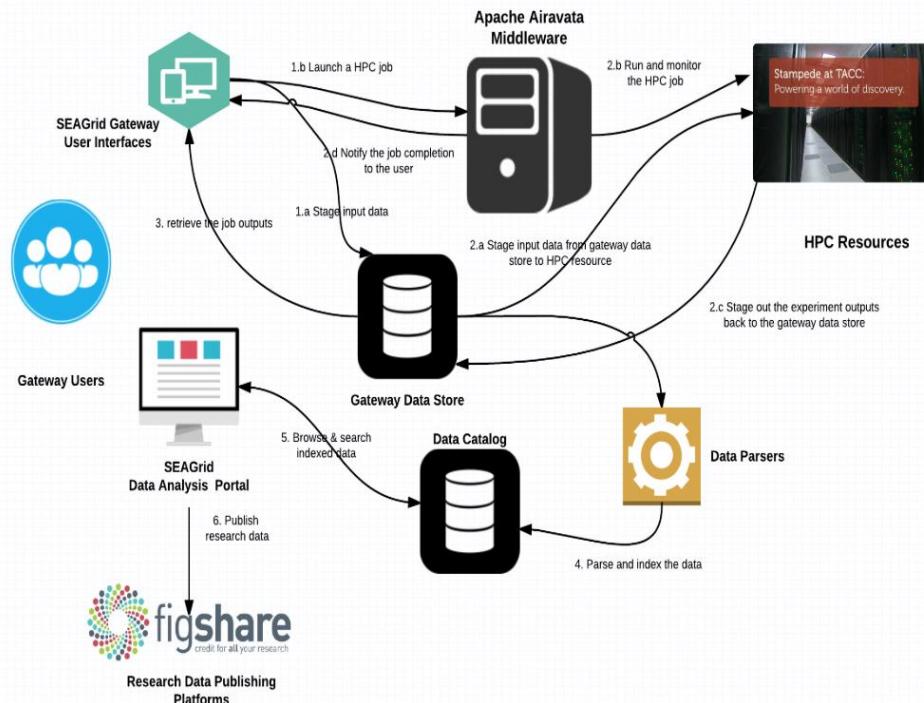
SEAGrid Admin Experiment Viewer

The screenshot shows the SEAGrid Admin Experiment Viewer interface. The left sidebar contains navigation links for Experiment Statistics, Users, Compute Resources, Storage Resources, App Catalog, Module, Interface, Deployment, Gateway Management, Credential Store, and Notices. The main content area is titled "Experiments". It features a search bar with the placeholder "Enter Experiment Id to View Summary:" and a "Get" button. Below the search bar is an "Overview" section with buttons for "Get Experiments from Last 24 h" and "Get Experiments from Last Week". It also includes a date range selector for "Select dates between which you want to review experiment statistics" and a "Get Statistics" button. The main summary section displays experiment statistics from March 1, 2018, to March 16, 2018, 9:06:00 PM. The statistics are presented in six colored boxes:

Total Experiments	Created Experiments	Running Experiments
3164	23	54
All	CREATED VALIDATED	SCHEDULED LAUNCHED EXECUTING

Successful Experiments	Canceled Experiments	Failed Experiments
3059	2	26
COMPLETED	CANCELLING CANCELLED	FAILED

SEAGrid Data Catalog



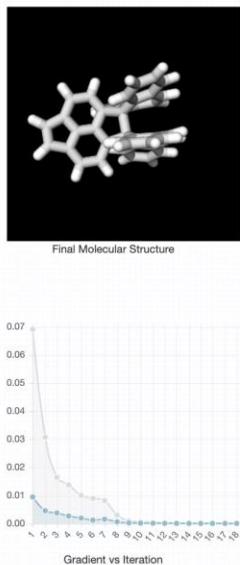
Experiment Name	Owner Name	Package	Formula	Finished Time	Basis Set	Number of Basis Functions	Energy
Clone_of_neopentanediol_G16_B5	spamidig	Gaussian 16, Revision B.01	C5H12O2	07-Mar-2019	CC-pVDZ	158	-348.2027935
synpentane24diol.153	spamidig	Gaussian 16, Revision B.01	C5H12O2	24-Jul-2018	CC-pVDZ	158	-348.2198212
npenatne12diol	spamidig	Gaussian 16, Revision B.01	C5H12O2	24-Jul-2018	CC-pVDZ	158	-348.2027935
NPENTANEDIOL.153	spamidig	Gaussian 16, Revision B.01	C5H12O2	24-Jul-2018	CC-pVDZ	158	-348.2174693
npenthane24diol.153	spamidig	Gaussian 16, Revision A.03	C5H12O2	23-Jul-2018	CC-pVDZ	158	-348.2174692
isopentane13diol	spamidig	Gaussian 16, Revision A.03	C5H12O2	20-Jul-2018	CC-pVDZ	158	-348.2018422
Clone_of_Neopentane	spamidig	Gaussian 16, Revision B.01	C5H12O2	20-Jul-2018	CC-pVDZ	158	-348.2027935
Clone_of_Clone_of_Neopentane	spamidig	Gaussian 16, Revision B.01	C5H12O2	20-Jul-2018	CC-pVDZ	158	-348.2027935
neopentanediol	spamidig	Gaussian 16, Revision B.01	C5H12O2	17-Jul-2018	CC-pVDZ	158	-348.2027935
npenthane24diol.153	spamidig	Gaussian 16, Revision B.01	C5H12O2	17-Jul-2018	CC-pVDZ	158	-348.2174692

