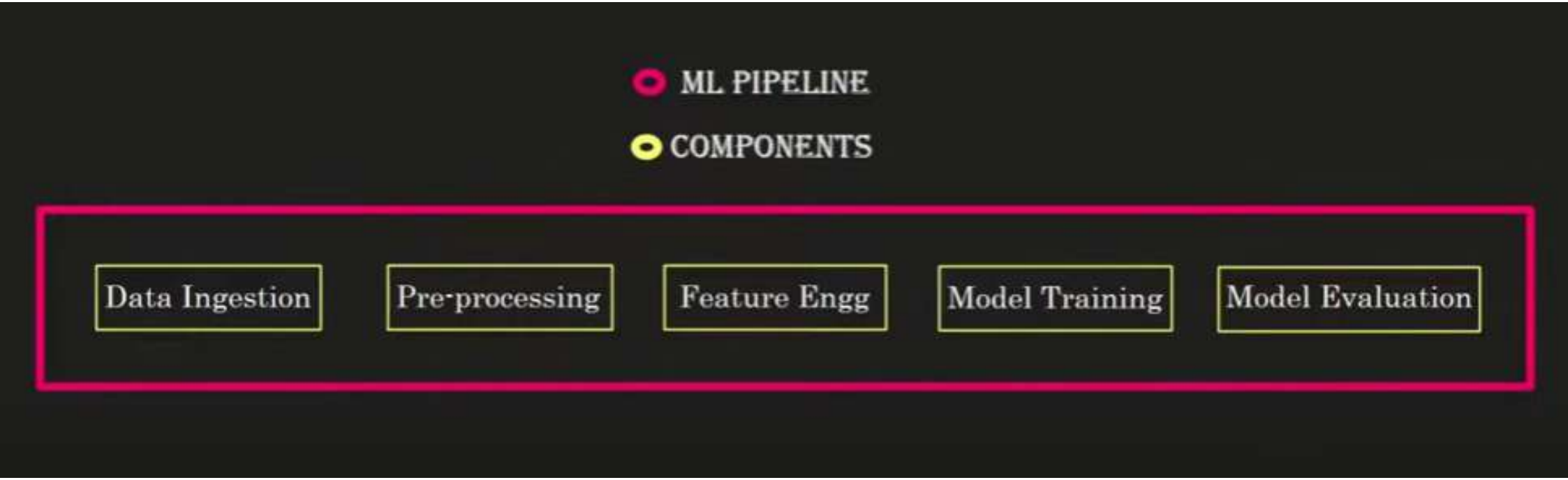
AGENDA

1. Create an end to end ML pipeline. (Logging, Exception etc)
2. Automate the pipeline using DVC. (YAML crash course)
3. Adding configurable params to pipeline.
4. Experiment tracking using dvclive.
5. AWS setup with DVC for data versioning.


ML PIPELINE
COMPONENTS
Data Ingestion
Pre-processing
Feature Engg
Model Training
Model Evaluation

