VERSION CONTROL WITH GIT AND GITHUB

CDIPS Data Science Pre-Workshop
Monday, July 7, 2014

PRE-WORKSHOP AGENDA

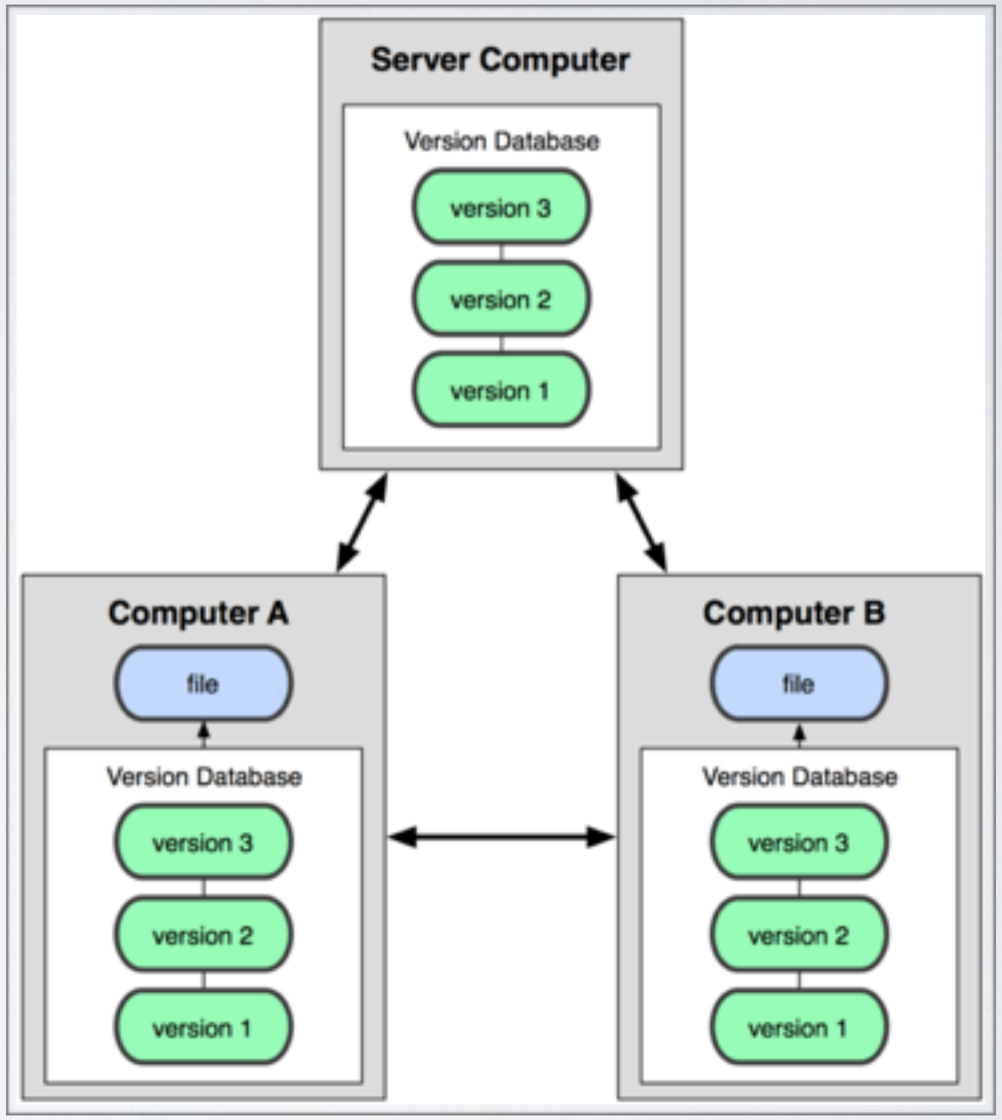
- 5 day agenda and recommended tutorials can be found on http://cdips.physics.berkeley.edu/agenda/
  - Day 1: Version Control with Git and GitHub
  - Day 2: Python Installation and Introduction
  - Day 3: Essential Python Libraries [numpy, scipy]
  - Day 4: **Managing Data with Pandas and SQL**
  - Day 5: **Machine Learning with Scikit-Learn**

