





AWS Workshop: GNSS RO and Cloud Computing

Stephen Leroy¹, Amy McVey¹, Mark Leidner¹, Hailing Zhang², Eric Wang³ OPAC-7 / IROWG-9, Leibnitz, Austria September 13, 2022

¹Verisk Atmospheric and Environmental Research, Lexington, Massachusetts, USA ²COSMIC, University Corporation for Atmospheric Research, Boulder, Colorado, USA ³Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

Agenda

1. Motivation

• Cloud computing with AWS

2. Getting started

- Logging in to EC2 instance (AWS virtual machine)
- · Jupyter notebooks

3. GNSS RO data in the AWS Open Data Registry

- Open Data Registry and Github support
- AWS command line interface

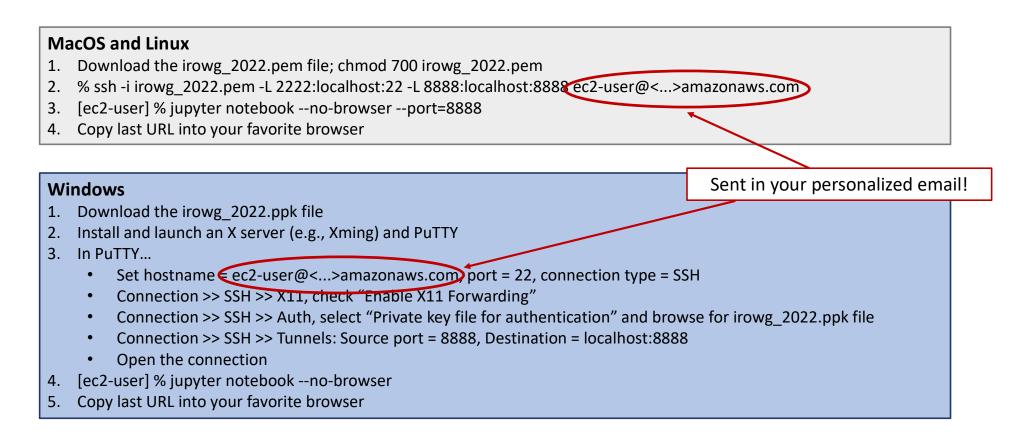
4. Data assimilation with JEDI-FV3

5. AWS Universe applied to GNSS RO data

Motivation

- Cloud computing with Amazon Web Services
- Cloud computing with Amazon Web Services using GNSS RO data
- Cloud computing with Amazon Web Services using GNSS RO data with Jupyter notebook and brief lectures
- Cloud computing with Amazon Web Services using GNSS RO data with Jupyter notebook and brief lectures without cost

Getting started with your virtual machine/EC2 instance



Jupyter notebooks

- Open the jupyter_intro.ipynb notebook and try a few python commands
- Open workshop.ipynb jupyter notebook
- Feel free to download any material directly to your laptop using "File >> Download". We will post the workshop.ipynb on the website after this workshop.

What is Amazon Web Services (AWS)?

- Cloud potential: Limitless compute and storage
- Cloud caveat #1: Set your own limits, or it will cost you!
- Cloud caveat #2: Develop a test plan, increase scale methodically.
- Services: Packaged resources with a specific purpose (over 200)

ll services		
Services by category		
📋 Compute	😣 Quantum Technologies	O Security, Identity, & Compliance
EC2	Amazon Braket	IAM
Lightsail Lambda	Management & Governance	Resource Access Manager Cognito
Batch Elastic Beanstalk	AWS Organizations CloudWatch	Secrets Manager GuardDuty
Serverless Application Repository AWS Outposts	AWS Auto Scaling CloudFormation	Inspector Amazon Macie
EC2 Image Builder AWS App Runner	Config OpsWorks	AWS IAM Identity Center (successor to AWS Single Sign-On)
🖮 Containers	Service Catalog Systems Manager	Certificate Manager Key Management Service
Elastic Container Registry	AWS AppConfig	CloudHSM
Elastic Container Service	Trusted Advisor	Directory Service WAF & Shield
Elastic Kubernetes Service Red Hat OpenShift Service on AWS	Control Tower AWS License Manager	AWS Firewall Manager
B Storage	AWS Well-Architected Tool AWS Health Dashboard	Artifact Security Hub
S3	AWS Chatbot	Detective
EFS	Launch Wizard	AWS Signer
FSx	AWS Compute Optimizer	AWS Network Firewall
S3 Glacier	Resource Groups & Tag Editor	AWS Audit Manager

