



Cloud Infrastructure Setup

Ashley Tung
Solution Engineer



Contents

- AWS
 - Quick Introduction to AWS Instances
 - What are Spot Instances?
 - What are Spot Interruptions?
- MMC
 - Architecture
 - OpCenter GUI
- How MMC works with Nextflow
- Submitting a pipeline
 - Via a template
 - Via command line
- Viewing a pipeline
 - Workflow view

AWS EC2 Instances

- What is AWS?
 - Cloud infrastructure platform
 - Provides resources for cloud computing
- EC2 Instances
 - Virtual server instances in the cloud
 - Most common instances are On-Demand



What are Spot Instances?

- Sold based off spare VM capacity
- Sold at up to 90% discount

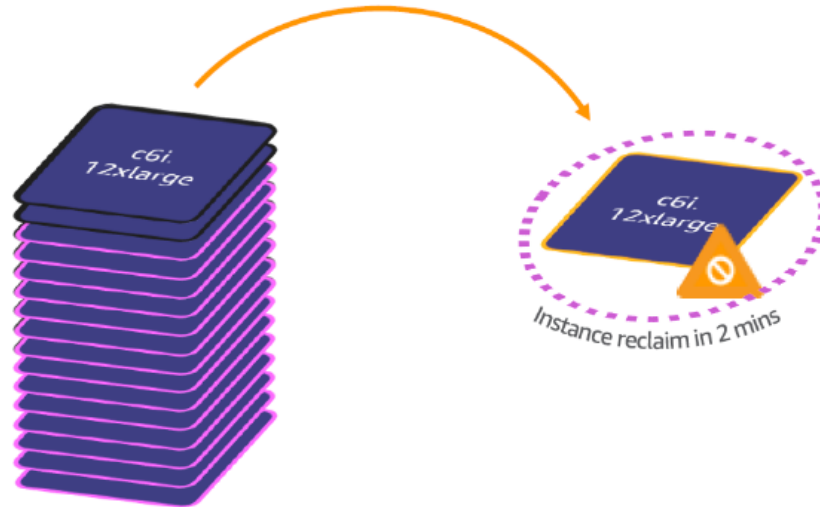


Instance name ▲	On-Demand hourly rate ▼
m4.10xlarge	\$2.0000
m4.16xlarge	\$3.2000

Instance Type ▲	Spot hourly rate ▼
m4.10xlarge	\$0.698
m4.16xlarge	\$1.0405

What are Spot Interruptions?

- At any point, Spot Instances can be reclaimed



- 2-minute grace period after notification
 - Inconvenience for long-running, fault-sensitive jobs
 - ex: cellranger, rnaseq, etc..

