Hands-on #11 : AWS EC2 & S3 : Word count & webapp

SRIJA BANDA

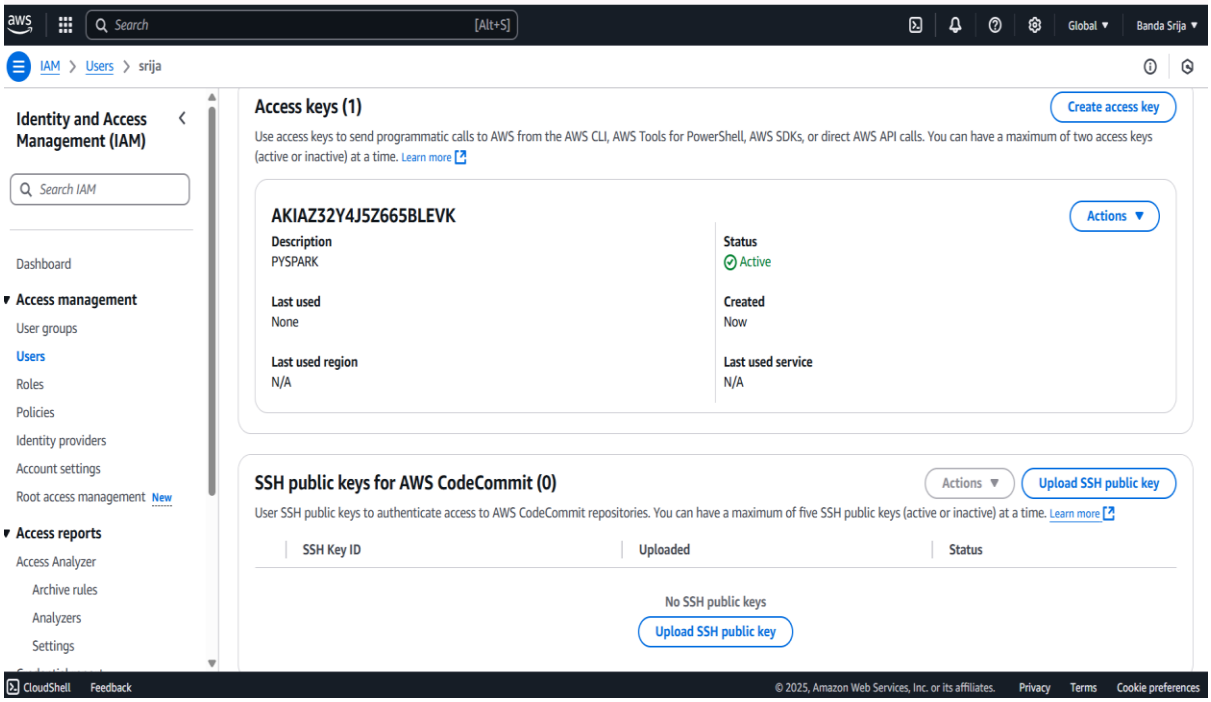