[Export to CSV](#)

next

Output Methods

N.B: This data is automatically extracted using set of configured parser and may contain errors. Please report any issues in the issue tracker



Canonical SMILES	c1ccc2c(c1)C=Cc1c([C@@]32c2ccc4c5c2c([C@@]23c3cccc3C=Cc3c2cccc3)ccc5C=C4)cccc1
Calculation	
Package	Gaussian 09, Revision E.01
Calculation Type	FOpt; Freq
Methods	RB3LYP; RB3LYP
Basis Set	6-31G(d,p)
Number of Basis Functions	760
Number of Molecular Orbitals in the Calculation	760
Keywords	# RB3LYP/6-31G(d,p) GFInput GFPrint lop(6/7=3) Opt Freq; #NGeom=AllCheck Guess=TCheck SCRF=Check Test GenChk RB3LYP/6-31G(d,p) Freq
Job Status	CalcDone
Calculated Properties	
Energy (au)	-1616.0971637
Dipole (debye)	-0.0328227,-0.1402467,0.0158595
HF (au)	-1616.0971637
Homos	138
Homo Eigenvalue (ev)	-5.340506429913
Homo Eigenvalues (ev)	Homo - 1 : -5.3832283044415, Homo - 2 : -5.78514046163, Homo - 3 : -5.8621486813215, Homo - 4 : -6.3320893011349995, Homo - 5 : -6.3739948341120005, Homo - 6 : -6.556311113946999, Homo - 7 : -6.667333564951, Homo - 8 : -6.744613898493, Homo - 9 : -6.965570345099,

Lumo Eigenvalue (ev)	-1.9453419172245
Lumo Eigenvalues (ev)	Lumo + 1 : -1.3630182771545, Lumo + 2 : -0.9167515623345, Lumo + 3 : -0.5586497350765, Lumo + 4 : -0.3235433682445, Lumo + 5 : -0.0383680529205, Lumo + 6 : 0.1820441659845, Lumo + 7 : 0.4051775233945, Lumo + 8 : 0.519737454455, Lumo + 9 : 0.573615996854,
Zero Point Energy Hartree/Particle)	0.5430401
lImag	0
Thermal (Hartree/Particle)	0.5721956
Enthalpy (Hartree/Particle)	0.573140;
Gibbs (Hartree/Particle)	0.485218;
Execution Environment	
Calculated By	GCOMMUNI
Calculated Machine	QINC-R677
Finished Time	14-Mar-2018
Job CPU Run Time	130491.1 seconds
Memory	3000 MB
Number of Shared Processors	28
Input File Configuration	
Link 0 Commands	%nproc=28;%mem=3000MB;%Chk=longccHC_CeN.chk
Route Commands	# RB3LYP/6-31G(d,p) GFInput GFPrint lop(6/7=3) Opt Freq

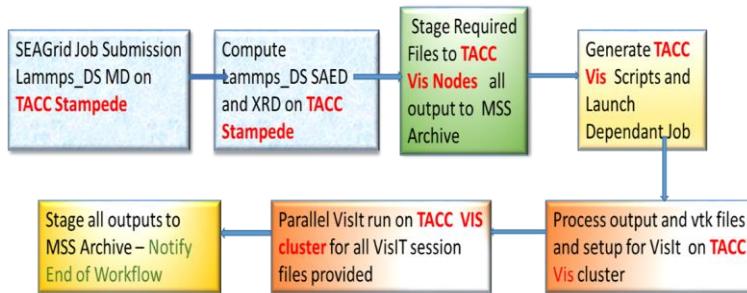
File Set	
Gaussian Input File	gaussian.in
Gaussian Output File	gaussian.log
Gaussian Checkpoint File	longccHC_CeN.chk
SDF Structure File	structure.sdf
PDB Structure File	structure.pdb
InChI File	inchi.txt
SMILES File	smiles.txt