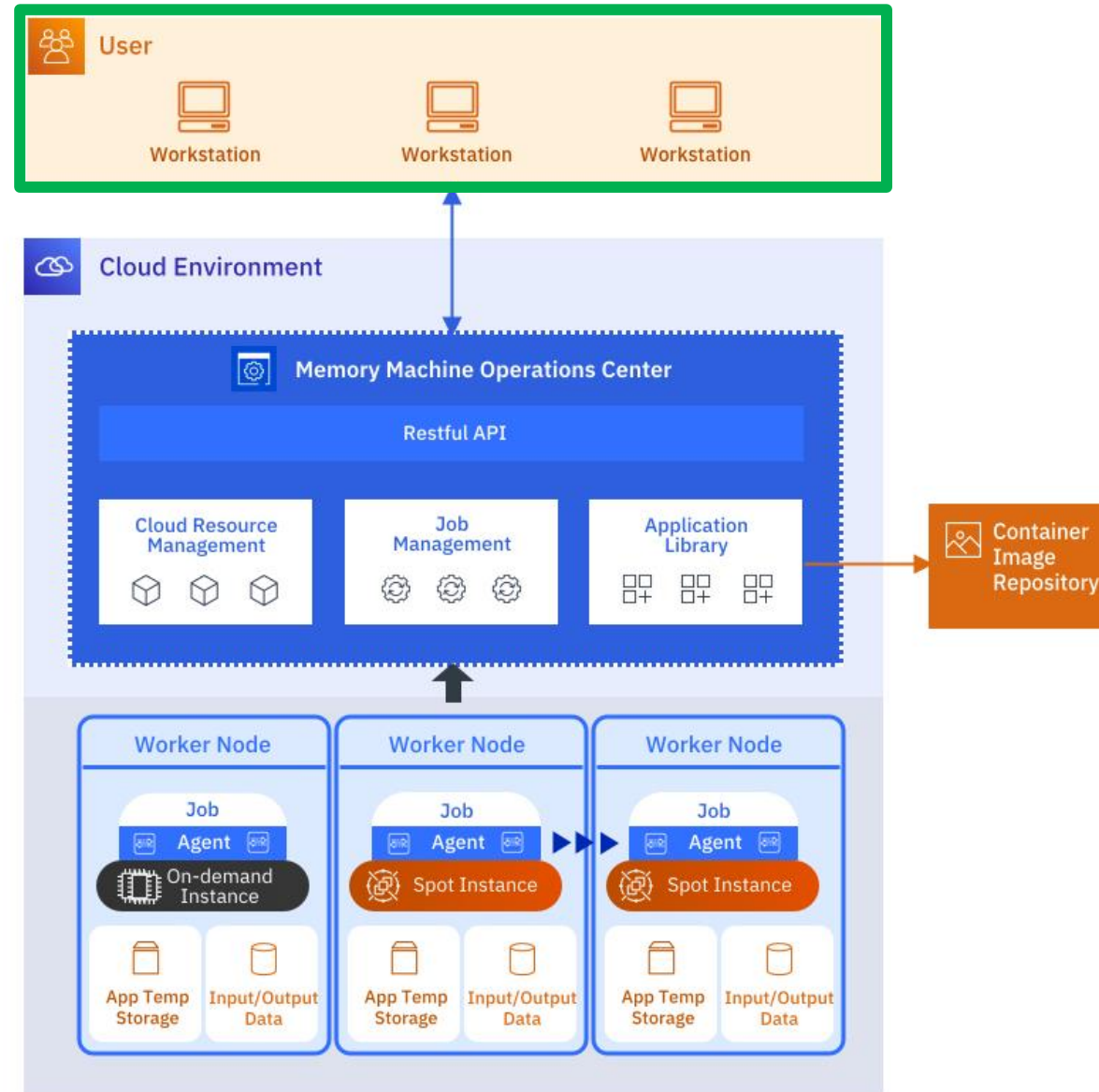


MMC – Memory Machine Cloud

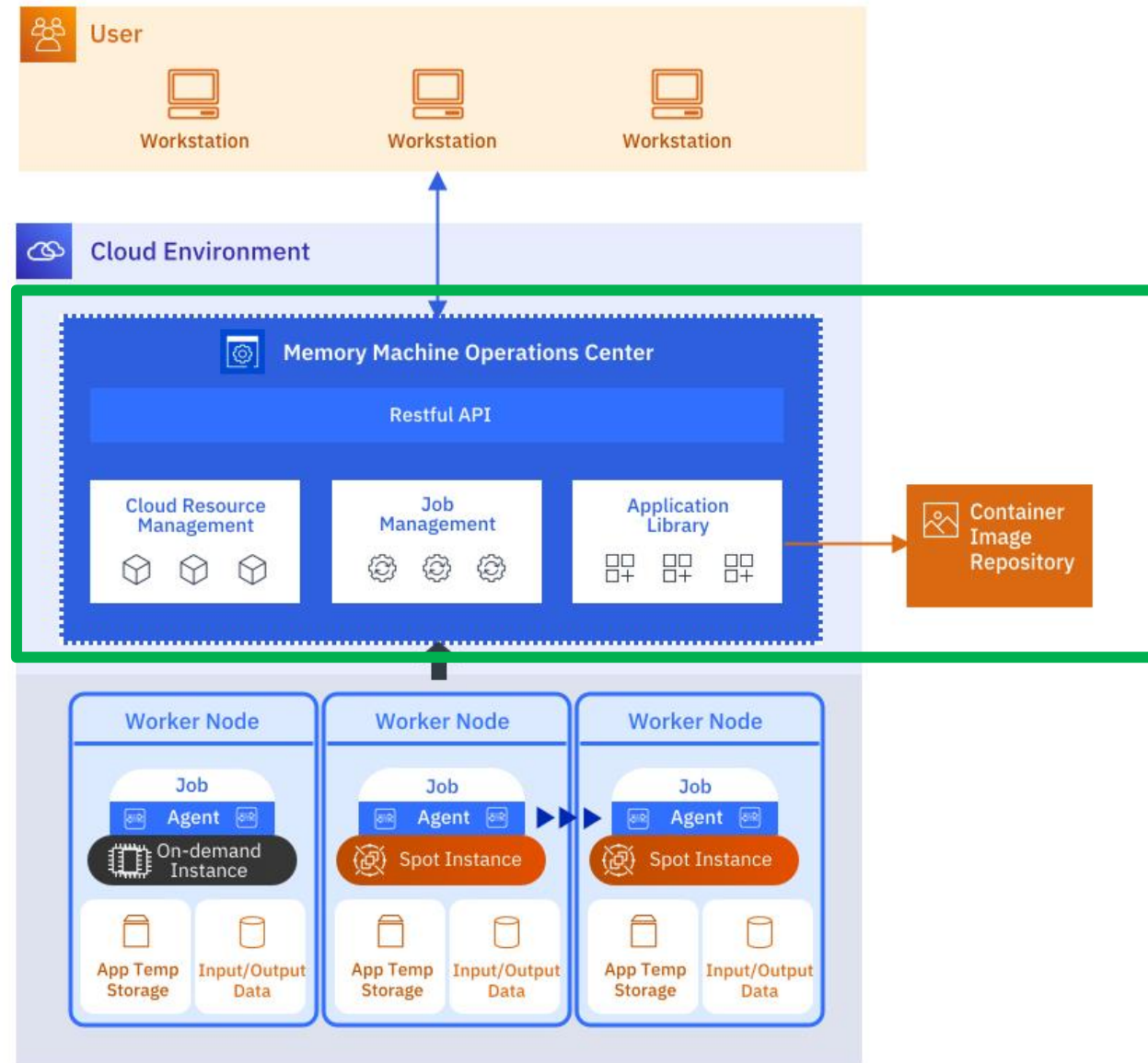
- Architect your application to be fault-tolerant
- Provide automated checkpoint recovery service
- Enables long-running jobs to run on multiple, low-cost Spot instances
- Includes setup specifically for Nextflow pipelines



