National Aeronautics and Space Administration

Overview of HECC Pay-for-Use AWS Cloud

Aug. 7, 2019

NASA Advanced Supercomputing Division

Outline

- The Whys?
  - Why offer commercial Cloud resources?
  - Why use Cloud through HECC?
  - Why AWS?
  - Why Pay-for-Use and How?
- AWS Basic Info
  - Cloud regions
  - Cloud services
  - Sample costs of AWS resources
- Integration of AWS into HECC
  - Cost conscious: dynamic on/off of resources
  - Flexible and familiar environment
  - Software stack

Question? Use the Webex chat facility to ask the Host

Why Offer Commercial Cloud Resources?

- Provide access to new hardware resources not currently available at HECC, such as newer GPUs or AMD CPUs.
- Provide an alternative environment for applications that do not compile or run on HECC on-premises resources (Pleiades, Electra, Merope or Endeavour).
- Provide faster turnaround when you do not want to wait in the on-premises queues for a long time.

Question? Use the Webex chat facility to ask the Host

Why Use Cloud Through HECC?

- You do not have to set it up yourself from a barebone IaaS Cloud. HECC Cloud experts provide you a PaaS environment that is ready to use.

  Infrastructure-as-a-Service (IaaS) -> Platform-as-a-Service (PaaS)

- HECC provides an easy and familiar hybrid environment where you can move your work between HECC on-premises resources and the HECC Cloud.

  HECC develops and maintains all the behind-the-scene infrastructure for accessing/bursting jobs to the Cloud

- HECC goes through NASA/EMCC, which may get some volume discount.
- HECC Cloud experts may be able to help resolve some of your issues and/or channel them to the Cloud provider(s).
- HECC follows strict government guidelines to ensure security.

Why AWS?

- NASA requirements for dealing with Cloud
  - **All federal agencies should use clouds with FedRAMP certification**: The Federal Risk and Authorization Management Program is a government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services.
  - **All NASA cloud access is controlled through EMCC**: NASA Enterprise Management of Cloud Computing enables NASA's cloud consumption to be consistent, safe, and compliant with industry best practices and Federal laws and requirements.
- Amazon Web Services is FedRAMP-compliant and currently the only cloud provider (with the majority of its services) approved by EMCC.
- Microsoft Azure will be on the EMCC-approved list soon.

Why Pay-for-Use and How?

- Why?
  - Guidance from the HQ HECC Portfolio Manager is that HECC will provide an easy mechanism for HECC users to access the diverse cloud offerings utilizing user funds and assist users as required to take advantage of the cloud as appropriate for their requirements.
- How?
  - PI
    - determine if ITAR/EAR99 environment is needed.
      - No -> proceed with Public Cloud
      - Yes -> wait till Gov Cloud is approved for HECC use
    - send NASA funding WBS to HECC to establish an ARC WBS billing account
    - provide a list of users allowed to use this HECC Cloud account
    - provide an initial desired Cloud configuration (can be adjusted later)
  - HECC
    - set up the account and configure the environment
    - track usage of all resources (front-end, PBS server, compute, filesystems, storage, network, support, …)
    - charge expenses against PI's funding and provide usage report and ensure the account is not overdrawn

AWS Basic Info – Cloud Regions

- Public Cloud Regions (available now)
  - Lower security requirements
  - Used for non-ITAR/EAR99
  - Cheaper
  - Multiple regions; HECC Cloud currently uses AWS US-West(Oregon); other public regions can be set up upon demand and/or request
- Government Cloud Regions (available in the near future)
  - High security requirements
  - Used for ITAR/EAR99
  - More expensive
  - AWS GovCloud (US-West) and (US-East)

Question? Use the Webex chat facility to ask the Host

AWS Basic Info – NASA Approved Services

- EC2: Elastic Compute Cloud
  - Compute resources to run applications on
  - Similar to the various Pleiades compute node types
- EBS: Elastic Block Store
  - Block devices that can be mounted on EC2 for faster I/O
  - Similar to the Pleiades filesystems ($HOME, $NOBACKUP)
- S3: Simple Storage Service
  - Object storage
  - Similar to NAS storage system Lou
- VPC: Virtual Private Cloud
  - Allows access to EC2 over a virtual private network
  - Similar to the HECC Enclave (where the access is controlled/blocked)
- IAM: Identity and Access Management
  - Provides authentication and policy-based control to various AWS services
  - Similar to an HECC account

