## **AWS Cloud for Atmospheric Scientists**

Fei Yao

School of GeoSciences

The University of Edinburgh

https://www.geos.ed.ac.uk/homes/s1855106/



## Contents

- What is Cloud Computing?
- What is AWS Cloud and Why?
- Access & Use AWS Cloud with a Joint Account
  - Basic usage
  - Advanced topics
- Summary & Prospects

## Contents

- What is Cloud Computing?
- What is AWS Cloud and Why?
- Access & Use AWS Cloud with a Joint Account
  - Basic usage
  - Advanced topics
- Summary & Prospects

## What is Cloud Computing?

Cloud Computing is on-demand delivery of IT resources and applications via the Internet with pay-as-you-go pricing.

- Usually costs, but always cost-effective
- Elastic and scalable (dynamically adjust hardware as needed)
- Replication, replication, replication!
- Globally available in seconds



### Contents

- What is Cloud Computing?
- What is AWS Cloud and Why?
- Access & Use AWS Cloud with a Joint Account
  - Basic usage
  - Advanced topics
- Summary & Prospects



### Cloud Computing with Amazon Web Services



Trade capital expense for variable expense



Increase speed and agility



Benefit from massive economies of scale



Stop spending money on running and maintaining data centers



Stop guessing capacity

Go global in minutes



All services

Cloud Computing with Amazon Web Services





### Contents

- What is Cloud Computing?
- What is AWS Cloud and Why?
- Access & Use AWS Cloud with a Joint Account
   Basic usage
  - Advanced topics
- Summary & Prospects

feiyao

My Account

My Organization

Sign In to the Console as root user





### Manipulate S3 in Graphical Console



Amazon Machine Image specifies the software side of an EC2 instance (computer). By replicating it, you avoid installing operating system, software, libraries, etc.

		-		
aws	Services 🗸	Resource Groups 🗸	<b>A</b>	¢
EC2 Dashboard	A R	esources		
Events	Yo	u are using the following Amazor	EC2 resources in the US East (N.	Virginia) region:
Tags		0 Running Instances	0	Elastic IPs
Reports		0 Dedicated Hosts	0	Snapshots
Limits		0 Volumes	0	Load Balancers
INSTANCES		1 Key Pairs	2	Security Groups
Instances		0 Placement Groups		
Launch Templates				
Spot Requests		Learn more about the latest in A	AWS Compute from AWS re:Invent	by viewing the EC2 V
Reserved Instances				
Dedicated Hosts	C	reate Instance		
Scheduled Instances	з То	start using Amazon EC2 you will	want to launch a virtual server, kno	own as an Amazon EC
Capacity Reservation	ns	aunch Instance 🔻		
IMAGES				
AMIs	Not	e: Your instances will launch in the US	S East (N. Virginia) region	

 Search for ami-06f4d4afd350f6e4c from Public images in the US East (N. Virginia) region

 that's the system with both the classic and the High-Performance versions of GEOS-Chem installed.

Click

aws	Services 🗸	Resource Gro	oups 🗸 🔥		$\bigtriangleup$	Administrator @	6198-8428-23	✓ N. Virginia ✓
EC2 Dashboard		aunch Actions	*					
Events								
Tags		Public images 👻	<b>Q</b> search : ami-06	6f4d4afd350f6e4c 💿 🛛 Add filt	ter			
Reports		Name	AMI Name	AMI ID	Source	Owner	Visibility	Status
Limits								
			GEOSChem	ami-06f4d4afd350f6e4c	470649987867/	470649987867	Public	available
INSTANCES								

Step 3. Launch from that AMI and choose an Instance Type

- Instance Type specifies hardware side of an EC2 instance, mostly about CPUs.
- For longer-term, higher-resolution runs, consider bigger ones like c5.9xlarge and c5.18xlarge.