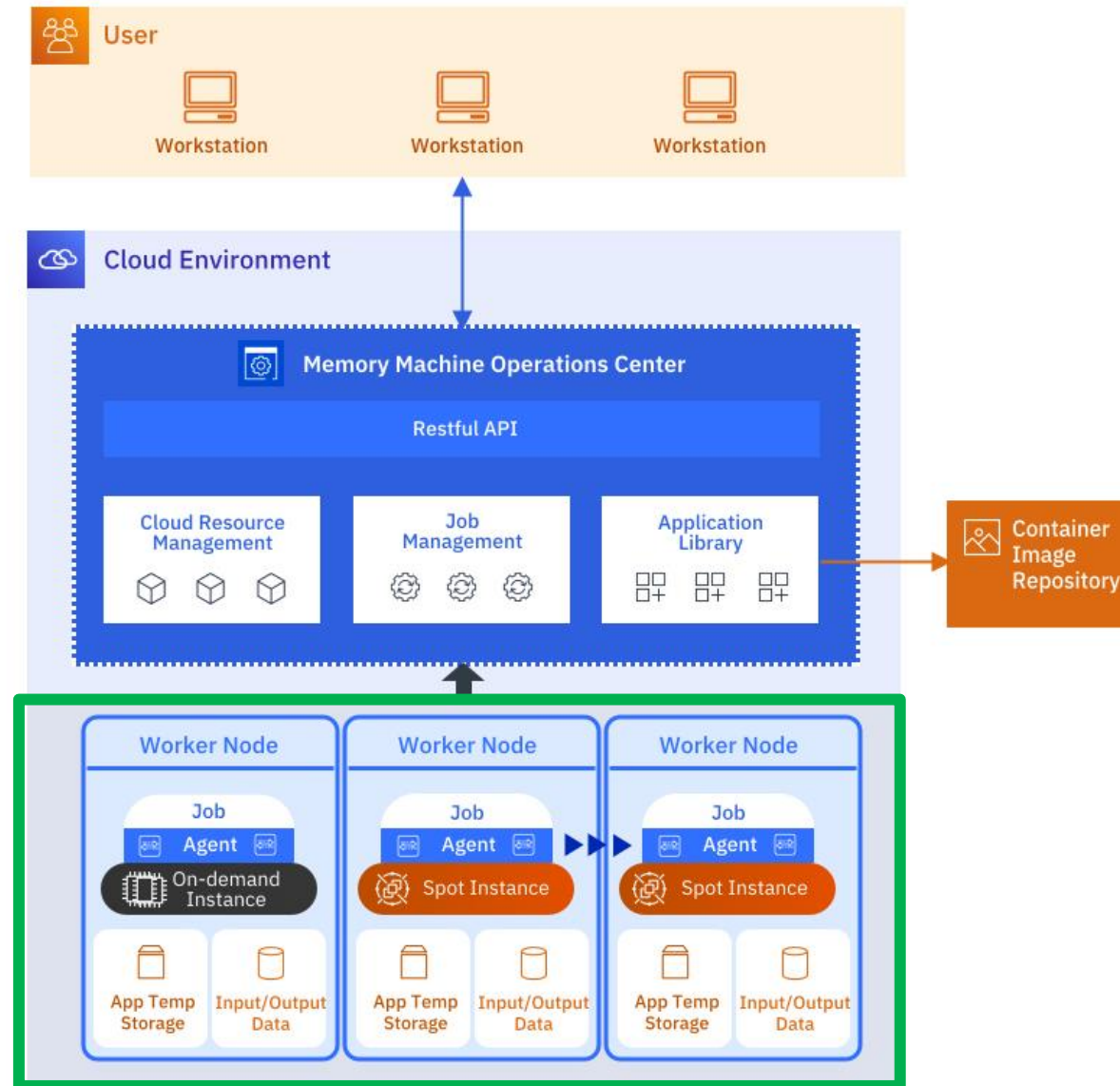
MMC – Architecture



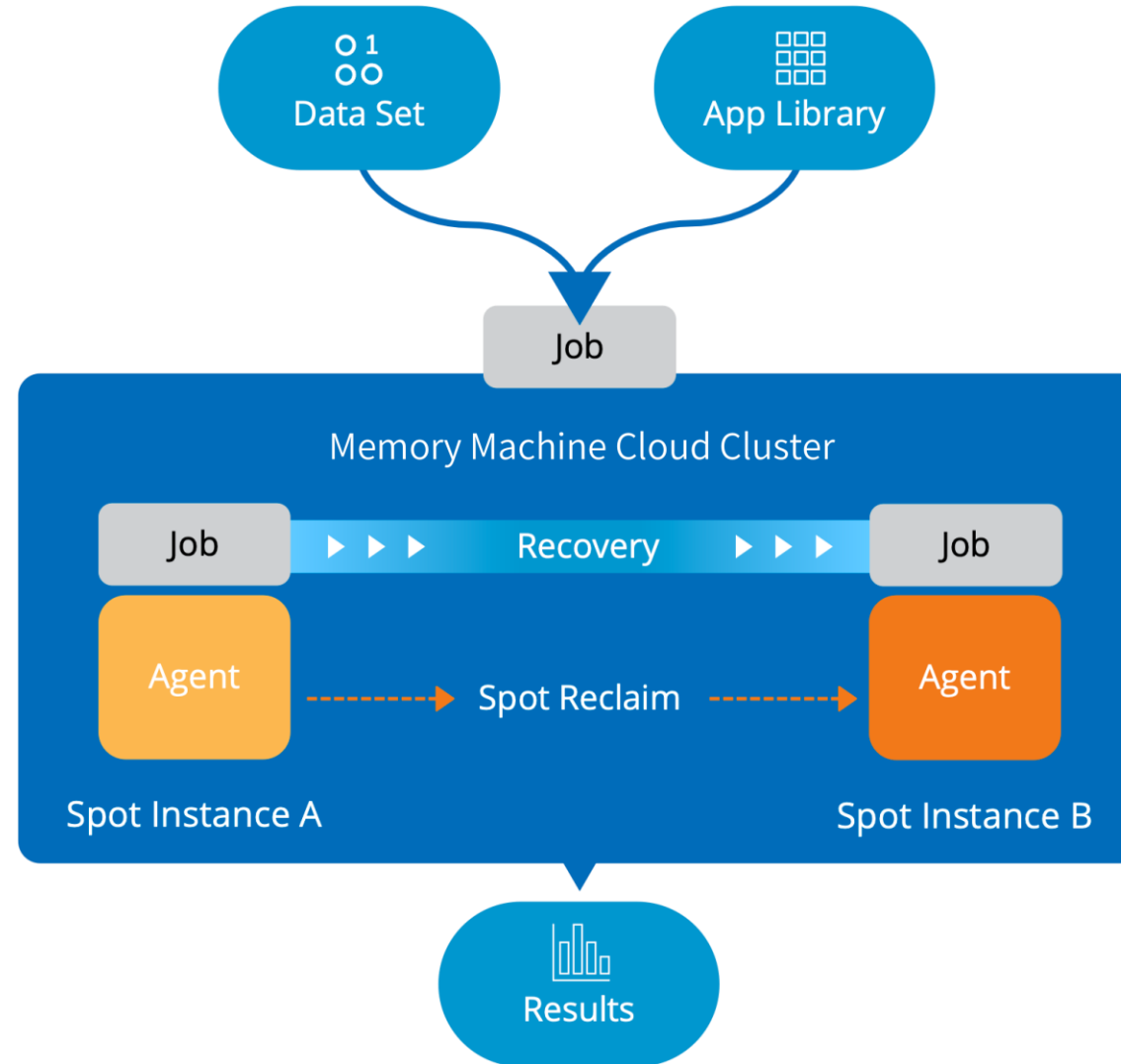
MMC – Architecture



MMC – Architecture



MMC – Architecture



OpCenter GUI

MMCloud OpCenter

+ Submit Job

Cost Summary

- Jobs
- Workflows
- Job Templates
- App Library
- SERVICE
- Gateways
- ADMINISTRATION
- SurfZone
- Users and Groups
- System Settings