September 13, 2022

AWS basic services you are using

- EC2 Elastic Cloud Compute
 - Physically, it is a set of compute nodes on a rack in a data center in a region.
 - Groups of machines are broken out into "families" each providing a characteristic ratio of CPU cores to RAM
 - CPU and RAM Examples

Туре	Family	vCPU	GB RAM	On-Demand \$/hr
c5.xlarge	C5	4	8	\$0.17
r5.24xlarge	R5	96	768	\$6.048
t3.medium	Т3	2	4	\$0.0416

- S3 Simple Storage Service (Unlimited Storage)
 - Can only store objects, so no symbolic links and no folders. Each object has a specific prefix.
 - Pricing tiers exist for cost savings potential
- Identity and Access Management (IAM)
 - IAM Roles allow users and services permission via IAM Policies to allow or deny actions or have condition-based permissions.

September 13, 2022

AWS command line interface

Local machine requirements: Python 3.8+ with awscli

"conda install -c anaconda awscli", or "pip install awscli"

aws --no-sign-request s3 ls s3://gnss-ro-data/contributed/v1.1/ucar/ aws --no-sign-request s3 ls \ s3://gnss-ro-data/contributed/v1.1/ucar/metop/atmosphericRetrieval/2009/06/01/ aws --no-sign-request s3 cp --recursive \ s3://gnss-ro-data/contributed/v1.1/romsaf/metop/atmosphericRetrieval/2009/06/01/./

See aws_commands.txt provided in email or aws_commands in EC2 instance.

September 13, 2022

How to access the AWS Open Data Registry in Python

- S3FS: Linux commands for AWS in Python
 - This allows you to link an s3 bucket as a file system.
 - Enables easy use of Is and cp commands to the files stored on s3.
- Boto3
 - Allows you to create, search, and modify any AWS resource and its metadata within they python language.

NASA ACCESS 2019

- GNSS RO data in the AWS Open Data Registry
 - <u>https://registry.opendata.aws/gnss-ro-opendata/</u>
 - s3://gnss-ro-data, region: "us-east-1", unsigned authentication
 - RO data in "contributed/"
 - Database data in "dynamo/"
- Support material in GitHub
 - <u>https://github.com/gnss-ro/aws-opendata/</u>
 - Documentation of RO data in repository, including detailed PDFs
 - Utilities to aid in manipulating the data, especially creating a private DynamoDB database for RO data
 - Tutorial demonstrations: Database queries, inter-center comparison, tropopause analysis

DynamoDB Service

- A serverless database, key-value NoSQL
- Must have a unique Partition Key and Sort Key combo
- Pay for what you use and Easy to query

	leo-ttt	~	date-time	⊽	gps_seconds ⊽	latitude 🔻	r local_time ▼	longitude 🛛	mission \triangledown	occid ⊽	receiver
	sacc-G03		2006-03-09-15-51		825954679	24. <mark>1</mark> 7	15.855	78.864	sacc	sacc-G03-2	sacc
	sacc-G03		2006-03-09-16-36		825957321	-54.438	16.589	-118.203	sacc	sacc-G03-2	sacc
	sacc-G03		2006-03-09-21-36		825975435	-23.238	21.621	170.068	sacc	sacc-G03-2	sacc
1)	sacc-G03		2006-03-09-23-19		825981613	-42.904	23.337	168.402	sacc	sacc-G03-2	sacc
	sacc-G03		2006-03-10-05-05		82 <mark>6</mark> 002 <mark>3</mark> 31	37.549	5.092	-102.621	sacc	sacc-G03-2	sacc
	sacc-G03		2006-03-10-10-51		826023079	-35.137	10.855	-11.389	sacc	sacc-G03-2	sacc
	sacc-G03		2006-03-10-15-40		826040295	-48.139	15.637	-82.931	sacc	sacc-G03-2	sacc
1)	sacc-G03		2006-03-10-16-36		826043814	31.29	16.615	78.109	sacc	sacc-G03-2	sacc
	sacc-G03		2006-03-10-17-22		826046433	-51.961	17.343	-146.332	sacc	sacc-G03-2	sacc

September 13, 2022

Create your own copy of the DynamoDB database

http://github.com/gnss-ro/aws-opendata/tree/master/utilities

Simple but slow...

