CDI Monthly Meeting 20180711

The Community for Data Integration (CDI) meetings are held the 2nd Wednesday of each month from 11:00 a.m. to 12:30 p.m. Eastern Time.

Meeting Recording

Meeting recordings are available to CDI Members approximately 24 hours after the completion of the meeting. Please log in to view the recording. If you would like to become a member of CDI, email cdi@usgs.gov.

During the call, you can ask and up-vote questions at slido.com, event code #T305.

Agenda (in Eastern time)

11:00a 2019 CDI Workshop Poll results

11:05a Welcome - Cheryl Morris - Director, Core Science Analytics, Synthesis, and Library

11:10a Collaboration Area Announcements [CDI_20180711_OpeningSlides.pdf]

11:20a **STEP-UP to Support Students & Science**, Chris Hammond, USGS STEP-UP Program Manager

11:30a **USGS Cloud Hosting Solutions Update**, Jennifer Erxleben, Chief CHS and Harry House, CHS Program Director

12:30p Adjourn

Abstracts

STEP-UP to Support Students & Science, Chris Hammond, USGS STEP-UP Program Manager

STEP-UP (Secondary Transition to Employment Program - USGS Partnership) provides employment training to young adults (ages 18-22) with cognitive and other disabilities. This overview will explain how the program works and describes several success stories. The CDI has recently heard from a number of groups that are looking for solutions to migrating legacy data or websites, we introduce the STEP-UP program as a possible solution to investigate further. Next month, we will hear more about a STEP-UP success story from CDI member Sue Kemp.

Read more:

- https://www.kqed.org/science/1922125/students-with-autism-excel-in-working-with-data-helping-scientists
- https://www.usgs.gov/news/step-science-engaging-young-adults-disabilities

Chris Hammond is the manager of the STEP-UP Program, which provides students with cognitive and other disabilities (ages 18-22) transitioning from school to work with volunteer opportunities that support USGS scientists. He has been on an extended detail to USGS from the U.S. Office of Personnel Management since October 2017. In addition to managing STEP-UP, Chris also coordinated this year's Take the Kids to Work Day events at the National Center in Reston.

USGS Cloud Hosting Solutions Update, Jennifer Erxleben, Chief CHS and Harry House, CHS Program Director

Cloud Hosting Solutions (CHS) is the required, supported, secure Cloud offering for USGS Science Centers and mission programs. CHS provides on-demand computing resources and services in the Cloud through Amazon Web Services (AWS), and has been available for USGS use 2015. This overview will provide an introduction to CHS, summary of current and planned CHS services, and discussion of cloud costs, lessons learned, and future directions. Join us for this presentation and Q&A on the latest developments in CHS.

USGS users on the USGS network can see more information at: USGS Cloud Hosting Solutions.

Presentations

Presentation: Slides are available to CDI Members. Please log in to download the slides. If you would like to become a member of CDI, email cdi@usgs.gov.

Highlights

- CDI FY18 Funded Projects to present at ESIP Summer Meeting on July 17 : https://my.usgs.gov/confluence/x/bg3SIw
- CDI FY17 Workshop Report on Enabling Integrated Science available to inspire you for the next workshop: https://doi.org/10.3133/ofr20181081
- More information about CDI Summer Reading “Get your Science Used” on July 25: https://my.usgs.gov/confluence/x/NQzSIw
- Sign up for a reminder for July 25’s reading discussion: https://goo.gl/2MjdpE
- The USGS STEP-UP program may be a solution for centers that are looking for detail-oriented student workers to contribute projects, especially those dealing with legacy data and website migration and archiving.
- The students can support work across USGS; some have worked on projects remotely.
- STEP-UP program site (on USGS network) https://sites.google.com/a/usgs.gov/step_program/
- Cloud Hosting Solutions is providing a variety of managed and custom services to meet the USGS needs.
- Contact: cloudservices@usgs.gov
- Services list: https://internal.usgs.gov/oei/cloud-hosting-solutions/chs-services/ (must be on USGS network)