Make Public

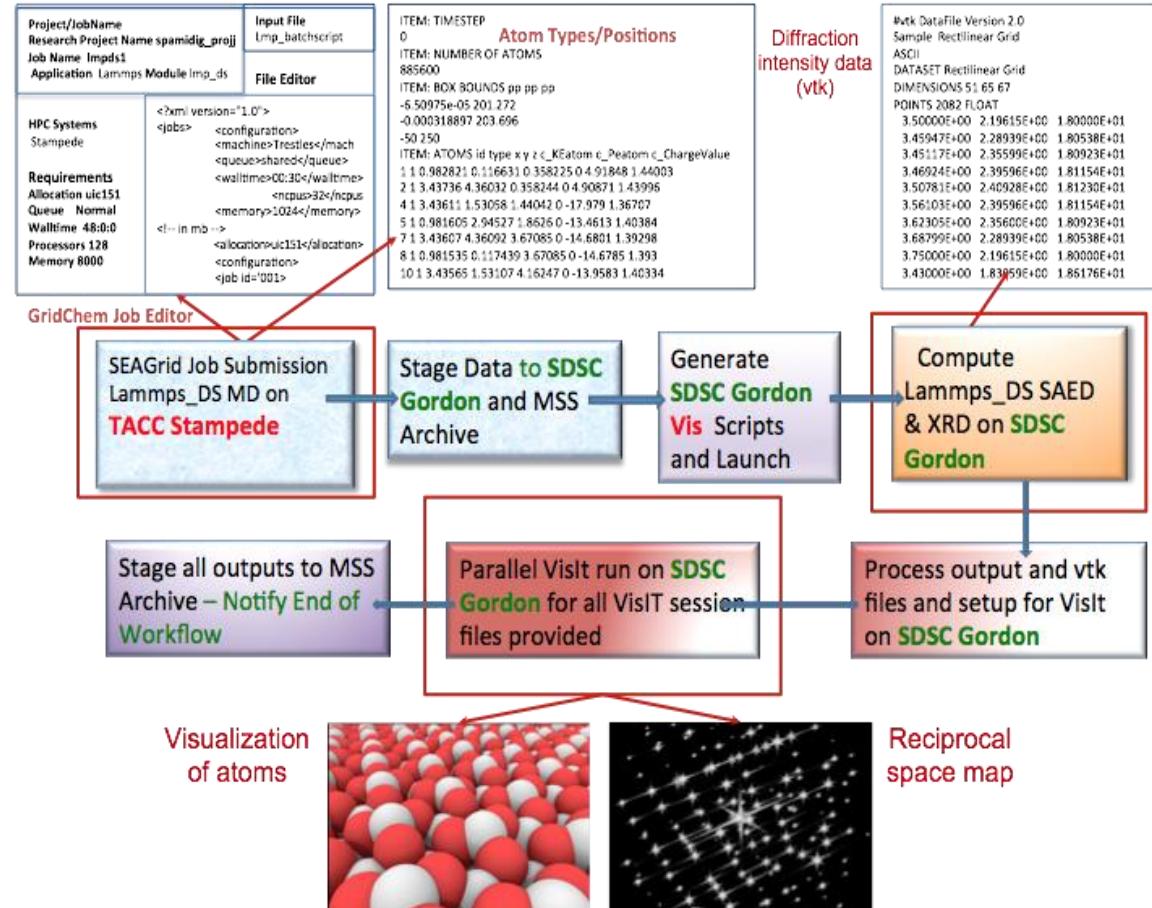
Workflow and Visualization

An XSEDE ECSS Project

- For large memory calculations a workflow is required to use the appropriate XSEDE resource
 - TACC Stampede: Atomistic simulation of alumina
 - SDSC Gordon: Calculation of diffraction intensities + Visualization
- Workflow implemented through SEAGrid Science gateway
 - Supports a private “DS” LAMMPS build
 - Supports single job ID handle for multi-resource job submission
 - Supports the development of a XML script for high throughput job submission
 - Compatible with parallel VisIt executions so that diffraction pattern generation is automated



