DVC Cheat Sheet

Basics

Initializing

Initialize a DVC environment

```
$ dvc init
```

Remote

Set up a remote to keep and share data files

```
$ dvc remote add -d myremote /path
```

Possible remotes include local, s3, gs, azure, ssh, hdfs and http.

Show all available remotes

```
$ dvc remote list
```

Modify remote settings

```
$ dvc remote modify myremote
```

Use if remote requires extra configuration

Adding Files

Add files under DVC control

```
$ dvc add filename
```

Use --no-commit to stop adding the file to the cache.

Share Data

Push all data files to the remote storage

```
$ dvc push
```

Push outputs of a specific .dvc file

```
$ dvc push filename.dvc
```

Retrieve Data

Download files from the remote storage

```
$ dvc pull
```

Download files from a specific .dvc file

```
$ dvc pull filename.dvc
```

Checkout files from cache into working space

```
$ dvc checkout
```

The Pipeline

Add transformations and generate a stage file from a given command

```
$ dvc run -d dependencyfile \
    -o outputfile python command.py
```

Use --file to specify the name of the generated .dvc file.
Use --metrics to output a file containing the metric.

Metrics

Collect and display project metrics

```
$ dvc metrics show
```

Use --all to show the metrics in all branches.

Visualizing

Show stages in a pipeline

```
$ dvc pipeline show --ascii file.dvc
```

Add --commands or --outs to show more detail.

Show connected pipelines of DVC stages

```
$ dvc pipeline list
```

Reproducing

Reproduce outputs defined in .dvc file

```
$ dvc repro filename.dvc
```

Name a .dvc file "Dvcfile" to be use by dvc repro by default

Other Commands

Set/unset cache directory location

```
$ dvc cache dir /path
```

Commit outputs to cache

```
$ dvc commit
```

Use if you specified --no-commit in dvc add/run/repro

Config repository or global options

```
$ dvc config
```

Config the default remote using core.remote myremote
Config core (loglevel, remote), cache and state settings

Fetch files from the remote to the local cache

```
$ dvc fetch file.dvc
```

Remove unused objects from cache

```
$ dvc gc
```

Import file from URL to local directory

```
$ dvc import url /path
```

Supported schemes include local, s3, gs, azure, ssh, hdfs and http.

Remove data files tracked by dvc

```
$ dvc remove filename.dvc
```

Show changed stages in the pipeline

```
$ dvc status
```