TODAY'S OUTLINE

- Welcome
- PreWS Overview
- **Introduction to Git**
- Install Git
- Basic Git Commands
- Cloning a GitHub Repo
- Sign-Up for GitHub Account
- Group Activities

INTRODUCTION TO GIT

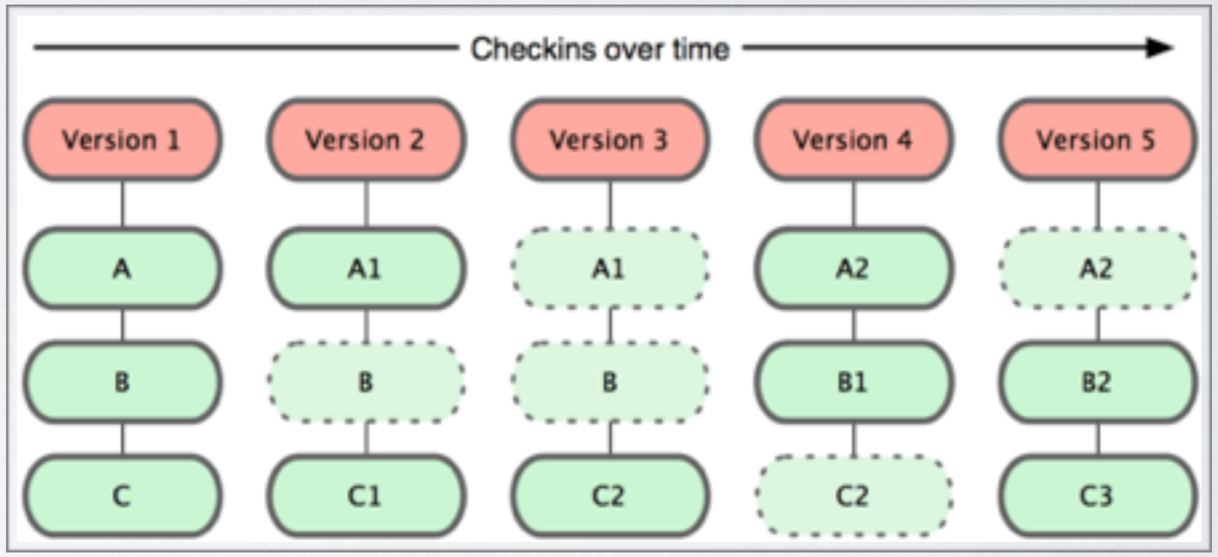
- What is **version control**?
  - "Version control is a system that records changes to a file or set of files over time so that you can recall specific versions later."
  - "A VCS [Version Control System] allows you to: revert files back to a previous state, revert the entire project back to a previous state, review changes made over time, see who last modified something that might be causing a problem, who introduced an issue and when, and more."
  - http://git-scm.com/book/en/Getting-Started-About-Version-Control
- Git Basics. What is Version Control?



Distributed Version Control System

INTRODUCTION TO GIT

- What is **git**, in a nutshell?
  - Version control
  - Fast
  - Simple
  - Strongly supports non-linear development (thousands of branches)
  - Fully distributed
  - Able to handle large projects efficiently
  - http://git-scm.com/book/en/Getting-Started-A-Short-History-of-Git
- Git Basics. What is Git?