801393582

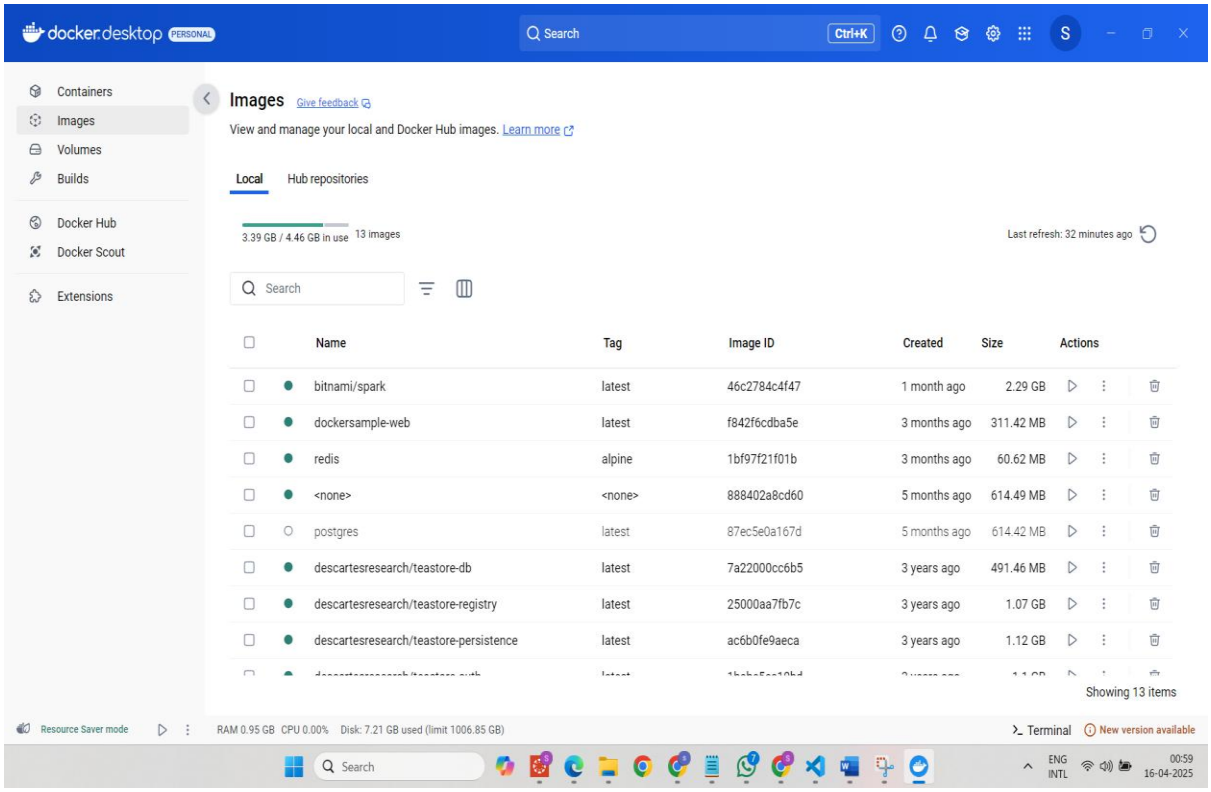
https://github.com/Srijabanda/word-count-spark.git