% python3 import_gnss-ro_dynamoDB.py --dynamodb_table_name my_ro_database \ --mission champ --date_str "2003"

Fast...

Follow instructions in DynamoDB_full_import_instructions.txt

September 13, 2022

Data Analysis with Python: DynamoDB structure

Partition key	Sort key	Information
leo-ttt	yyyy-mm-dd-hh-mm	
metopa-G01	2020-01-01-00-47	Satellite information: transmitter, receiver, mission
	2020-01-01-01-51	<i>Geolocation information:</i> longitude, latitude, time, local_time, setting
	2020-01-01-02-41	deblocation injornation. longitude, latitude, time, local_time, setting
		Paths: ucar_calibratedPhase, ucar_refractivityRetrieval, ucar_atmosphericRetrieval,
metopa-G02	2020-01-01-01-10	romsaf_refractivityRetrieval, romsaf_atmosphericRetrieval
	2020-01-01-02-02	
	2020-01-01-03-14	
metopb-G01	2020-01-01-00-17	
	2020-01-01-01-21	
	2020-01-01-02-11	
metopc-G32	2020-01-01-00-19	
	2020-01-01-21-31	
	2020-01-01-22-24	
	2020-01-01-23-06	

- Querying requires one specific value for a partition key (receiver-transmitter).
- Querying requires sort key specification, but it can be loose ("between" two values).
- Querying can optionally filter results of query by RO "information" (metadata); compound filters allowed.

September 13, 2022

Some things to explore...

To discover what methods are available to you, ...

>> dir(s3)

>> dir(Key)

>> dir(Attr)

>> from Resources import valid_missions

Change the processing center...

>> processing_center = "romsaf"

Try rising occultations...

>> filters = filters & Attr("rising").eq("False")

September 13, 2022

Running NWP on AWS instance – an example of JEDI

 The goal is to conduct NWP experiments on AWS using the RO open data.



- It is not difficult!
- We use JEDI for an example: Joint Effort for Data assimilation Integration (JEDI), developed by JCSDA with partners. <u>https://github.com/JCSDA-internal/</u>
- Config and launch AWS instance
 - Select appropriate instance: cost vs. requirement
 - Use existing Amazon Machine Image(AMI)
 - etc.
- Conduct NWP experiment
 - Activate the modules from AMI
 - Compile/build you system
 - Fetch data from s3 and prepare data for your NWP system.
 - Converter/reader from RO open data to JEDI IODA format
 - Conduct NWP experiments as you usually do on your HPC clusters!
- Make your work sharable Create your own AMI