Question? Use the Webex chat facility to ask the Host

AWS EC2

- Current Generation of Instance types (July 2019)
  - General Purpose: a1, t2, t3/t3a, m4, m5/m5a
  - Compute optimized: c4, c5/c5d/c5n/c5.metal
  - Memory optimized: x1/x1e, r4, r5/r5a/r5d/r5ad, z1d
  - Accelerated computing: p2, p3/p3dn, g3/g3s
  - Storage optimized: h1, i3/i3en, d2

a: use AMD processors instead of Intel Xeon
d: with local NVMe SSD on the instance
n: more memory, higher EBS/network bandwidth
metal: use non-virtualized environment

- On-demand instances
  - Full price
  - Price varies with AWS regions, OS, instance types
- Spot instances
  - Significant discount (~70% discount is common) except for newly introduced instances whose price remains high for some time
  - Price can fluctuate every 5 minutes
  - May take a long time to get them
  - Can be interrupted by EC2 with 2 min of notification

Some compute instance info

Instance Type	Processor Model	CPU Base Frequency (GHz)	# Physical Cores	Memory (GiB)	Max EBS Bandwidth (Gbps)	Network Bandwidth (Gbps)	On-demand Price ($)	Sample Spot Price ($)
c4.large	Haswell (E5-2666 v3)	2.9	1	3.75	0.5	Moderate	0.1	0.0307
c4.xlarge	Haswell (E5-2666 v3)	2.9	2	7.5	0.75	High	0.199	0.0712
c4.2xlarge	Haswell (E5-2666 v3)	2.9	4	15	1	High	0.398	0.1711
c4.4xlarge	Haswell (E5-2666 v3)	2.9	8	30	2	High	0.796	0.2708
c4.8xlarge	Haswell (E5-2666 v3)	2.9	18	60	4	10	1.591	0.5017
c5.large	Skylake (Platinum 8124M)	3.0	1	4	3.5	Up to 10	0.085	0.0322
c5.xlarge	Skylake (Platinum 8124M)	3.0	2	8	3.5	Up to 10	0.17	0.0644
c5.2xlarge	Skylake (Platinum 8124M)	3.0	4	16	3.5	Up to 10	0.34	0.1288
c5.4xlarge	Skylake (Platinum 8124M)	3.0	8	32	3.5	Up to 10	0.68	0.2576
c5.9xlarge	Skylake (Platinum 8124M)	3.0	18	72	7	10	1.53	0.5795
c5.18xlarge	Skylake (Platinum 8124M)	3.0	36	144	14	25	3.06	1.159
c5.12xlarge	Cascade Lake (Platinum 8275CL)	3.0	24	96	7	12	2.04	0.7727
c5.24xlarge	Cascade Lake (Platinum 8275CL)	3.0	48	192	14	25	4.08	1.5454
c5.metal	Cascade Lake (Platinum 8275CL)	3.0	48	192	14	25	4.08	1.5454

- c4 and c5 instances come with many different configurations (nxlarge). They are for compute intensive traditional HPC workload.
- AWS web site lists vCPU number, which is double the number of physical CPU cores shown in this table.
- Max EBS bandwidth (dedicated for I/O into EBS volumes) info from https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSOptimized.html
- Max network bandwidth of 25 Gbps of regular c5 instances is much smaller than the 100 Gbps of HECC Skylake/Cascade Lake. Expect worse cross-node performance than Pleiades/Electra/NFE. One Nov 26, 2018, c5n with 100 Gbps became available.
- Pricing based on US-West(Oregon) region, basic Linux (Amazon Machine Image) OS.
  On-demand https://aws.amazon.com/ec2/pricing/on-demand/; More powerful instances more expensive.
  Spot https://aws.amazon.com/ec2/spot/pricing/ (July 26, 2019 record; may change), usually ~1/3 of On-demand
- Cascade c5 instances were made available starting June 18, 2019. Ahead of NAS (Late Aug, 2019) – 1152 nodes (2.5 GHz, 40 cores, 192 GiB).

