Kaggle Tutorial

Git Basics

Pandas Basics

Visualization Basics

Practice along with Kernel

https://www.kaggle.com

We use cookies on kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using kaggle, you agree to our use of cookies. Got it Learn more

kaggle Search Competitions Datasets Kernels Discussion Learn ••• Sign In

Online Community
For Data Science

Kaggle is the place to do data science projects

See how it works

Register with just one click:

We won't share anything without your permission

Sign up with Google

Sign up with Facebook

Manually create an account:

Email

Password

Register

Start a new project

Explore projects created by others

Join one of our competitions

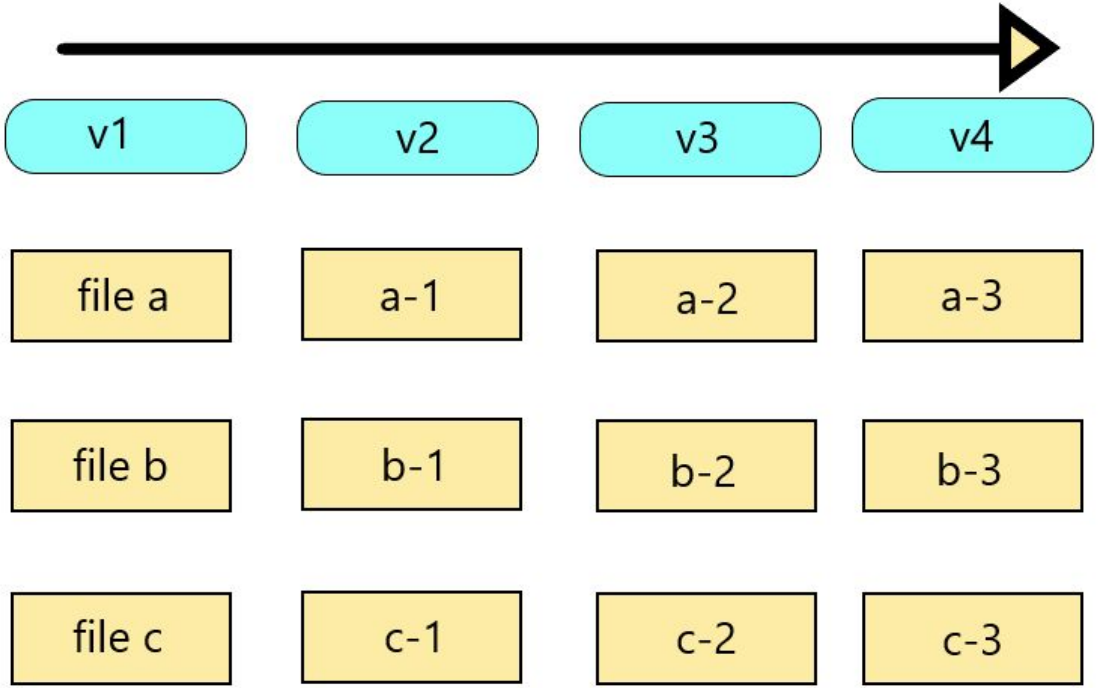
git

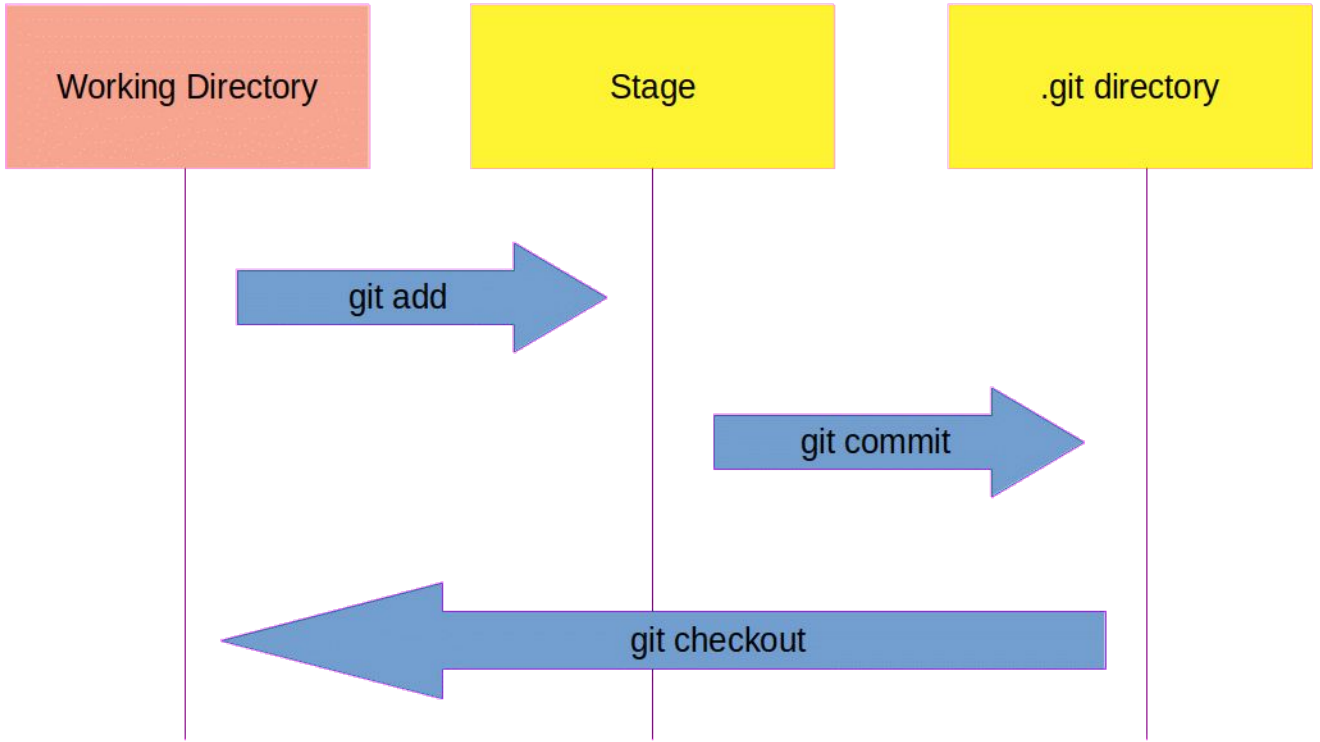
from 2005

Linus Torvalds

github, gitlab,
bitbucket등의 서비스가

git은 아님.


v1
v2
v3
v4
file a
a-1
a-2
a-3
file b
b-1
b-2
b-3
file c
c-1
c-2
c-3


Working Directory
Stage
.git directory
git add
git commit
git checkout

git config --global user.name “Cornelii Son”

git config --global user.email “xxxxxxxx@gmail.com”

git init

git clone {address}

git status

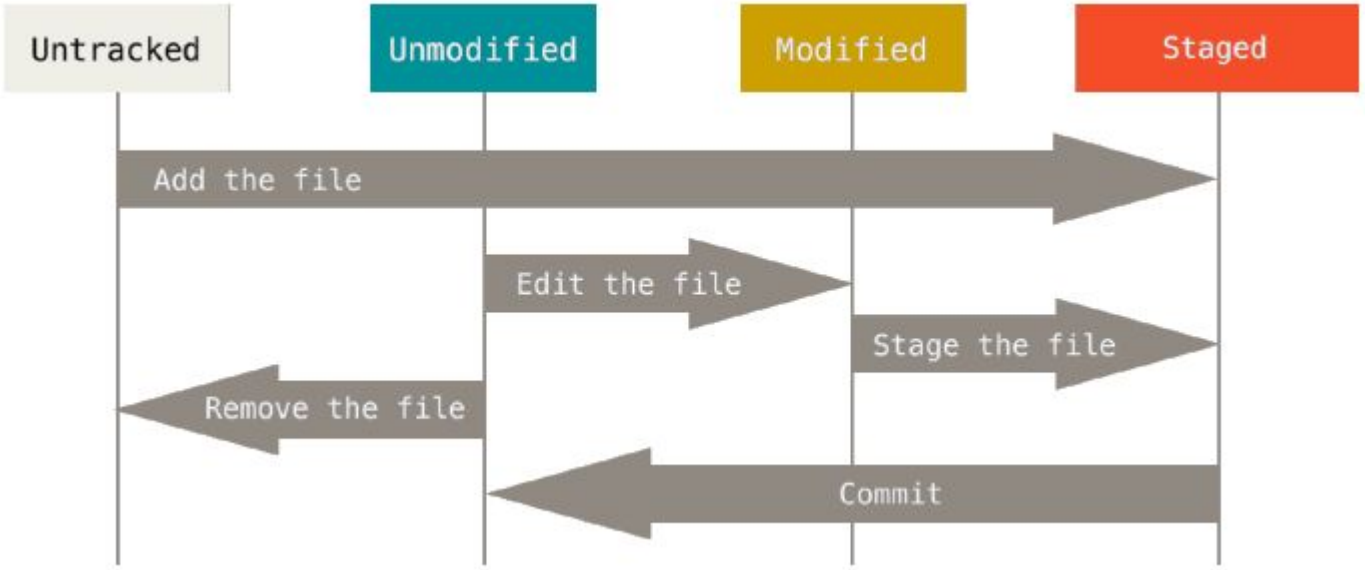
git add (<=> git reset)