Workflow to run on TACC
resources only

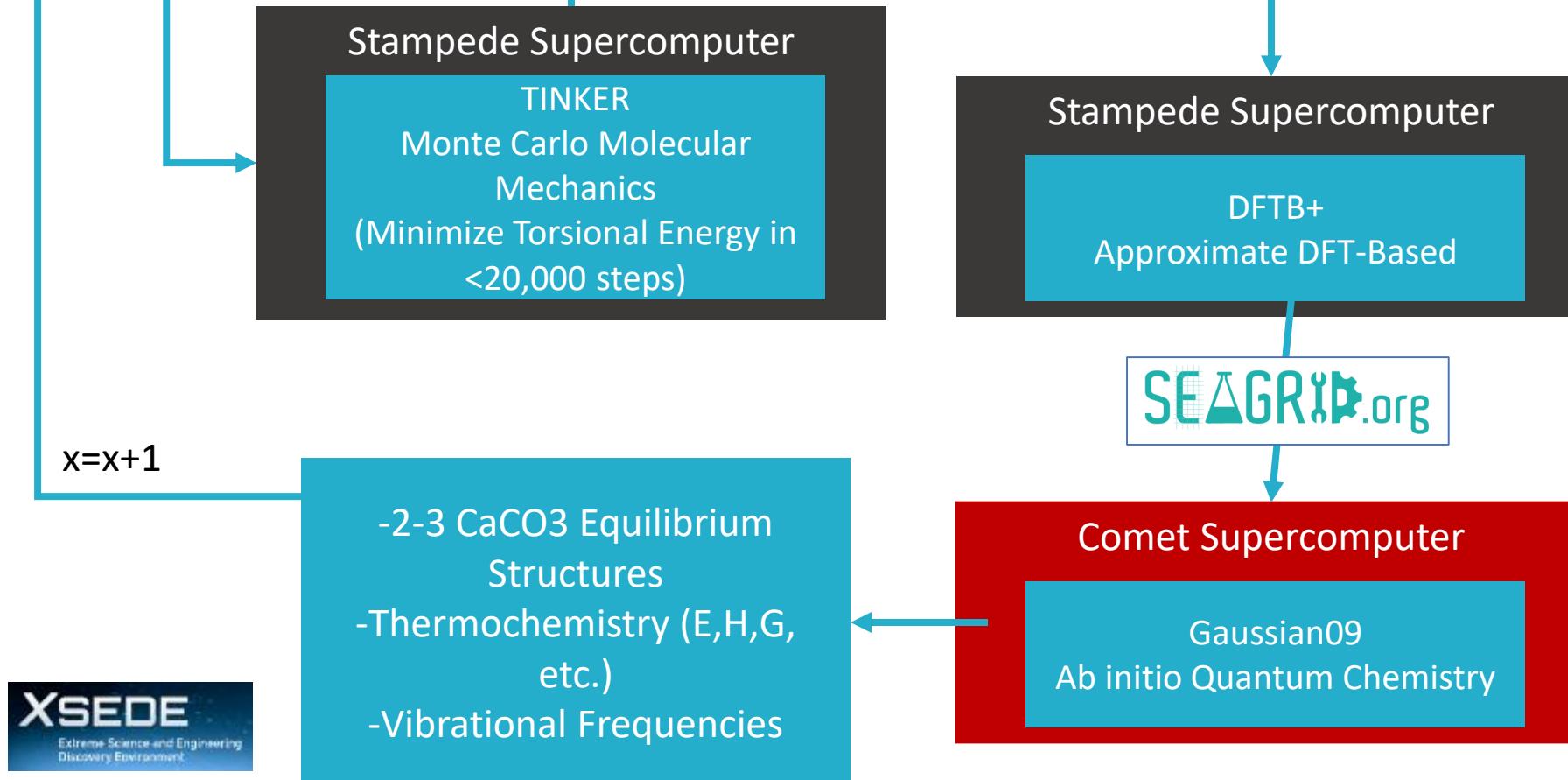


Workflow running on TACC and
SDSC Resources

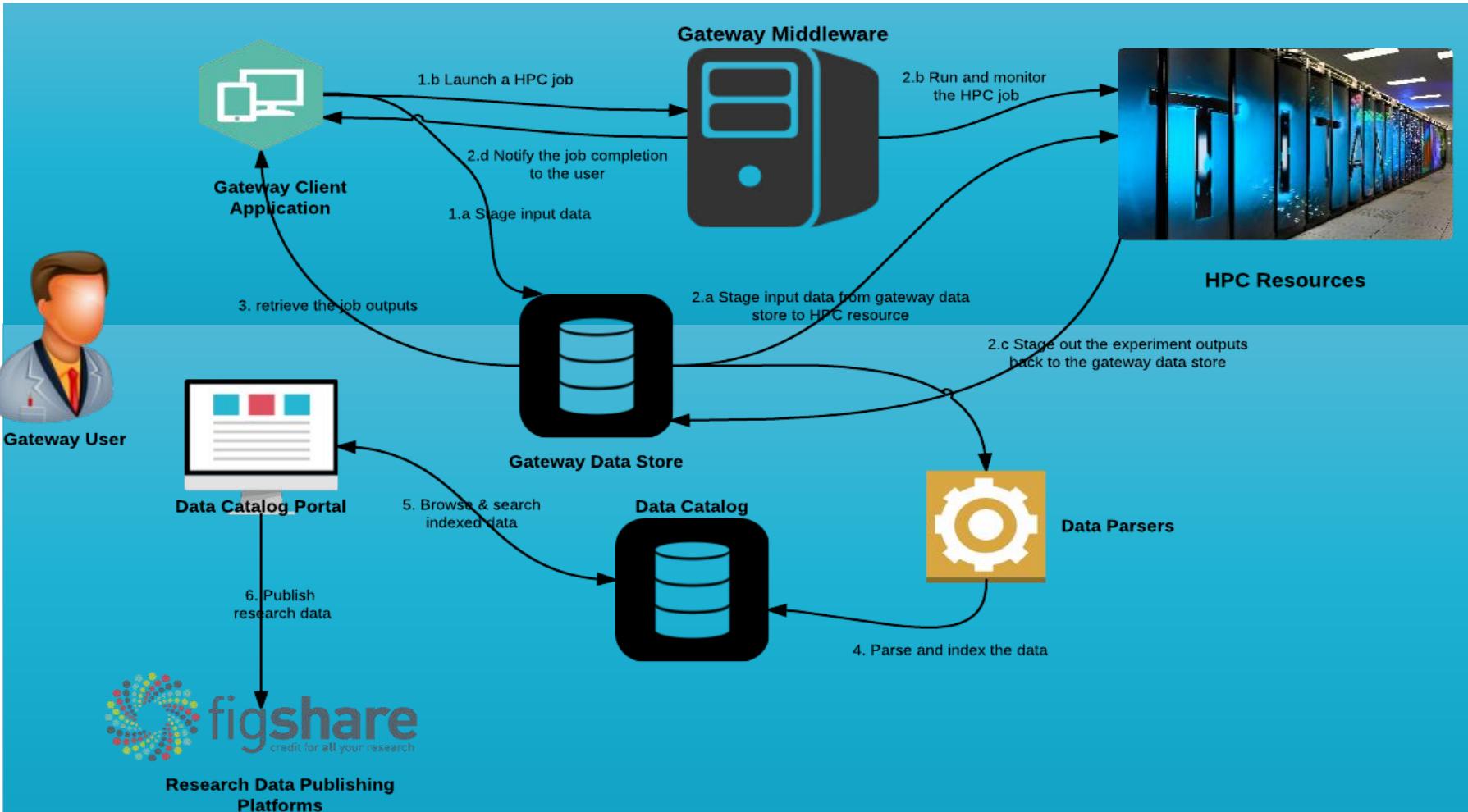
$\text{CaCO}_3 \cdot x\text{H}_2\text{O}$ Initial
guess