Cur	<b>irrently selected:</b> r5.large (10 ECUs, 2 vCPUs, 2.5 GHz, Intel Xeon Platinum 8175, 16 GiB memory, EBS only)													
	Family	Туре 🗸	vCPUs (i) 👻	Memory (GiB) 👻	Instance Storage (GB) (i)	EBS-Optimized Available (i)	Network Performance	IPv6 Support						
	Memory optimized	r5.large	2	16	EBS only	Yes	Up to 10 Gigabit	Yes						
71			har and a second a					har when he has a start whe has a start when he has a start when he has a start when h						

Click

#### Access & Use AWS Cloud with a joint account – Basic Usage – EC2 Step 4. Configure Instance Details Number of instances Launch int • You can launch several instances per time Purchasing option Request Spot instances Current price Availability Zone Current price Spot instances can always save ~70% money compared to us-east-1a \$0.0356 us-east-1b \$0.0356 on-demand ones. \$0.0356 us-east-1c

					T	
<ul> <li>Choose</li> </ul>	e this role to enable following a	awscli on EC2		us-east-1f	\$0.0356	
Persistent request (j)	Persistent request	$\neg$ $\land$	Maximum price 🧻	\$ e.g. 0.045 = 4.5	5 cents/hour (Optio	nal)
Interruption behavior (j)	Hibernate	•	Persistent request 🧃	Persistent requ	uest	
		M role 🧃	full_S3_access_from_	EC2		•
						-~~

us-east-1d

\$0.0356

### Step 5. Add Storage

- The volume attached to EC2 instance is called Elastic Block Storage (EBS) and can be used by EC2 instance directly. S3 is independent of any EC2 instances. AWS provides a series of commands to help transfer data between S3 and EC2 very efficiently (>100MiB/s).
- You need to use up/down arrow key to adjust volume size attached to the EC2 instance.



### Step 6. Configure Security Group

 "Security group" controls what IPs are allowed to access your server. Use an pre-defined one allowing all IPs to access is generally fine because we also need to have the EC2 Key Pair in order to access to a specific server.

Assign a security group: OCreate a new security group

Select an existing security group

Security Group ID	Name	Description	Actions
sg-66551026	default	default VPC security group	Copy to new
sg-07edc093f7e83ac94	launch-wizard-1	launch-wizard-1 created 2019-01-14T15:37:22.887+00:00	Copy to new

Inbound rules for sg-07edc093f7e83ac94 (Selected security groups: sg-07edc093f7e83ac94)

	Type (j)	Protocol (i)	Port Range (i)	Source (j)	Description (j)
SSH TCP 22 0.0.0/0	SSH	ТСР	22	0.0.0/0	

Cancel Previous Review and Launc

Step 7. Review & Launch

EC2 Key Pair is equivalent to the password you enter to ssh to your local server, but much more secure.

Select an existing key pair or create a new key pair

×

A key pair consists of a **public key** that AWS stores, and a **private key file** that you store. Together, they allow you to connect to your instance securely. For Windows AMIs, the private key file is required to obtain the password used to log into your instance. For Linux AMIs, the private key file allows you to securely SSH into your instance.

Note: The selected key pair will be added to the set of keys authorized for this instance. Learn more about removing existing key pairs from a public AMI.

Choose an existing key pair	۳
Select a key pair	
fei-aws-keypair	۳

I acknowledge that I have access to the selected private key file (fei-aws-keypair.pem), and that without this file, I won't be able to log into my instance.