September 13, 2022

EC2 instance

Services (9) Features (46)	Search results for 'ec2'	sole.aws.amazon.com/console/home?region=us-east-1			G Q 🖞 🛠 🕷 🖠			
		Search for services, features, blogs, docs, and more	[Option+S]	D & (2 N. Virginia hailingz @			
Features (46)	Services	See				azon.com/ec2/v2/home?region=us-east-1#Hom	e:	G Q 🖄 🛠 B 🎗
	EC2 ☆	Console Home Info	Reset to def	US East (N. Virginia)	aws III Services Q Search	for services, features, blogs, docs, and mo	ne [Option+S] D	🕘 N. Virginia ▼ hailingz@j
Blogs (1,781) Documentation (130,680)	Virtual Servers in the Cloud			US East (Ohio)	New EC2 Experience			
Knowledge Articles (30)	EC2 Image Builder 🚖	Recently visited Info		US West (N. California)	Tell us what you think	Resources	EC2 Global view 🖾 🛛 💿	Account attributes
Tutorials (18)	A managed service to automate build, customize and deploy OS images			US West (Oregon)	EC2 Global View	You are using the following Ama	zon EC2 resources in the US East (N. Virginia) Region:	Supported platforms
Events (8)		EC2			Events	Instances (running)	19 Dedicated Hosts 0	VPC Default VPC
Marketplace (1,547)	AWS Compute Optimizer ☆ Recommend optimal AWS Compute resources for your workloads	CodeBuild		Africa (Cape Town)	Tags Limits	Elastic IPs	12 Instances 78	vpc-4a9db231
		S 3		Asia Pacific (Hong Kong	▼ Instances	Key pairs	59 Load balancers 1	Settings EBS encryption
	KwS Firewall Manager ☆ Central management of firewall rules				Instances New	Placement groups	5 Security groups 107	Zones
		CloudFormation		Asia Pacific (Jakarta)	Instance Types Launch Templates	Snapshots	86 Volumes 150	EC2 Serial Console
	Features s	-		Asia Pacific (Mumbai)	Spot Requests			Default credit specification Console experiments
	Dashboard	4		Asia Pacific (Osaka)	Savings Plans		deploy Microsoft SQL Server Always On X 5 using the AWS Launch Wizard for SQL	
	C2 feature			Asia Pacific (Seoul) Reserved Instances New Dedicated Hosts		Server. Learn more	Explore AWS	
	Limits			Asia Pacific (Singapore)	Scheduled Instances			
	C2 feature	Vie	ew all services	Asia Pacific (Sydney)	Capacity Reservations	Launch instance To get started, launch an Amazon EC2	Service health	Get Up to 40% Better Price Performance
	AMIs			Asia Pacific (Tokyo)	▼ Images	instance, which is a virtual server in th cloud.	e C	T4g instances deliver the best price performance for burstable general
C2 feature		Welcome to AWS	AWS Heal		AMI Catalog		AWS Health Dashboard	purpose workloads in Amazon EC
	Elastic IPs	4		Canada (Central)	▼ Elastic Block Store	Launch instance Migrate a server	Region	Save up to 90% on EC2 with Spot
	Elastic IPs	Getting started with AWS 🖸	Open issues	Europe (Frankfurt)	Volumes New	rigiace a server [5]	US East (N. Virginia)	Instances
		Learn the fundamentals and find valuable information to get the most out of AWS.	0		Snapshots New Lifecycle Manager New	Note: Your instances will launch in the East (N. Virginia) Region	US Status This service is operating normally	Optimize price-performance by combining EC2 purchase options in
	Blass		Scheduled change	Europe (Ireland)	▼ Network & Security			single EC2 ASG. Learn more 🔀
		Training and certification	0	Europe (London)	Security Groups	Scheduled events	C Zones	Save Up to 45% on ML Inference EC2 Inf1 instances provide high

EC2 instance

Services Q Search for services, features, blogs, docs, and more	e [Option+S]	Services Q Search for services,	features, blogs, docs, and m [Option+S]	🗘 ၇ N. Virginia	▼ hai	ilingz @ jcsda-
EC2 > Instances > Launch an instance Launch an instance Info Amazon EC2 allows you to create virtual machines, or instances, tha following the simple steps below.	t run on the AWS Cloud. Quickly get started by	Choose an Amazon	Machine Image (AMI) software configuration (operating system, appli an select an AMI provided by AWS, our user com	cation server, and application	ns)	
Name and tags info		Q aws-opendata-reader-ucar			×	•
Name hz_nwp_ro_opendata	Add additional tags		AMIs (0) AWS Marketplace AMIs (2551 ted by me AWS & trusted third-party AMIs) Community AMIs (1 Published by anyone	Ð	
 Application and OS Images (Amazon Machine An AMI is a template that contains the software configuration (operatin launch your instance. Search or Browse for AMIs if you don't see what you aws-opendata-reader-ucar 	g system, application server, and applications) required to	Refine results	aws-opendata-reader-ucar (1 filtered, 0 u aws-opendata-r	eader-ucar	< 1	1 >
Recents My AMIs Quick Start Owned Owned Shared by me Owned Shared Amazon Machine Image (AMI) Owned	Q Browse more AMIs Including AMIs from AWS, Markeplace and the Community	 Clear all filters ▼ Owner ✓ Owned by me Shared with me ▼ OS category All Linux/Unix All Windows 	Owner: 4692053540 Publish date: 2022-0	uilt based on the at-demo. All s to the x Architecture: x86_64	Select	
skylab-1.0.0-ubuntu20-bufr-11.7.1 ami-0b24d818d257651f6 2022-08-30T20:29:47.000Z Virtualization: hvm ENA enabled: tr September 13, 2022		□ All Windows ▼ Publish date range Workshop: GNSS RO in the Cl	name: aws-opendata			nriec 18

Config instance

▼ Instance type Info	EC2 > Instances > Launch an instance
Instance type t2.micro Free tier eligible Family: t2 1 vCPU 1 GiB Memory On-Demand Linux pricing: 0.0116 USD per Hour ▼ Compare instance types On-Demand Windows pricing: 0.0162 USD per Hour ▼ Compare instance types	Success Successfully initiated launch of instance (i-08d7bc407b4168327) Launch log
Key pair (login) Info You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.	Next Steps
Key pair name - required	Get notified of estimated charges
hz_nwp_ro_opendata	Create billing alerts to get an email notification when estimated charges on your AWS bill exceed an amount you define (for example, if you exceed the free usage tier)
▼ Network settings Info Edit	you define (for example, if you exceed the nee daage def)
Network Info	How to connect to your instance
vpc-4a9db231 cdash-server Subnet Info	Your instance is launching and it might be a few minutes until it is in the running state, when it will be ready for you to use
No preference (Default subnet in any availability zone) Auto-assign public IP Info Enable	Click View Instances to monitor your instance's status. Once your instance is in the 'running' state, you can connect to it from the Instances screen. Find out how to connect to your instance
Firewall (security groups) Info A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance. Create security group Select existing security group	View more resources to get you started
Common security groups Info	
Select security groups and group sub-	View all instances
Global SSH sg-0f715b41f30e6d3aa X VPC:vpc-4a9db231 Security groups that you add or remove here will be added to or removed from all your network interfaces.	

AWS Workshop: GNSS RO in the Cloud

September 13, 2022

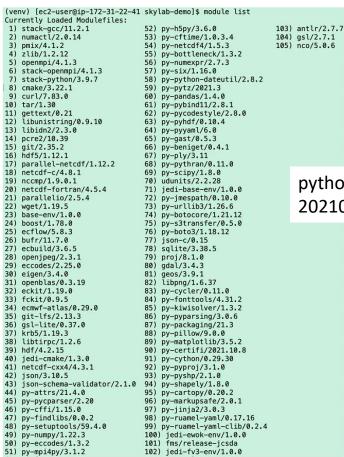
Connect to instance

Instances (1/1) Info	C Connect Inst	tance state Actions Launch instances	
Q Search		< 1	
hz X Clear filters			
✓ Name ♥ Instance ID	Instance state v Instance type v Status check	Alarm status Availability Zone 🔻 Public	c IPv4 DNS
hz_nwp_ro_op i-0ca2c2aa701ab6442	⊘ Running ⊕ Q t2.micro Ø 2/2 checks participation	assed No alarms + us-east-1a ec2-1	07.20.109
			EC2 > Instances > i-Oca2c2aa701ab6442 > Connect to instance
Instance: i-0ca2c2aa701ab6442 (hz_nwp_ro_c	= opendata)		
i-0ca2c2aa701ab6442 (hz_nwp_ro_opendata)	107.20.109.189 open address	□ 172.31.22.41	Connect to instance Info
IPv6 address	Instance state	Public IPv4 DNS	Connect to your instance i-0ca2c2aa701ab6442 (hz_nwp_ro_opendata) using any of these options
-	⊘ Running	ec2-107-20-109-189.compute-1.amazonaws.com address	
Hostname type	Private IP DNS name (IPv4 only)		EC2 Instance Connect Session Manager SSH client EC2 serial console
IP name: ip-172-31-22-41.ec2.internal	D ip-172-31-22-41.ec2.internal		
Answer private resource DNS name	Instance type	Elastic IP addresses	
IPv4 (A)	t2.micro	-	Instance ID
Auto-assigned IP address	VPC ID	AWS Compute Optimizer finding	
D 107.20.109.189 [Public IP]	vpc-4a9db231 (cdash-server)	Opt-in to AWS Compute Optimizer for recommend	问 i-0ca2c2aa701ab6442 (hz_nwp_ro_opendata)
		Learn more 🖸	1. Open an SSH client.
IAM Role	Subnet ID	Auto Scaling Group name	2. Locate your private key file. The key used to launch this instance is hz_nwp_ro_opendata.pem
	🗇 subnet-ff51d0b5 🖸	-	
▼ Instance details Info			 Run this command, if necessary, to ensure your key is not publicly viewable. chmod 400 hz_nwp_ro_opendata.pem
Platform	AMIID	Monitoring	
Other Linux (Inferred)	ami-03ef6e6e1ecb63692	disabled	4. Connect to your instance using its Public DNS:
Platform details	AMI name	Termination protection	□ ec2-107-20-109-189.compute-1.amazonaws.com
Red Hat Enterprise Linux with High Availability	D aws-opendata-reader-ucar	Disabled	Example:
Stop protection	Launch time	AMI location	
Disabled	Wed Aug 31 2022 23:24:04 GMT-0600 (Mountain	469205354006/aws-opendata-reader-ucar	ssh -i "hz_nwp_ro_opendata.pem" root@ec2-107-20-109-189.compute-1.amazonaws.com
	Daylight Time) (7 minutes)		
Instance auto-recovery	Lifecycle	Stop-hibernate behavior	(3) Note: In most cases, the guessed user name is correct. However, read your AMI usage instructions to check if
Default	normal	disabled	the AMI owner has changed the default AMI user name.
AMI Launch index	Key pair name	State transition reason	the Arit owner has changed the default Arit user halfie.
0	D hz_nwp_ro_opendata		