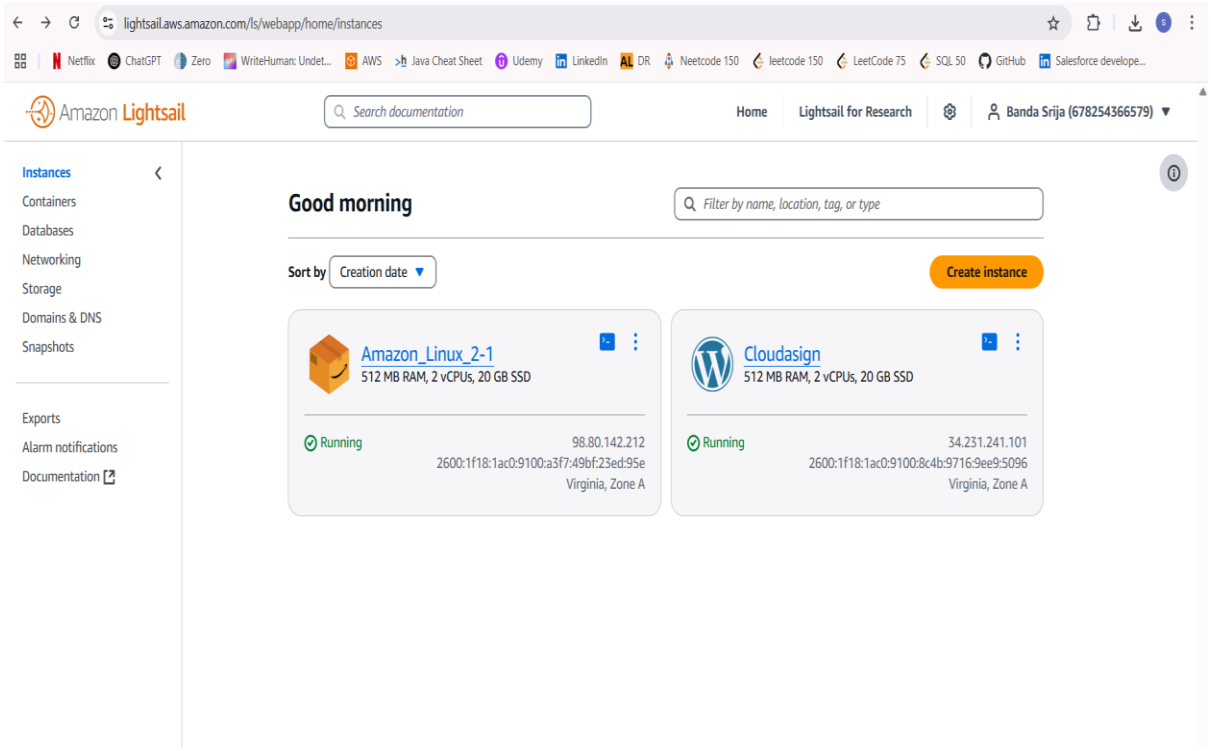
TASK - 1

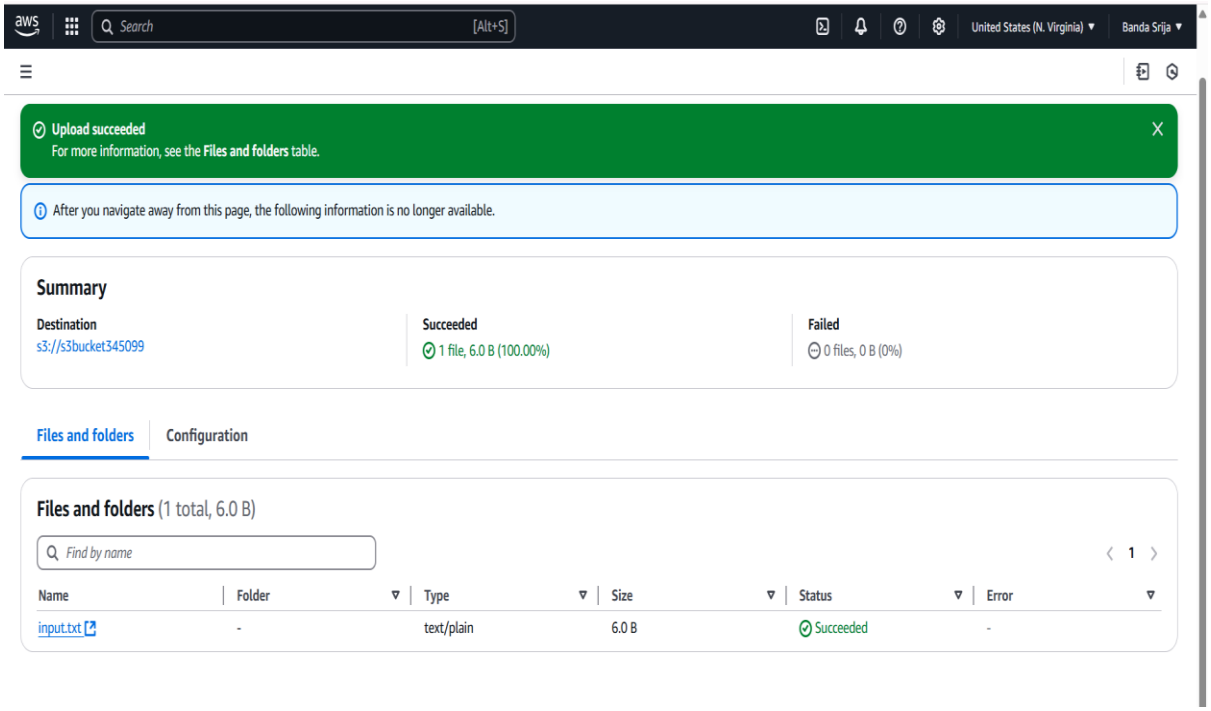
Wordcount.py

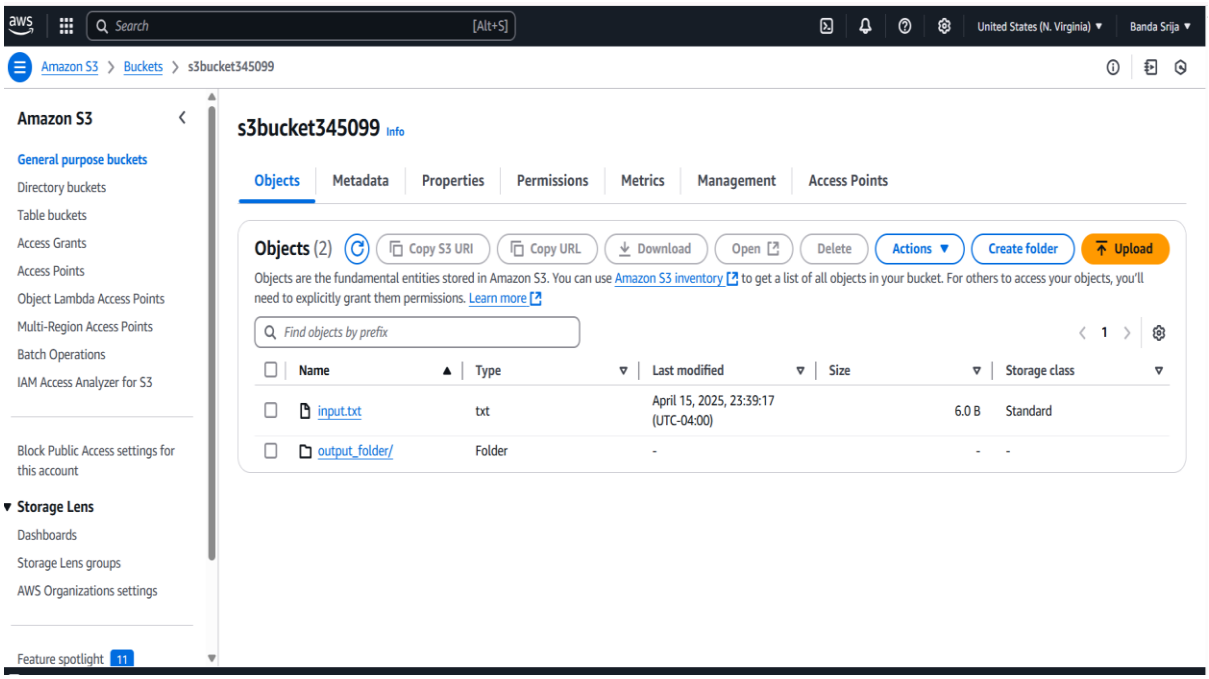
```
from pyspark.sql import SparkSession


AWS_ACCESS_KEY_ID = 'AKIAZ32Y4J5Z665BLEVK'

AWS_SECRET_ACCESS_KEY = '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'


S3_INPUT = 's3a://s3bucket345099/input.txt'

S3_OUTPUT = 's3a://s3bucket345099/output_folder/'


spark = SparkSession.builder.appName("WordCount").config("spark.jars.packages",
"org.apache.hadoop:hadoop-aws:3.3.1,com.amazonaws:aws-java-sdk-
bundle:1.11.901").getOrCreate()


hadoop_conf = spark.sparkContext._jsc.hadoopConfiguration()

hadoop_conf.set("fs.s3a.access.key", AWS_ACCESS_KEY_ID)

hadoop_conf.set("fs.s3a.secret.key", AWS_SECRET_ACCESS_KEY)

hadoop_conf.set("fs.s3a.endpoint", "s3.amazonaws.com")


text_file = spark.sparkContext.textFile(S3_INPUT)

counts = text_file.flatMap(lambda line: line.split()).map(lambda word: (word,
1)).reduceByKey(lambda a, b: a + b)

counts.saveAsTextFile(S3_OUTPUT)

spark.stop()
```