### Step 8. Connect

Launch Instance Connect Actions V Q Filter by tags and attributes or search by keyword 0 < 1 to 1 of 1 > > Instance Type v Availability Zone v Instance State v Status Checks v Instance ID Alarm Status Public DNS (IPv4) IPv4 Public IP IPv6 IPs Key Name Monitoring Launch Ti i-034e1ea98c4c62009 r5.large us-east-1a running 🛣 Initializing None ec2-3-88-233-21.comp. 3.88.233.21 feivao-ams-ke. disabled February 4 \$ ssh -i fei-aws-keypair.pem ubuntu@ec2-34-205-50-14.compute-1.amazonaws.com Welcome to Ubuntu 18.04.1 LTS (GNU/Linux 4.15.0-1021-aws x86\_64) Connect To Your Instance × \* Documentation: https://help.ubuntu.com \* Management: https://landscape.canonical.com Recommend \* Support: https://ubuntu.com/advantage A Java SSH Client directly from my browser (Java required) 1 System information as of Mon Feb 11 14:00:08 UTC 2019 Git Bash To access your instance: 267 System load: 0.18 Processes: Usage of /: 35.2% of 387.70GB Users logged in: 0 Open an SSH client. (find out how to connect using PuTTY) IP address for ens5: 172.31.9.6 Memory usage: 1% 2. Locate your private key file (fei-aws-keypair.pem). The wizard automatically detects the key you Swap usage: 0% root -> ubuntu used to launch the instance. Your key must not be publicly viewable for SSH to work. Use this command if needed: Get cloud support with Ubuntu Advantage Cloud Guest: http://www.ubuntu.com/business/services/cloud chmod 400 fei-aws-keypair.pem Something 121 packages can be updated. 4. Connect to your instance using its Public DNS: 0 updates are security updates. already here ec2-34-205-50-14.compute-1.amazonaws.com \*\*\* System restart required \*\*\* 船船 Example: Last login: Sat Dec 15 20:01:59 2018 from 65.112.8.207 ancu@าp-172-31-9-6:~\$ ไร 😒 -i "fei-aws-keypair.pem" root@ec2-34-205-50-14.compute-1.amazonaws.com ∬ExtData gchp.ubuntu.en∨ miniconda tutorial ubuntu@ip-172 Please note that in most cases the username above will be correct, however please ensure that you read your AMI usage instructions to ensure that the AMI owner has not changed the 

### Manipulate S3 with AWSCLI in EC2

- Create bucket
   S aws s3 mb s3://bucket-name
- See bucket
- Upload/download files/directories
- Remove files/directories

- \$ aws s3 ls s3://bucket-name
- \$ aws s3 cp [--recursive] source target
- \$ aws s3 rm [--recursive] target
- Source & target can be any combination of S3, EC2, Local and S3.
- These commands together with aws ec2 ... are also applicable in local machines provided you properly install and configure awscli.



### Get source code and checkout model versions

- Make a separate folder
- Enter that folder
- Download GEOS-Chem source code & Unit tester
- Do version control if needed

- \$ mkdir ~/GC
- \$ cd ~/GC
- \$ git clone https://github.com/geoschem/geos-chem Code.GC
- § git clone https://github.com/geoschem/geos-chem-unittest.git UT
- \$ cd Code.GC (UT)
- \$ git log --tags --simplify-by-decoration --pretty="format:%ci %d"
- \$ git checkout 12.1.0; git branch
- git checkout master

### Configure unit tester and generate run directory



make realclean

### Configure Makefile and Compile the source code

- Change source code path
- Compile
- # Remove all pre-compiled ones
- # -j4 denotes 4 jobs simultaneously
- # mpbuild denotes using multiple processors
- # NC\_DIAG denotes exporting netCDF diagnostics
- # BPCH\_DIAG denotes exporting BPCH diagnostics
- # TIMERS?

make -j4 mpbuild NC\_DIAG=y BPCH\_DIAG=n TIMERS=1

endif

### Configure input.geos and HISTORY.rc

- Change simulating period
- Create a directory named OutputDir under the run directory

%%% SIMULATION MENU %%% : Start YYYYMMDD, hhmmss End YYYYMMDD, hhmmss Run directory Root data directory Global offsets IO, JO : 0 0

: 20160701 000000 : 20160701 010000 : /home/ubuntu/ExtData

# EXPID allows you to specify the beginning of the file path corresponding to each diagnostic collection. For example:

EXPID: ./GEOSChem Will create netCDF files whose names begin "GEOSChem", in this run directory.

EXPID: ./OutputDir/GEOSChem Will create netCDF files whose names begin with "GEOSChem" in the OutputDir sub-folder of this run directory.