Question? Use the Webex chat facility to ask the Host

Some GPU instance info

Instance Type	Processor Model	CPU Frequency (GHz)	# of Physical CPU Cores	# of GPUs	CPU Memory (GiB)	Max EBS Bandwidth (Gbps)	Network Bandwidth (Gbps)	Oregon On-demand Price ($)	Oregon Sample Spot Price ($)	Gov On-demand ($)
p2.xlarge	Broadwell (E5-2686 v4) + Tesla K80	2.3	2	1	61	0.75	High	0.9	0.285	1.08
p2.8xlarge	Broadwell (E5-2686 v4) + Tesla K80	2.3	16	8	488	5	10	7.2	2.160	8.64
p2.16xlarge	Broadwell (E5-2686 v4) + Tesla K80	2.3	32	16	768	10	25	14.4	4.320	17.28
p3.2xlarge	Broadwell (E5-2686 v4) + Tesla V100	2.3	4	1	61	1.75	Upto 10	3.06	0.918	3.67
p3.8xlarge	Broadwell (E5-2686 v4) + Tesla V100	2.3	16	4	244	7	10	12.24	3.672	14.69
p3.16xlarge	Broadwell (E5-2686 v4) + Tesla V100	2.3	32	8	488	14	25	24.48	7.344	29.38

- p2 and p3 instances contain GPUs and are ideal for machine learning, HPC, data processing, and cryptography.
- Pricing data based on July 26, 2019 AWS Pricing page. p3.16xlarge is expensive ($24.48/hour).
- Gov price is slightly higher than US-West(Oregon) price.
- p3 (with V100) instances were made available starting Oct 25, 2017. Ahead of NAS (June 2019).
- NAS GPU resources:
  - 64 san_gpu nodes: Sandy Bridge CPU and 1 Nvidia K40 GPU card
  - 19 sky_gpu nodes: Skylake CPU + V100 GPUs (17 nodes with 4 V100, 2 with 8 V100 cards)

Some general/memory optimized instance info

Instance Type	Processor Model	Processor Frequency (GHz)	# Physical CPU Cores	Memory (GiB)	Max EBS Bandwidth (Gbps)	Network Bandwidth (Gbps)	On-demand Price ($)	Sample Spot Price ($)
m5a.large	AMD EPYC 7000	2.50	1	8	2.12	Upto 10	0.086	0.0338
m5a.xlarge	AMD EPYC 7000	2.50	2	16	2.12	Upto 10	0.172	0.0676
m5a.2xlarge	AMD EPYC 7000	2.50	4	32	2.12	Upto 10	0.344	0.1352
m5a.4xlarge	AMD EPYC 7000	2.50	8	64	2.12	Upto 10	0.688	0.2704
m5a.12xlarge	AMD EPYC 7000	2.50	24	192	5	10	2.064	0.8113
m5a.24xlarge	AMD EPYC 7000	2.50	48	384	10	20	4.128	1.6226
r5a.large	AMD EPYC 7000	2.50	1	16	2.12	Upto 10	0.113	0.0354
r5a.xlarge	AMD EPYC 7000	2.50	2	32	2.12	Upto 10	0.226	0.0708
r5a.2xlarge	AMD EPYC 7000	2.50	4	64	2.12	Upto 10	0.452	0.1417
r5a.4xlarge	AMD EPYC 7000	2.50	8	128	2.12	Upto 10	0.904	0.2833
r5a.12xlarge	AMD EPYC 7000	2.50	24	384	5	10	2.712	0.8499
r5a.24xlarge	AMD EPYC 7000	2.50	48	768	10	20	5.424	1.6999

- m5a is general purpose; used for web servers, app servers, gaming, etc.
- r5a is memory optimized; used for data mining, data analytics, etc.
- Pricing data based on July 26, 2019 AWS online pricing page.
- Slightly lower-cost (10%) than comparable Intel-based instances.
- Availability announced by AMD and AWS on Nov 6, 2018.
- Currently, there are no AMD processors at NAS.

Sample EBS, S3, and Network Costs

- EBS (US-West) provision – pay even if you have no data there
  - General Purpose SSD (gp2) Volumes: $0.10 per GB-month
  - Provisioned IOPS SSD (io1) Volumes: $0.125 per GB-month storage, $0.065 per provisioned IOPS-month
  - Throughput Optimized HDD (st1) Volumes: $0.045 per GB-month
  - Cold HDD (sc1) Volumes: $0.025 per GB-month
- S3 (US-West) – pay only for what you use
  - Standard storage: first 50 TB/month $0.023 per GB; next 450 TB/month $0.022 per GB
  - Standard infrequent access: all storage/month $0.0125 per GB
- Data transfer out from AWS S3 to Internet

