Private Cloud Environment within OpenStack

Jingchao Yang (presenter)

Keith Clarke, Wendy Guan, Mengchao Xu, Fei Hu, Chaowei Yang, and others

NSF Spatiotemporal Innovation Center

George Mason University, UC-Santa Barbara, and Harvard University

Spatiotemporal Thinking, Computing and Applications

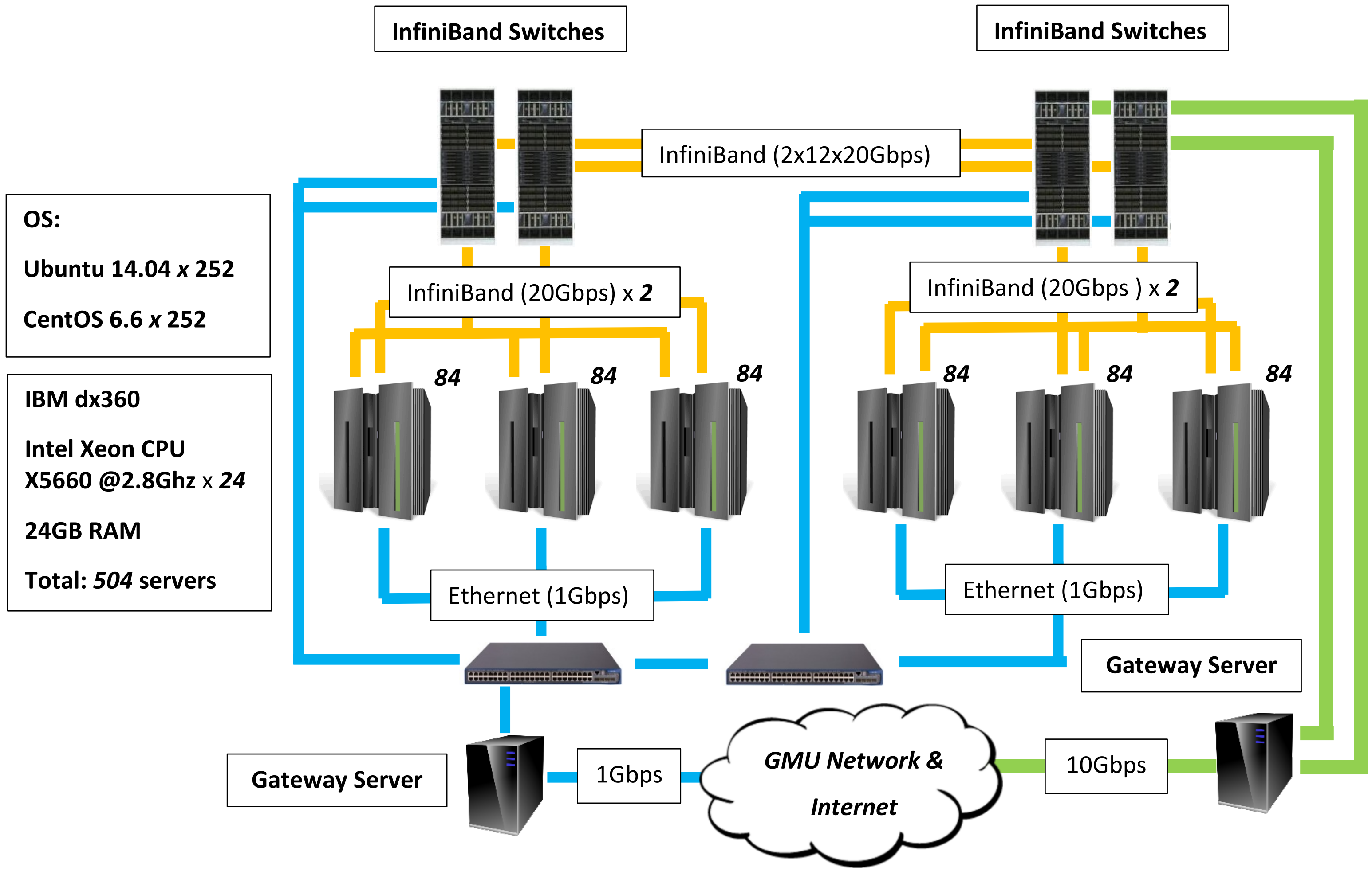
Outline

- Project Overview
- Cloud Infrastructure and Network Configuration
- Service Communities
- Accomplishments and Sample Projects
- Impact
- Plans for the Next Reporting Period