git commit -m “comment”

git log

git log -p -5 --graph

File classification in Git



Remote Repository

git remote add [name] [url]

git remote -v

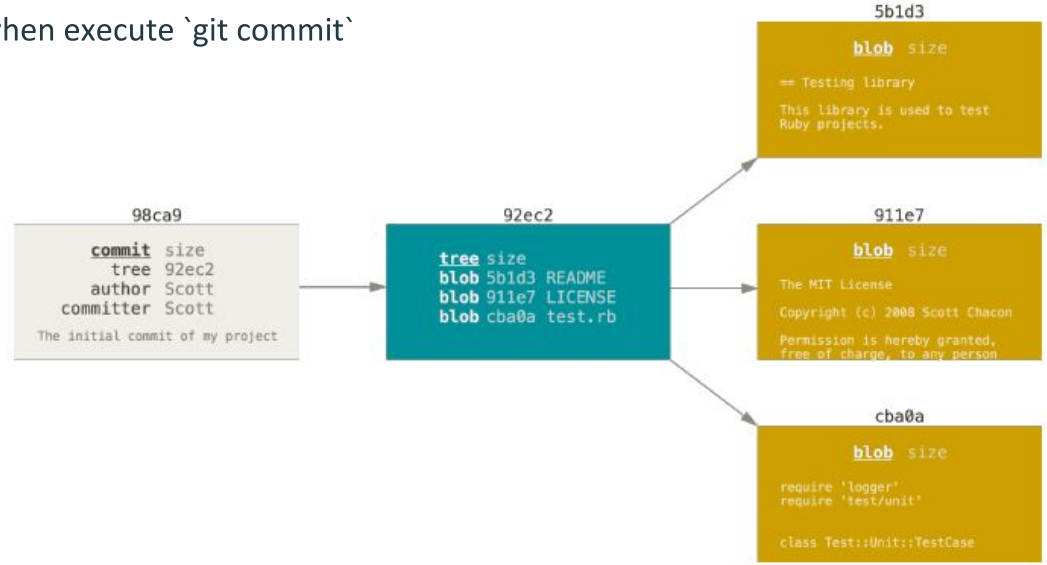
git fetch & git merge

git pull

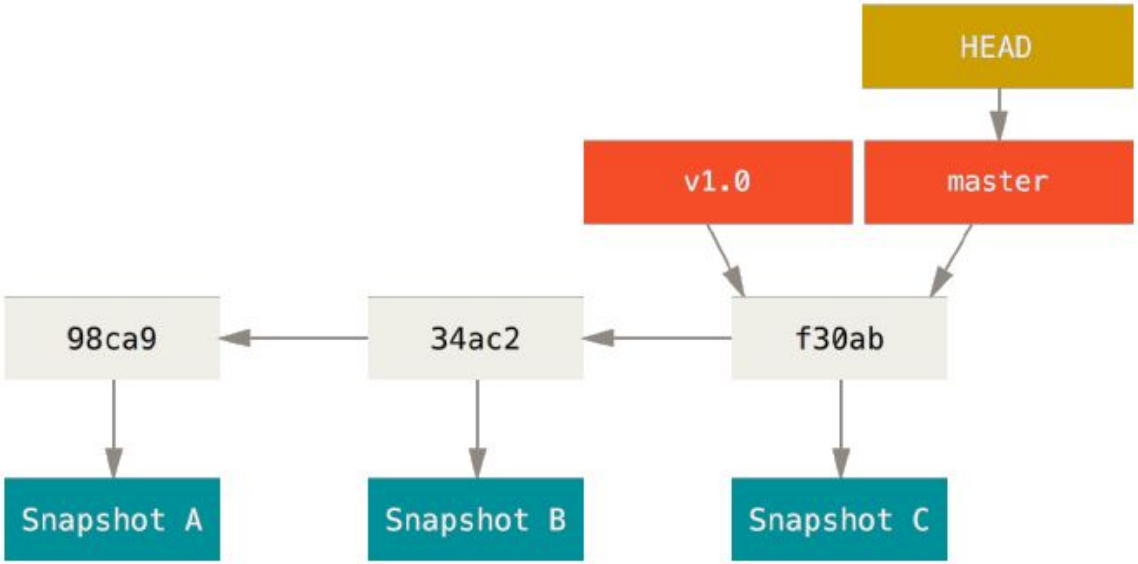
git push -u [remote repo-name] [branch name]