Git's approach is to take snapshots of the working directory at every commit

SETTING UP GIT

TODAY'S OUTLINE

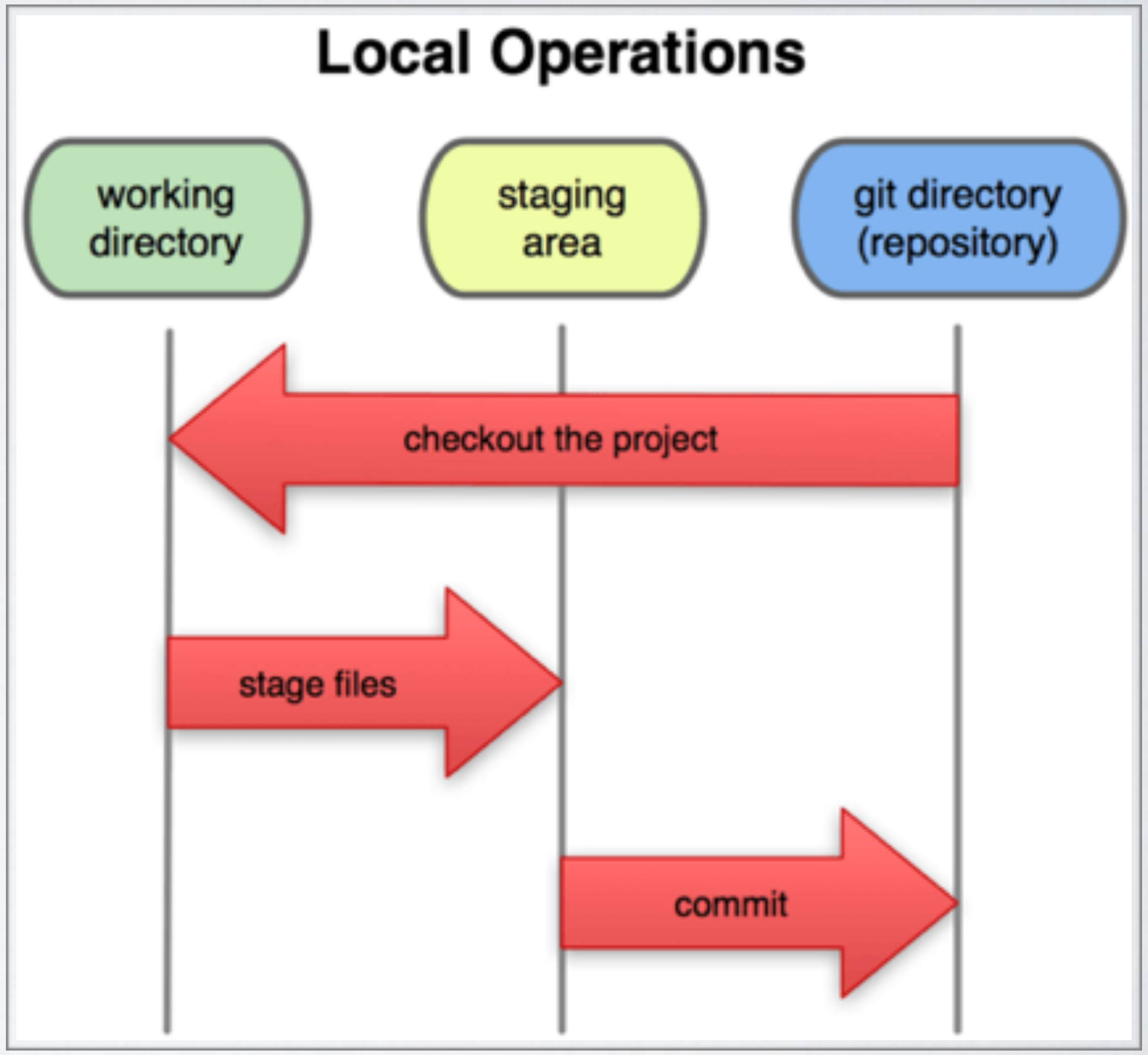
- Welcome
- PreWS Overview
- Introduction to Git
- Install Git
- **Basic Git Commands**
- Cloning a GitHub Repo
- Sign-Up for GitHub Account
- Group Activities

BASIC GIT COMMANDS 1/4

- Repository Basics
  - git init:
    - Initialize git repository
  - git config -l:
    - List configured git/repo parameters
  - git status:
    - Check status of the repository
  - git log:
    - See list of commits

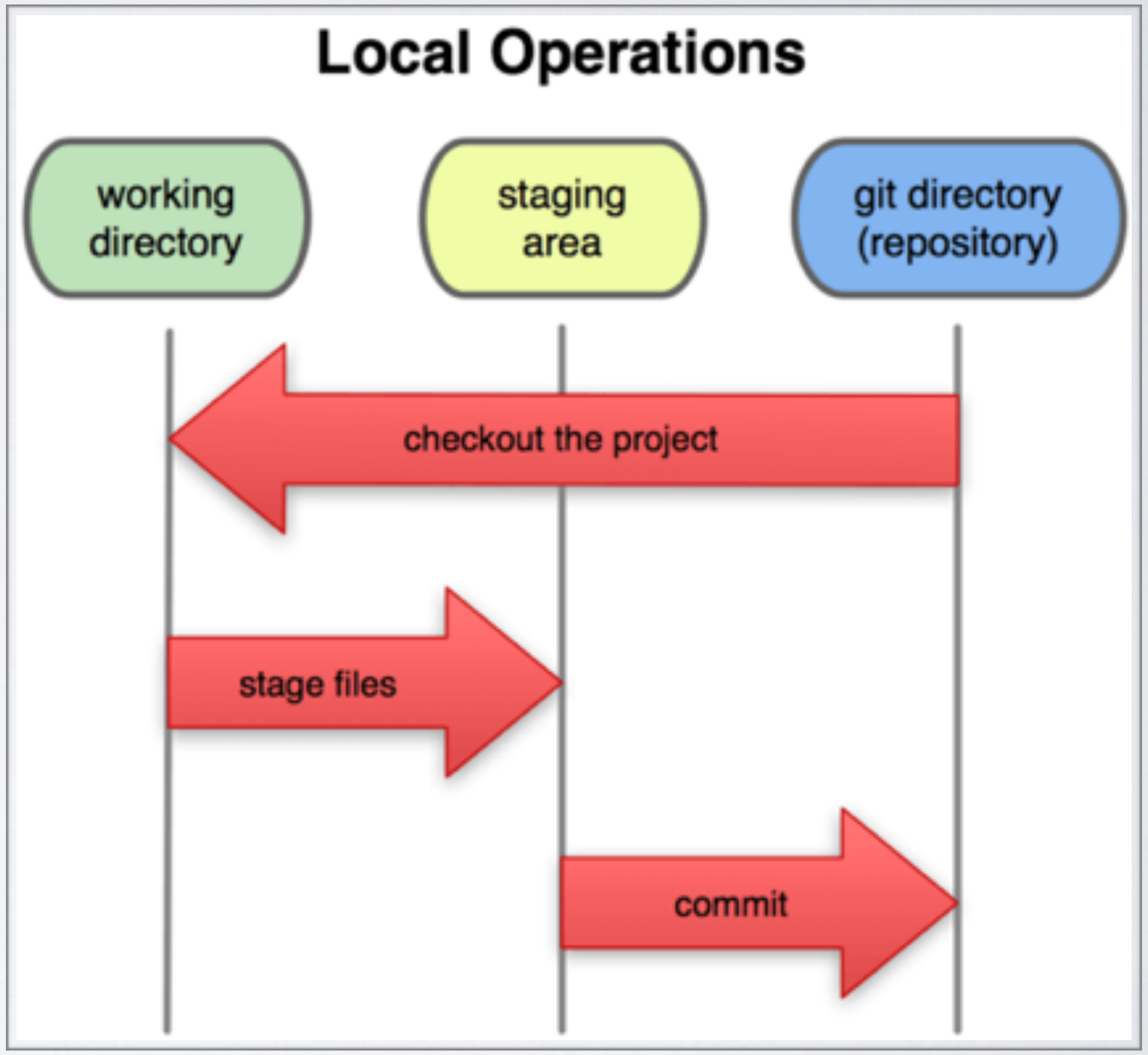
MAKING CHANGES WITH GIT

- We use Git in three stages.
  - Working directory
  - Staging area
  - Git Directory
  - http://git-scm.com/book/en/Getting-Started-Git-Basics



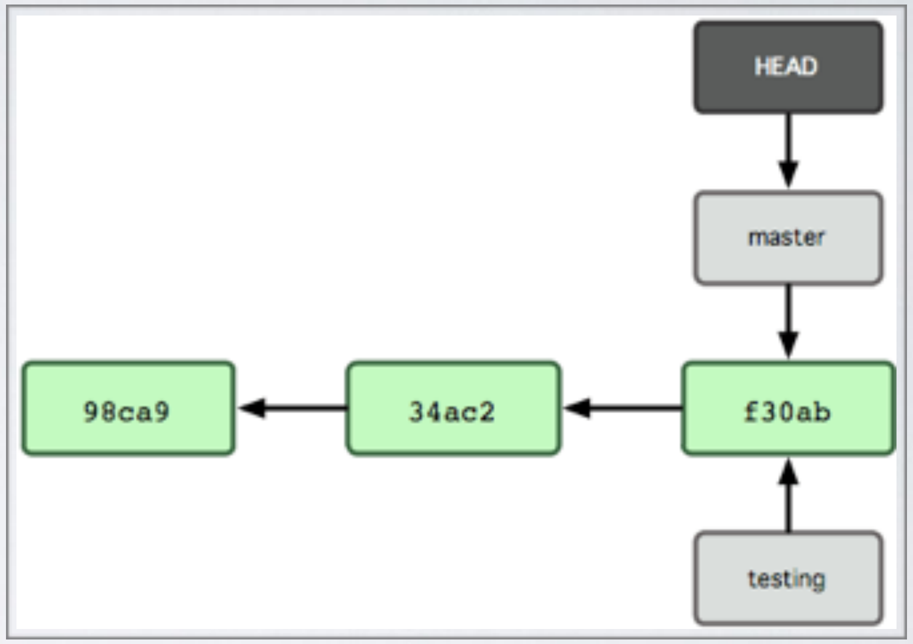
BASIC GIT COMMANDS 2/4

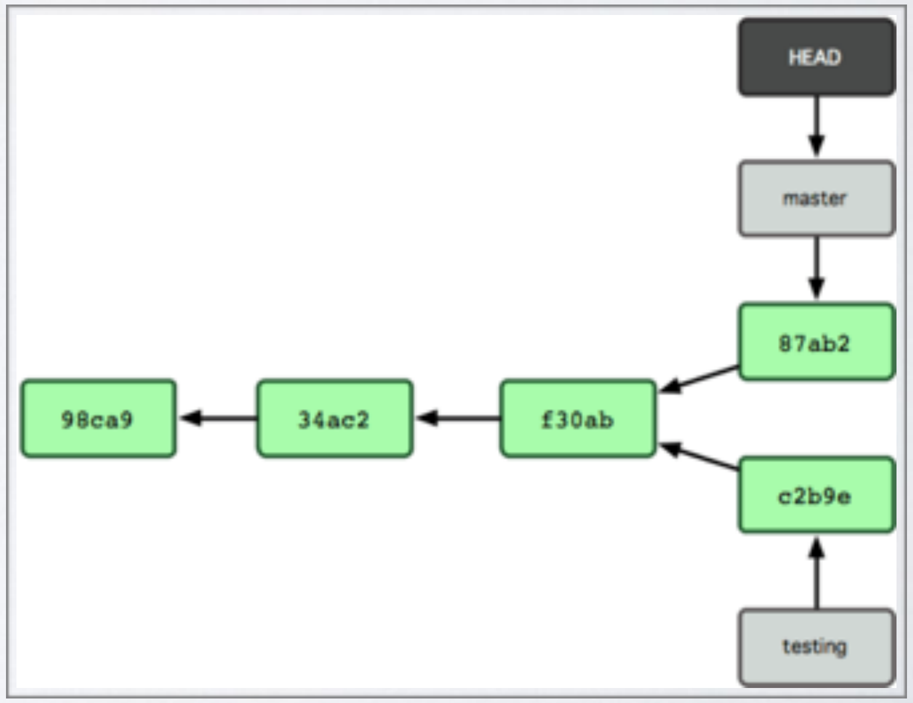
- Making Changes to the Repo
  - git add <file>:
    - Stage modified (or new) files for commit
  - git rm <file>:
    - Stage file removal for commit
  - git reset HEAD -- :
    - Unstage all changes
    - git reset can also restore a repo after a bad commit…
  - git commit:
    - Commit staged changes