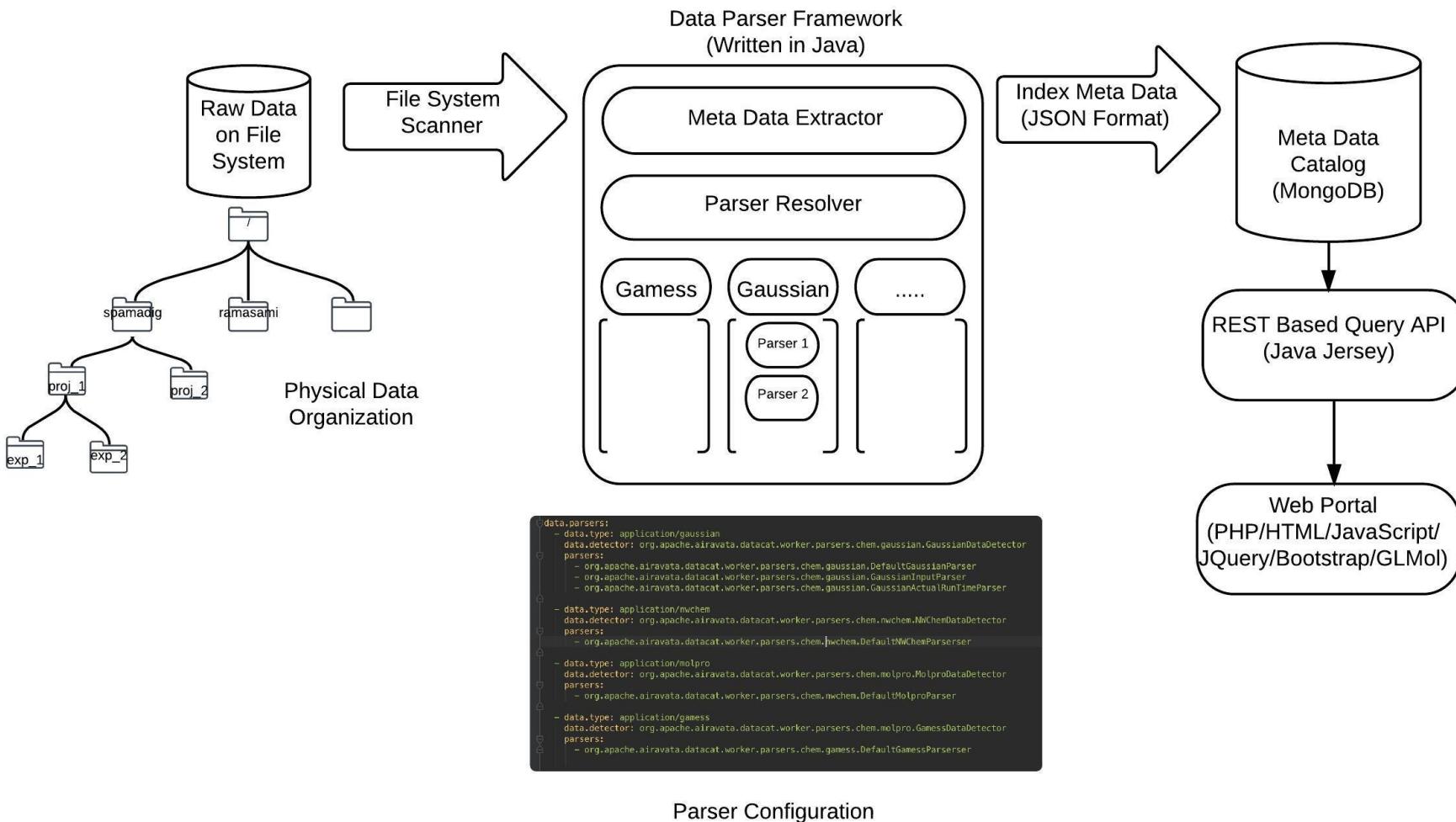
SEAGrid.org enabled workflow



SEAGrid Data Catalog



SEAGrid Data System Architecture



Searching for Cataloged Data

SEAGrid Data Catalog Search Directory Browser

✓ Experiment Name
Project Name
Package
Formula
InChI
SMILES
Calculation Type
Calculation Methods
Basis Sets
Number of Atoms
Actual Job Run Time
Indexed Time