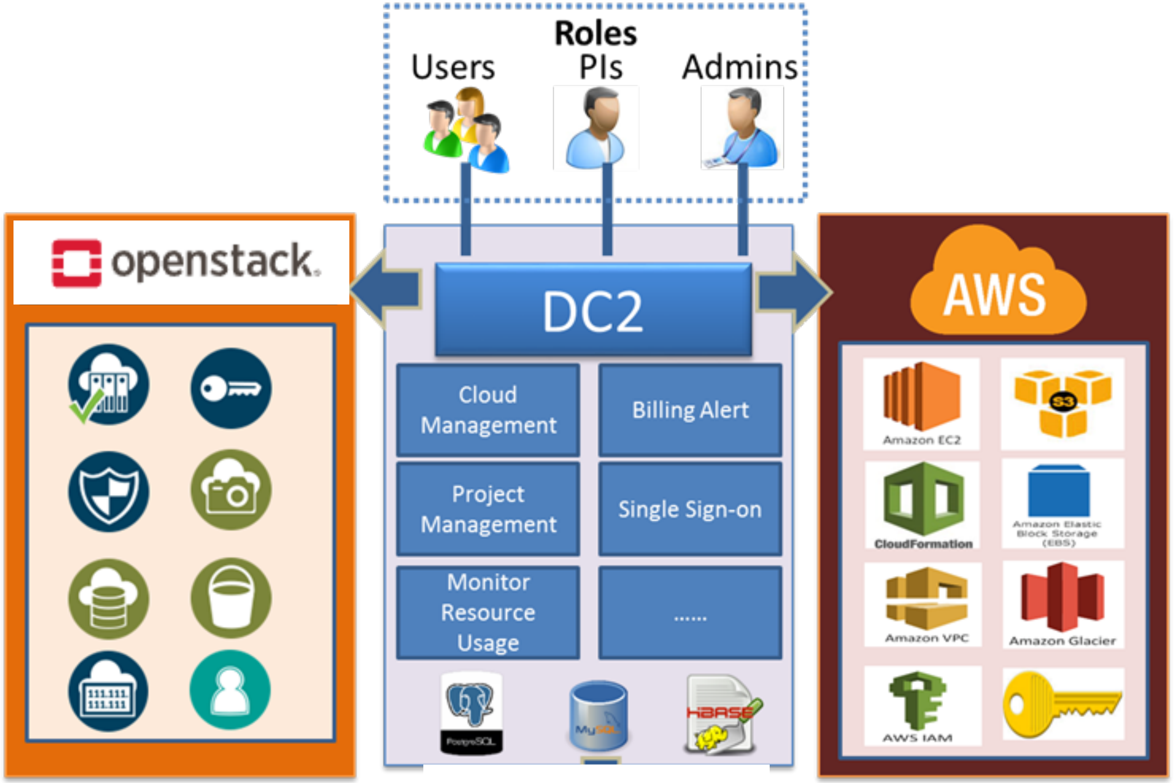
Project Overview

- Big data analytics requires significant computing support for model simulations, e.g. dust, flooding, fire, storm, require High Performance Computing (HPC)
- Most center projects need computing support
- The community suggested the building of a cross site center wide computing infrastructure to remove duplication and improve computing capability to meet center advancement needs.
- Develop and maintain a spatiotemporal computing infrastructure.
- Provide spatiotemporal computing coordination to all projects that need computing by
  - Acquiring a high performance computing facility
  - Developing data mining components to detect security, usage and other patterns of interest
  - Maintaining a highly capable research staff support for computing optimization