DATA SCIENCE

1. Pre-processing
2. Feature Engineering
3. Model Hypertuning
4. Grid search etc

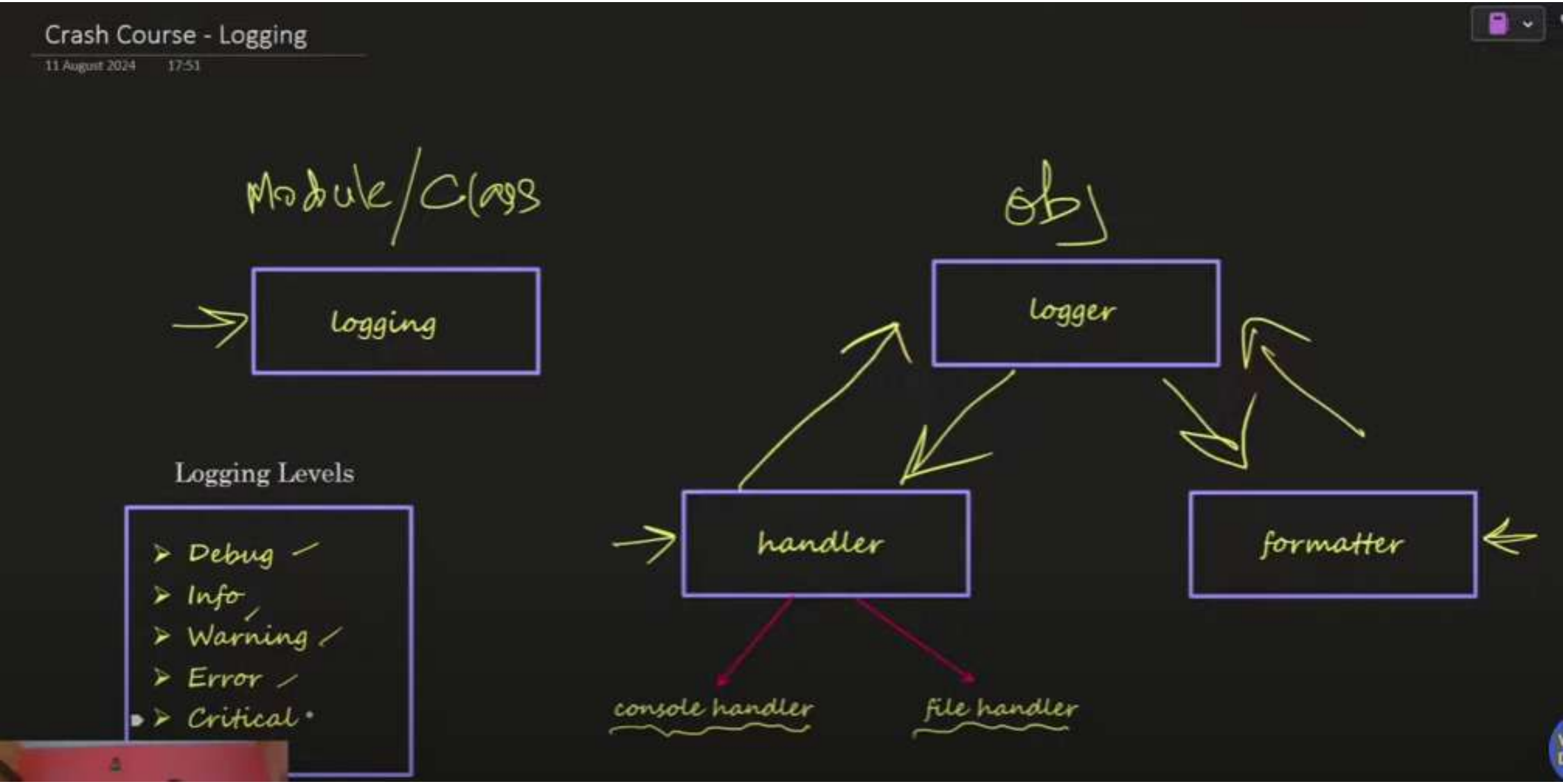
VS

MLOPS PRACTICE

1. Coding practices
2. Robust pipeline
3. Experiments
4. AWS, YAML etc

Crash Course - Logging

11 August 2024 17:51



File Edit View