AND OR

Formula contains C17

AND OR

InChI contains H1

+ Add rule + Add group X Delete

+ Add rule + Add group X Delete

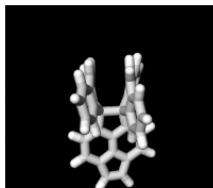
+ Add rule + Add group X Delete

Reset Search

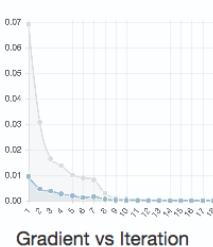
Experiment Name	Project Name	Package	Formula	Indexed Time
bm28.comet.sdsc.xsede.org.738999.150801	ramasami_proj	Gaussian 09, Revision D.01	C17H11NO2S	2016-03-03 23:53:15
h_dv1.trestles.sdsc.teragrid.org.2532797.150303	ramasami_proj	Gaussian 09, Revision D.01	C17H15NO	2016-03-03 23:28:46
ms95.gordon.sdsc.edu.2048727.150513	ramasami_proj	Gaussian 09, Revision D.01	C17H11NO2S	2016-03-03 23:21:40
mb20.trestles.sdsc.teragrid.org.2583124.150422	ramasami_proj	Gaussian 09, Revision D.01	C17H11NO2S	2016-03-03 23:16:17
bm23.comet.sdsc.xsede.org.738994.150801	ramasami_proj	Gaussian 09, Revision D.01	C17H11NO2S	2016-03-03 23:12:08
mre3.comet.sdsc.xsede.org.678254.150709	ramasami_proj	Gaussian 09, Revision D.01	C17H13Br2ClFeN3O2S(2)	2016-03-03 23:09:35

Output Metadata

Organization	
Experiment	LongCC_HC1.1521059306868
Project	sharingtest
Owner	spamidig
Indexed Time	2018-03-14 17:50:40
Molecule	
Formula	C42H26
Number of Atoms	68
Electron Symmetry	1-A
Multiplicity	1
Charge	0
Orbital Symmetry	Occupied A
Identifiers	
InChI	InChI=1S/C42H26/c1-5-13-33-27(9-1)17-18-28-10-2-6-14-34(28)41(33)37-25-23-31-21-22-32-24-26-38(40(37)39(31)32)42(41)35-15-7-3-11-29(35)19-20-30-12-4-8-16-36(30)42/h1-26H
InChI Key	QQYNKBIOZSXWD-UHFFFAOYSA-N
SMILES	c12cccc1C=Cc1c(cccc1)[C@@]12[C@]2(c3cccc3C=Cc3c2cccc3)c2cccc3c4c2c1ccc4C=C3



Final Molecular Structure



Gradient vs Iteration

Canonical SMILES	<chem>c1ccc2c(c1)C=Cc1c([C@@]32c2cccc4c5c2c([C@@]32c3cccccc3C=Cc3c2cccc3)cccc5C=C4)cccc1</chem>
Calculation	
Package	Gaussian 09, Revision E.01
Calculation Type	FOpt; Freq
Methods	RB3LYP; RB3LYP
Basis Set	6-31G(d,p)
Number of Basis Functions	760
Number of Molecular Orbitals in the Calculation	760
Keywords	# RB3LYP/6-31G(d,p) GFInput GFPrint lop(6/7=3) Opt Freq; #NGeom=AllCheck Guess=TCheck SCRF=Check Test GenChk RB3LYP/6-31G(d,p) Freq
Job Status	CalcDone
Calculated Properties	
Energy	-1616.0971637
Dipole	-0.0328227,-0.1402467,0.0158595
HF	-1616.0971637
Homos	[138]

Execution Environment	
Calculated By	GCOMMUNI
Calculated Machine	GINC-R677
Finished Time	14-Mar-2018
Job CPU Run Time	130491.1 seconds
Memory	3000 MB
Number of Shared Processors	28
Input File Configuration	
Link 0 Commands	%nproc=28;%mem=3000MB;%Chk=longccHC_CeN.chk
Route Commands	# RB3LYP/6-31G(d,p) GFInput GFPrint lop(6/7=3) Opt Freq
File Set	
Gaussian Input File	gaussian.in
Gaussian Output File	gaussian.log
Gaussian Checkpoint File	longccHC_CeN.chk
SDF Structure File	structure.sdf
PDB Structure File	structure.pdb
InChI File	inchi.txt
SMILES File	smiles.txt

