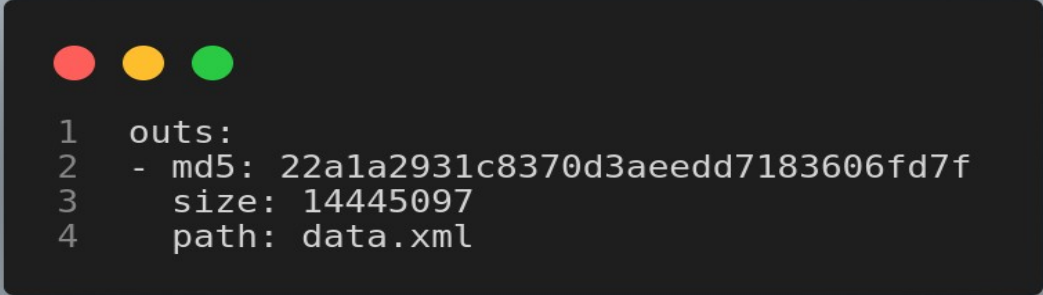


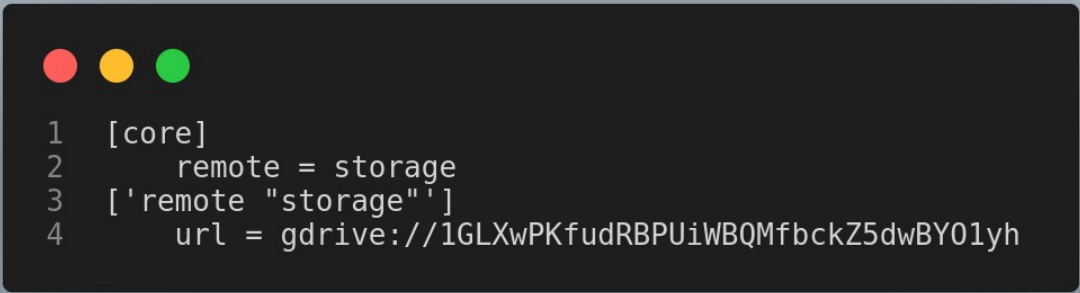
----DVC Walkthrough----

1) **data/data.xml.dvc** --> contains address of dataset where it is stored in cloud



```
1  outs:
2  - md5: 22a1a2931c8370d3aedd7183606fd7f
3    size: 14445097
4    path: data.xml
```

2) **.dvc/config** --> tells where our storage is



```
1  [core]
2    remote = storage
3  ['remote "storage"']
4    url = gdrive://1GLXwPKfudRBPUIWBQMfbckZ5dwBY01yh
```

3) Initializing git inside the working directory.

- **git init**

4) Initializing dvc inside the working directory.

- **dvc init**

5) Creating data folder

- **mkdir data**

6) Grab the data set from the registry and bring it to local machine

- **dvc get** <https://github.com/iterative/dataset-registry> get-started/data.xml -o data/data.xml

- 7) Add data to dvc tracking
 - **dvc add data/data.xml**
- 8) Follow given commands given in terminal then commit
 - **git add data/data.xml.dvc data/.gitignore**
 - **git commit -m "Add raw data"**
- 9) Adding dvc remote to add data into gdrive storage
 - **dvc remote add -d storage gdrive://folderlink**
- 10) Commit dvc config file and push
 - **git commit .dvc/config -m "Configure remote storage"**
 - **dvc push**
- 11) New file will appear inside gdrive remote folder with random name.
- 12) **data.xml.dvc** will go inside the git repository that will help us to point where the file is in our dvc repository i.e google drive.
- 13) Check git logs
 - **git log --oneline**
- 14) Checkout (removing) of old (first) version of dvc file
 - **git checkout HEAD^1 data/data.xml.dvc**
 - **dvc checkout**
 - **git commit data/data.xml.dvc -m "Revert dataset updates"**