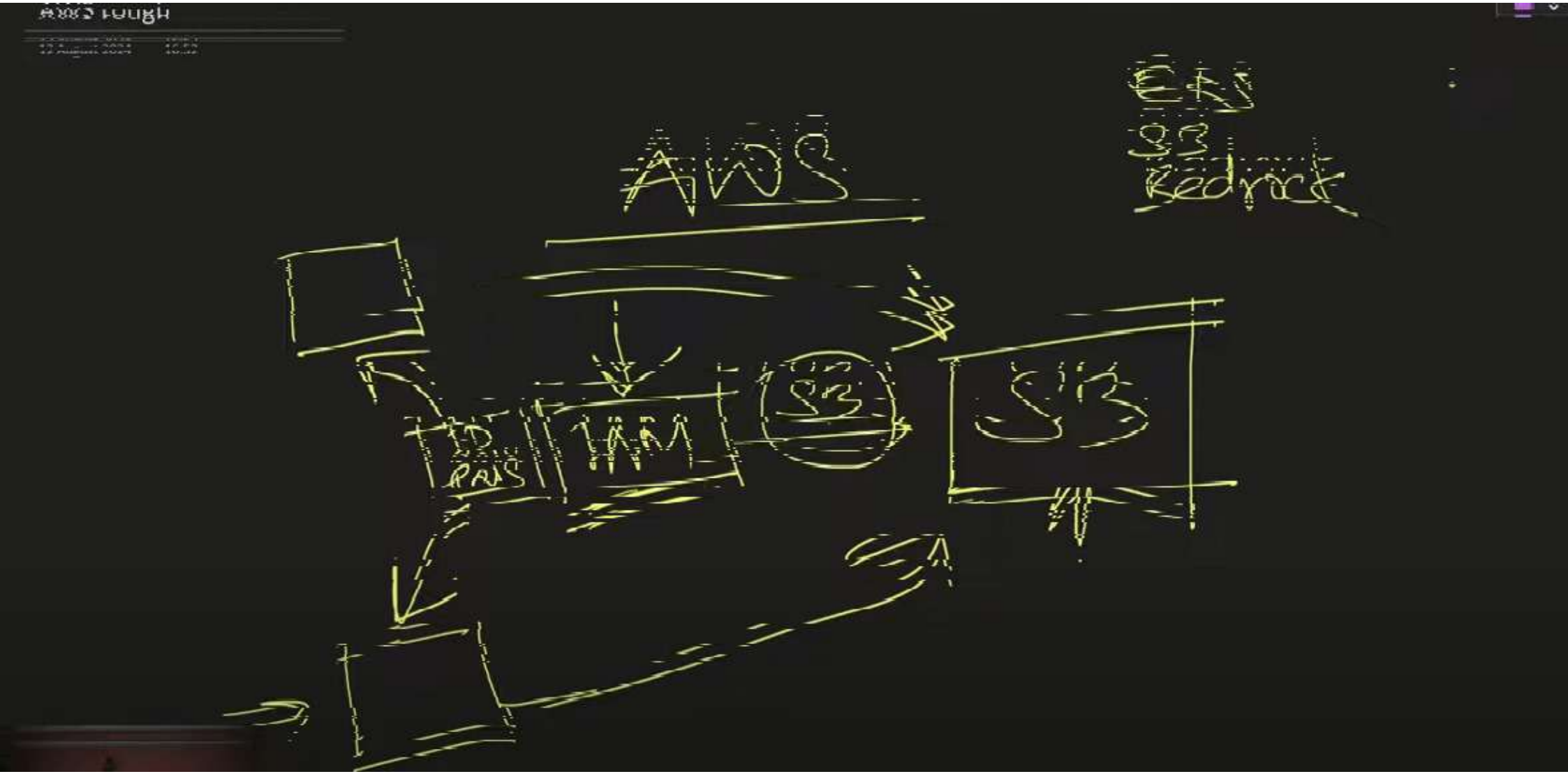
Building Pipeline:
1> Create a GitHub repo and clone it in local (Add experiments).
2> Add src folder along with all components(run them individually).
3> Add data, models, reports directories to .gitignore file
4> Now git add, commit, push