  GovCloud pricing info
  - first 10 TB/month $0.155 per GB
  - next 40 TB/month $0.115 per GB
  - next 100 TB/month $0.090 per GB

Question? Use the Webex chat facility to ask the Host

Sample Costs

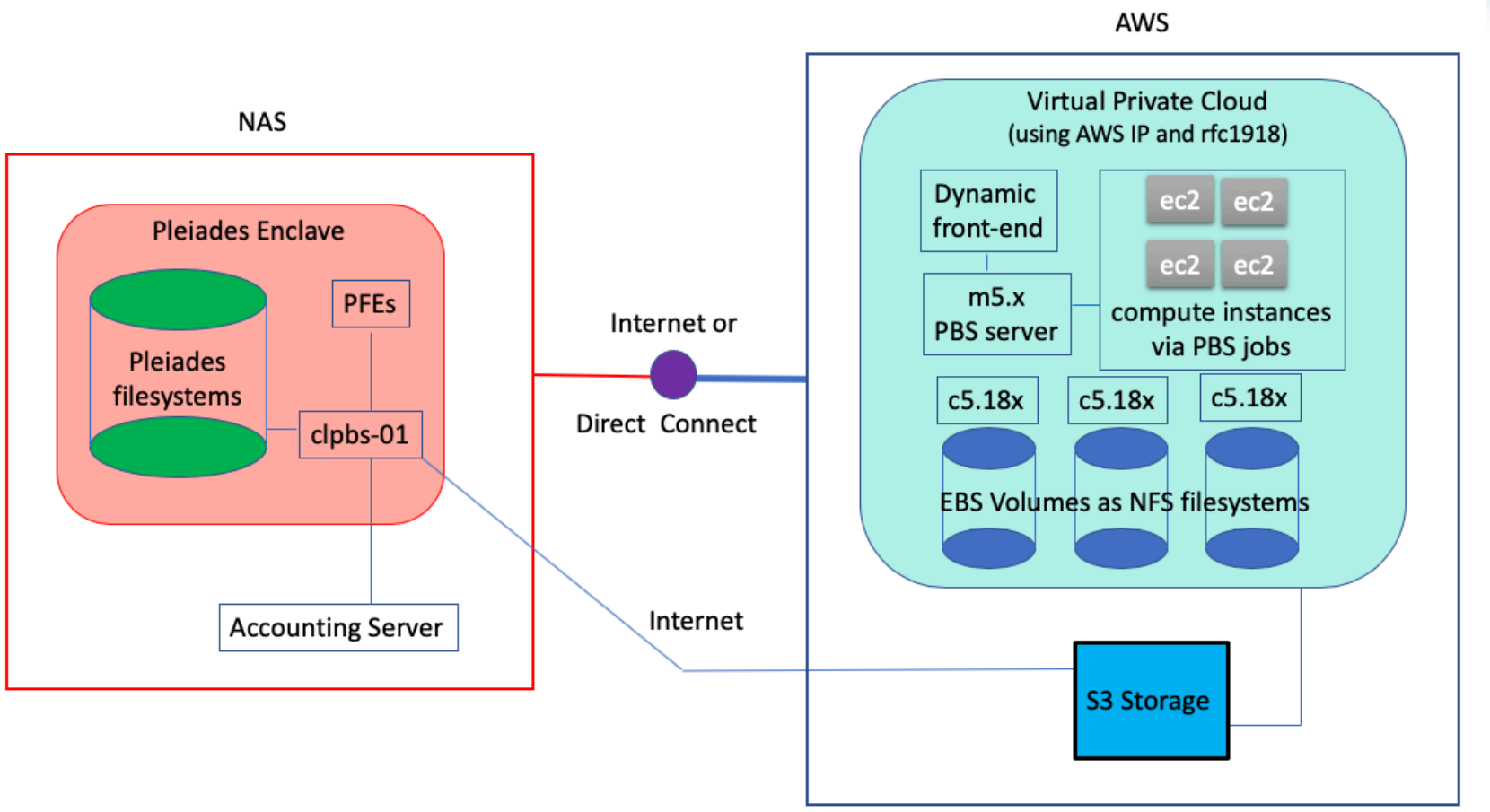
- EC2 1 instance use for 1 month continuously

c5.18xlarge (Skylake)	~$2,200/month
p3.16xlarge (GPU)	~$17,625/month
r5a.24xlarge (AMD)	~3,900/month

- EBS gp2: provision 10 TB costs $1,000/month
- S3 standard: use 10 TB costs $230/month
- 10 TB transfer out from AWS to Internet ~$920/month

Question? Use the Webex chat facility to ask the Host

Integration of AWS into HECC



For security and accounting purposes,
each project has its own "isolated" HECC Cloud environment.