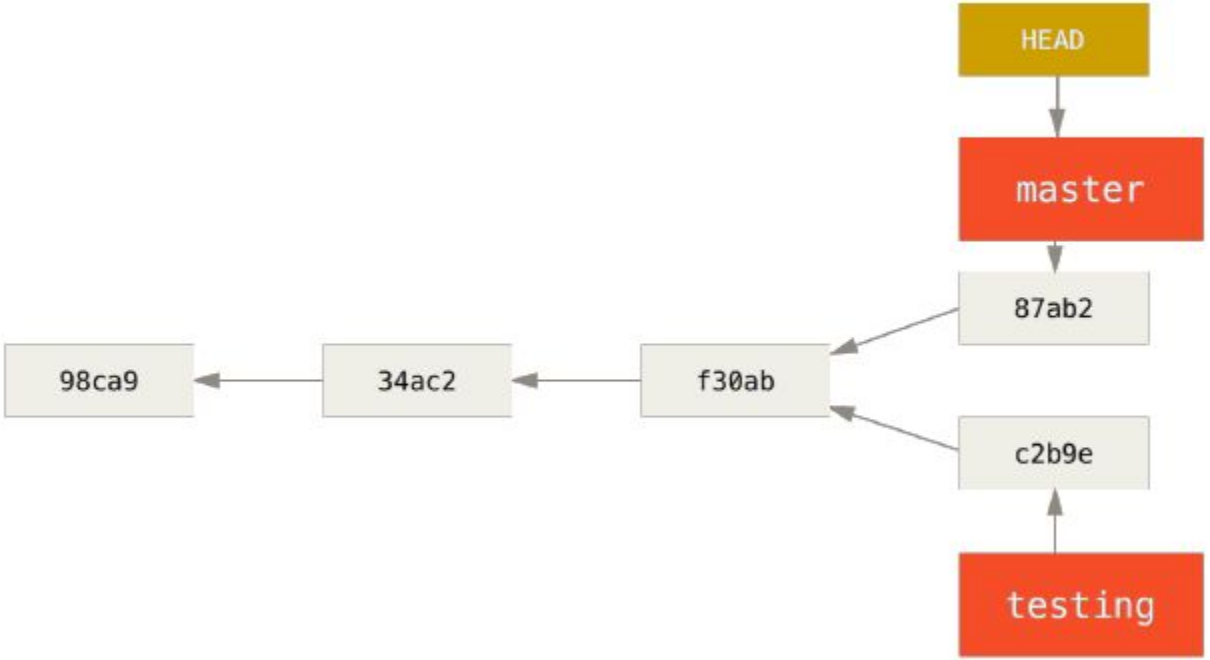
Understanding Branch

- What happens when execute `git commit`



Branch is like pointer to commits




HEAD
master
87ab2
98ca9
34ac2
f30ab
c2b9e
testing

git branch [branch-name]

git branch

git checkout [branch-name]

git branch -d [branch-name]

git merge [branch2_name]

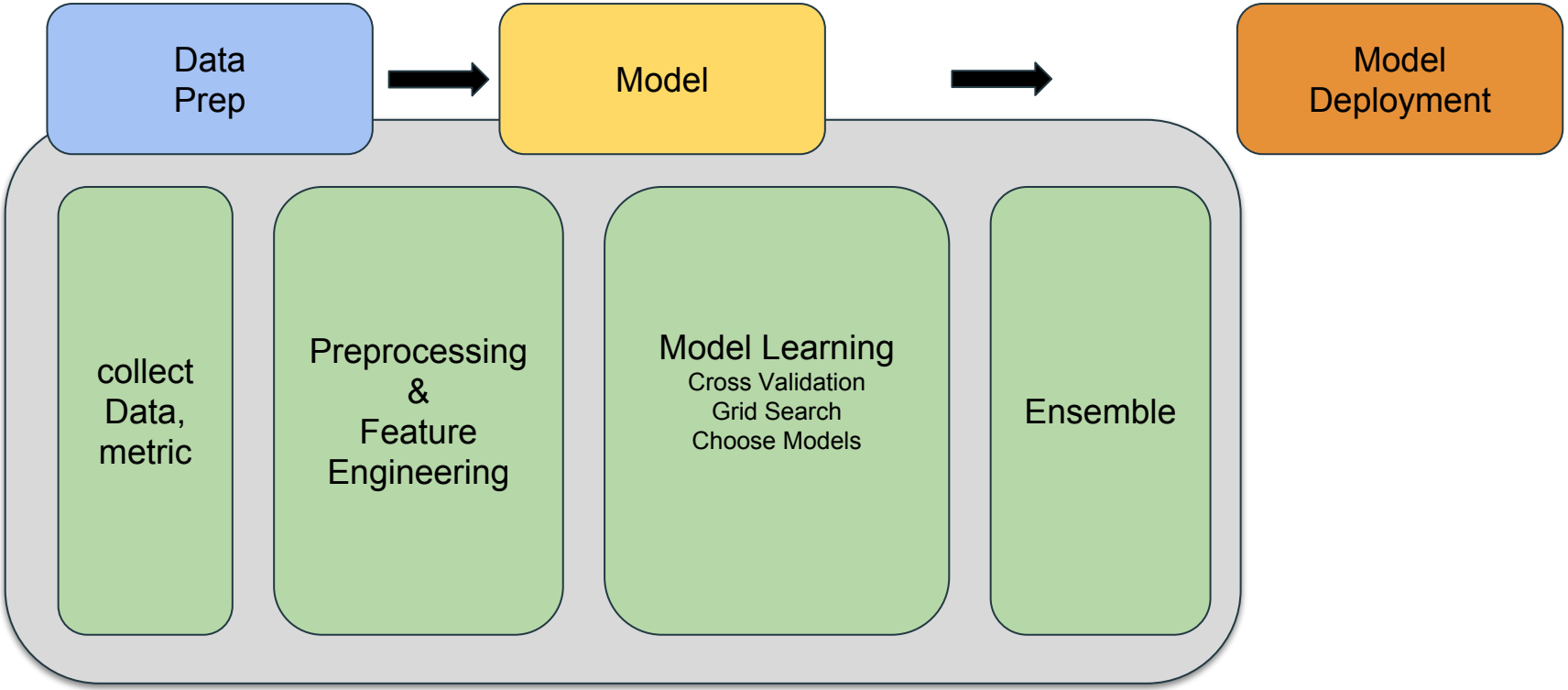
at the branch1 (merge branch2 -> branch1)

git branch -d [branch-name]