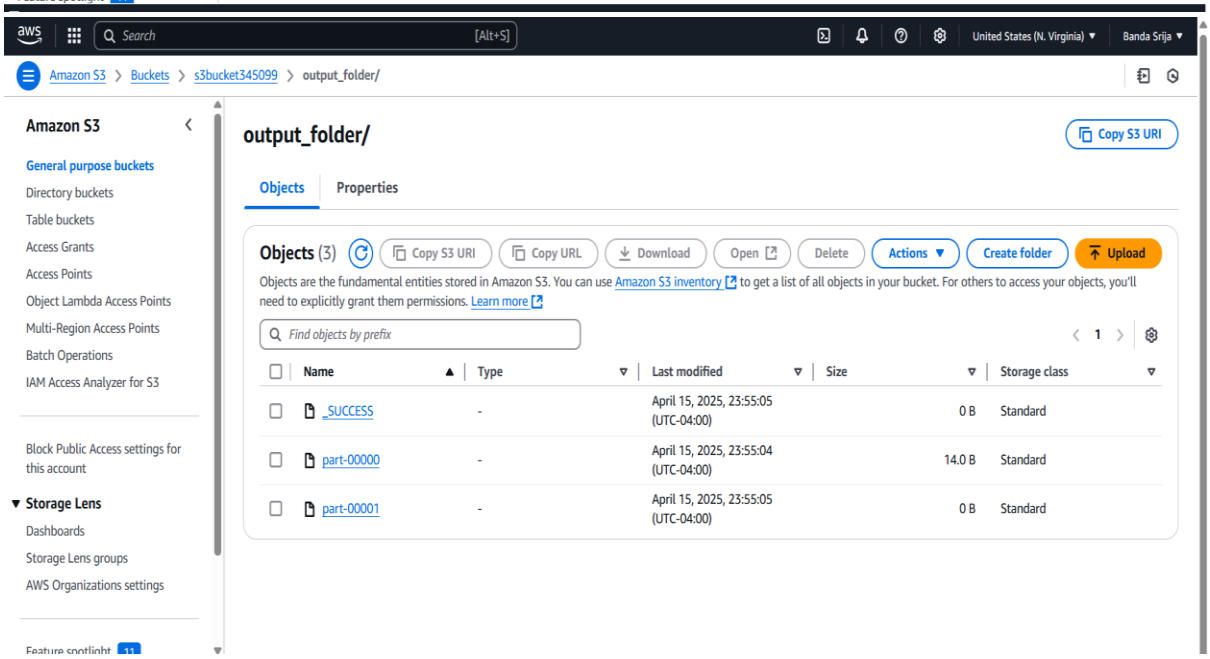

Search [Alt+S]
Global
Banda Srija
IAM > Users > srija
Identity and Access Management (IAM)
Search IAM
Dashboard
Access management
User groups
Users
Roles
Policies
Identity providers
Account settings
Root access management New
Access reports
Access Analyzer
Archive rules
Analyzers
Settings
Access keys (1)
Create access key
Use access keys to send programmatic calls to AWS from the AWS CLI, AWS Tools for PowerShell, AWS SDKs, or direct AWS API calls. You can have a maximum of two access keys (active or inactive) at a time. Learn more
AKIAZ32Y4J5Z665BLEVK
Actions
Description
PYSPARK
Status
Active
Last used
None
Created
Now
Last used region
N/A
Last used service
N/A
SSH public keys for AWS CodeCommit (0)
Actions
Upload SSH public key
User SSH public keys to authenticate access to AWS CodeCommit repositories. You can have a maximum of five SSH public keys (active or inactive) at a time. Learn more
SSH Key ID
Uploaded
Status
No SSH public keys
Upload SSH public key
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

