Terraform Apply:


```
module.eks.module.eks_managed_node_group["one"].aws_eks_node_group.this[0]: Sti
l creating... [01m50s elapsed]
module.eks.module.eks_managed_node_group["two"].aws_eks_node_group.this[0]: Crea
tion complete after 1m51s [id=eks-g85pk5LY:node-group-2-2025111714125422430000000
1b]
module.eks.module.eks_managed_node_group["one"].aws_eks_node_group.this[0]: Sti
l creating... [02m00s elapsed]
module.eks.module.eks_managed_node_group["one"].aws_eks_node_group.this[0]: Crea
tion complete after 2m2s [id=eks-g85pk5LY:node-group-1-2025111714125422370000001
9]

Apply complete! Resources: 68 added, 0 changed, 0 destroyed.

Outputs:

cluster_endpoint = "https://6FBA74FD2D0B83737D4CCB24983F0832.gr7.us-west-1.eks.a
mazonaws.com"
cluster_name = "eks-g85pk5LY"
cluster_security_group_id = "sg-0ca184c09971432cb"
region = "us-west-1"
```

Authenticate Docker to ECR:

BV1VEQkFFdU1CRUVESn1SQmE4cUlPQTNybHZtd3dJQkVJQTdLb01PaXJueFVCYmo3REpwVmhxMzRMMlo
vWEJaQml5MHpUREM0d1lEL3hjaAGhSYmVpUmVPRHNGejQ1YmV0QWJ6cWQ5bGxocDgrb29QaWWgOPSIsInZ
lcnNpb24iOiIyIiwidHlwZSI6IkRBEFfSOVZIiwiZXhwaXJhdGlvbiI6MTc2MzQzMzAzMn0= 838240
021663.dkr.ecr.us-west-1.amazonaws.com
```
WARNING! Using --password via the CLI is insecure. Use --password-stdin.
Login Succeeded

AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
```

Local tests result:

report.html

Environment

Python	3.11.0
Platform	Windows-10-10.0.26200-SP0
Packages	• pytest: 9.0.1 • pluggy: 1.6.0
Plugins	• html: 4.1.1 • metadata: 3.1.1
JAVA_HOME	C:\User Program Files\Java\jdk-17.0.2

Summary

13 tests took 00:00:07.

(Un)check the boxes to filter the results.

☑ 0 Failed, ☑ 13 Passed, ☑ 0 Skipped, ☑ 0 Expected failures, ☑ 0 Unexpected passes, ☑ 0 Errors, ☑ 0 Reruns   Show all details / Hide all details

Result ▲	Test	Duration	Links
Passed	src/test/etl_tests.py::TestGeocodingAPI::test_geocode_nyc_address	828 ms	
Passed	src/test/etl_tests.py::TestGeocodingAPI::test_geocode_london_address	590 ms	
Passed	src/test/etl_tests.py::TestGeocodingAPI::test_geocode_tokyo_address	374 ms	
Passed	src/test/etl_tests.py::TestLatLngWithAPI::test_get_lat_lng_with_real_api	437 ms	
Passed	src/test/etl_tests.py::TestLatLngWithAPI::test_get_lat_lng_london_with_real_api	393 ms	
Passed	src/test/etl_tests.py::TestGeohash::test_generates_geohash_from_valid_coordinates	1 ms	
Passed	src/test/etl_tests.py::TestGeohash::test_returns_none_for_null_latitude	0 ms	
Passed	src/test/etl_tests.py::TestGeohash::test_returns_none_for_null_longitude	0 ms	
Passed	src/test/etl_tests.py::TestGeohash::test_returns_none_for_both_null	0 ms	
Passed	src/test/etl_tests.py::TestGeohash::test_handles_invalid_input_gracefully	0 ms	
Passed	src/test/etl_tests.py::TestGeohash::test_different_locations_produce_different_hashes	1 ms	
Passed	src/test/etl_tests.py::TestEncryption::test_encrypts_specified_pii_fields	00:00:04	
Passed	src/test/etl_tests.py::TestEncryption::test_raises_error_for_invalid_fields	552 ms	