September 13, 2022

Start your work

source ./activate.sh





python awsro2ioda.py -i ref* -o output.nc4 -d 2021080502

September 13, 2022

Create your own AMI

Instances (1/1) Inf	fo		C	Connect Instance state 🔻	Actions 🔺 Launch	instances 🔻
Q Search					Connect	1 > ©
hz X Clea	ar filters				View details	
hz X Clea	al inters				Manage instance state	
		Instance state	pe ⊽ Status check Ala	arm status Availability Zone ⊽	Instance settings	ic IPv4 ▽
aws III Services Q Search	n for services, features, blogs, docs, and more	[Option+S]		אַ אָ		20.109.189
New EC2 Experience	Amazon Machine Images (AMIs) (1/1) Info		C Z Recycle Bin Z EC2 Image Build	der Actions V Launch instance from AMI	Networking	•
EC2 Dashboard	Owned by me 🔻 Q Search			< 1 > 6	Security	E
EC2 Global View	aws-opendata X Clear filters				Image and templates	•
Events	Name V		⊽ Source	♥ Owner ♥ Visibility eader 469205354006 Public	Monitor and troubleshoot	▶ ⊚ ×
Tags 	2 -	ami-03ef6e6e1ecb63692 aws-opendata	a-reader-ucar 469205354006/aws-opendata-re	eader 469205354006 Public		
▼ Instances						
Instances New	AMI ID: ami-03ef6e6e1ecb63692))))	© >		
Instance Types	AMI ID: ami-0366666 160063692			0 /		
Launch Templates	Details Permissions Storage Tags					
Spot Requests	AMI ID	Image type	Platform details	Root device type		
Savings Plans	ami-03ef6e6e1ecb63692	machine	Red Hat Enterprise Linux with High Availability	EBS		
Reserved Instances New	AMI name	Owner account ID	Architecture	Usage operation		
Dedicated Hosts Scheduled Instances	🗇 aws-opendata-reader-ucar	☐ 469205354006	x86_64	RunInstances:1010		
Capacity Reservations	Root device name	Status Ø Available	Source 3469205354006/aws-opendata-reader-ucar	Virtualization type hvm		
		-		MARADO		
Images	Boot mode	State reason	Creation date Sun Aug 28 2022 00:52:49 GMT-0600 (Mountain	Kernel ID		
AMIS New			Daylight Time)			
AMI Catalog	Block devices	Description	Product codes	RAM disk ID		
Elastic Block Store	/dev/sda1=snap- 032e6f0ca27177e78:995:true:gp2	This image was built based on the skylab-1.0.0- redhat-demo. All copyright belongs to the	-			
Volumes New	052e010ca2/17/0/6:995:true:gp2	JCSDA/UCAR.				
					COSMIC -	╬UCA
					COSIVIL	

September 13, 2022