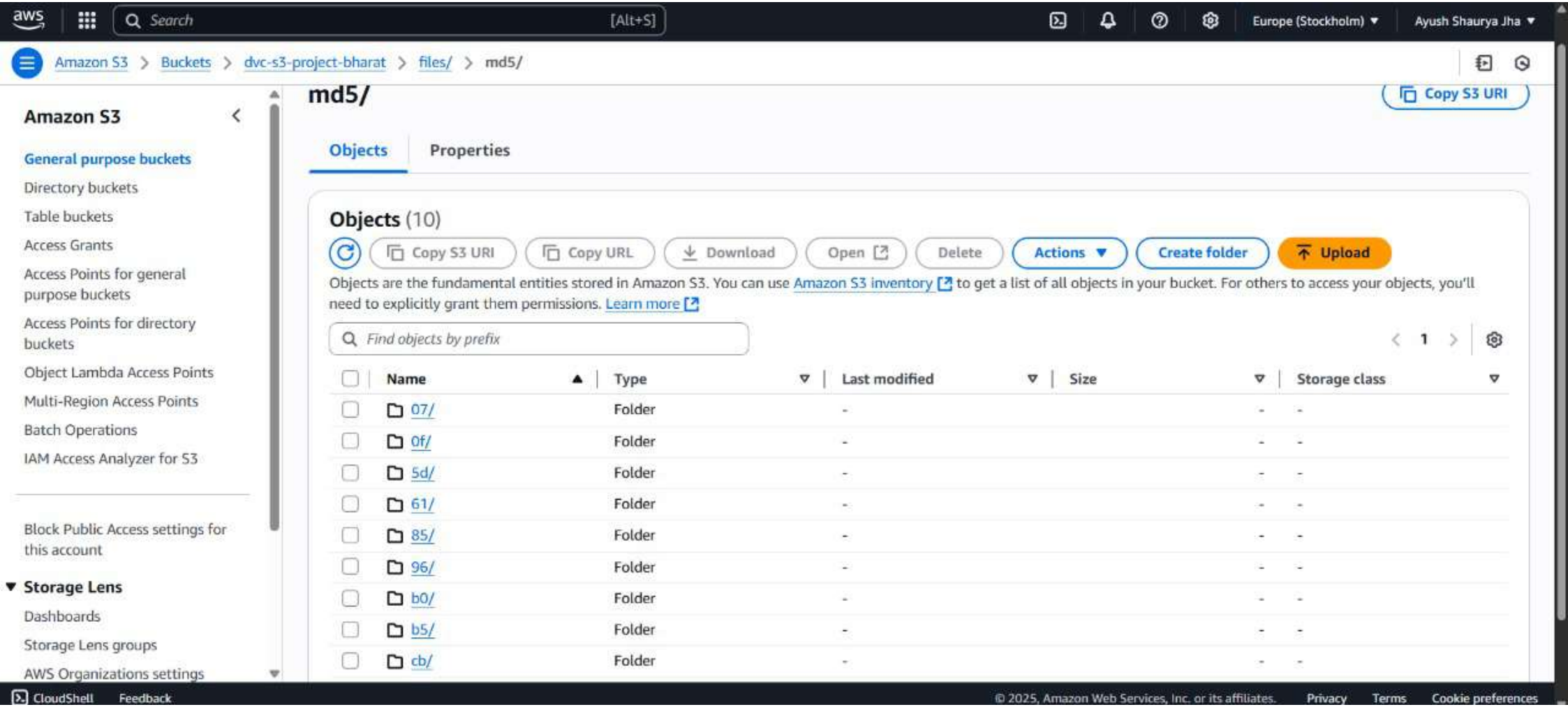
Setting up dcv pipeline (without params)
5> Create dvc.yaml file and add stages to it.
6> dvc init then do dvc repro to test the pipeline automation. (check dvc dag)
7> Now git add, commit, push

Setting up dcv pipeline (with params)
8> add params.yaml file
9> Add the params setup (mentioned below)
10> Do "dvc repro" again to test the pipeline along with the params
11> Now git add, commit, push

Expermients with DVC:
12> pip install dvclive
13> Add the dvclive code block (mentioned below)
14> Do "dvc exp run", it will create a new dvc.yaml(if already not there) and dvclive directory (each run will be considered as an experiment by DVC)
15> Do "dvc exp show" on terminal to see the experiments or use extension on VSCode (install dvc extension)
16> Do "dvc exp remove {exp-name}" to remove exp (optional) | "dvc exp apply {exp-name}" to reproduce prev exp
17> Change params, re-run code (produce new experiments)
18> Now git add, commit, push

Adding a remote S3 storage to DVC:
19> Login to AWS console
20> Create an IAM user (straight forward process)
21> Create S3 (enter unique name and create)
22> pip install dvc[s3]
23> pip install awscli
24> aws configure
25> dvc remote add -d dvcstore s3://bucketname
26> dvc commit-push the exp outcome that you want to keep
27> Finally git add, commit, push


AWS
S3
S3
Redirect


aws
Search
[Alt+S]
Europe (Stockholm)
Ayush Shaurya Jha
Amazon S3 > Buckets > dvc-s3-project-bharat > files/ > md5/
Amazon S3
General purpose buckets
Directory buckets
Table buckets
Access Grants
Access Points for general purpose buckets
Access Points for directory buckets
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
md5/
Copy S3 URI
Objects
Properties
Objects (10)
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
1
Name
Type
Last modified
Size
Storage class
07/ Folder
0f/ Folder
5d/ Folder
61/ Folder
85/ Folder
96/ Folder
b0/ Folder
b5/ Folder
cb/ Folder
CloudShell
Feedback
© 2025, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences