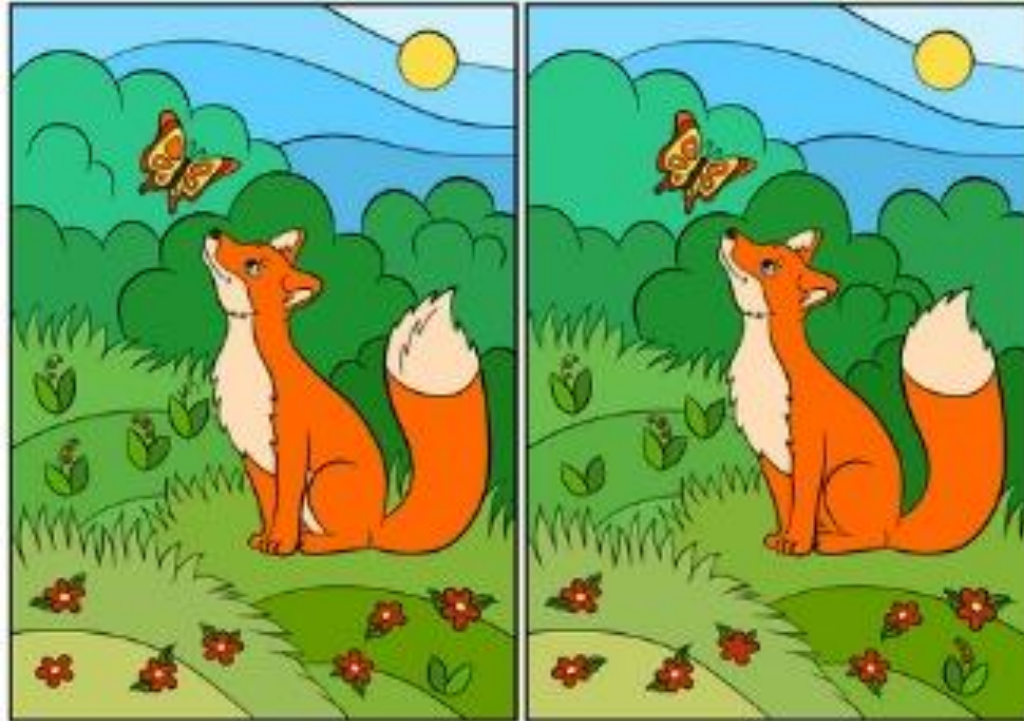


Version control



The Data Scientist's Toolbox

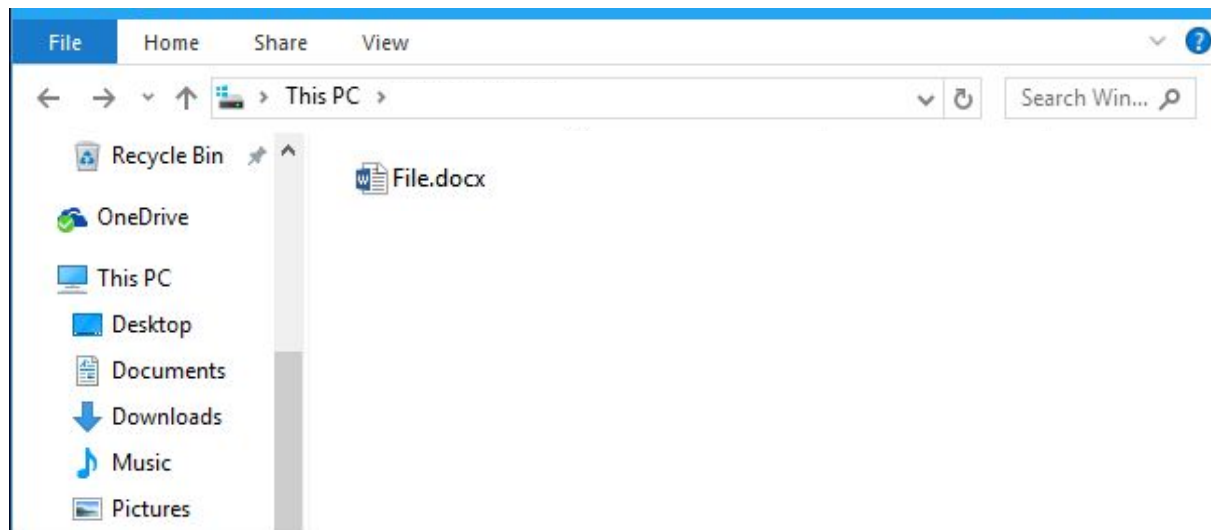
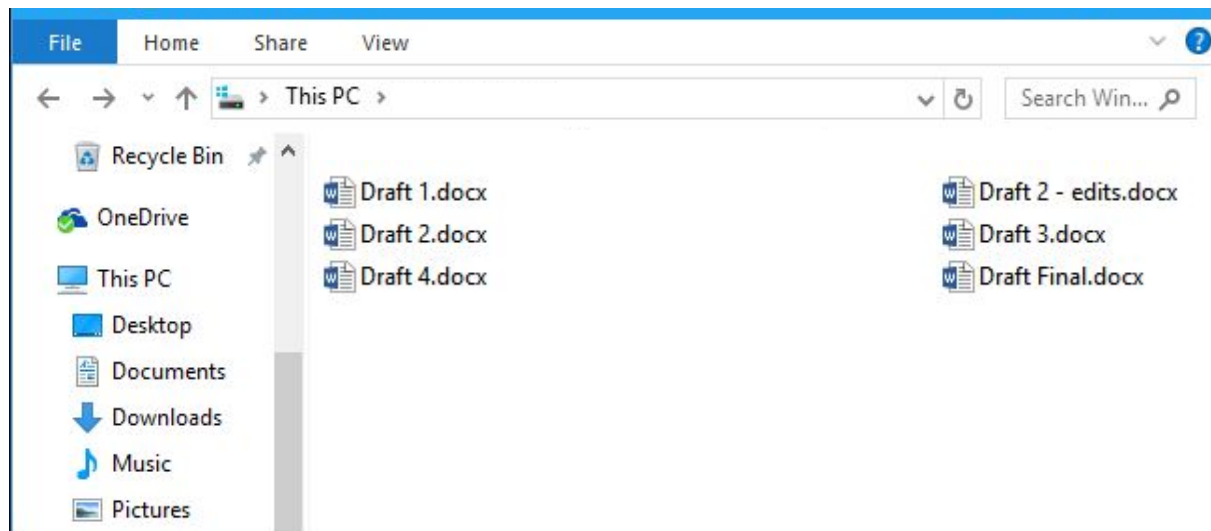
Find 12 differences



amet, consectetur adipiscing elit. In suscipit est in dolor adipiscing, vitae dictum
sequat id mi vel ultrices. Maecenas ~~varius augue in dui~~ tincidunt volutpat.
us auctor tristique quis id enim. Nam laoreet pharetra nibh, sit amet consequat
Aenean eu eros sit amet lorem ultrices ~~lorem ipsum condimentum~~ eu ut
nvalis augue et enim placerat aliquet. Nunc in eros pulvinar, viverra quam ac,
elit nisl. Sed eu pellentesque tellus, nec aliquet ante.

n accumsan. Sed in vulputate arcu, quis faucibus dui. Etiam vitae mauris quam.
mentum nisi. Phasellus nec eleifend nisl. Vestibulum ~~quis~~ libero quis urna
is orci. Nam pharetra lacus nec mi ullamcorper pellentesque. Etiam a placerat
suada fames ac ante ipsum primis in faucibus.

que elit, sed sodales elit egestas id. ~~Nulla aliquet, turpis sit amet malesuada
a tempor dui, vel scelerisque nibh urna eu nisl.~~ Nullam iaculis cursus leo non
a tortor, pretium ut risus in, rutrum vehicula magna. Donec ornare felis odio,
tor tincidunt et. Curabitur nec tellus nisl. Cras vel commodo nisl, eget facilis