Cost Summary

admin

Cloud Cost Month-to-date

\$ 5190.2415 USD

Net Saving Month-to-date

\$ 5287.9617 USD

Cloud Cost Year-to-date

\$ 14595.4015 USD

Net Saving Year-to-date

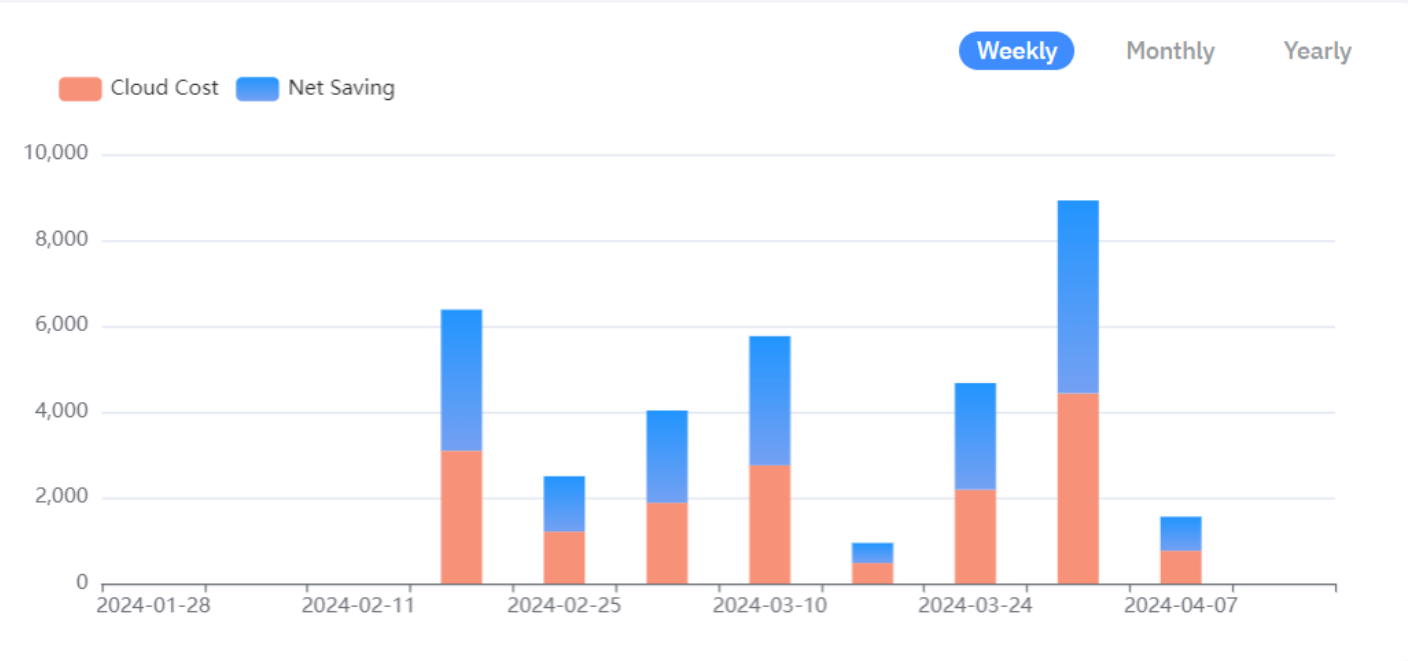
\$ 15583.5366 USD

Net Saving Summary

This Week

\$ 0.0974

Jobs	2
Cloud Compute Time	1h 31m 0s
Spot Instances	2
On-demand Instances	0
Total Cloud Cost	\$ 0.0965



Top Applications by Cost Weekly

Top Applications by Jobs Weekly

Top User by Cost Weekly

OpCenter GUI

MMCloud OpCenter

+ Submit Job

Cost Summary

Jobs

Workflows

Job Templates

App Library

SERVICE

Gateways

ADMINISTRATION

SurfZone

Users and Groups

System Settings

Cost Summary ?

Cloud Cost Month-to-date

\$ 5190.2415

USD

Net Saving Month-to-date

\$ 5287.9617

USD

Cloud Cost Year-to-date

\$ 14595.4015

USD

Net Saving Year-to-date

\$ 15583.5366

USD

Net Saving Summary

This Week

\$ 0.0974

Jobs2

Cloud Compute Time1h 31m 0s

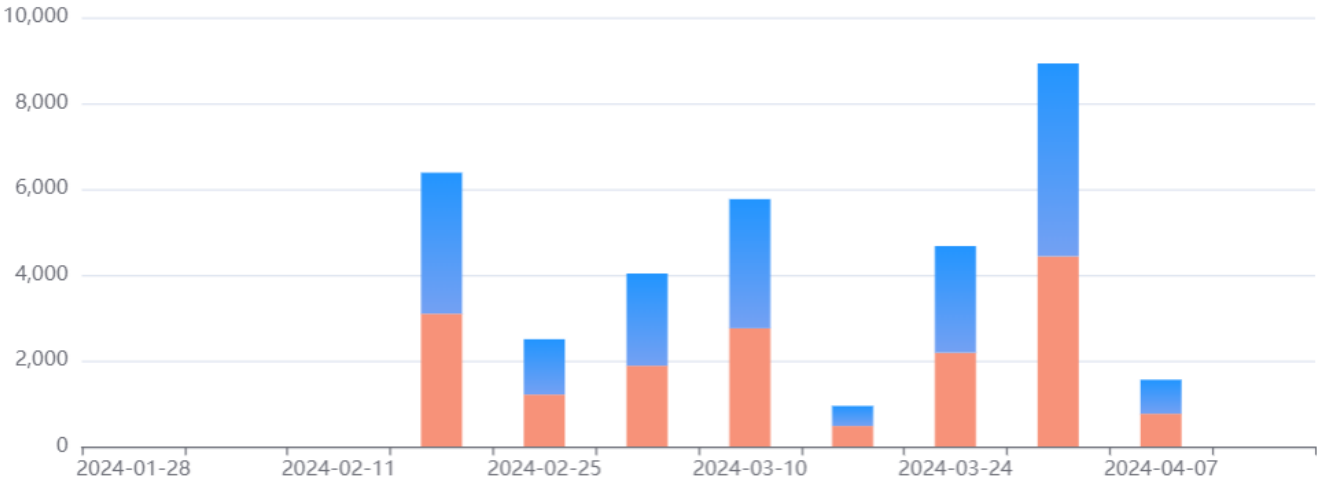
Spot Instances2

On-demand Instances0

Total Cloud Cost\$ 0.0965

Weekly Monthly Yearly

Cloud Cost Net Saving



Top Applications by Cost Weekly

Top Applications by Jobs Weekly

Top User by Cost Weekly

Job Tab

















- See Executing/Completed/Failed jobs submitted

Jobs

Category: Pending Running Success Failed Image Search Envs (e.g. proj=mmc,proc=s*)