Docker build, tag, push:

```
=> CACHED [stage-0  6/15] RUN pip3 install --no-cache-dir --upgrade pip    0.0s
=> CACHED [stage-0  7/15] COPY ./entrypoint.sh /opt/                       0.0s
=> [stage-0  8/15] COPY ./dist/sparkbasics-*.egg /opt/                     0.0s
=> [stage-0  9/15] COPY --from=extra-source ./src /opt/src                 0.0s
=> [stage-0 10/15] RUN chmod +x /opt/*.sh                                  0.3s
=> [stage-0 11/15] RUN ln -s /usr/bin/tini /sbin/tini                      0.3s
=> [stage-0 12/15] RUN groupadd -g 999 spark &&      useradd -m -u 185 -g  0.3s
=> [stage-0 13/15] WORKDIR /opt/spark/work-dir                             0.1s
=> [stage-0 14/15] RUN mkdir -p /opt/spark/work-dir &&      chown -R spar  0.2s
=> [stage-0 15/15] RUN chown spark:spark /opt/sparkbasics-*.egg && chmod   0.3s
=> exporting to image                                                      0.1s
=> => exporting layers                                                     0.1s
=> => writing image sha256:9046262460926755ba33b7ba962528dc5bd59085e6ec8  0.0s
=> => naming to docker.io/library/m06-spark-image                         0.0s
(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main
$ docker tag m06-spark-image:latest 838240021663.dkr.ecr.us-west-1.amazonaws.com
/spark-python-06:latest
(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main
$ docker push 838240021663.dkr.ecr.us-west-1.amazonaws.com/spark-python-06:lates
t
The push refers to repository [838240021663.dkr.ecr.us-west-1.amazonaws.com/spar
k-python-06]
3e78a17574c9: Pushed
5491a2c83852: Pushed
0181a1c1dcbc: Pushed
5b75bd9b9b9d: Pushed
eeb5fc5d4231: Pushed
f596e771c530: Pushed
86170511e037: Pushed
d88664949ebd: Pushed
7b4762837fdc: Pushed
3f0fc864f267: Pushed
6bbfe40fe5d5: Pushed
eddb715c2785: Pushed
39b2ff005b05: Pushed
6686d8a0d20c: Pushed
73974f74b436: Pushed
latest: digest: sha256:ae223e62d9fd517b13e7025683a390e24ec02162f75485c580b655dbf
92dfb06 size: 3454
(.venv)
```

Kubernetes Service Account Setup for Spark:

```
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main
$ aws eks list-clusters --region us-west-1
{
    "clusters": [
        "eks-g85pk5LY"
    ]
}

(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main
$ aws eks update-kubeconfig --name eks-g85pk5LY --region us-west-1
Added new context arn:aws:eks:us-west-1:838240021663:cluster/eks-g85pk5LY to C:\
Users\YuriiHordiichuk\.kube\config
(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main
$ kubectl create clusterrolebinding spark-rolebinding \
  --clusterrole=edit \
  --serviceaccount=default:spark
clusterrolebinding.rbac.authorization.k8s.io/spark-rolebinding created
(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main
$ kubectl create serviceaccount spark
serviceaccount/spark created
(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main
$ eksctl create iamserviceaccount --name spark --namespace default --cluster eks
-g85pk5LY --attach-policy-arn arn:aws:iam::838240021663:policy/S3SparkAccessPoli
cy --approve --override-existing-serviceaccounts
2025-11-17 15:48:55 [ℹ]  1 iamserviceaccount (default/spark) was included (based
 on the include/exclude rules)
2025-11-17 15:48:55 [!]  metadata of serviceaccounts that exist in Kubernetes wi
ll be updated, as --override-existing-serviceaccounts was set
2025-11-17 15:48:55 [ℹ]  1 task: {
    2 sequential sub-tasks: {
        create IAM role for serviceaccount "default/spark",
        create serviceaccount "default/spark",
    } }2025-11-17 15:48:55 [ℹ]  building iamserviceaccount stack "eksctl-eks-g85
pk5LY-addon-iamserviceaccount-default-spark"
2025-11-17 15:48:55 [ℹ]  deploying stack "eksctl-eks-g85pk5LY-addon-iamserviceac
count-default-spark"
2025-11-17 15:48:55 [ℹ]  waiting for CloudFormation stack "eksctl-eks-g85pk5LY-a
ddon-iamserviceaccount-default-spark"
2025-11-17 15:49:26 [ℹ]  waiting for CloudFormation stack "eksctl-eks-g85pk5LY-a
ddon-iamserviceaccount-default-spark"
2025-11-17 15:49:28 [ℹ]  serviceaccount "default/spark" already exists
2025-11-17 15:49:28 [ℹ]  updated serviceaccount "default/spark"
(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main/spark_config
$ kubectl apply -f spark-role.yaml
role.rbac.authorization.k8s.io/spark-role created
(.venv)
AzureAD+YuriiHordiichuk@EPPLWROW0218 MINGW64 ~/Desktop/m06_sparkbasics_python_aw
s-main/m06_sparkbasics_python_aws-main/spark_config
$ kubectl apply -f spark-rolebinding.yaml
rolebinding.rbac.authorization.k8s.io/spark-rolebinding created
(.venv)
```

spark-submit run:

```
Administrator: Windows PowerShell                                    —    □    ✕

PS C:\WINDOWS\system32> spark-submit `
>>    --master "k8s://https://6FBA74FD2D0B83737D4CCB24983F0832.gr7.us-west-1.eks.amazonaws.com" `
>>    --deploy-mode cluster `
>>    --name sparkbasics `
>>    --conf spark.kubernetes.container.image=838240021663.dkr.ecr.us-west-1.amazonaws.com/spark-python-0
6:latest `
>>    --conf spark.kubernetes.driver.request.cores=500m `
>>    --conf spark.kubernetes.driver.request.memory=2g `
>>    --conf spark.kubernetes.executor.request.memory=1g `
>>    --conf spark.kubernetes.executor.request.cores=500m `
>>    --conf spark.kubernetes.namespace=default `
>>    --conf spark.kubernetes.authenticate.driver.serviceAccountName=spark `
>>    --conf spark.kubernetes.executor.instances=1 `
>>    --conf spark.pyspark.python=/usr/bin/python3 `
>>    --conf spark.pyspark.driver.python=/usr/bin/python3 `
>>    --py-files local:///opt/sparkbasics-1.0.0-py3.11.egg `
>>    local:///opt/src/main/python/main_etl.py
25/11/17 16:09:44 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using
 builtin-java classes where applicable
25/11/17 16:09:45 INFO SparkKubernetesClientFactory: Auto-configuring K8S client using current context f
rom users K8S config file
25/11/17 16:09:46 INFO KerberosConfDriverFeatureStep: You have not specified a krb5.conf file locally or
 via a ConfigMap. Make sure that you have the krb5.conf locally on the driver image.
25/11/17 16:09:49 INFO LoggingPodStatusWatcherImpl: State changed, new state:
        pod name: sparkbasics-58230d9a925d62bd-driver
        namespace: default
        labels: spark-app-name -> sparkbasics, spark-app-selector -> spark-5c2eacb5aa154a7393a384e8dd9b
1f23, spark-role -> driver, spark-version -> 3.5.7
        pod uid: d96019bd-79dc-43cd-b6ae-1634e502dd3b
        creation time: 2025-11-17T15:09:47Z
        service account name: spark
        volumes: aws-iam-token, spark-local-dir-1, spark-conf-volume-driver, kube-api-access-zh2jc
        node name: ip-10-0-3-137.us-west-1.compute.internal
        start time: 2025-11-17T15:09:47Z
        phase: Pending
        container status:
                container name: spark-kubernetes-driver
                container image: 838240021663.dkr.ecr.us-west-1.amazonaws.com/spark-python-06:latest
```

```
              container started at: 2025-11-17T15:09:49Z
              container finished at: 2025-11-17T15:18:23Z
              exit code: 0
              termination reason: Completed
25/11/17 16:18:24 INFO LoggingPodStatusWatcherImpl: Application status for spark-5c2eacb5aa154a7393a384e
8dd9b1f23 (phase: Running)
25/11/17 16:18:25 INFO LoggingPodStatusWatcherImpl: State changed, new state:
       pod name: sparkbasics-58230d9a925d62bd-driver
       namespace: default
       labels: spark-app-name -> sparkbasics, spark-app-selector -> spark-5c2eacb5aa154a7393a384e8dd9b
1f23, spark-role -> driver, spark-version -> 3.5.7
       pod uid: d96019bd-79dc-43cd-b6ae-1634e502dd3b
       creation time: 2025-11-17T15:09:47Z
       service account name: spark
       volumes: aws-iam-token, spark-local-dir-1, spark-conf-volume-driver, kube-api-access-zh2jc
       node name: ip-10-0-3-137.us-west-1.compute.internal
       start time: 2025-11-17T15:09:47Z
       phase: Succeeded
       container status:
              container name: spark-kubernetes-driver
              container image: 838240021663.dkr.ecr.us-west-1.amazonaws.com/spark-python-06:latest
              container state: terminated
              container started at: 2025-11-17T15:09:49Z
              container finished at: 2025-11-17T15:18:23Z
              exit code: 0
              termination reason: Completed
25/11/17 16:18:25 INFO LoggingPodStatusWatcherImpl: Application status for spark-5c2eacb5aa154a7393a384e
8dd9b1f23 (phase: Succeeded)
25/11/17 16:18:25 INFO LoggingPodStatusWatcherImpl: Container final statuses:


              container name: spark-kubernetes-driver
              container image: 838240021663.dkr.ecr.us-west-1.amazonaws.com/spark-python-06:latest
              container state: terminated
              container started at: 2025-11-17T15:09:49Z
              container finished at: 2025-11-17T15:18:23Z
              exit code: 0
              termination reason: Completed
25/11/17 16:18:25 INFO LoggingPodStatusWatcherImpl: Application sparkbasics with application ID spark-5c
2eacb5aa154a7393a384e8dd9b1f23 and submission ID default:sparkbasics-58230d9a925d62bd-driver finished
25/11/17 16:18:25 INFO ShutdownHookManager: Shutdown hook called
25/11/17 16:18:25 INFO ShutdownHookManager: Deleting directory C:\Users\YuriiHordiichuk\AppData\Local\Te
mp\spark-2cf92429-0efb-4f33-8ba4-02aa787b9be7
```