Make Public

Cyberinfrastructure Integration Research Center

- Science Gateways Communities Institute EDS Activities
- XSEDE ECSS-Gateways Collaborations
- XSEDE ECSS ESRT/NIP Consulting
- NSF Grants – OGCE, SciGaP, XSEDE, JetStream, Cyberwater, DELTA_Topo, Graph-theory-based molecular fragmentation methods
- NASA Grants – GeoGateway
- Private Sector Partnerships

Acknowledgements and Contacts

- NSF Grants ACI-1547611, ACI-1339774
- Apache Software Foundation
- Indiana University
- GSOC Programs
- Science Gateways Research Center <https://circ.iu.edu/>
 - Center email: sgrc-iu-group@iu.edu
 - Marlon Pierce: marpierc@iu.edu, Director
 - Apache Airavata Open Source Science Gateway Software <http://airavata.apache.org/>
 - Sudhakar Pamidighantam pamidigs@iu.edu SEAGrid
 - Suresh Marru: smarru@iu.edu Apache Airavata

SEAGrid Desktop Client CSD Database

Import Structure

Atom Database My Files Function-Group Ion Molecule

Indiana Database CSD

Search Text : ibuprofen Limit : 5 search

Formula Chemical Name

0.68(C13 H18 O2),0.32(C8 H9 N1 O2)	ibuprofen
0.56(C13 H18 O2),0.44(C8 H9 N1 O2)	ibuprofen
0.62(C13 H18 O2),0.38(C8 H9 N1 O2)	ibuprofen
0.75(C13 H18 O2),0.25(C8 H9 N1 O2)	Ibuprofen
(C64 H44 Fe8 O44)n,n(C13 H18 O2)	

Nanocad Editor

Summary of Nanocad Commands:

- Rotate: drag gray space
- Translate: Shift-drag gray space
- Zoom: Ctrl-drag gray space
- Move Atom: drag atom
- Add Atom: Shift-click gray space
- Delete Atom: Shift-click atom
- Add Bond: Shift-drag atom to atom
- Delete Bond: Ctrl-drag atom to atom
- Select Atom: Alt-click atom
- Select Group: Ctrl-Alt-click atom
- Add double bond: Shift-drag between bonded atoms

About

View with RasMol

View XYZ

View Native

Gaussian Input

GAMESS Input

NWChem Input

PSI4 Input

Molcas Input

Group Geometry Forces Help Structure Clear Undo Add H Get Potential --Minimize-- --Force Field-- --Input/Output Menu-- Structure may have many sub-structures from unit cell