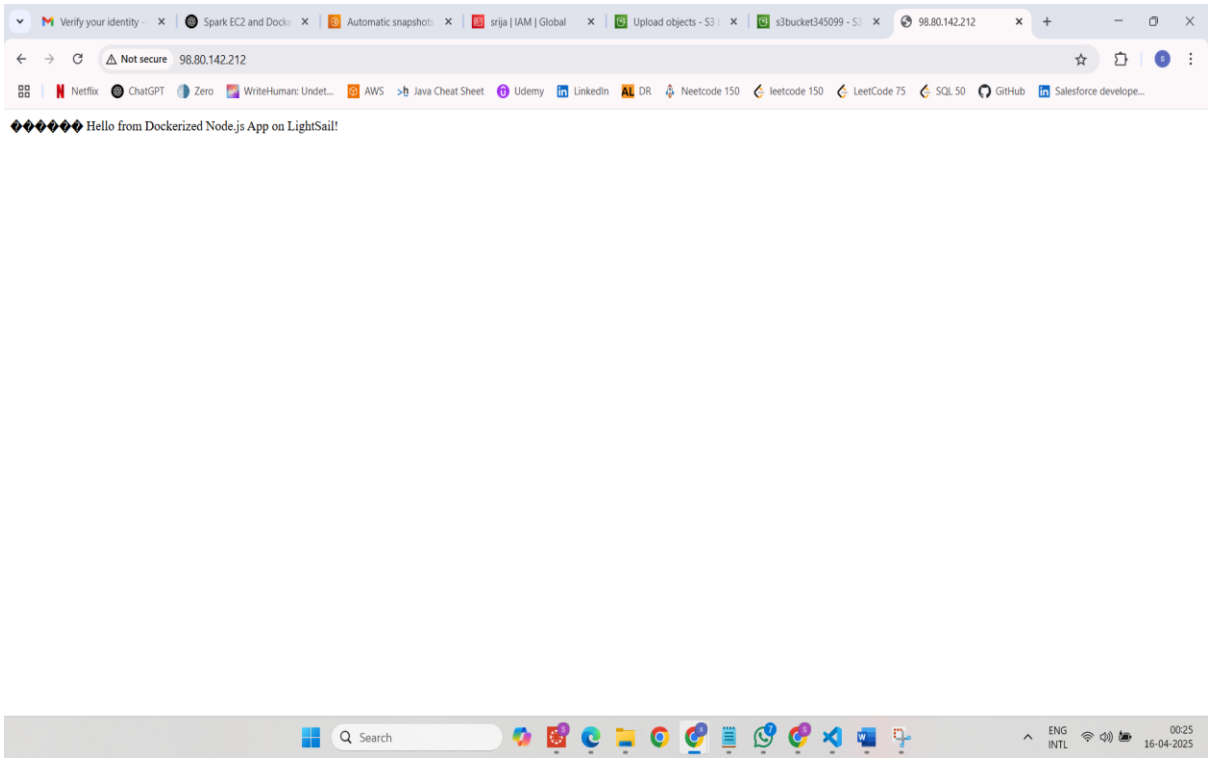

docker desktop PERSONAL
Search Ctrl+K
Containers
Images
Volumes
Builds
Docker Hub
Docker Scout
Extensions
Images
Give feedback
View and manage your local and Docker Hub images. Learn more
Local
Hub repositories
3.39 GB / 4.46 GB in use
13 images
Last refresh: 32 minutes ago
Search
Name
Tag
Image ID
Created
Size
Actions
bitnami/spark latest 46c2784c4f47 1 month ago 2.29 GB
dockersample-web latest f842f6cdba5e 3 months ago 311.42 MB
redis alpine 1bf97f21f01b 3 months ago 60.62 MB
<none> <none> 888402a8cd60 5 months ago 614.49 MB
postgres latest 87ec5e0a167d 5 months ago 614.42 MB
descartesresearch/teastore-db latest 7a22000cc6b5 3 years ago 491.46 MB
descartesresearch/teastore-registry latest 25000aa7fb7c 3 years ago 1.07 GB
descartesresearch/teastore-persistence latest ac6b0fe9aeca 3 years ago 1.12 GB
Showing 13 items
Resource Saver mode
RAM 0.95 GB CPU 0.00% Disk: 7.21 GB used (limit 1006.85 GB)
Terminal
New version available
Search
ENG INTL
00:59 16-04-2025


lightsail.aws.amazon.com/ls/webapp/home/instances
Amazon Lightsail
Search documentation
Home
Lightsail for Research
Banda Srija (678254366579)
Instances
Containers
Databases
Networking
Storage
Domains & DNS
Snapshots
Exports
Alarm notifications
Documentation
Good morning
Filter by name, location, tag, or type
Sort by
Creation date
