Last Edited: 10/4/21

2021 Cloud Infrastructure Working Group Charter

Version 1.2

Distribution: Public

https://ci-compass.org/about/working-groups/

Table of Contents

1 Working Group Purpose, Goals, and Objectives **1**

2 Timeline: Events and Milestones **2**

3 Working Group Members **3**

CI Compass 3

Major Facilities 4

Community 4

4 Risks and Mitigations **5**

5 Metrics of Success **5**

6 Products **5**

7 Version Log **5**

1 Working Group Purpose, Goals, and Objectives

In 2020, the CI CoE Pilot (https://cicoe-pilot.org) conducted a number of interviews with Major Facilities Managers and Staff regarding their Data Lifecycle. From these interviews, one major theme that trended throughout conversations was that of Cloud Infrastructure. Topics of cost, vendor lock-in, and storage were among the most mentioned as being challenging for Major Facilities to navigate. True to the CI CoE Pilot mission statement, an effort to aid Major Facilities in tackling these hot topics was started.

"CI CoE provides expertise and active support to cyberinfrastructure practitioners at NSF

Major Facilities in order to accelerate the data lifecycle and ensure the integrity and effectiveness of the cyberinfrastructure upon which research and discovery depends. *- CI CoE Pilot Mission"*

In December 2020, the CI CoE Pilot launched a new Cloud Infrastructure Working Group. This working group set out to (1) understand the current practices for Cloud Infrastructure used by MFs, (2) research alternative solutions and keep up to date with emerging cloud technologies, and (3) develop a general set of best practices that can inform the MFs and how those can be adapted for specific facilities.

This working group will continue its effort under CI Compass (https://ci-compass.org/) as phase two of the CI CoE Pilot.

The working group will hold a virtual workshop to dive deeper into the challenges that Major Facilities face. The workshop content will be informed by MF representatives and the workshop report will be published as a freely-available report by the end of 2021. We will use community events and available social media channels to disseminate this report.

The CI Compass is responsible for project management of the WG. The working group will generally meet virtually once every two weeks.

2 Timeline: Events and Milestones

Timeframe: July 15, 2021 - June 30, 2022 (to be extended as needed). This working group started under CI CoE Pilot and is continuing under CI Compass.

Q1	• Organize bi-monthly meetings and solicit topics.
Q2	• Document considerations for using clouds for MF DLC. • 2021
Q3	• Organize a cloud-focused session at CI for MFs workshop. • Present findings at the CI for MFs workshop. • Solicit additional group members. • 2022
Q4	• Summarize findings from CI for MFs cloud session • Conduct workshop planning for workshop in Y2. • 2023

3 Working Group Members

CI Compass

Karan Vahi. (Computer Scientist in the Science Automation Technologies group, at the Advanced Systems Division, ISI. Lead developer of Pegasus WMS) will lead the Working Group. Email: vahi@isi.edu

Chuck Vardeman (Computational Scientist in Notre Dame's Center for Research Computing (CRC) and a Research Assistant Professor in the Department of Computer Science and Engineering) is participating in the working group. cvardema@nd.edu

Rob Ricci (Research Associate Professor in the Flux Research Group at the School of Computing, University of Utah) is participating in the working group. ricci@cs.utah.edu

Erik Scott (Senior Research Software Developer at RENCI and expert in Data Science and External Engagement) is participating in the working group. escott@renci.org

Angela Murillo (Assistant Professor at the School of Informatics and Computing at Indiana University-Purdue University Indianapolis and Program Director of the Applied Data and Information Science Program) is participating in the working group. apmurill@iu.edu

Mats Rynge (Computer scientist in the Science Automation Technologies group at the USC Information Sciences Institute) is participating in the working group. rynge@isi.edu

Loic Pottier (Computer Scientist in the Science Automation Technologies group at the USC Information Sciences Institute) is participating in the group work. lpottier@isi.edu

Rajiv Mayani (Research Programmer in the Science Automation Technologies group at the USC

Information Sciences Institute) is participating in the group work. mayani@isi.edu

Major Facilities

Benedikt Riedel (Computing Manager at IceCube) is participating in the group work. briedel@icecube.wisc.edu

Frank Wuerthwein (Executive Director of the Open Science Grid and Distributed High-Throughput Computing Lead - SDSC; LHC-CMS) is participating in the group work. fkw@ucsd.edu

Julio Alvarado Negron (Program Manager, Big Data; Arecibo (AO)) is participating in the group work. julio.alvarado@ucf.edu

Brian Dobbins (NCAR) is participating in the group work. bdobbins@ucar.edu

Jeff De La Beaujardiere aka 'Jeff DLB' (NCAR) - Director of NCAR's Information Systems Division, cloud focus is on data/analysis in the cloud is participating in the group work. jeffdlb@ucar.edu

Jeff Glatstein (Senior Cyberinfrastructure and Data Manager; OOI) is participating in the group work. jglatstein@whoi.edu

Chad Trabant (Deputy Director, Products & Chief Architect for IRIS; SAGE/GAGE) Project manager of Common Cloud Platform focussed on a joint-facility migration to the cloud is participating in the group work. chad@iris.washington.edu

Community

Jeremy Fischer (Manager, Jetstream Cloud) is participating in the group work. jeremy@iu.edu

David Hahn (JCSDA/RAL) - lots of varied cloud work, from HPC to serverless apps for science is participating in the group work. hahnd@ucar.edu

Nirav Merchant is Director of Data Science Institute at UArizona and is CoPI of CyVerse. nirav@arizona.edu

Tyson Swetnam is a research assistant professor of geoinformatics at UArizona and is CoPI of CyVerse. tswetnam@arizona.edu

Bruce Berriman (IPAC) is a Senior Research Scientist at IPAC, Caltech. gbb@ipac.caltech.edu

Alumni

Wendy Whitcup (Project Manager for the CI Compass and Science Automation Technologies group at the USC Information Sciences Institute) will facilitate communication and logistics for the working group. Email: wwhitcup@isi.edu

Mark Miesch (JCSDA) is the software infrastructure lead for the JEDI project, a next-generation data assimilation system for Earth system prediction. He manages JCSDA cloud computing resources and activities, which include HPC cluster computing with the JEDI & FV3 models, data hosting, analysis, and visualization, and automated CI/CD workflows is participating in the group work. miesch@ucar.edu

4 Risks and Mitigations

- Risk: The working group may not achieve consensus on its scope. What falls under the Cloud Infrastructure umbrella? How can we pare down topics that are most meaningful to MFs?
    - Mitigations: Recruit specific representatives from MFs to join in the discussions and planning of our workshop. Share the workshop report widely and see what topics others choose to engage with us on.
- Risk: The working group may not impact the broader community.
    - Mitigations: Communicate and actively participate within NSF ecosystem to relay the workshop report information

5 Metrics of Success

- Number virtual workshop attendees
- Number of community members providing feedback
- Number of report downloads

6 Products

- Document - document describing considerations when using clouds to support the MF DLC.
- Workshop Session Report - The group discussions will inform a session at the MF CI workshop in Q3 and a related report.
- Workshop - Based on the wg discussions and session held at the MF CI workshop, the group will also plan for a topical workshop in Year 2.

7 Version Log

- Version 1.0 in Development
- Version 1.1 updated timeline and products
- Version 1.2 updated from CI CoE Pilot to CI Compass