RStudio: Review Changes

ChangesHistorymaster (all commits)G

SearchPull

Initial commit of lecture 18, Big data (no script)	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	ec9f5a78
Initial commit of lecture 16, Data science questions (no script)	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	e01f81d0
Initial commit of Lecture 14, Projects under version control (no script)	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	ab1d8e53
Initial commit of Lecture 13, Linking Git, GitHub, and RStudio	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	f98618ab
Initial commit of Lecture 12, GitHub and Git (no script)	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	6f2a13e9
Initial commit of Lecture 10, Projects in R	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	6dfd7958
Updated with finished lectures to date	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	135b736e
Exchanging XKCD images for links + small edits	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	9a157dfb
Small update to summary	Jane Everyday Doe <Jane.Everyday.Doe@gn 2018-05-02	84e4aebc

Commits 1-100 of 102

SHAec9f5a78

AuthorJane Everyday Doe <Jane.Everyday.Doe@gmail.com>

Date2018-05-02 21:08

SubjectInitial commit of lecture 18, Big data (no script)

Parente01f81d0

manuscript/18_Big_data.md

manuscript/18_Big_data.md

View file @ ec9f5a78

@@ -0,0 +1,58 @@

1# Big Data

2

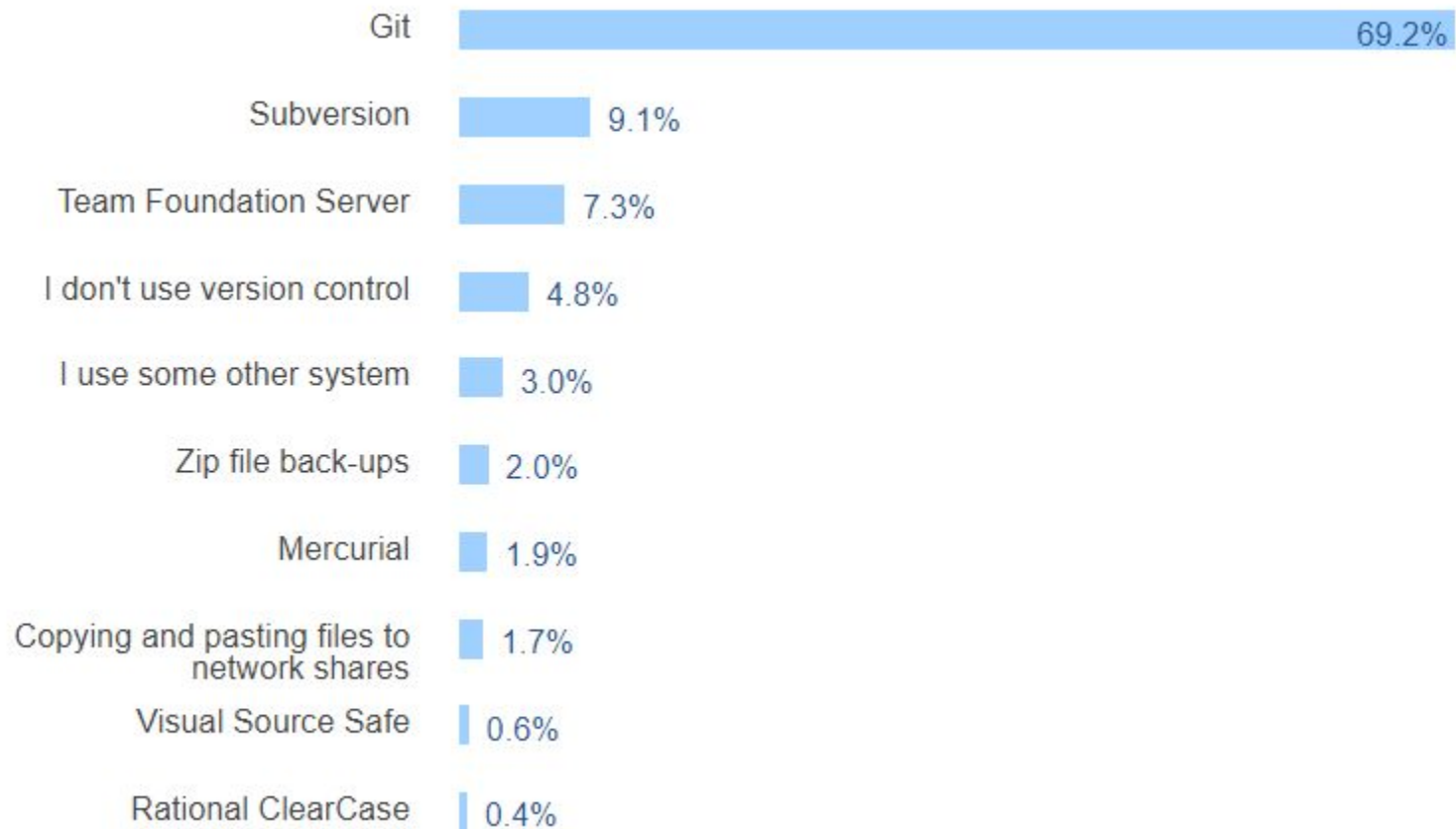
3A term you may have heard of before this course is "Big Data" - there have always been large datasets, but it seems like lately, this has become a buzzword in data science. But what does it mean?