BASIC BRANCHING

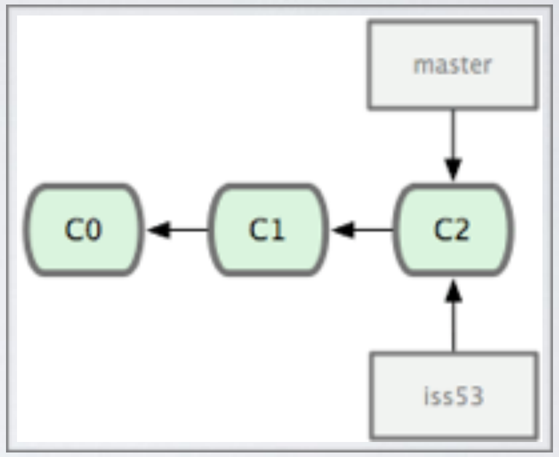
- What is a **branch** in git?
  - A branch is a pointer to a particular commit (snapshot of the repository)
  - The **master** branch is the canonical distribution branch (shared and up-to-date)
  - http://git-scm.com/book/en/Git-Branching-What-a-Branch-Is

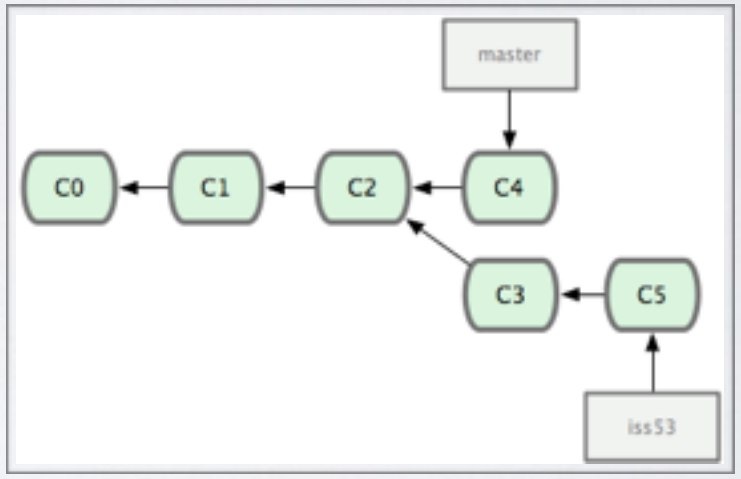


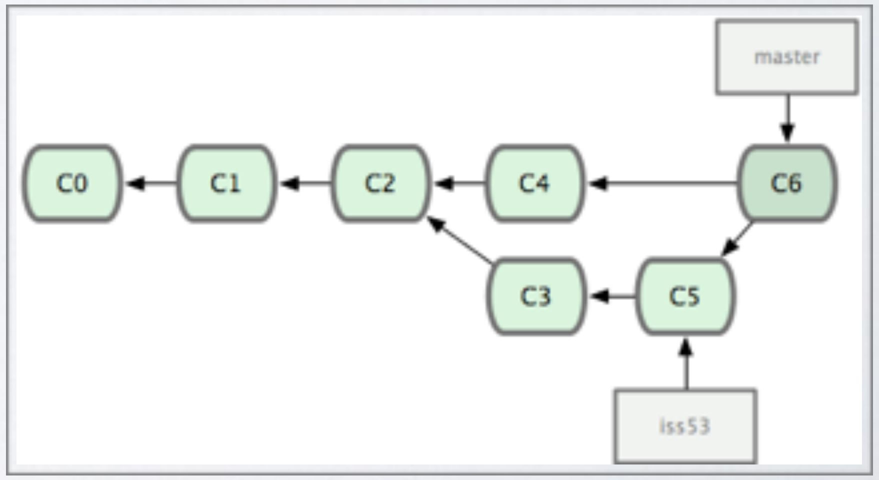


BASIC MERGING

- What is a **merge** in git?
  - A merge is a commit that includes changes made in two branches.
  - You can merge your test branch with the master to add your changes to the main distribution.
  - If the same data was edited in both branches, the merge must be resolved by a user.
  - http://git-scm.com/book/en/Git-Branching-Basic-Branching-and-Merging

