

# **AIOPS Assignment 3**

## **1. What is DVC, and why is DVC used?**

DVC stands for data version control. dvc file is a small text file that points to your actual data files in remote storage. The .dvc file is lightweight and meant to be stored with your code in GitHub. When you download a Git repository, you also get the.

DVC is built to make ML models shareable and reproducible. It is designed to handle large files, data sets, machine learning models, and metrics as well as code.

## **2. How is DVC different from git and GitHub?**

While Git is used to store and version code, DVC does the same for data and model files. Git can store code locally and also on a hosting service like GitHub, Bitbucket, or GitLab. Likewise, DVC uses a remote repository to store all your data and models.

## **3. Which command can be used to initialise a DVC project?**

To initialize DVC command is => `dvc init`

## **4. In What all use cases DVC can be used?**

Below are the use cases of DVC

1. Versioning data and models
2. CI/CD for machine learning
3. Fast and Secure Data Caching Hub
4. Experiment Tracking
5. Data Registry

## **5. Which command can be used to reproduce the entire pipeline?**

Command to reproduce pipeline => `dvc repro`

## **6. Which DVC command can be used to check metrics?**

Below are the 2 commands used for dvc metrics

```
dvc metrics show  
dvc metrics diff
```

## **7. Can we store a large amount of Data on GitHub? Justify.**

No, we cannot add large data files in github. Files that you add to a repository via a browser are limited to 25 MB per file. You can add larger files, up to 100 MB each, via the command line.