Confluence Documentation

AWS Architecture

November 16, 2021

Nikki Tebaldi (ntebaldi@umass.edu)

TABLE OF CONTENTS
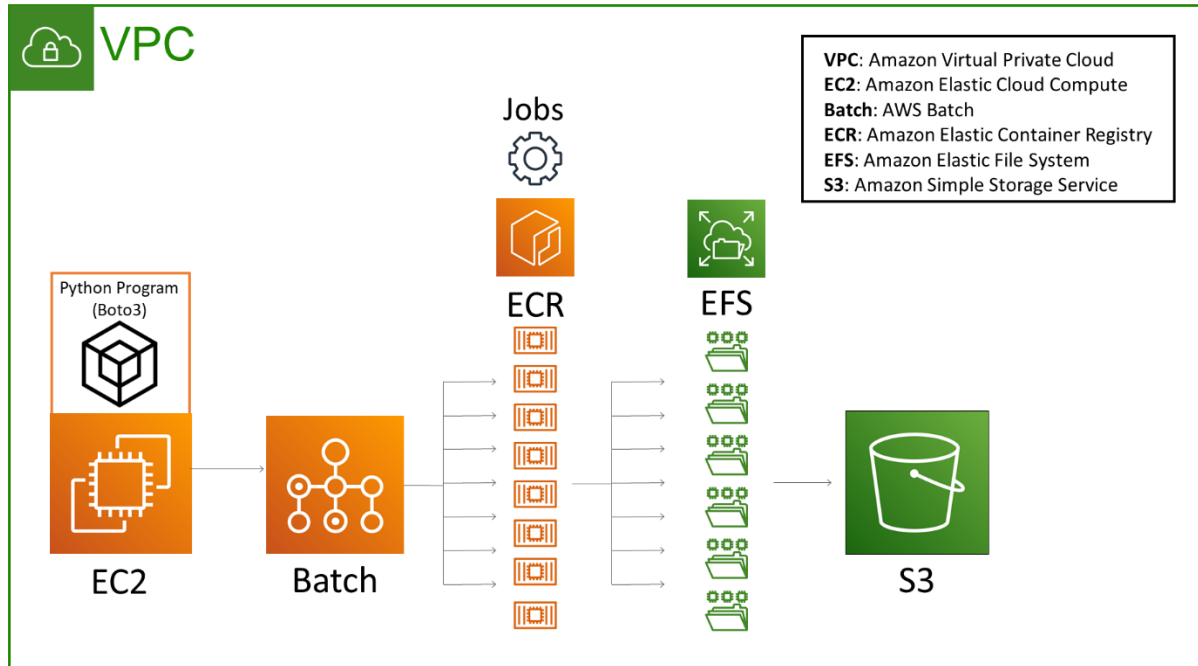
TABLE OF FIGURES

AWS ARCHITECTURE OVERVIEW

Figure 1 AWS Architecture

Confluence as a platform operates within several AWS architectural components. AWS Batch is used to submit and execute containerized jobs for each module in a stage. The containers are hosted in the AWS Elastic Container Registry (ECR). The Batch infrastructure is made up of several compute environments that are backed by AWS Fargate. A job queue exists for each stage and job definitions are created for each module. Intermediate storage is provided through Amazon Elastic File System (EFS) and is provided for each stage of processing. The results of an execution of the Confluence workflow and prior data information is stored in an Amazon Simple Storage Service (S3) bucket.

VPC

Name	Description	Region	Subnets in Availability Zones
confluence	Virtual Private Cloud network to execute Batch jobs in	us-west-2 (Oregon)	us-west-2a (private), us-west-2b (private), us-west-2c (private and public)

IAM Resources

Name	Description	Component Notes
ecsTaskExecutionRole	Batch Fargate Role that allows ECS task execution	Grants ECS container and Fargate agents to make AWS API calls
BatchEFSJobRole	Batch Fargate Role that allows job access to EFS	Allows ECS container access to EFS

Compute Environments

Name	Description	Max vCPUs
data	Managed, Fargate compute environment	1000
diagnostics	Managed, Fargate compute environment	1000
flpe	Managed, Fargate compute environment	1000
moi	Managed, Fargate compute environment	1000
offline	Managed, Fargate compute environment	1000
validation	Managed, Fargate compute environment	1000

Job Queues

Name	Compute environment
input	data
flpe	flpe
moi	moi
priors	data
prediagnostics	diagnostics
postdiagnostics-flpe	diagnostics
postdiagnostics-moi	diagnostics
offline	offline
output	data
validation	validation

