July 21, 2020 Introduction to AWS for Researchers

Q&A from Session

- Will we be provided with the recording?
    - Yes - this is being recorded. Both the recording and slides will be made available in the coming days.
    - People who registered will receive a follow up email with the links to the video and the slides.
- Can you take advantage of the AWS Marketplace under NET+ contract?
    - Yes, the AWS Marketplace is provided as part of the NET+ program.
- Does AWS have any bioinformatics services? Sequence alignment, etc.
    - AWS does not provide a native bioinformatics service. However, we support a large number of bioinformatics, genomics, protoemics, etc pipelines and workflow engines (https://aws.amazon.com/health/, https://aws.amazon.com/government-education/cromwell-on-aws/, https://aws.amazon.com/quickstart/biotech-blueprint/nextflow/). We also have many partners (e.g., Illumina, Seven Bridges) that provide solutions in this field.
- What is a point of presence?
    - The POPs are used for both AWS CloudFront to deliver content to end users at high speeds, and Lambda@Edge to run Lambda functions with the lowest possible latency.
- Can I get a location of the availability locations in the US?
    - The US regions are :
        - Northern Virginia (us-east-1)
        - Ohio (us-east-2)
        - Northern California (us-west-1)
        - Oregon (us-west-2)
        - https://aws.amazon.com/about-aws/global-infrastructure/
- Are all instance types available from the Canada Region?
    - A complete list is available on this page if you select "Canada (Central)" as your region: https://aws.amazon.com/ec2/pricing/on-demand/
- Are there whitepapers or articles that discuss some of the design decisions for using on premise GPU or HPC clusters vs. AWS GPU EC2 instances?   Beyond steady-state / fully utilized scheduling observations?
    - This whitepaper might be helpful in thinking through some design decisions: https://d1.awsstatic.com/HPC2019/Challenging-Barriers-to-HPC-in-the-cloud-Oct2019.pdf
- Is it possible to benchmark my code so that I can get an estimate of how much it might cost using EC2?
    - There are Proof-of-concept credits available for you to benchmark/optimize your workload. You can get access to those through your AWS account team

- Is there a date set for when AWS will end support for 2003R2?
    - Currently no, there is not an end of support for 2003R2 instances. That being said, we do not offer a 2003 image, so to create a new 2003 EC2 instance an image will need to be imported.
    - https://aws.amazon.com/ec2/vm-import/
- What distro is Amazon Linux based on?
    - Amazon Linux 2 is developed by Amazon Web Services (AWS). You can find additional information here: https://aws.amazon.com/amazon-linux-2/
- Is slurm (job scheduler software) supported on AWS (in addition to other schedulers)?
    - Yes. SLURM is supported by ParallelCluster.
- Is hyperthreading work supported?
    - Yes, you can turn on/off HT on ParallelCluster
- Could you give an example of an aws credit?
    - https://aws.amazon.com/research-credits/
- With the RONIN license purchase, is the annual license purchase matched every year with $20K AWS credits or is the $20K AWS credit a singular event?
    - The AWS credit is a singular event.
- Is S3 Intelligent Tiering ever more expensive than S3 standard? For instance, if everything was frequently accessed?
    - You can see the pricing of all tiers here: https://aws.amazon.com/s3/pricing/
- What exactly is amazon-cash (it was mentioned now), hand what is it used for?
    - I checked with the internal team and we're unsure of the Amazon-cash reference. Danyell may have mentioned cache in a memory context but we're unable to find the reference to Amazon-cash. The other reference could have been around AWS credits too which could be find here: https://aws.amazon.com/research-credits/
- 8PB per database?
    - 8 PB of compressed data per Redshift cluster, which can host multiple databases.
- Does AWS provide data publishing service?
    - Yes, Amazon Data Exchange. https://aws.amazon.com/data-exchange/
- Will these questions and answers be a part of the recording, or will they otherwise be made available, as well?
    - Yes we can make the Q&A file available.
- Can someone share a link to the researcher-specific learning pathways, please? I'd like to distribute those to our researcher community.
    - https://aws.amazon.com/blogs/publicsector/no-cost-online-aws-training-pathway-researchers-research-it/
- Is storing the regulated data more expensive than the regular data? Can you provide me the information on the related cost? Thanks
    - Regulated data is the same price as any other data. All data in S3 can take advantage of strong encryption and access controls at no additional cost.

- Is Globus supported to transfer data to/from AWS?
  - Yes, globus has a connector to move data to and from S3: https://www.globus.org/connectors/s3
- Is there more information about the Redshift claim that it is "50% less expensive than all other cloud data warehouses?" This is surprising to me in a context like Cloud where how you use it and how much is what usually drives cost. Particularly interested in comparisons to Snowflake which is popular on our campus.
  - The claim comes from: https://aws.amazon.com/redshift/
  - ...which focuses on the pay-for-what-you-use nature of Redshift, but doesn't detail the exact configurations being compared. You may want to reach out to your AWS account team for a deeper conversation around Redshift.
- When transferring regulated data, how do I ensure that AWS is able to receive the data?
  - AWS has a shared responsibility model, that describes customer and AWS responsibilities for security and compliance. You as the customer are responsible for the secure configuration of your AWS resources, while AWS is responsible for the underlying security of the platform.
  - That said I, think this is about HIPAA PHI regulated data. AWS has a standard BAA agreement that we use with customers. For more information:
    - https://aws.amazon.com/compliance/hipaa-compliance/
    - https://aws.amazon.com/compliance/shared-responsibility-model/
- Can we get access to the cloudformation templates that you are using? For Rstudio etc. .
  - You can use this one https://docs.rstudio.com/rstudio-team/cloudformation/
- Why are tags used?
  - To help you manage your instances, images, and other Amazon EC2 resources, you can assign your own metadata to each resource in the form of tags. Tags enable you to categorize your AWS resources in different ways, for example, by purpose, owner, or environment. This is useful when you have many resources of the same type—you can quickly identify a specific resource based on the tags that you've assigned to it.
- Is the recipient of the shared notebook required to have the AWS login etc. and access permission to the instance?
  - Yes, the recipient needs to have access to that account and proper IAM permissions
- Does it support SAML2 for connecting to institutional IdPs?
  - Yes, it does through AWS SSO integration with your IdP
- Any free services for college students who are just getting into AWS?
  - Yes. https://aws.amazon.com/free/ and https://aws.amazon.com/education/awseducate/