Cost Conscious

- We dynamically turn on/off AWS front-end, servers for you
  - On Pleiades, PFEs, PBS and Lustre servers are up year-round
  - On HECC AWS Cloud, they are turned on only when you need them
  - AWS front-end: cost varies depending on the instance type you choose
  - AWS filesystem server: c5.18xlarge 1 instance 24 hour x 30 days, ~$2,200/month
    Depending on filesystems, cheaper servers may be used
  - AWS PBS server: m5.xlarge 1 instance 24 hour x 30 days, ~$140/month
- Go with job-time filesystems if you do not need persistent filesystems
  - Example: EBS gp2, provision 10 TB costs $1,000/month
  - Persistent: data stays there after batch job ends (you pay $1000/month)
    When accessed, need a filesystem server to support it (additional cost)
  - Job-time: data is removed when batch job ends (you pay much less)
    #CLOUD –volume_type, -volume_size, -volume_mount directives
    Most job-time filesystem types (except “shared”) use the same instances for compute and as filesystem server(s)

Cost Conscious

- Do not transfer data out to internet with an expensive instance (for example, a GPU p3.16xlarge which costs $24.48/hour)
  - File transfer to internet has a limited bandwidth and can take a long time
  - If you have to transfer data out of a filesystem at end of a job which uses expensive compute instances, better to transfer them to S3 inside the job.

    You can transfer data between S3 and Pleiades afterwards.

    See a later slide for file transfer options.

- Delete data you don't need in S3
  - pfe% nas_s3_ls, nas_s3_del
  - aws% nas_s3_ls, nas_se_del
  - #CLOUD –volume_list, -volume_delete

Flexible and Familiar Environment (Front-End)

- Request an AWS dynamic front-end
  - Use /u/scicon/tools/bin/aws_fe [options] from a PFE
  - Highly recommend adding /u/scicon/tools/bin to your $PATH
  - To request 1 node with 2 hours time

    pfe% aws_fe –t 2

    The front end is booting, when the instance is ready an email will be sent to your NAS email with login instructions.
  - To check the status of your front-end request

    pfe% aws_fe –l

    Front End Queued

    pfe% aws_fe –l

    env SSH_AUTH_SOCK="" ssh -i ~/.ssh/id_rsa_yours your_nas_username@iii.jjj.kkk.lll
  - To connect to the front-end node

    pfe% env SSH_AUTH_SOCK="" ssh -i ~/.ssh/id_rsa_yours your_nas_username@iii.jjj.kkk.lll
  - To properly terminate the front-end node

    pfe% aws_fe –k

Question? Use the Webex chat facility to ask the Host

Flexible and Familiar Environment (File Transfer)

- Between PFE and an AWS Dynamic Front-End
  - scp from a PFE

    pfe% scp -i ~/.ssh/id_rsa_yours ~/file1 iii.jjj.kkk.lll:/nobackup/your_user_name
  - Set up SUP on AWS and use sup/shift from AWS

    aws% sup shiftc pfe21.nas.nasa.gov:~/file1 .
  - Set up SSH passthrough on AWS and use scp from AWS

    aws% scp pfe21.nas.nasa.gov:~/file1 .
- Between a Front-End and your AWS S3 space (nas_s3_xxx)
  - PFE or an AWS front-end

    pfe (or aws)% nas_s3_put file1 /
    pfe (or aws)% nas_s3_get /file1
- Staging files from Pleiades in a PBS job
  - #CLOUD –stagein_file, -stagein_dir,
  - #CLOUD -stageout_file, -stageout_dir, -stageout_file_del, -stageout_dir_del
- Transferring data from/to S3 inside a PBS job
  - Persistent filesystem: #CLOUD –get_file, -get_dir, -put_file, -put_dir
  - Job-time filesystem: #CLOUD –volume_get, -volume_put

Flexible and Familiar Environment (PBS)