JOB DEFINITIONS

Name	Memory	vCPU	EFS Volume: Mount Point
geobam	8192	4	**input**: /mnt/data/input **flpe**: /mnt/data/output
hivdi	4096	2	**input**: /mnt/data/input **flpe**: /mnt/data/output
metroman	8192	4	**input**: /mnt/data/input **flpe**: /mnt/data/output
momma	0.25	512	**input**: /mnt/data/input **flpe**: /mnt/data/output
sad	2048	1	**input**: /mnt/data/input **flpe**: /mnt/data/output
sic4dvar	2048	1	**input**: /mnt/data/input **flpe**: /mnt/data/output
input	8192	1	**input**: /mnt/data
moi	2048	1	**input**: /mnt/data/input **flpe**: /mnt/data/flpe **moi**: /mnt/data/output
offline	2048	1	**input**: /mnt/data/input **offline**: /mnt/data/output
output	2048	1	**input**: /mnt/data/input **flpe**: /mnt/data/output **moi**: /mnt/data/moi **diagnostics**: /mnt/data/diagnostics **offline**: /mnt/data/offline **validation**: /mnt/data/validation **output**: /mnt/data/output
postdiagnostic_flpe	1024	0.5	**input**: /mnt/data/input **flpe**: /mnt/data/flpe **diagnostics/reach**: /mnt/data/output
postdiagnostics_moi	1024	0.5	**input**: /mnt/data/input **flpe**: /mnt/data/flpe **moi**: /mnt/data/moi

			diagnostics/basin: /mnt/data/output
prediagnostics	1024	0.5	**input**: /mnt/data
priors	8192	1	**input**: /mnt/data
validation	2048	1	**input**: /mnt/data/input **offline**: /mnt/data/offline **validation**: /mnt/data/output

ECR CONTAINERS

Name	URI	Visibility	Dockerfile Link	Size (MB)
hivdi	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/hivdi	private	HiVDI	1542.66
input	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/input	private	Input	339.07
metroman	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/metroman	private	MetroMan	185.91
moi	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/moi	private	MOI	403.75
momma	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/momma	private	MOMMA	440.61
offline	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/offline	private	Offline	93.27
output	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/output	private	Output	115.93
postd-flpe	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/postd-flpe	private	Postdiagnostics FLPE	403.75
postd-moi	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/postd-moi	private	Postdiagnostics MOI	402.71
prediagnostics	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/prediagnostics	private	Prediagnostics	344.59
priors	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/priors	private	Priors	1170.18
sic4dvar	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/sic4dvar	private	n/a	273.96
validation	xxxxxxxxxxxx.dkr.ecr.us-west-2.amazonaws.com/validation	private	Validation	176.94
geobam	public.ecr.aws/s8t4s7g5/geobam	public	geoBAM	590.51
sad	public.ecr.aws/s8t4s7g5/sad	public	SAD	314.47

EFS

Name	ID	Contents
input	fs-60e5fd67	basin.json continent.json reaches.json sets.json gage/ sos/ sword/ swot/
flpe	fs-b6f6eeb1	hivdi/ geobam/ metroman/ momma/ sad/ sic4dvar/
moi	fs-89c6048f	
diagnostics	fs-92ea2894	basin/ reach/
offline	fs-39f13f3f	
validation	fs-b55580b3	figs/ stats/
output	fs-ea26ebec	version/
logs	fs-1d71a91b	sic4dvar/

S3 BUCKETS

Name	URI	Region	Contents
confluence-sos	s3://confluence-sos	US West (Oregon) us-west-2	constrained/ figs/ unconstrained/
confluence-swot	s3://confluence-swot	US West (Oregon) us-west-2	pass249/ pass264/ pass527/

LIST OF ACRONYMS

AWS: Amazon Web Services

EC2: Amazon Elastic Compute Cloud

ECR: Amazon Elastic Container Registry

EFS: Amazon Elastic File System

FLPE: Flow Law Parameter Estimation

IAM: AWS Identity and Access Management (IAM)

JPL: Jet Propulsion Laboratory

PO.DAAC: Physical Oceanography Distributed Active Archive Center

S3: Amazon Simple Storage Service

SDS: Science Data System

SoS: SWORD of Science

SWORD: SWOT River Database

SWOT: Surface Water and Ocean Topography

VPC: Amazon Virtual Private Cloud