Cluster:

eks-g85pk5LY

Delete cluster   Upgrade version   **Monitor cluster**

ⓘ ▶ End of standard support for Kubernetes version 1.32 is March 23, 2026.   Upgrade

▼ **Cluster info** Info

Status	Kubernetes version Info	Support period	Provider
⊘ Active	1.32	ⓘ Standard support until March 23, 2026	EKS

Cluster health	Upgrade insights	Node health issues	
⊘ **0**	⊘**4** ⊗**1**	⊘ **0**	

Overview   Resources   **Compute**   Networking   Add-ons   Access   Observability   Update history & backups   Tags

Nodes (6) Info

🔍 Filter Nodes by property or value   ‹ **1** 2 ›

Node name ▲	Instance type ▽	Compute ▽	Managed by ▽	Created ▽	Status ▽
ip-10-0-1-43.us-west-1.compute.internal	t3.medium	Node group	node-group-1-20251117141254223700000019	📋 an hour ago	⊘ Ready
ip-10-0-2-108.us-west-1.compute.internal	t3.medium	Node group	node-group-2-20251117141254224300000001b	📋 an hour ago	⊘ Ready
ip-10-0-2-139.us-west-1.compute.internal	t3.medium	Node group	node-group-1-20251117141254223700000019	📋 an hour ago	⊘ Ready
ip-10-0-2-64.us-west-1.compute.internal	t3.medium	Node group	node-group-1-20251117141254223700000019	📋 an hour ago	⊘ Ready
ip-10-0-3-137.us-west-1.compute.internal	t3.medium	Node group	node-group-2-20251117141254224300000001b	📋 an hour ago	⊘ Ready

⚠ Amazon EKS will no longer publish EKS-optimized Amazon Linux 2 (AL2) AMIs after November 26th, 2025. Additionally, Kubernetes version 1.32 is the last version for which Amazon EKS will release AL2 AMIs. From version 1.33 onwards, Amazon EKS will continue to release AL2023 and Bottlerocket based AMIs.   Learn more ⧉

Node groups (2) Info

Edit   Delete   **Add**

Node groups implement basic compute scaling through EC2 Auto Scaling groups.