Q&A

1. can you define "managed services?"
    a. The CHS team invests staff time to provide support for the service. Examples are offsite backup, HPC, GIS, data visualization: we grab a commercial product but need to configure it for our environment, provide some support and perhaps advocacy.
2. Is the sandbox environment shared by all users or does each sandbox user get their own environment?
    a. It is a custom AWS account labeled as sandbox, it’s a single environment where everyone is in a playground with everyone else. We haven’t had any complaints about that yet.
3. Is there a listserv for the CHS monthly user meeting?
    a. Not currently but it is something the communications team is investigating. If you are interested in attending the customer user group meeting, email cloudservices@usgs.gov and we’ll add you to the list.
4. You mentioned Tableau. How does that fit in CHS?
    a. We’re looking at standing up Tableau Server in our secure managed environment. We would handle the infrastructure in the cloud, and we would provision sites and access based on teams or projects.
5. if one uses software that is already part of ALCES, such as open-source GIS, would that be considered a "non-managed" service or resource?
    a. ALCES flight is a managed service, any software application available on alces flight is a component of the managed service.
6. How are you supporting DevOps approaches on CHS, particularly healthy continuous deployment from development teams using repos on code.chs.usgs.gov?
    a. Development groups with production accounts can do it the way they want. Within CHS we adopted a methodology to make iterative changes. More @1:02:10 on the recording. (and see question 13).
7. As USGS employees, are we required to use CHS to access AWS? Or can we can use AWS directly?
    a. What is meant when we say CHS is required is that if you are going to use a cloud solution in USGS it has to go through CHS first, unless it is something we cannot accommodate. We can help guide you to other options, but CHS is first line for anything cloud. You cannot go directly to Amazon and just get your own service.
8. Is AWS Lambda supported/available yet? Thought I heard Harry say "if"...
    a. Lamda within CHS is turned on for our region. These are available to Dev/Pro accounts, and it can be used to explore what’s in the sandbox.
9. What are the options for establishing geographically redundant solutions within CHS?
    a. We’re exploring those right now. CHS is authorized to use AWS region in Oregon. We’re looking at solutions to fail-over or have redundancy in other regions. (But there is redundancy within the Oregon region, just not across regions.)
10. Follow-on from question about direct AWS access: noticed ESRI cloud options just got listed in FedRAMP. Do we go through CHS to explore that?
    a. It is on our list for current review. We’re looking into it and it would be something similar to the Tableau solution, potentially a managed solution. We’d like to know who is interested in that, please contact the CHS communications team cloudservices@usgs.gov.
11. Comment: can't find CHS on the internal site A-Z index. Might be good to get link added there!
12. Comment from Kevin Wood: I’m with NGTOC and The National Map. A few comments on our use with CHS: we have AD accounts up there so it is easier to work with AD. We are looking at working with other federal agencies to share resources (mostly data), we are also looking to share applications. We spent a lot of time in trial and error in CHS, we’ve had a good experience. When talking about the expenses you need to have good practices in place. We were waiting to get in there for quite some time so there was a high demand at first. A lot of functionality is there that has made our lives a little easier especially with data downloads. We’ve looked at some increased cost from downloads, we worked with CHS to create a user pay bucket, some customers have much larger download needs. We can work with users to get access to data if they stay in the same AWS region. We are trying to create images that others can use without having to create a new virtual server every time. Now we have delivery services and lidar point cloud data in CHS, we are trying to move more applications out there, hopefully all in the next several years.
13. From zoom: Can you explain how "backups" of data maintained in the cloud are handled? ie - how would we access them; how long would they stick around?
    a. You configure one of the backup software solutions, we use Cloudberry server. It is installed on premise and managed there locally. Configured and managed through CHS. Once that is established, everything is controlled on premise. (more on the recording, around 1:15:00).
14. From zoom: Who on the CHS team is the go-to person to discuss maintaining "project data" in the cloud - including backups of this data?
    a. The entry point would be the communications team. It depends what you want to do with it. cloudservices@usgs.gov
15. Can you all (CHS team) post your recipe for CI/CD, including the yml config you mentioned for the battery of tests you are running?
    a. https://code.chs.usgs.gov/CHS-IaC/Custom-Account
16. In establishing a CHS account, can you elaborate on what the security review process looks like?
    a. 1:18:02. It depends on what you want to do with CHS. For managed services there is more responsibility on CHS and the customer needs to agree with our security rules of behavior. With a custom environment, there is more security responsibility on the customer end. Fill out an information types workbook so that the data and app are FISMA moderate or low. The customer need to update their assessment and authorization documentation to reflect what has moved into the cloud. Customers have ~6 months to complete that, but right now there is no formal security review process for that.
17. interested in other folks' experiments w/the sandbox. Is there any kind of record about that stuff that we could browse? E.g., what really _doesn't_ work?
    a. That’s something we don’t have but we can look into that. We have customers fill out a Google form on why they are going into the sandbox, that is something we can share. Multiple customers have started in the sandbox and moved to production environments.
18. How long does it typically take to get a new CHS account approved?

    a. Part of the onboarding process is the financials. Then it also depends on the ability to provide custom accounts depending on your use case. We may sometimes deny service until a managed service is available to you. Otherwise once financing is available, the account can be deployed within minutes. It also depends on which CHS service you want to use.
19. can you provide info/training or direct potential users to things like that to help them assess how well "cloud" might support a given project?
    a. CHS is working on a customer knowledge base to provide some of this information, it is in the works. There are so many variables and many ways to meet a certain use case. It's a hard question to provide information on.
20. Kind of sounds like everything is golden with CHS. What's not working? Where are your major hurdles to getting more of our USGS apps online?
    a. Right now it is scalability. Providing resources and support to let everyone come in with their custom accounts is very difficult to manage. We are trying to find the best managed services that will serve the most groups so that we can have less custom environments.
    b. Not a lot of people have experience with cloud, it's not their day job, people need to learn new technology and develop with that.
21. can you give a quick example of a non-managed use case? Would "load my data and code, compile, crunch data, offload results" be an example?
    a. Yes it would. That would be a small workflow, does not need custom account, can use the sandbox. That could be done in a single quarterly sandbox.
22. Is there a possibility that a different cloud vendor, other than AWS, would be selected at the end of this current Period of Performance?

    1. We are planning for a new open competition acquisition that will provide continued cloud services at the conclusion of the current task order. It is the responsibility of the contracting officer and the evaluation panel to determine an award decision based on evaluation factors and basis for award that are tailored to the acquisition, following the Federal Acquisition Regulation (FAR).

Attendees

A Participant Report is available to CDI Members. Please log in to download the report. If you would like to become a member of CDI, email cdi@usgs.gov.