Create instance
Amazon_Linux_2-1
512 MB RAM, 2 vCPUs, 20 GB SSD
Running
98.80.142.212
2600:1f18:1ac0:9100:a3f7:49bf:23ed:95e
Virginia, Zone A
Cloudasign
512 MB RAM, 2 vCPUs, 20 GB SSD
Running
34.231.241.101
2600:1f18:1ac0:9100:8c4b:9716:9ee9:5096
Virginia, Zone A


Search
[Alt+S]
United States (N. Virginia)
Banda Srija
Upload succeeded
For more information, see the Files and folders table.
After you navigate away from this page, the following information is no longer available.
Summary
Destination
s3://s3bucket345099
Succeeded
1 file, 6.0 B (100.00%)
Failed
0 files, 0 B (0%)
Files and folders
Configuration
Files and folders (1 total, 6.0 B)
Find by name
Name
Folder
Type
Size
Status
Error
input.txt
-
text/plain
6.0 B
Succeeded
-


Search [Alt+S]
United States (N. Virginia)
Banda Srija
Amazon S3 > Buckets > s3bucket345099
Amazon S3
General purpose buckets
Directory buckets
Table buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight 11
s3bucket345099 Info
Objects Metadata Properties Permissions Metrics Management Access Points
Objects (2)
Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload
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 them permissions. Learn more
Find objects by prefix
Name Type Last modified Size Storage class
input.txt txt April 15, 2025, 23:39:17 (UTC-04:00) 6.0 B Standard
output_folder/ Folder - - -