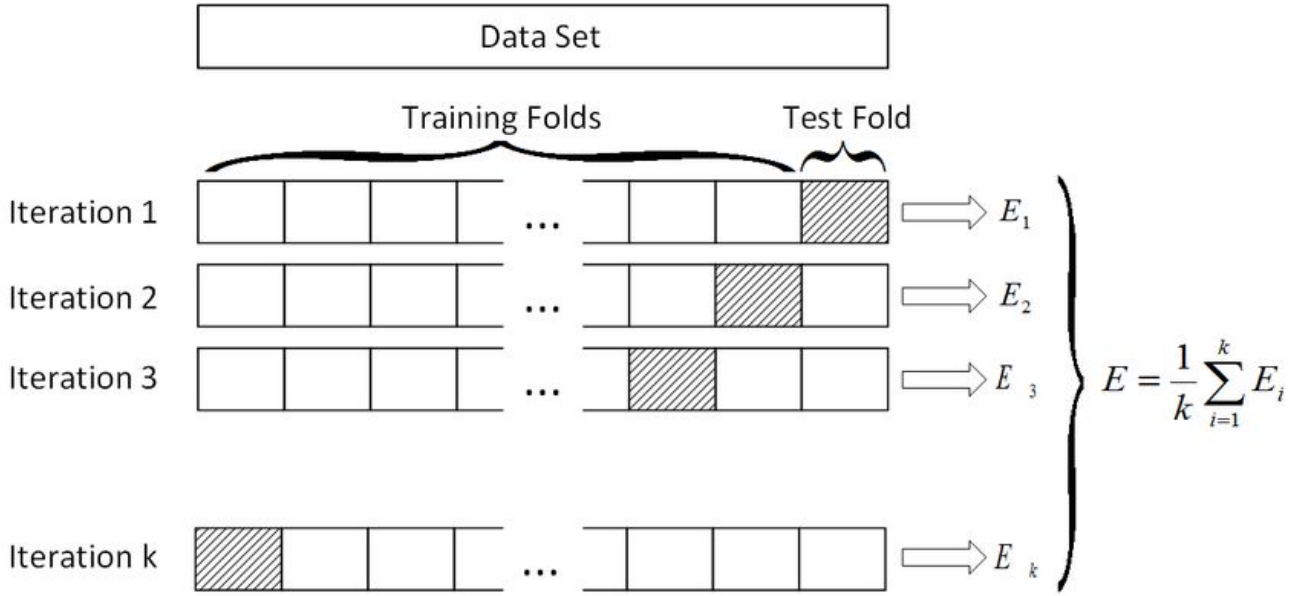
git push [remote-name] [branch-name]

Pandas and Visualization
Hands-on Practice!

Data Science?!

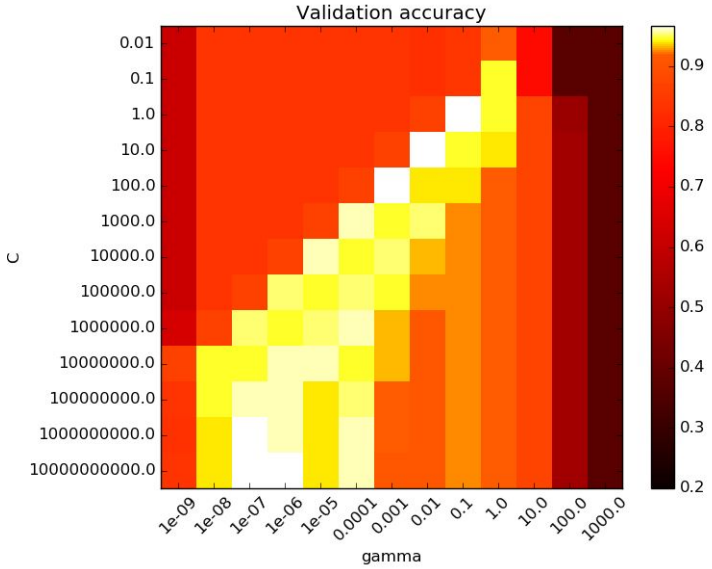


Cross Validation



Grid Search

Procedure to find optimized parameters.



You have to consider Computational costs.

Hands-on Practice

1. EDA and Data preprocessing using pandas, matplotlib.pyplot, seaborn, etc.
2. Use of sklearn
3. Submit data to the kaggle competition.

Data: Home Sales.