Understanding Git with Alloy

Milestone 3

Cláudio Lourenço Renato Neves

University of Minho
Formal Methods in Software Engineering

July 10, 2012

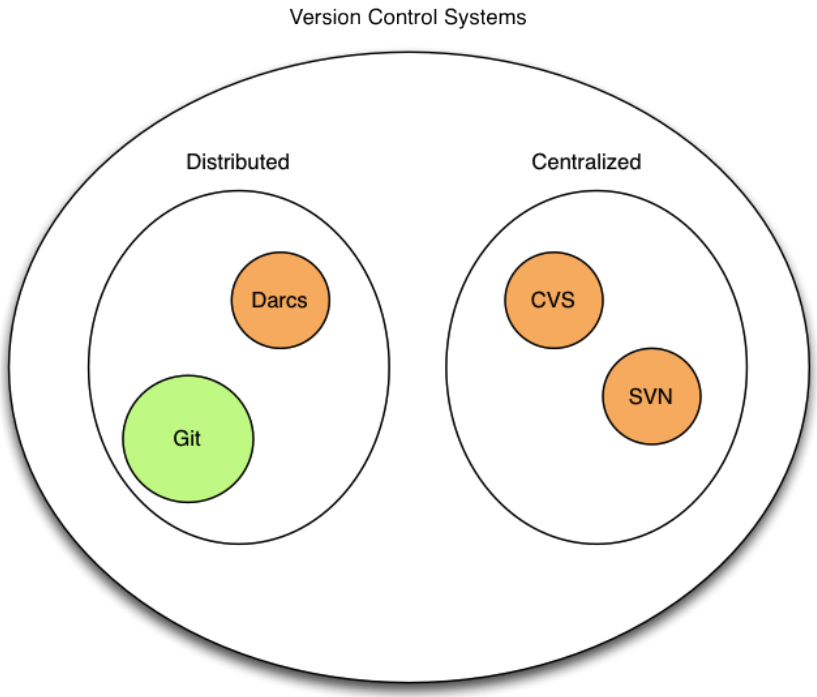
Table of contents

Git as VCS

Git is one of many Version Control Systems

- Fast
- Efficient
- Oriented to snapshots, not differences
- Widely used

Git as VCS


Version Control Systems
Distributed
Centralized
Darcs
Git
CVS
SVN


CSAIL

Motivation for this project

Gap in the understanding of Git

- Lack of precise descriptions
- Contradictions in some manuals

An opportunity appears

- Developers could benefit from a manual that is precise and rigorous

The dark world of Git

The common (and above average) knowledge of Git

- "if there are any uncommitted changes when you run git checkout, Git will behave very strangely." [1]
- "When you create a branch, it will contain everything committed on the branch you created it from at that given point. So if you commit more things on the master branch like you have done (after creating b), then switch to branch b, they won't appear. This is the correct behavior. Does that answer your question?" [2]

[1] "Understanding Git" Manual

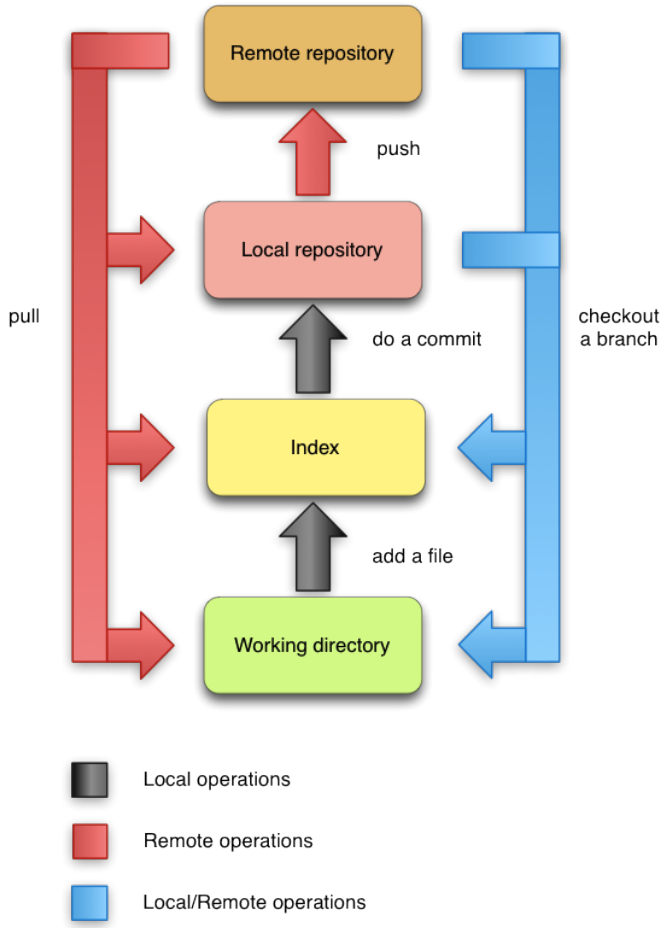
[2] An user of Git development mailing list