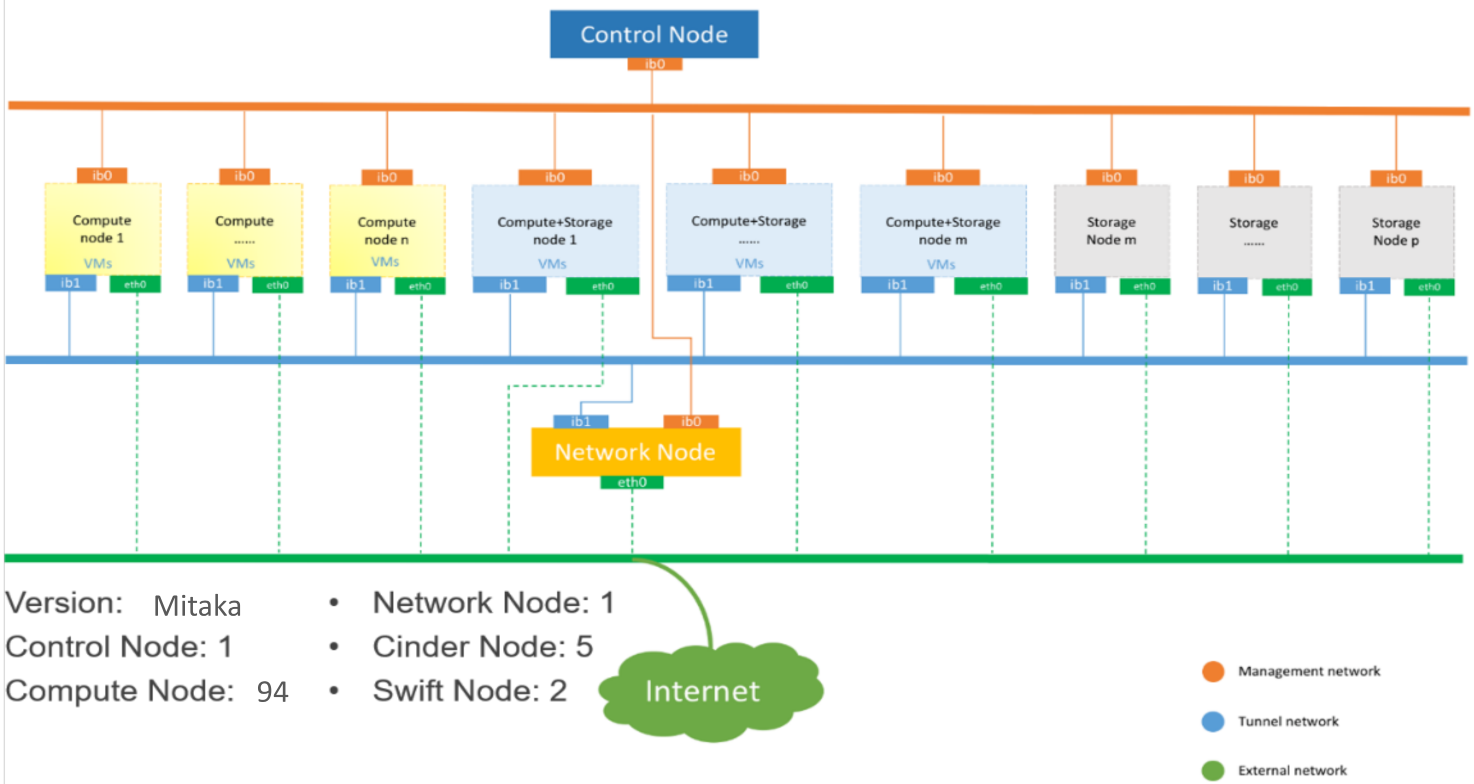
Infrastructure setup and network configuration



Cloud software management architecture



Cloud configuration example - Openstack



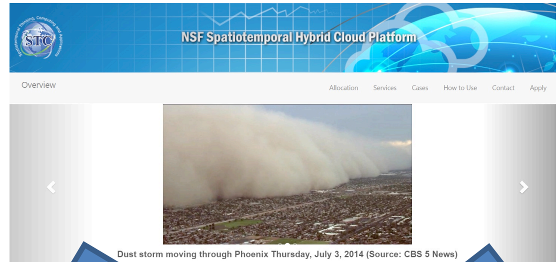
- Version: Mitaka
- Control Node: 1
- Compute Node: 94
- Network Node: 1
- Cinder Node: 5
- Swift Node: 2

Planned Software/data infrastructure

Data sources → Data preprocessing → Data storage & management → Data access → Operations

Roles	Users, Admin
Application	data mining, data analysis, …
Software	GIS/RS software, GIS servers, other common software, …
Data Management	MySQL, PostgreSQL, HIVE, HBASE, mongoDB, rasdaman raster data management
Data Storage	Data Storage/ High Performance Network File System (e.g., Lustre)
Hardware	

Service Communities



EARTHCUBE

NSF EarthCube Integration and Testing Environment (ECITE):

Dashboard : Project Management > Project

Create Project | Operate | Refresh | Delete Project Tag

	Action	Project title	Description	Date Created	Cloud Account	Allocated Funding ($)
		BCube	BCube in ECITE	Apr 22, 2016 2:19:06 PM	aws-ecite-operate	2000.00
		emily-project	For testing	Apr 19, 2016 5:52:13 PM	aws-ecite-operate	500.00
		STCtestbed	STCtestbed	Apr 11, 2016 2:27:28 PM	aws-ecite-operate	2000.00
		EarthCube-DAsHER	NSF EarthCube DAsHER CD	Apr 7, 2016 4:27:49 PM	aws-ecite-operate	500.00
		CINERGI	Community Inventory of EarthCube Resources for Geosciences Interoperability	Feb 29, 2016 10:26:16 AM	aws-ecite-operate	2000.00
		CHORDS	Cloud-Hosted Real-time Data Services for the Geosciences	Feb 22, 2016 11:01:28 AM	aws-ecite-operate	2000.00

ESIP
The Federation of Earth Science Information Partners
Fostering connections to make data matter

AAG
ASSOCIATION of AMERICAN GEOGRAPHERS

CHORDS

Accomplisments

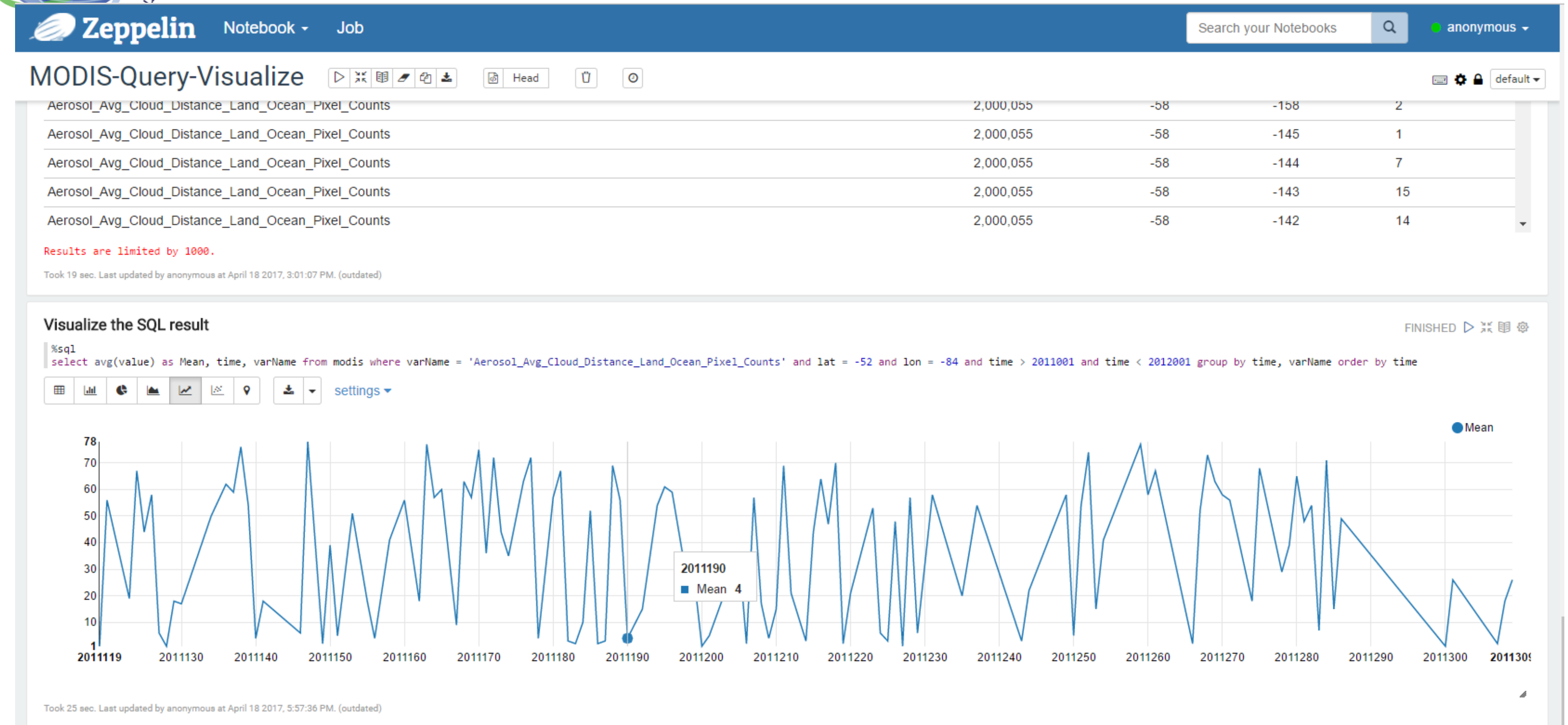
- Hardware: an NCCS (NASA Center for Climate Simulation), donation of 504-node computer cluster with two 20Gbps internal connections and a 1Gbps management connection
- Websites:
  - Information portal: cloud.gmu.edu
- Documents:
  - Maintenance documents
  - User manuals
  - Operation specifications

Impact

- Consolidate and reduce cost spent on computing resources at the center
- Provide a computing testbed for center projects and potential IAB directed research
- Accumulate advanced computing operation experiences
- Serve NSF EarthCube program through ECITE
- Serve ESIP through its cloud computing cluster
- Serve the community through AAG, Geo4All, etc.

Sample projects on the cloud

GMU-15-01 ClimateSpark: An In-memory Distributed Computing Framework for Big Climate Data Analytics

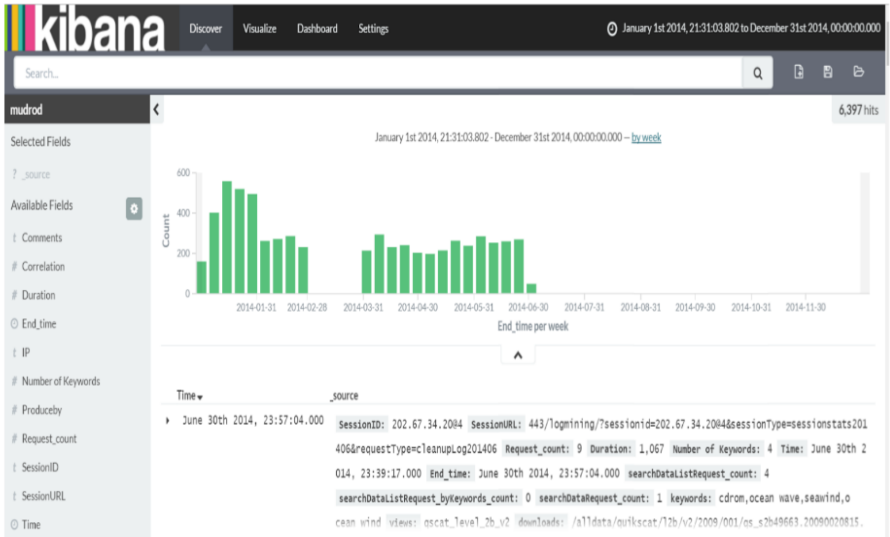


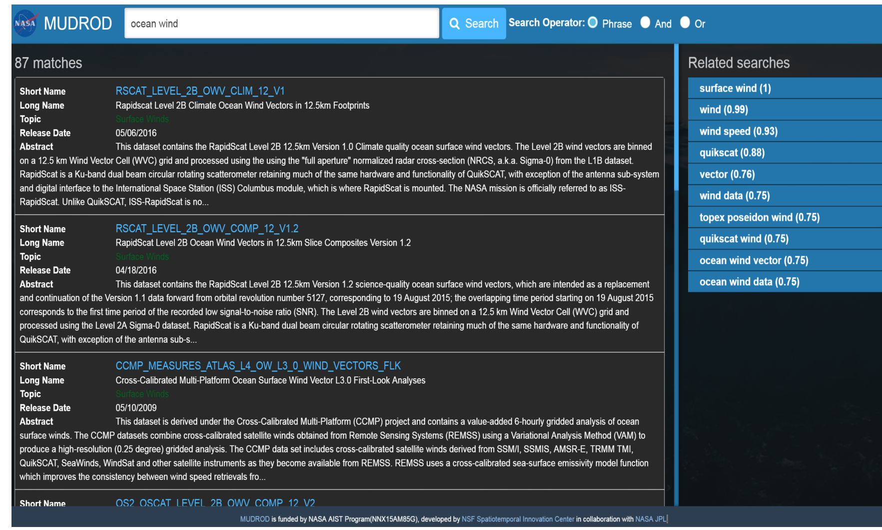
A web portal to ease the remote interaction between climatologists, climate data, analytic operations, and computing resources (e.g. using SQL query and Scala/Python notebook)

AWS: http://10.192.20.172:8080/#/notebook/2C172AC58

Supported by NCCS (NASA Center for Climate Simulations/Daniel Duffy)

MUDROD, NASA AIST 2014, Managed by Mike Little





- Analyze web logs to discover user knowledge
- Data management
- Search engine
- **OpenStack:**
- http://199.26.254.151:8080/mudrod-service/

GMU-15-09 Planetary Defense Funded by NASA PDCO

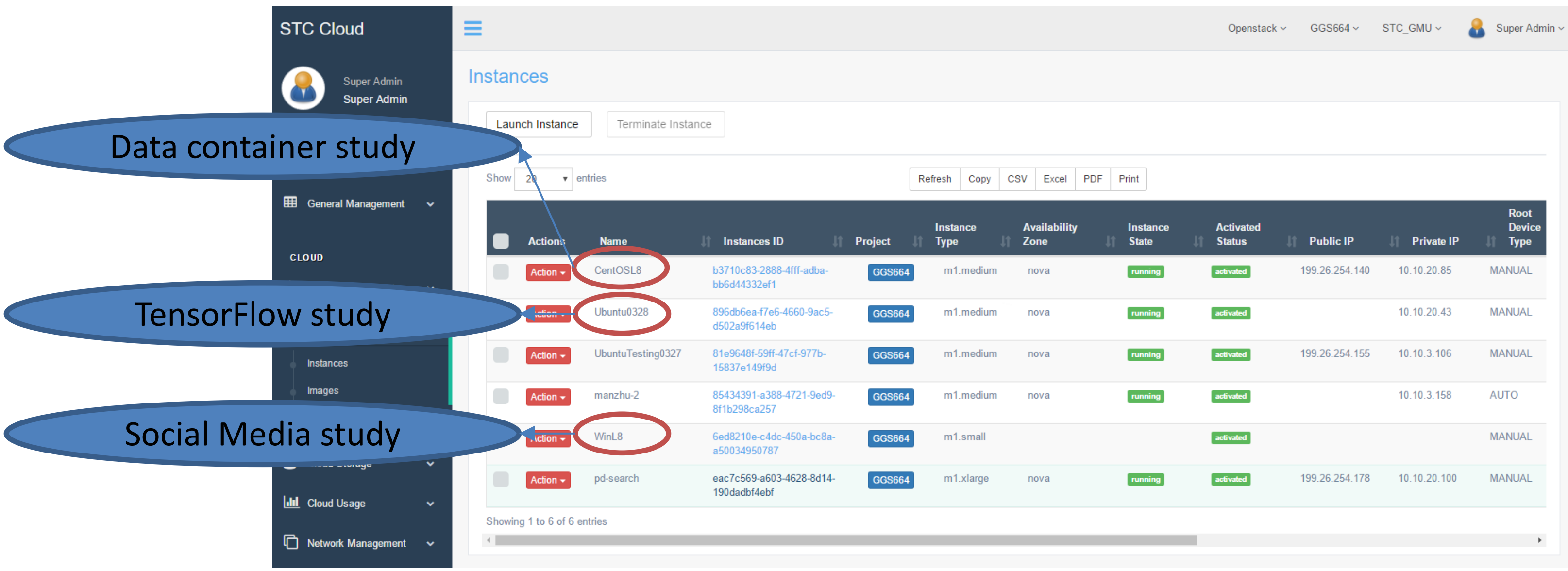
- Domain Crawler
- Search Ranking
- Thanks to Myra Bambacus and Bernie Serry

OpenStack:

  - **http://pd.cloud.gmu.edu/**

Platform for Education

- GGS 664 Spatiotemporal Data Structure
  - Easy the teaching processes by using platform to customize instances according to needs.