WEXPID: ./OutputDir/GEOSChem

### Configure input.geos and HISTORY.rc



Pull shared data from S3 to EC2

- s3://gcgrid is where GEOS-Chem shared data reside.
  - \$ for month in {07 08}
  - > do
  - > aws s3 cp --request-payer=requester --recursive \
  - > s3://gcgrid/GEOS\_2x2.5/GEOS\_FP/2016/\$month \
  - > ~/ExtData/GEOS\_2x2.5/GEOS\_FP/2016/\$month
  - > done

Obtain additional files

#### # GEOSFP 2x2.5 CN metfield

- \$ aws s3 cp --request-payer=requester --recursive s3://gcgrid/GEOS\_2x2.5/GEOS\_FP/2011/01/ ~/ExtData/GEOS\_2x2.5/GEOS\_FP/2011/01/
- # 2x2.5 restart file
- \$ aws s3 cp --request-payer=requester s3://gcgrid/GEOSCHEM\_RESTARTS/v2018-11/initial\_GEOSChem\_rst.2x25\_standard.nc ~/ExtData/GEOSCHEM\_RESTARTS/v2018-11/
- *# fix the softlink in run directory*
- \$ In -s ~/ExtData/GEOSCHEM\_RESTARTS/v2018-11/initial\_GEOSChem\_rst.2x25\_standard.nc ~/GC/geosfp\_2x25\_standard/GEOSChem.Restart.20160701\_0000z.nc4

*# run GEOS-Chem (To monitor running time, type time ./geos.mp)* 

\$ ./geos.mp

- Subject to: \*\*\*\*\*\*\*\*\*\*\* \*\*\*\*\*\*\*\*\*\* END 0 F GEOS--CHEM Instance type real 8m51.286s \*\*\*\*\*\*\*\*\*\*\* de END -- C H E M 0 F 49m27.746s time steps user Om55.446s sys 4m24.605s real "Warm-up" 49m20.661s user Om54.415s sys
- Denotes time-averaged value between 20160701-0000z to 20160701-0020z

ubuntu@ip-172-31	L-9-6:~/GC/g	geosfp_2x25	j_stand	dard\$ 11	l ./OutputDir/	
total 718020						
drwxrwxr-x 2 ubu	untu ubuntu	4096	Feb 11	1 16:29	-/	
drwxrwxr-x 3 ubu	untu ubuntu	4096	Feb 11	1 16:29	/	
/	untu ubuntu	25402406	Feb 11	1 16:28	GEOSChem.AerosolMass.	20160701_0000z.nc4
-rw-rw-r 1 ubu	untu ubuntu	25346481	Feb 11	1 16:29	GEOSChem.AerosolMass.	.20160701_0020z.nc4 ~~
👌 -rw-rw-r 1 ubι	untu ubuntu	25312359	Feb 11	1 16:29	GEOSChem.AerosolMass.	20160701_0040z.nc4
🚟-rw-rw-r 1 ubι	intu ubuntu	659173261	Feb 11	1 16:29	GEOSChem.Restart.2016	50701_0100z.nc4 🛛 🗧

Analyze model outputs using Jupyter

- # Re-connect using port-forwarding
- \$ ssh -i "your-key-name.pem" ubuntu@xxx.amazonaws.com -L 8999:localhost:8999
- # Activate geo python environment
- \$ source activate geo
- # Start a jupyter
- \$ jupyter notebook --NotebookApp.token=" --no-browser --port=8999 --notebook-dir ~/
- # Open the url and write analyzing codes
- \$ http://localhost:8999/
- *# Deactivate geo environment*
- \$ source deactivate

Save files to S3, terminate server, and start over whenever needed

- # Save files (most time run directories) to S3
- \$ aws s3 cp --recursive ~/GC s3://fei-geoschem-run-directory/GC
- # Terminate server (to do this in the graphical console is convenient enough)
- # Start an EC2 Instance again (steps have been described before)
- # Retrieve run directories from S3 back to EC2
- \$ aws s3 cp --recursive s3://fei-geoschem-run-directory/GC ~/GC
- # Restore execution permission for geos.mp
- \$ chmod u+x ~/GC/geosfp\_2x25\_standard/geos.mp
- # Obtain shared data and GEOSFP 2x2.5 CN metfield (refer to p.28, 29)