- Sample PBS Script

```
#PBS –lselect=1:ncpus=2:mpiprocs=2
#PBS –lwalltime=2:00:00
#PBS –q cloud@clpbs-01          (if submitting from PFE, specify this queue)
#PBS –W group_list=xxxxx
#CLOUD –stagein_file=gridfile   (if you need gridfile to be stagedin)
#CLOUD –volume_type=shared      (both nodes will be able to access this shared filesystem)
#CLOUD –volume_size=10G         (requesting 10 GB for this job-time filesystem)
#CLOUD -volume_mount=/shared_data
#CLOUD –volume_get=/binary     (assume that your S3:/binary folder has a.out)
#CLOUD –volume_put=/results    (everything under /shared_data will be put to this S3 folder)

module load intel-mpi/2019.3.062
cd $PBS_O_WORKDIR
mpiexec –np 4 ./a.out > output
```


Question? Use the Webex chat facility to ask the Host

Flexible and Familiar Environment (PBS)

- PBS Commands
  - qsub, qstat, and job-deletion can be done from either PFE or an AWS dynamic front-end

    pfe% qsub –q cloud@clpbs-01 jobscript

    pfe% qstat –a @clpbs-01 or *qstat -a @`aws_pbs_host`* (after job moved)

    pfe% aws_qdel jobid

    or

    aws% qsub jobscript

    aws% qstat -a

    aws% qdel jobid

  - PBS output/error files are sent back to your $PBS_O_WORKDIR (PFE/AWS)
  - PBS output file contains job cost summary (see next)

Question? Use the Webex chat facility to ask the Host

Per-Job Cost Summary

```
Job Resource Usage Summary for 3552.clpbs-01.nas.nasa.gov
Total Runtime : 00:03:03
Job Stage In Time (free) : 00:00:00
Job Startup Time : 00:01:03
Time Spent In PBS Script : 00:02:00
Job Stage Out Time : 00:00:00
Walltime Requested : 02:00:00
Execution Queue : AWS Cloud
Charged To : cstaff
Job Finished : Wed Jul 3 10:38:09 2019
Instance Types (ondemand): 2 m5.xlarge
EBS Usage : 8577331200 bytes
S3 Usage : 0 bytes
Charged Bandwidth Usage : 0 bytes
NAS overhead charge : 0.000 percent
Job Costs : $0.00981639861027
```


Question? Use the Webex chat facility to ask the Host

Flexible and Familiar Environment (Accounting)

- HECC accounting tools
  - *pfe% acct_ytd –caws* or *acct_ytd –ccloud-all*

```
                                                                                  Linear
                            Fiscal                                                 YTD    Project
Project   Host/Group         Year        Used     % Used      Limit      Remain     Usage   Exp Date
--------- -------------- ------ ----------- ------- ----------- ----------- -------- ------------
cstaff    cloud              2019   19675.517    39.35   50000.000   30324.483    48.56    09/30/19
```

  - *pfe% acct_query –caws –u username –b 07/01/2019 –e 07/19/2019*
  - *pfe% acct_query –ccloud-all –b 07/01/2019 –e 07/19/2019*

```
Printing information for all the users supplied.
REPORT FROM 07/01/19 TO 07/19/19
ACCT_QUERY: Generating data ...
GRAND TOTAL FOR 07/01/19 TO 07/19/19
CLIENT       USER                 PROJECT              QUEUE             SBU Hrs/Cloud BU
------------ ------------------ ------------------- ----------------- -----------
aws          staff1               cstaff               cloud_exec              2.337
aws          staff1               cstaff               drc                     0.000
aws          staff1               cstaff               frontend                5.821
TOTAL FOR                    aws.cstaff                                        8.158
TOTAL FOR ALL PROJECTS FOR CLIENT: aws                                         8.158
TOTAL FOR PROJECT ON ALL CLOUD: cstaff                                      8.158
```

Software Stack under /nasa at AWS

- Compilers
  - Intel compilers
  - PGI compilers
- MPI
  - Intel MPI
  - OpenMPI
- GPU related
  - CUDA
- Python (can be installed upon request)
  - Intel Distribution for Python 3
  - Cuda Python

Install other software packages and bring licenses yourself.

For More Information

- PI on-boarding or questions/feedbacks
  - Send email to support@nas.nasa.gov

- AWS Cloud KB (ID: 581 to 596)
  - HECC AWS Cloud Overview
    http://www.nas.nasa.gov/hecc/support/kb/entry/581
  - HECC AWS Cloud File Transfer Overview
    http://www.nas.nasa.gov/hecc/support/kb/entry/582
  - Cloud Billing Units and Job Accounting
    http://www.nas.nasa.gov/hecc/support/kb/entry/596