Filters: update<=12w

Total 2441 items < 1 2 3 4 5 ... 245 > 10 / page Go to Page


ID	Name	Host	User	Status	Submitted	Updated	Wall Time	Cost	Actions
i2thjxtrga06h5wvp4mf	seurat-ukn42y-t2.large	54.242.227.201 (2Core8GB/Spot)	workshop-user20	Executing	2024-10-09 10:23:34	2024-10-09 10:31:17	7m46s	0.0046 USD	   
onycpdh7whcgtcjvdcg8t	seurat-ukn42y-t2.large		workshop-user20	Completed	2024-10-09 10:12:59	2024-10-09 10:23:29	10m30s	0.0067 USD	   
2fgs0kgw5wg8drn64gb5x	seurat-ukn42y-r5.large	52.90.216.11 (2Core16GB/Spot)	workshop-user20	Executing	2024-10-09 10:06:04	2024-10-09 10:31:17	25m15s	0.0220 USD	   
ei1vfg1neudadh2adl503	seurat-ukn42y-t2.large		workshop-user20	Completed	2024-10-09 10:02:54	2024-10-09 10:12:55	10m1s	0.0063 USD	   


Job Tab

- Each job will have further details

Job Details - **seurat-ukn42y-r5.large** Executing

docker.io/mdibbiocore/seurat:latest

 Refresh

 Migrate

 Modify

More Actions ▾

ID	Host	User	Submitted Time	Wall Time	Cost
2fgs0kgw5wg8drn64gb5x	52.90.216.11 (2Core16GB/Spot)	workshop-user20	2024-10-09 10:06:04	26m12s (Queue Time: 0s)	0.0229 USD

Output
23

Instances (2)Storage Volumes (2)SettingsAttachmentsWaveWatcher

[Download Log Bundle](#)

▼ Instance ID: i-0a2c0a40a0dbc300f [Logs](#)

Zone: us-east-1b | Subnet: subnet-0ee170e07276518ad

Spot t2.large | 2 Core 8 GB

Address	Public IP	Status	Start Time	End Time	Cost
10.0.147.198:443	54.209.166.136	reclaimed	2024-10-09 10:06:06	2024-10-09 10:21:18	0.0067 USD

▼ Instance ID: i-059286c327b51ce3e [Logs](#)

Zone: us-east-1b | Subnet: subnet-0ee170e07276518ad

Spot r5.large | 2 Core 16 GB

Address	Public IP	Status	Start Time	End Time	Cost
10.0.155.6:443	52.90.216.11	normal	2024-10-09 10:17:25		0.0097 USD

Job Tab

- Each job will have further details
 - Instance cost

Job Details - seurat-ukn42y-r5.large

Executing

Refresh

Migrate

Modify

More Actions

[docker.io/mdiblbio/seurat:latest](#)

ID	Host	User	Submitted Time	Wall Time	Cost
2fgs0kgw5wg8drn64gb5x	52.90.216.11 (2Core16GB/Spot)	workshop-user20	2024-10-09 10:06:04	26m12s (Queue Time: 0s)	0.0229 USD

Output

23

Instances (2)

Storage Volumes (2)

Settings

Attachments

WaveWatcher

Download Log Bundle

Instance ID: i-0a2c0a40a0dbc300f

Logs

Zone: us-east-1b | Subnet: subnet-0ee170e07276518ad

Spot

t2.large

2 Core 8 GB

Address	Public IP	Status	Start Time	End Time	Cost
10.0.147.198:443	54.209.166.136	reclaimed	2024-10-09 10:06:06	2024-10-09 10:21:18	0.0067 USD

Job Tab

- Each job will have further details
 - Private and public IP

Job Details - seurat-ukn42y-r5.large Executing

[docker.io/mdiblbicore/seurat:latest](#)

Refresh Migrate Modify More Actions ▾

ID	Host	User	Submitted Time	Wall Time	Cost
2fgs0kgw5wg8drn64gb5x	52.90.216.11 (2Core16GB/Spot)	workshop-user20	2024-10-09 10:06:04	26m12s (Queue Time: 0s)	0.0229 USD

Output

23

[Instances \(2\)](#) [Storage Volumes \(2\)](#) [Settings](#) [Attachments](#) [WaveWatcher](#)

[Download Log Bundle](#)

▼ Instance ID: i-0a2c0a40a0dbc300f [Logs](#)

Zone: us-east-1b | Subnet: subnet-0ee170e07276518ad

Spot t2.large | 2 Core 8 GB

Address	Public IP	Status	Start Time	End Time	Cost
10.0.147.198:443	54.209.166.136	reclaimed	2024-10-09 10:06:06	2024-10-09 10:21:18	0.0067 USD

Job Tab

- Each job will have further details
 - WaveWatcher tab shows stats on resource usage



Job Tab

- Each job will have further details
 - WaveWatcher tab shows stats on resource usage



Workflow Tab

- Workflow view allows the user to examine all the tasks grouped in a single workflow
 - Feature applies to Nextflow pipelines

Workflows

admin

Search workflow name

User

Status

Refresh

Name	Pipeline	User	Submitted	Updated	Status				Wall Time	Cost
					Pending	Running	Success	Failed		
suspicious-lavoisier	nf-core-atacseq	ashley_tung	2024-10-01 13:37:54	2024-10-01 13:48:14	0	0	27	1	10m20s	\$ 0.0406
awesome-waddington	nf-core-atacseq	ashley_tung	2024-10-01 13:18:14	2024-10-01 13:30:22	0	0	24	1	12m8s	\$ 0.0384
astonishing-bassi	nf-core-atacseq	ashley_tung	2024-10-01 12:22:38	2024-10-01 12:50:21	0	0	116	0	27m43s	\$ 0.1400
fervent-mcclintock	nf-core-atacseq	ashley_tung	2024-10-01 11:51:31	2024-10-01 11:57:24	0	0	19	5	5m53s	\$ 0.0306

Workflow Tab

- Workflow view allows the user to examine all the tasks grouped in a single workflow
 - Feature applies to Nextflow pipelines
 - Includes a summary of the entire pipeline

Workflow Details - awesome-murdock

Refresh

Pipeline

mdibl-scscope

User

workshop-user20

Submitted

2024-10-09 07:01:44

Updated

2024-10-09 11:19:55

Wall Time

4h18m11s

CPU Time

110.45 Core Hours

Cost

\$ 2.5078

Jobs

Success

151

Running

3

Pending

0

Failed

2

Migrations

10

Cloud Instances

1

On-Demand

165

Spot

Jobs

Timeline

ID

Name

Process

Host

Status

Submitted Time

Wall Time

Cost

2zkoassqpbxq3l02d9j5n

seurat-ukn42y-r4.large

NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER

54.237.240.15 (2Core16GB/Spot)

Executing

2024-10-09 10:56:19

23m37s

0.0179 USD

6tfzkapn51iwjk0r3k12y

seurat-ukn42y-r6in.large

NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER

Completed

2024-10-09 10:39:49

30m56s

0.0170 USD

i2thjxtrga06h5wvp4mf

seurat-ukn42y-r5.large

NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER

FailToComplete

2024-10-09 10:23:34

32m41s

0.0289 USD

onycpdh7whcgtcjvdcg8t

seurat-ukn42y-t2.large

NFCORE-SCSCAPE-SCSCAPE-INTEGRATION

Completed

2024-10-09 10:12:59

10m30s

0.0067 USD

2fgs0kgw5wg8drn64gb5x

seurat-ukn42y-r5.large

NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER

FailToComplete

2024-10-09 10:06:04

33m43s

0.0300 USD

ei1vfg1neudadh2adl503

seurat-ukn42y-t2.large

NFCORE-SCSCAPE-SCSCAPE-PCA-MULT

Completed

2024-10-09 10:02:54

10m1s

0.0063 USD

Workflow Tab

- Workflow view allows the user to examine all the tasks grouped in a single workflow
 - Feature applies to Nextflow pipelines
 - Includes a summary of the entire pipeline

Workflow Details - awesome-murdock

Pipeline

mdibl-scscap

User

workshop-user20

Submitted

2024-10-09 07:01:44

Updated

2024-10-09 11:19:55

Wall Time

4h18m11s

CPU Time

110.45 Core Hours

Cost

\$ 2.5078

Jobs

Success

151

Running

3

Pending

0

Failed

2

Migrations

10

Cloud Instances

1

On-Demand

165

Spot

Jobs

Timeline

ID	Name	Process	Host	Status	Submitted Time	Wall Time	Cost
2zkoassqpbxq3l02d9j5n	seurat-ukn42y-r4.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER	54.237.240.15 (2Core16GB/Spot)	Executing	2024-10-09 10:56:19	23m37s	0.0179 USD
6tfzkapn51iwjk0r3k12y	seurat-ukn42y-r6in.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER		Completed	2024-10-09 10:39:49	30m56s	0.0170 USD
i2thjxtrga06h5wvp4mf	seurat-ukn42y-r5.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER		FailToComplete	2024-10-09 10:23:34	32m41s	0.0289 USD
onycpdh7whcgtcjvdcg8t	seurat-ukn42y-t2.large	NFCORE-SCSCAPE-SCSCAPE-INTEGRATION		Completed	2024-10-09 10:12:59	10m30s	0.0067 USD
2fgs0kgw5wg8drn64gb5x	seurat-ukn42y-r5.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER		FailToComplete	2024-10-09 10:06:04	33m43s	0.0300 USD
ei1vfg1neudadh2adl503	seurat-ukn42y-t2.large	NFCORE-SCSCAPE-SCSCAPE-PCA-MULT		Completed	2024-10-09 10:02:54	10m1s	0.0063 USD

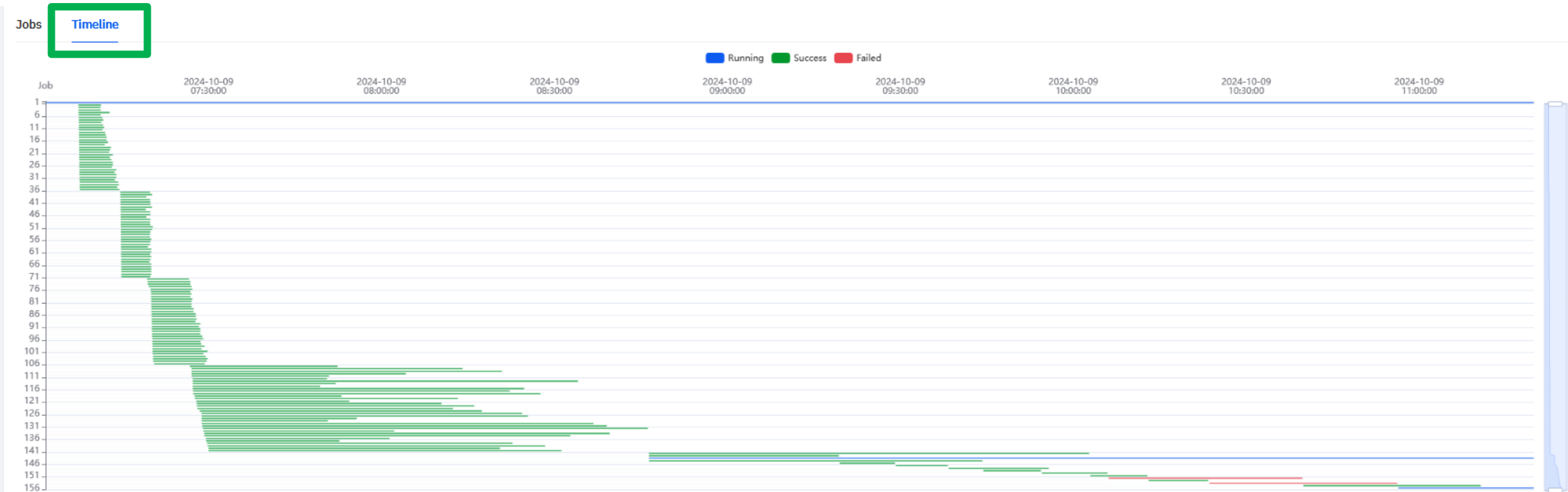
Workflow Tab

- Workflow view allows the user to examine all the tasks grouped in a single workflow
 - Feature applies to Nextflow pipelines
 - Includes a summary of the entire pipeline

Workflow Details - awesome-murdock										Refresh
Pipeline	User	Submitted	Updated	Wall Time	CPU Time	Cost				
mdibl-scscap	workshop-user20	2024-10-09 07:01:44	2024-10-09 11:19:55	4h18m11s	110.45 Core Hours	\$ 2.5078				
Jobs	Success 151	Running 3	Pending 0	Failed 2	Migrations 10	Cloud Instances 1 On-Demand	165 Spot			
ID	Name	Process	Host	Status	Submitted Time	Wall Time	Cost			
2zkoassqpbxq3l02d9j5n	seurat-ukn42y-r4.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER	54.237.240.15 (2Core16GB/Spot)	Executing	2024-10-09 10:56:19	23m37s	0.0179 USD			
6tfzkapn51iwjk0r3k12y	seurat-ukn42y-r6in.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER		Completed	2024-10-09 10:39:49	30m56s	0.0170 USD			
i2thjxtrga06h5wvp4mf	seurat-ukn42y-r5.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER		FailToComplete	2024-10-09 10:23:34	32m41s	0.0289 USD			
onycpdh7whcgtcjvdcg8t	seurat-ukn42y-t2.large	NFCORE-SCSCAPE-SCSCAPE-INTEGRATION		Completed	2024-10-09 10:12:59	10m30s	0.0067 USD			
2fgs0kgw5wg8drn64gb5x	seurat-ukn42y-r5.large	NFCORE-SCSCAPE-SCSCAPE-FIND-NN-CLUSTER		FailToComplete	2024-10-09 10:06:04	33m43s	0.0300 USD			
ei1vfg1neudadh2adl503	seurat-ukn42y-t2.large	NFCORE-SCSCAPE-SCSCAPE-PCA-MULT		Completed	2024-10-09 10:02:54	10m1s	0.0063 USD			

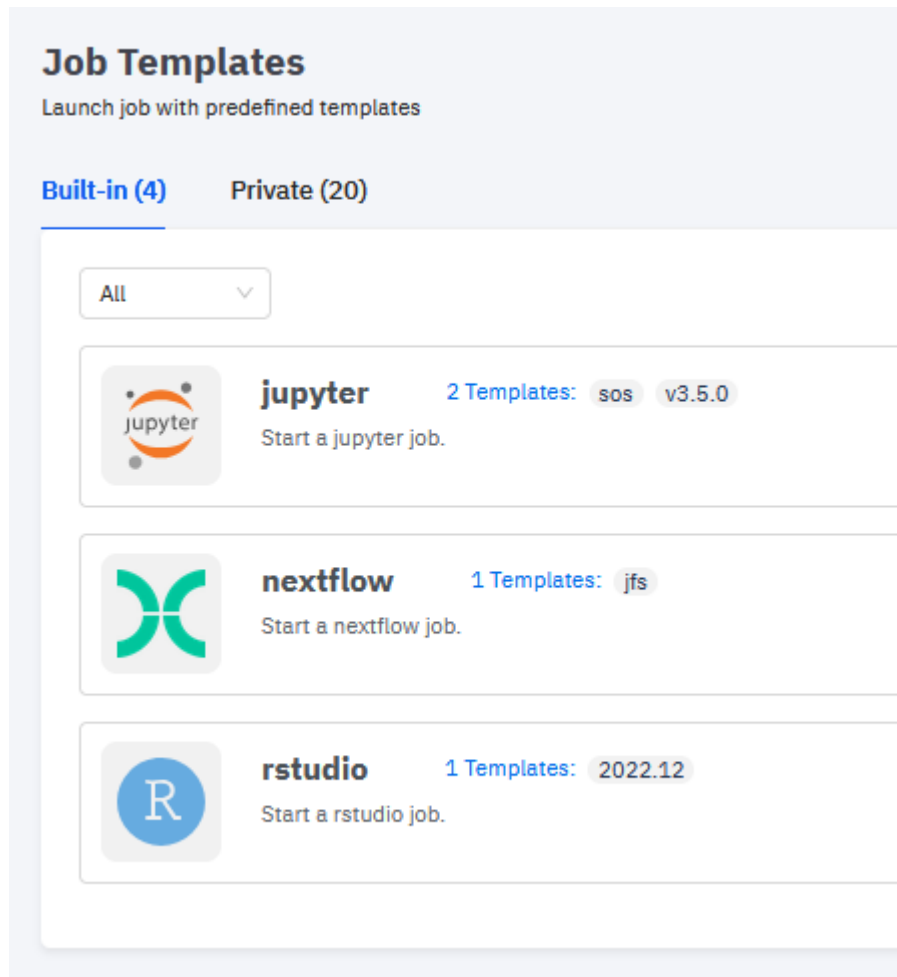
Workflow Tab

- Workflow view allows the user to examine all the tasks grouped in a single workflow
 - Feature applies to Nextflow pipelines
 - Includes a summary of the entire pipeline



Submitting Pipelines – Job Templates

- Templates simplify the procedure for submitting jobs by having preconfigured parameters
 - OpCenter contains a library of built-in templates



Submitting Pipelines – Job Templates

- Build private templates by saving previously-submitted job for future resubmissions

Job Templates
Launch job with predefined templates

Build-in (4)

Private (20)