Satellite remotely sensed data processing

- Launch an EC2 instance
- Upload satellite data to S3 (bear in mind that some satellite data are already in S3)
- Pull satellite data from S3 to EC2
- Write Python or similar codes to process satellite data
- Export processed data to S3 and shut down EC2

### Contents

- What is Cloud Computing?
- What is AWS Cloud and Why?
- Access & Use AWS Cloud with a Joint Account
  - Basic usage
  - Advanced topics
- Summary & Prospects

### Simplify login process via scripts

• Create Access Key under My security credentials dashboard



### Simplify login process via scripts

- Install AWSCLI on your own computer (i.e. pip install awscli)
- Run aws configure

[s1855106@burn ~]\$ aws configure AWS Access Key ID [None]: AKIAICXBCQO AWS Secret Access Key [None]: 9pRSkCA Default region name [None]: us-east-1 Default output format [None]: json

• Configure files are stored in ~/.aws directory

- It's in the accessKeys file you downloaded
- us-east-1 is an alias to the "US East (N. Virginia)" region
- Specify configure output format as json

### Simplify login process via scripts

- Write a script and store it in a directory contained in \$PATH
- Run the script from terminal
- Export the AWSCLI path under Windows
- If require spot instance

```
#!/bin/bash
  # == often change ==
  TYPE=c5.4xlarge # EC2 instance type
 6 # == set it once and seldom change ==
 7 AMI=ami-06f4d4afd350f6e4c # AMI to launch from
 8 COUNT=1 # how many instances to launch
 9 IAM=full_S3_access_from_EC2 #EC2 IAM role name
10 EBS_SIZE=400 # root EBS volume size (GiB)
11 SG=sg-07edc093f7e83ac94 # security group ID
12 KEY=fei-aws-keypair # EC2 key pair name
14 #
    == almost never change; just leave it as is ==
  aws ec2 run-instances --image-id $AMI \
       --instance-type $TYPE \
16
       --count $COUNT \
       --iam-instance-profile Name=$IAM
       --block-device-mapping DeviceName="/dev/sda1", Ebs={VolumeSize=$EBS_SIZE
19
       --security-group-ids $SG \
       --kev-name $KEV
       --instance-market-options '{"MarketType":"spot"
```

### Keep program running after logging out using tmux

- Create a new session
- Split a panel
- Switch between panels
- Detach from a session
- List all sessions
- Attach to a session

- \$ tmux new -s session\_name
- Ctrl-b %/"
- Ctrl-b  $\leftarrow / \rightarrow / \uparrow / \downarrow$
- Ctrl-b d
- \$ tmux list
- \$ tmux attach -t session\_name

## What if I want to change Instance Type of a launched EC2 instance?

- On-demand instances support this kind of demand.
- Stop the instance -> Change its instance type
   -> start it again



Actions A

Get Windows Password

Launch More Like This

Instance State

Create Template From Instance

ility Zone 👻

1c

Instance State

terminated

Status

What if I need more volume size to a launched EC2 instance?



What if I need more volume size to a launched EC2 instance?

• Attach that volume to an EC2 instance

Actions A	Volume	i) vol-0ff168420cc18a5b2 i	n us-east-1c				
Modify Volume	Instance	i-06a0846e364bb108f		-east-1c			
Create Snapshot	Device	l) /dev/sdf					
Delete Volume		Linux Devices: /dev/sdf t	hrough /dev/sdp				
Attach Volume							
Detach Volume	te: Newer Lin	ix kernels may rename your devi	ces to /dev/xvdf through /dev/xv	dp internally, even wher	n the device name entered here (	and shown in the details) is /dev	v/sdf through /dev/sdp.
Force Detach Volu	me						
Change Auto-Enab	le						
Add/Edit Tags							Cancel Attach
All a station							1 - hours

What if I need more volume size to a launched EC2 instance?

- Make that volume usable
- # Show that volume to be mounted
- \$ Isblk
- # Create a file system for that volume
- \$ sudo mkfs -t ext4 /dev/nvme1n1
- # Mount that volume to a directory
- \$ mkdir new\_disk
- \$ sudo mount /dev/nvme1n1 new\_disk

- # Show the mounted volume
- \$ df --h
- # Change permission
- \$ sudo chown ubuntu new\_disk
- # Touch a new file
- \$ touch ~/new\_disk/new\_text

What if I need more volume size to a launched EC2 instance?