Objectives of this project

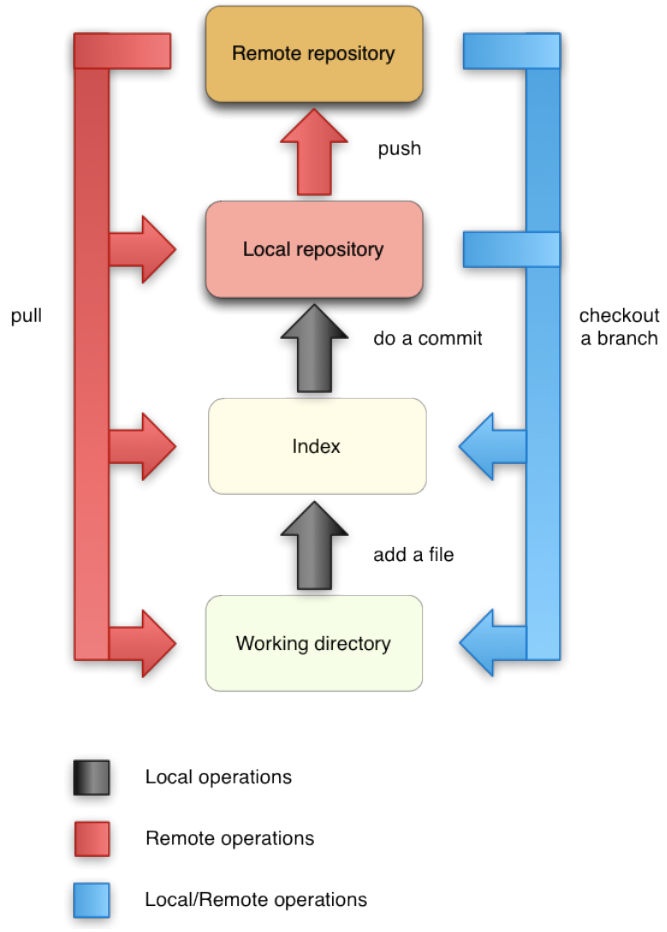
Shine some light in Git internals

- Build a precise model of how Git works, using Alloy
- Analyze the model and verify which properties does (not) guarantee
- Help others to understand Git better, building a manual with public access

The Git Structure

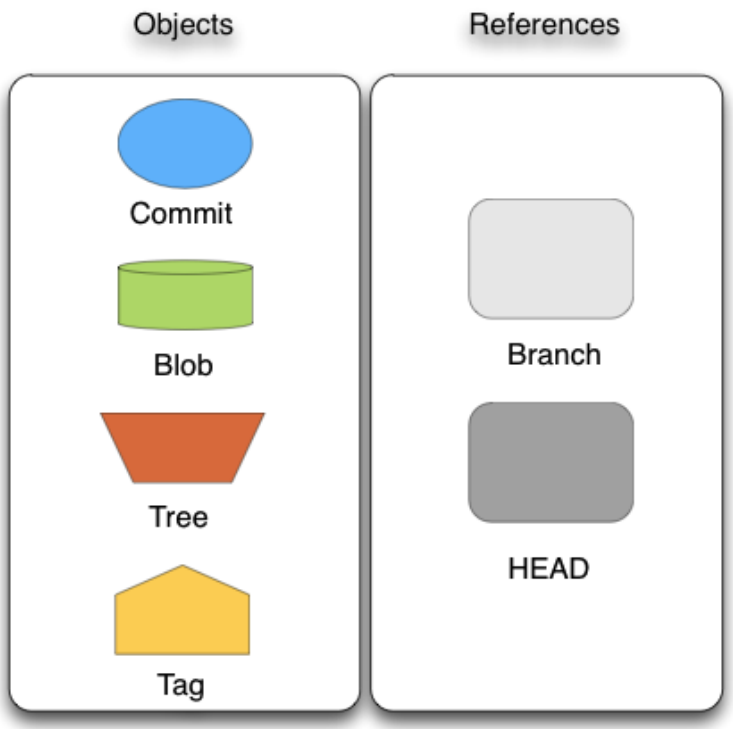


Repository


Remote repository
push
Local repository
pull
do a commit
checkout
a branch
Index
add a file
Working directory
Local operations
Remote operations
Local/Remote operations