Search Templates

Status

User

Refresh

Tag	Status	Owner	Usage Count	Last Updated	Last Used	Actions
– user01-scrnaseq 1						
scrnaseq	Available	admin	0	2024-10-08 08:15:31		
– user02-scrnaseq 1						
scrnaseq	Available	admin	0	2024-10-08 08:15:57		
– user03-scrnaseq 1						
scrnaseq	Available	admin	0	2024-10-08 08:16:19		

Submitting Pipelines – Job Templates

Submit Job

Apply a Template

Start from Scratch

Built-in 4Private 20

Search Templates

user01-scrnaseq

scrnaseq

user02-scrnaseq

scrnaseq

user03-scrnaseq

scrnaseq

user04-scrnaseq

scrnaseq

user05-scrnaseq

scrnaseq

user06-scrnaseq

scrnaseq

user07-scrnaseq

scrnaseq

user08-scrnaseq

scrnaseq

user09-scrnaseq

scrnaseq

user01-scrnaseq

scrnaseq

Parameters

Image:docker.io/memverge/juiceflow

CPU:4

Memory:16

Environment Variables:BUCKET=https://workshop-work.s3.us-east-1.amazonaws.com

Data Volume:[size=60]:/mnt/jfs_cache

Directory Mappings:/mnt/jfs:/mnt/jfs

Host Init:#!/bin/bash ## This script will be submitted as a containerInit hook script and do following things: ## - Setup a temporary JuiceFS filesystem as the working directory. # - Install nextflow software and its dependencies on the working directory. # - Generate the nextflow configuration file for MMCloud. # - Setup the nextflow working directory and environment. # # All the needed parameters will be passed in as environment variables. ## BUCKET: The bucket for JuiceFS storage. # AWS_BUCKET_ACCESS_KEY: The bucket access key for JuiceFS. (Optional) # AWS_BUCKET_SECRET_KEY: The bucket secret key for JuiceFS. (Optional) # JFS_CACHE_DIR: The cache directory for JuiceFS. (default is /mnt/jfs_cache) # JFS_MOUNT_POINT: The mount point for JuiceFS. (default is /mnt/jfs) #

Additional Options

Job Name

Job Name

Instance Type

Instance Type

vCPU (min)

Core

Memory (min)

GB

Security Group

Security group ID or Network tag in GCP

+ Add

Gateway ID

Target Port

Environment Variables

Key = Value

+ Add

Submitting Pipelines – Job Templates

Submit Job

Apply a Template

Start from Scratch

Built-in 4 Private 20

Search Templates

user01-scrnaseq
scrnaseq

user02-scrnaseq
scrnaseq

user03-scrnaseq
scrnaseq

user04-scrnaseq
scrnaseq

user05-scrnaseq
scrnaseq

user06-scrnaseq
scrnaseq

user07-scrnaseq
scrnaseq

user08-scrnaseq
scrnaseq

user09-scrnaseq
scrnaseq

user01-scrnaseq scrnaseq

Parameters

Image:

docker.io/memverge/juiceflow

CPU:

4

Memory:

16

Environment Variables:

BUCKET=https://workshop-work.s3.us-east-1.amazonaws.com

Data Volume:

[size=60]:/mnt/jfs_cache

Directory Mappings:

/mnt/jfs:/mnt/jfs

Host Init:

```
#!/bin/bash ## This script will be submitted as a containerInit hook script and do following things: ## - Setup a temporary JuiceFS filesystem as the working directory. # - Install nextflow software and its dependencies on the working directory. # - Generate the nextflow configuration file for MMCloud. # - Setup the nextflow working directory and environment. # # All the needed parameters will be passed in as environment variables. ## BUCKET: The bucket for JuiceFS storage. # AWS_BUCKET_ACCESS_KEY: The bucket access key for JuiceFS. (Optional) # AWS_BUCKET_SECRET_KEY: The bucket secret key for JuiceFS. (Optional) # JFS_CACHE_DIR: The cache directory for JuiceFS. (default is /mnt/jfs_cache) # JFS_MOUNT_POINT: The mount point for JuiceFS. (default is /mnt/jfs) #
```

Additional Options

Job Name

Job Name

Instance Type

Instance Type

vCPU (min)

Core

Memory (min)

GB

Security Group

Security group ID or Network tag in GCP

+ Add

Gateway ID

Target Port

Environment Variables

Key = Value

+ Add

Submitting Pipelines – Job Templates

Submit Job

Apply a Template

Start from Scratch

Load a Template

private::user01-scrnaseq:scrnaseq

Basic

WaveRider

Environment Variables

Network

Misc.

Job Content

* Image

docker.io/memverge/juiceflow

Name

scrnaseq

Job Script

Enter URL

Input Content

Load File...

#!/bin/bash

---- User Configuration Section ----

Instance

Instance Preferences

Instance Type

* vCPU (min)

4

vCPU (max)

* Memory (min GB)

16

Memory (max GB)

GPU

VM Policy

VM instance creation policy

Spot First

Spot Only

On-Demand Only

Retry Limit

5

Retry Interval


300

s

Price Limit

Generated command line

float submit --template private::user01-scrnaseq:scrnaseq --overwriteTemplate ""



Submitting Pipelines – CLI

- Download the CLI to your terminal
- All job submission on the GUI can be done with CLI equivalent

```
[ashleyt@ashleyt2 ~]$ float -h
"float" is a tool supplied by MemVerge to control float service

Usage:
  float [flags]
  float [command]

Job Management
  list      show all jobs
  show      show information about a job or host
  submit    submit a new job
  cancel    cancel one job
  migrate   migrate workload
  modify    modify job attributes
  suspend   suspend a running job
  resume    resume a suspended/stopped job
  snapshot  manage job snapshots
  rerun     rerun a job
  promote   promote a job to a higher priority
  hosts     show information about hosts
```

```
[ashleyt@ashleyt2 ~]$ float submit --hostInit ~/memverge_solutions/juiceflow/aws/transient_JFS_AWS.sh \
> -c 2 -m 4 \
> -i docker.io/memverge/juiceflow \
> --dataVolume [size=60]:/mnt/jfs_cache \
> --migratePolicy [disable=true] \
> --vmPolicy [onDemand=true] \
```

Questions or Comments?