Share files between EC2 instances via EBS

- # Umount the mounted volume
- \$ sudo umount /dev/nvme1n1
- # Rmove new\_disk (Optional)
- \$ rmdir new\_disk
- # Detach that volume

Change Auto-Enable IO Setting

Actions A Modify Volume Create Snapshot Delete Volume Attach Volume Eorce Detach Volume

- # Attach to another instance
- # Mount that volume to a directory
- \$ mkdir new\_disk
- \$ sudo mount /dev/nvme1n1 new\_disk
- # Show contents in new\_disk
- \$ Is new\_disk/

What if I need more volume size to a launched EC2 instance?

- Sweep the battlefield
- # Save files to S3 if applicable
- # Umount the mounted volume
- \$ sudo umount /dev/nvme1n1
- # Rmove new\_disk (Optional)
- \$ rmdir new\_disk
- # Detach that volume
- # Delete that volume

#### Actions A

Modify Volume Create Snapshot Delete Volume Attach Volume Detach Volume Force Detach Volume Change Auto-Enable IO Setting Add/Edit Tags

What if I want to customize my own EC2 instance and save it in order to facilitate later use?

• Launch an instance from a clean AMI

- Install compliers, libraries, and etc. by apt or yum dependent on which system you launch
- Save that instance as your AMI

Launch Instance	Connect	Actions A										Δ,	Ð	<b>•</b> (	?
Q Filter by tags ar	nd attributes or search	Connect Get Windows Password									0	K ≺ 1 to 6	of 6	> >	
Name	<ul> <li>Instance ID</li> </ul>	Create Template From Insta Launch More Like This	ance pility Zone	✓ Instance State ✓ State	atus Checks 🔻	Alarm Status		Public DNS (IPv4)	IPv4 Public IP	- IPv6 IPs	- Key Name -	Monitoring	- I	aunch	Ti
	i-018417e9ae06	Instance State	-1a	🥚 terminated		None	<b>\</b> @		-	-	fei-aws-keypair	disabled	F	ebruar	y 1
	i-06a0846e364b	Instance Settings	-1c	🥚 terminated		None	<b>\</b> _		-	-	fei-aws-keypair	disabled	F	ebruar	y 1
	i-09d15d18fc02	Image	Create	Image	Checks	None	<b>\</b> @	ec2-3-85-243-92.comp	3.85.243.92	-	fei-aws-keypair	disabled	F	ebruar	у 1
	i-0a4ecc63738b	Networking	Bundle	Instance (instance store AMI)	) checks	None	<b>\</b> _	ec2-3-208-16-166.com	3.208.16.166	-	fei-aws-keypair	disabled	F	ebruar	y 1
	i-0f2e404dcb7b	CloudWatch Monitoring	- <b>1</b> a	🥚 terminated	_	None	<b>\</b> @		-	-	fei-aws-keypair	disabled	F	ebruar	y 1
	i-0f9ffc692acf14	a03 r5.large	us-east-1a	🥚 terminated		None	æ		-	-	fei-aws-keypair	disabled	F	ebruar	у 1

### Contents

- What is Cloud Computing?
- What is AWS Cloud and Why?
- Access & Use AWS Cloud with a Joint Account
  - Basic usage
  - Advanced topics
- Summary & Prospects

## **Summary & Prospects**

**GEOS-Chem research workflow on the AWS cloud** 



## Summary & Prospects

### High Performance GEOS-Chem

• Already on AWS deserving exploration.

### More flexible permission controls within group

• Will surely come in the near future with your effort.

### Prepayment

• e.g. Reserved Instance

## Materials mainly from:

### **GEOS-Chem on AWS cloud tutorial**

• <u>https://cloud-gc.readthedocs.io/en/latest/index.html</u>

### AWS Document

• <u>https://docs.aws.amazon.com/index.html#lang/en\_us</u>



### Thanks! Your Response is appreciated!