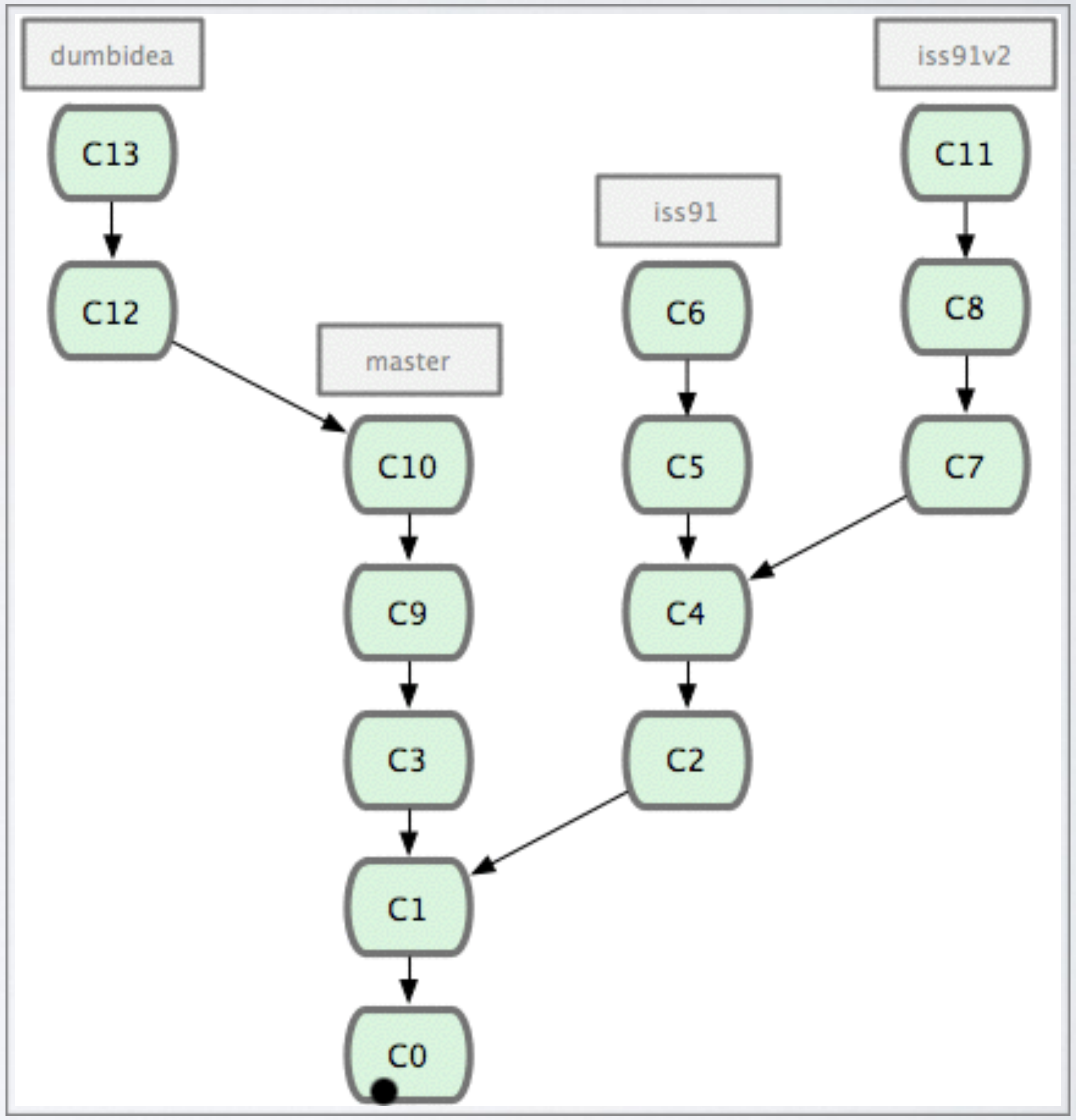






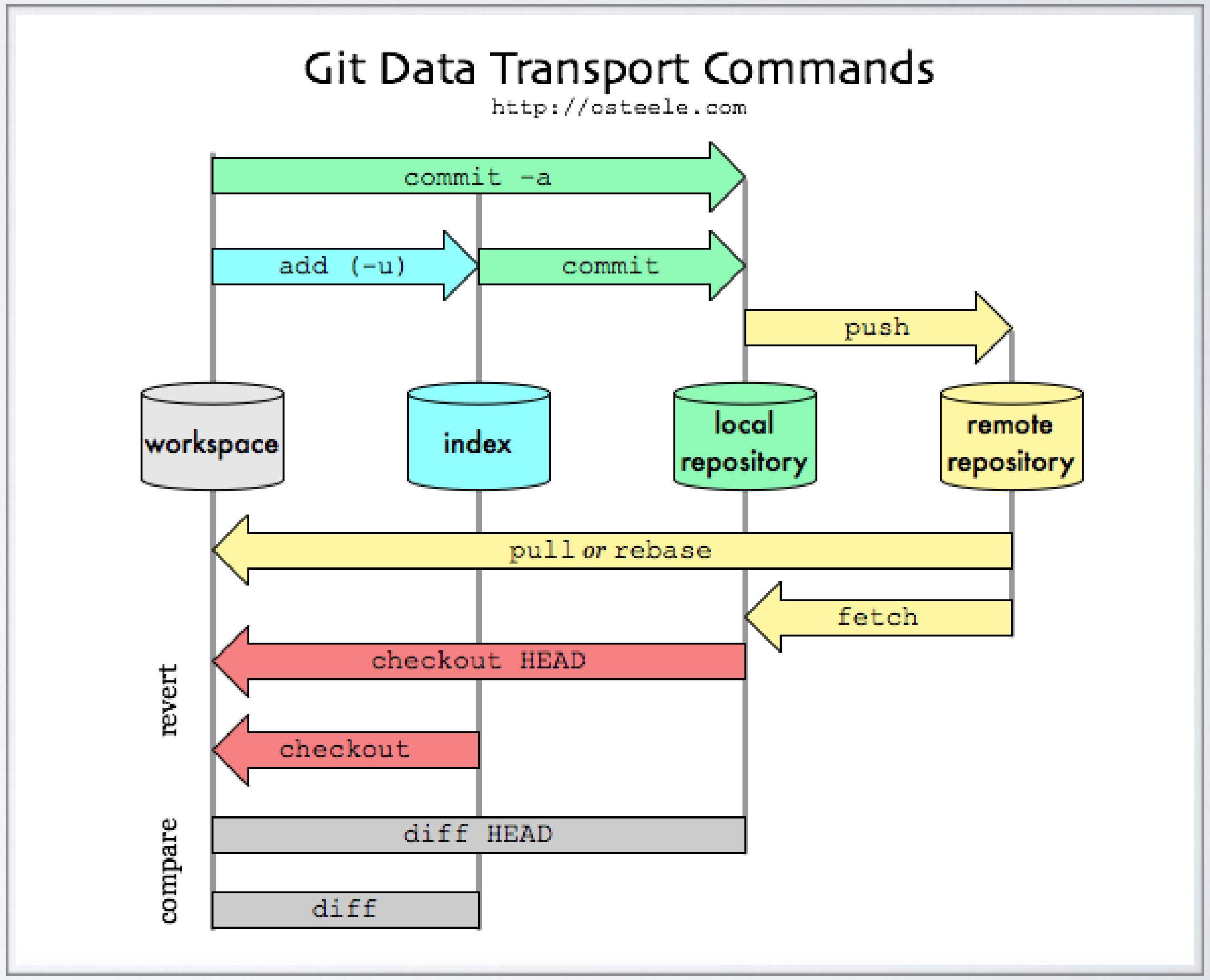
BASIC GIT COMMANDS 3/4

- Branching and Merging
  - git branch <branch>:
    - Create new branch
  - git checkout <branch>:
    - Move to another branch
  - git branch -d <branch>:
    - Delete branch
  - git merge <branch>:
    - Merge with other branch



BASIC GIT COMMANDS 4/4

- Working with Remote Repos
  - git clone <url>:
    - Clone a remote repository
  - git fetch:
    - Fetch all changes from remote repo
  - git pull:
    - Fetch and merge
  - git push:
    - Push commits to remote repo



GITHUB REPOSITORIES