SEAGrid Desktop Client

File Help

Create Project Create Experiment Storage Nanocad G03 Log Out

--Input/Output Menu--

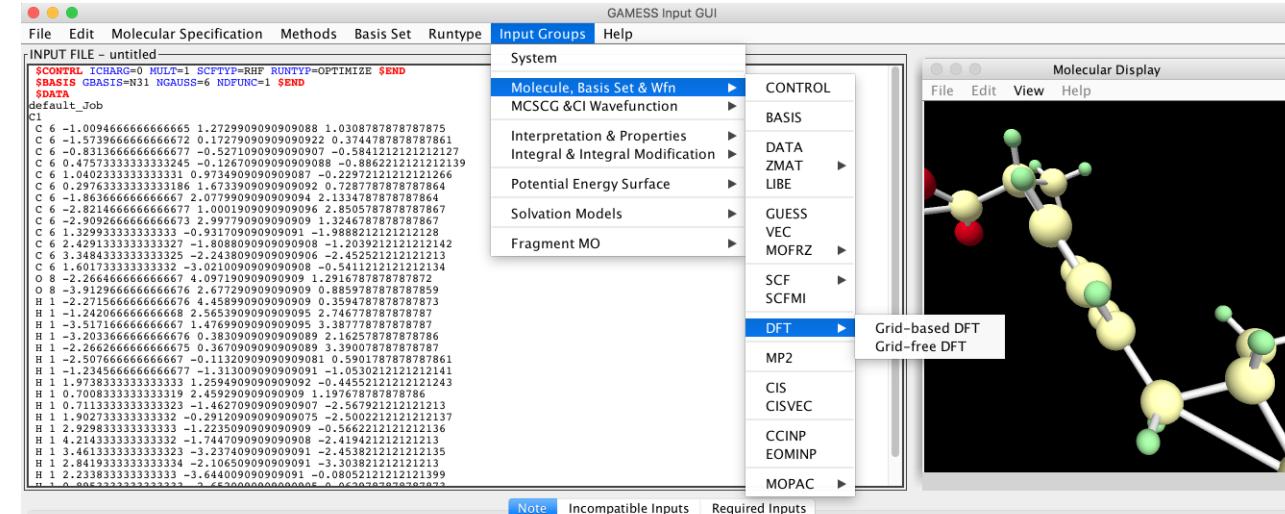
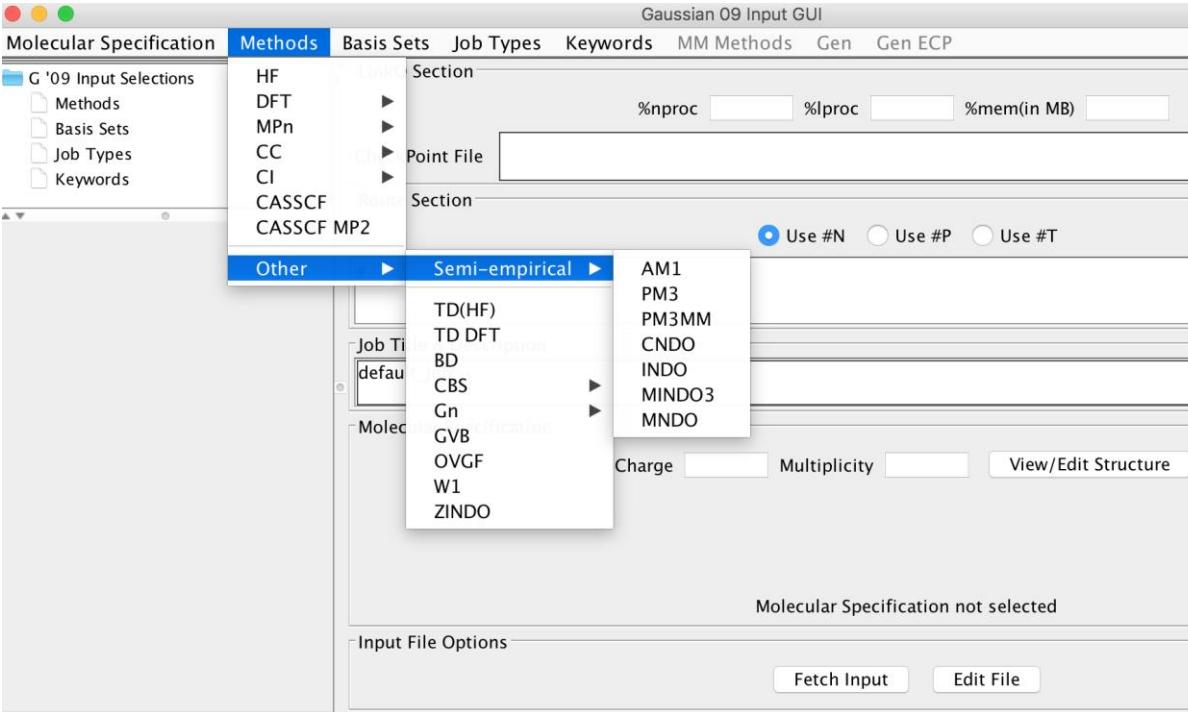
Recent Experiments Projects

Recent Experiments Filter Experiments:

Experiment	Application	Host	Experiment	Created Time
Clone of c...	Gaussian	comet.sds...	CREATED	2016-04-...
cysteine.inp	Gaussian	comet.sds...	LAUNCHED	2016-04-...
test	Gaussian	comet.sds...	COMPLETED	2016-04-...
test	Gaussian	comet.sds...	FAILED	2016-04-...
hydrogen	Gaussian	comet.sds...	CREATED	2016-03-...
test	Gaussian	comet.sds...	COMPLETED	2016-03-...

Launch Selected Delete Selected

Graphical Interfaces to Application Software



SEAGrid Desktop Client

Storage Browser

Filename	Filesize	Last Modified
bioinfo		2016-03-26T09:37:20
build.xml	50277 bytes	2015-03-03T15:47:38
Cd16Se16_Cu2_relax...	1523 bytes	2015-09-02T16:01:09
conf.d.tar	81920 bytes	2015-04-10T09:09:29
Creative Cloud Files		2015-06-24T22:42:39
Desktop		2016-04-21T20:45:12
Documents		2016-03-01T20:54:12
Downloads		2016-04-20T07:46:16
Dropbox		2016-04-15T08:29:51
etst.sh	195 bytes	2015-02-12T15:20:06
ExperimentData		2016-04-08T13:58:23
ext_cyl_20150904_ti...	683785 bytes	2015-12-09T14:21:56
git		2016-01-08T10:17:20
gridchem		2015-02-23T17:26:23
i-134-filledNS.pdf	113838 bytes	2015-06-08T08:48:10
i-134.pdf	78744 bytes	2015-05-28T12:28:53
iccs2016_2.docx	605054 bytes	2016-01-14T00:17:50
j3d		2015-04-24T10:40:28
jobslist.xml	1832598 bytes	2016-04-18T21:42:29
jogl		2015-04-24T11:09:20

Filename	Filesize	Last Modified
..		2016-03-09T11:30...
test1457541036		2016-03-18T13:06...
hydrogen.14583...		2016-04-06T16:30...
test.1459978243...		2016-04-07T15:37...
test.1460061432...		2016-04-07T16:18...
cysteine_inp.146...		