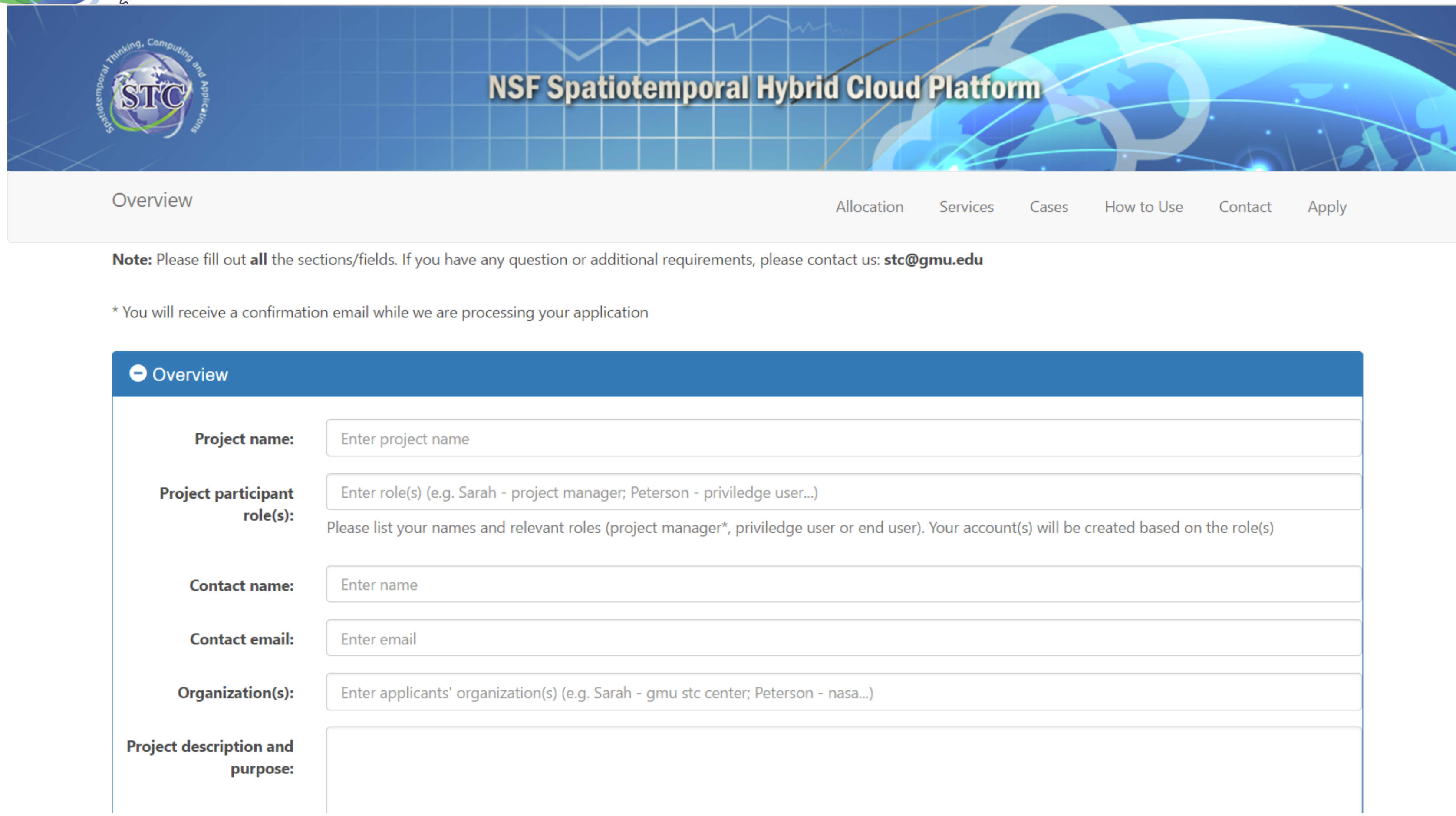


  - Different instances assigned to students, providing different operation system environments and preinstalled with different software

Functionalities being added

- One-button Hadoop
- Auto Scaling
- Load Balancing
- Develop big data management platform

Apply for Platform Resources



NSF Spatiotemporal Hybrid Cloud Platform

Overview

Allocation | Services | Cases | How to Use | Contact | Apply

Note: Please fill out **all** the sections/fields. If you have any question or additional requirements, please contact us: **stc@gmu.edu**

* You will receive a confirmation email while we are processing your application

Overview

Project name: Enter project name

Project participant role(s): Enter role(s) (e.g. Sarah - project manager; Peterson - priviledge user...)

Please list your names and relevant roles (project manager*, priviledge user or end user). Your account(s) will be created based on the role(s)

Contact name: Enter name

Contact email: Enter email

Organization(s): Enter applicants' organization(s) (e.g. Sarah - gmu stc center; Peterson - nasa...)

Project description and purpose:

- http://sites.cloud.gmu.edu/sthcp/form.php

Contacts

- Jingchao Yang, jyang43@gmu.edu
- Phil Yang, cyang3@gmu.edu