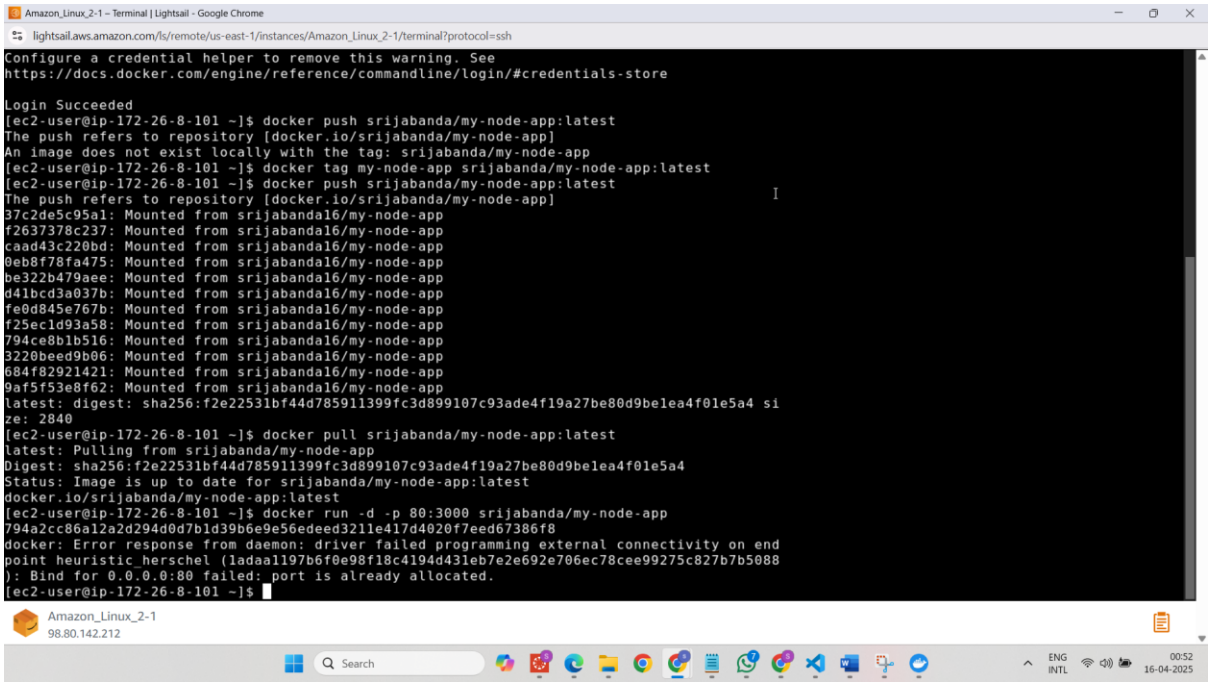

Search [Alt+S]
United States (N. Virginia)
Banda Srija
Amazon S3 > Buckets > s3bucket345099 > output_folder/
Amazon S3
General purpose buckets
Directory buckets
Table buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
output_folder/
Copy S3 URI
Objects Properties
Objects (3)
Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload
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 them permissions. Learn more
Find objects by prefix
Name Type Last modified Size Storage class
_SUCCESS - April 15, 2025, 23:55:05 (UTC-04:00) 0 B Standard
part-00000 - April 15, 2025, 23:55:04 (UTC-04:00) 14.0 B Standard
part-00001 - April 15, 2025, 23:55:05 (UTC-04:00) 0 B Standard

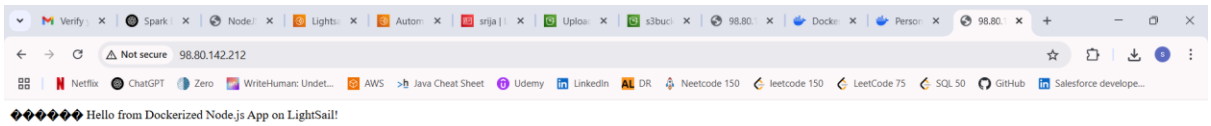


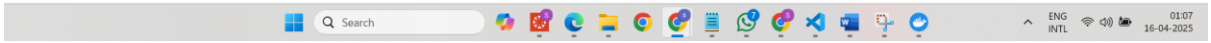
TASK-2

server.js

```
const express = require('express');
const app = express();
const port = 3000;
app.get('/', (req, res) => {
  res.send(' 🚀 Hello from Dockerized Node.js App on LightSail!');
});
app.listen(port, () => {
  console.log(`Server running at http://localhost:${port}`);
});
```







PORT URL - http://98.80.142.212/

Docker Repository URL:
https://hub.docker.com/r/srijabanda/my-node-app

Docker Image: https://hub.docker.com/r/srijabanda/my-node-app

Pull Command: docker pull srijabanda/my-node-app