🔍 Filter node groups by property or value   ‹ **1** ›

Group name ▲	Desired size ▽	AMI release version ▽	Launch template ▽	Status ▽
⚪ node-group-1-20251117141254223700000019	3	1.32.9-20251112	one-20251117141246027400000017 (1)	⊘ Active
⚪ node-group-2-20251117141254224300000001b	3	1.32.9-20251112	two-20251117141246024200000015 (1)	⊘ Active

Results:

☰   Amazon S3 > Buckets > m06awsbucket-g85pk5ly > m06sparkbasics/ > result/

result/

Objects   Properties

Objects (2)

↻   📋 Copy S3 URI   📋 Copy URL   ⬇ Download   Open ⧉   Delete   **Actions ▾**

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory ⧉ to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant the

🔍 Find objects by prefix

☐	Name ▲	Type ▽	Last modified ▽	Size ▽	Storage class
☐	📄 _SUCCESS	-	November 17, 2025, 17:17:22 (UTC+01:00)	0 B	Standard
☐	📄 part-00000-1bb67209-39b6-4e42-88ae-9ce96eb63cc8-c000.snappy.parquet	parquet	November 17, 2025, 17:17:21 (UTC+01:00)	729.1 KB	Standard

result_df:

```
result_s3_path = "s3a://m06awsbucket-g85pk5ly/m06sparkbasics/result/"

df_result = spark.read.option("recursiveFileLookup", "true").parquet(result_s3_path)
```
✓ 3.4s

```
df_result.show(10)
```
✓ 2.3s

```
+-------+-------------+--------------------+--------------------+--------------------+--------------------+--------------------+---------+---------+----------------+----------------+
|geohash|           Id|                Name|             Country|                City|             Address| Latitude| Longitude|       avg_tmpr_f|       avg_tmpr_c|
+-------+-------------+--------------------+--------------------+--------------------+--------------------+--------------------+---------+---------+----------------+----------------+
|  9yqg|1477468749829|PhICmvZwu97R5N27y...|zBs9dL9DhsW5CmJSn...|yeBLLCCWOYRGKl30v...|WlRpJjeBrjeXPv/SK...| 35.79068| -91.63828|71.40808757819484|21.89186773905272|
|  9zwx| 592705486848|UYMPj+p+uI2000rVK...|rGxtAXtIprvK0d6JL...|CE4kHk3La11xv/k1t...|/xl1Fj+2KWAwPB2+u...| 43.56077|-92.010704|60.20548007246376|15.669429347826076|
|  dp88|1142461300736|J6OTlgjoX4S92S/XP...|cI/xkN+KC0+uS8mLs...|gaqYqtn2DOuoWRnWg...|ThIInyfrQiJCgjGo0...| 42.27388|  -88.9672|63.04345598509549|17.247275267815553|
|  dr3b|1073741824001|8bjBmF76za0+HjycK...|rQIkaMaOVpmbvIk6/...|riFeDLik/7xOpVxm8...|oyJmiGF5/8voifouy...| 40.814148| -76.13433|60.52094685127289|15.84452880750335|
|  gcpv|2052994367489|7ncaPd58s5YKRbLYg...|OQVy9CmeuBgBdE3Qx...|q4B+KYIEapNbbs4Oh...|yv/pkjefnPpKCwTok...|51.5066991|-0.0280208|56.02065217391304|13.345652173913045|
|  drv5| 506806140929|nk9OcNRvYbP8RJzEY...|O+tZ5r0hd0/MIXaUa...|gCTcZDhti6TOLvTVq...|x0/uApR2bdr9k+bIj...| 44.266968| -71.56098|54.13619660082683|12.29834634818558|
|  gcpv|2671469658114|IsQoB3rZ99lAuyfQU...|hU1tuIk9Beu+PDaqH...|AnheLA+rIu6arTgeg...|T12UqIp6lrTXMR3OL...|51.5065578| -0.004514|56.02065217391304|13.345652173913045|
|  gcpv|2748779069440|y1piXSFza4Xygleai...|37LuENS4A31PmRahm...|ZT9ey6554+MF5u4uM...|JYmZqNj+0OPFOby3/...|51.5426054|-0.0072708|56.02065217391304|13.345652173913045|
|  sp3e|2972117368834|Ix487rT/mbhv+7lJ5...|gKRp0sFz3KtOR9eUB...|X199V5eGe4uuJxQbH...|fSmjn4GxG/v+OYjCm...|41.3944029|  2.150194|70.22282608695652|21.233152173913044|
|  dpc0| 773094113281|Sibfw19x90l3obP3u...|kAhdd1BsCw2VU7g81...|/aGsBAuWcE8OSuNI8...|zE6EdBno8QoM007ps...| 43.75521| -88.47649|60.44180253623188|15.80108695652174|
+-------+-------------+--------------------+--------------------+--------------------+--------------------+--------------------+---------+---------+----------------+----------------+
only showing top 10 rows
```

```
df_result.count()
```
✓ 1.5s

2494