4

5### What is big data?

6

7We talked a little about big data in the very first lecture of this course. As the name suggests, big data are very large data sets. We previously discussed three qualities that are commonly attributed to big data sets: Volume, Velocity, Variety. From these three adjectives, we can see that big data involves large data sets of diverse data types that are being generated very



GitHub



Repository, AKA Repo



Repository, AKA Repo



Commit



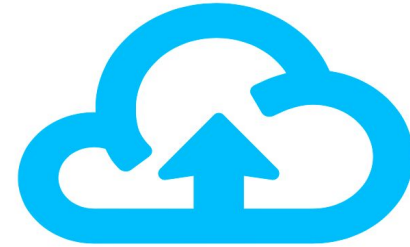
Repository, AKA Repo



Commit



Push



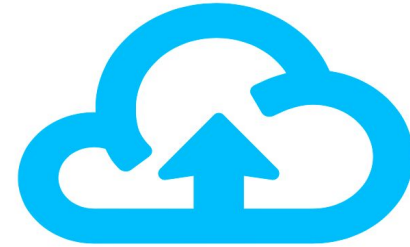
Repository, AKA Repo



Commit



Push



Pull



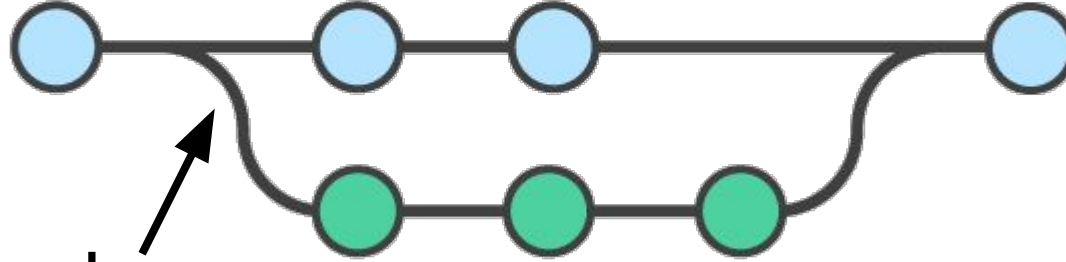
Files are hosted in a **repository** that is shared online with collaborators. You **pull** the repository's contents so that you have a local copy of the files that you can edit. Once you are happy with your changes to a file, you **stage** the file and then **commit** it. You **push** this commit to the shared repository. This uploads your new file and all of the changes and is accompanied by a message explaining what changed, why and by whom.



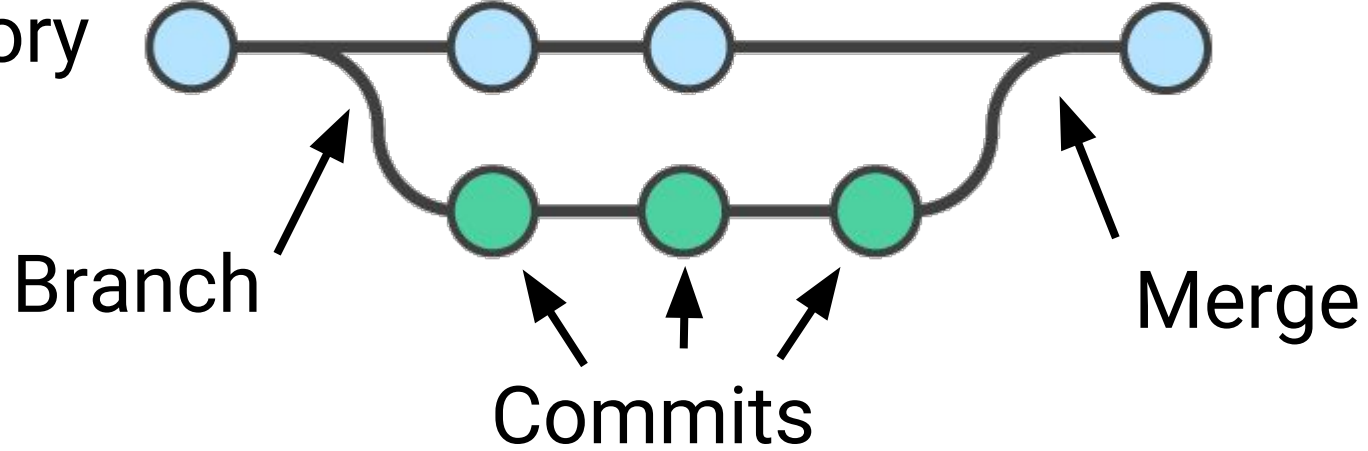
Repository

Branch

Commits



Repository

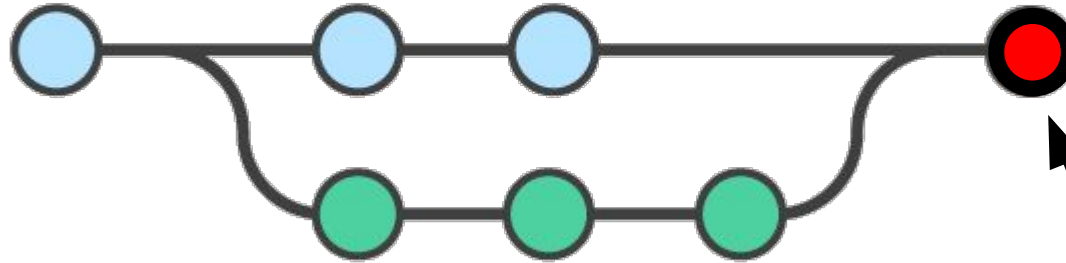
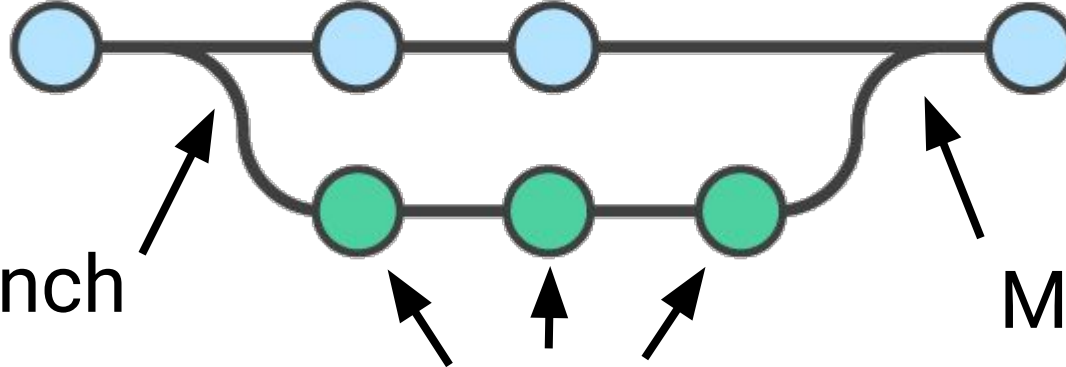


Repository

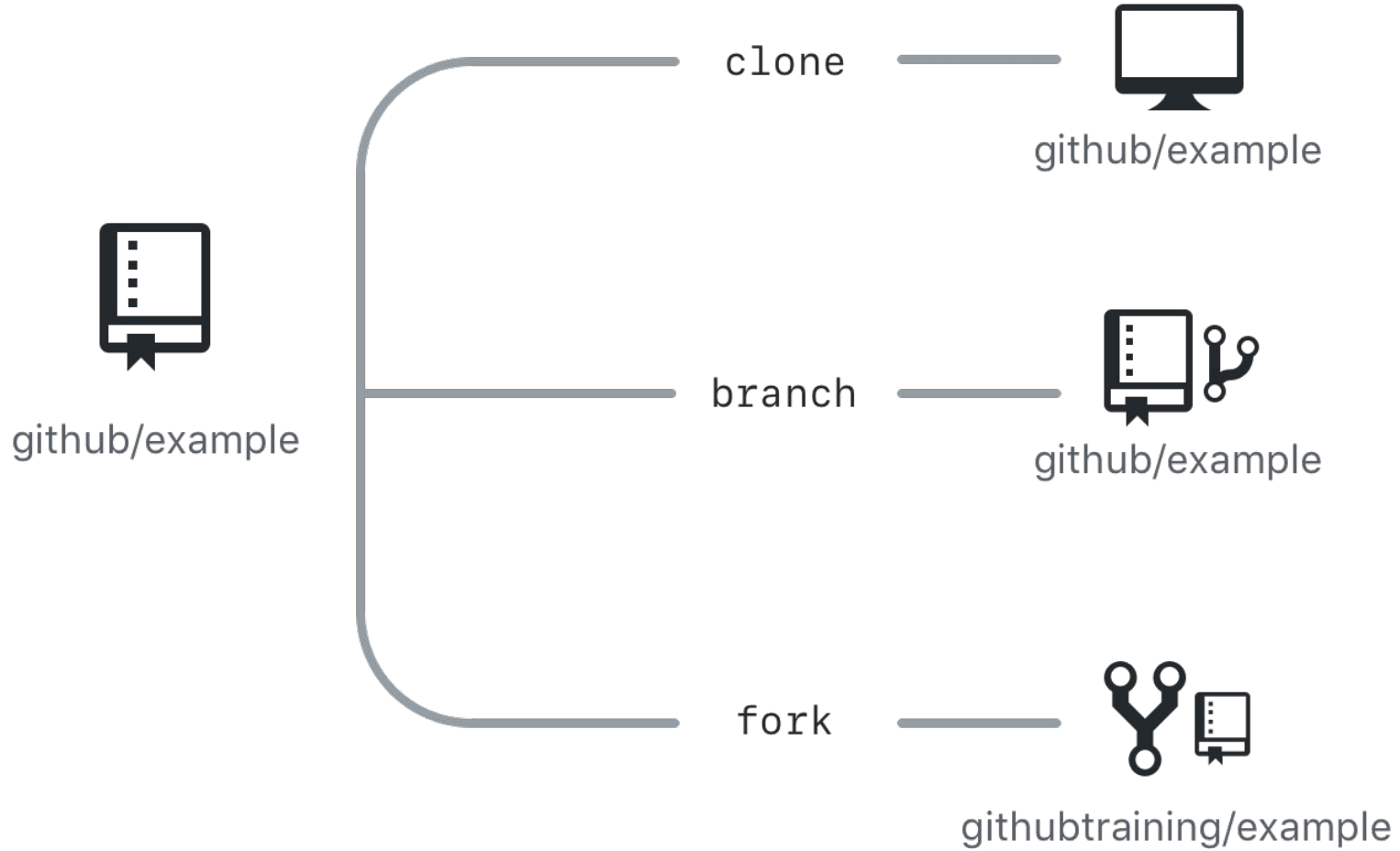
Branch

Commits

Merge

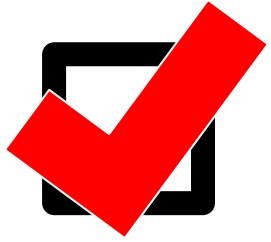


Conflict

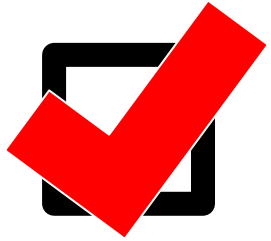




Purposeful, single
issue commits



Informative commit
messages



Pull and push often

Summarizing: Version control



The Data Scientist's Toolbox