CSAIL

Repository



Blob and Tree

Blob

- Represents the content of a file
- The name is calculated from its content

```
sig Blob extends Object {}
```

Tree

- Relation from names to Blobs or/and Trees
- Used to represent the file system structure

```
sig Tree extends Object {
   contains: Name -> lone(Tree+Blob)
}
```

Commit

- It is like a snapshot of the project on a certain moment in time
- Author, Committer, Comment - Not important for us
- Parent - The Commit which originated the current
- Tree - Pointer to a Tree Object

```
sig Commit extends Object {
    points : Tree,
    parent : set Commit,
    abs: Path -> Object,
    merge : set State
}

sig RootCommit extends Commit {}
```

Branch and HEAD

Branch

- It is just a pointer to a commit

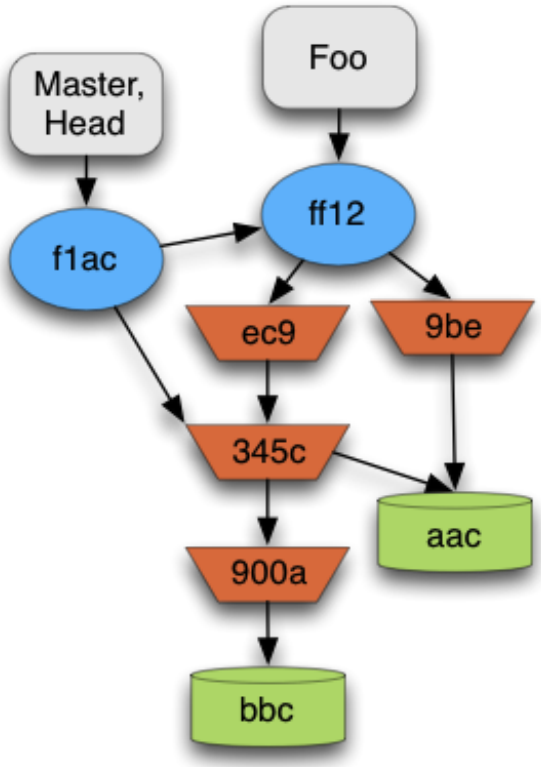
HEAD

- Special reference that identifies the current Branch

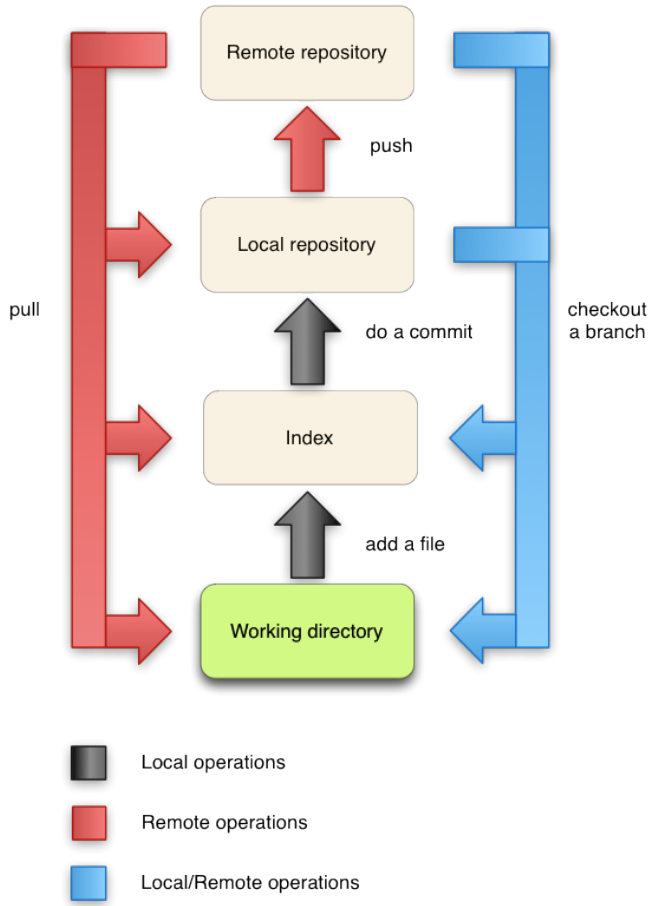
```
sig Branch{
    marks: Commit lone -> State,
    branches: set State,
    head: set State
}

lone sig Master extends Branch{}
```

Repository



Working Directory



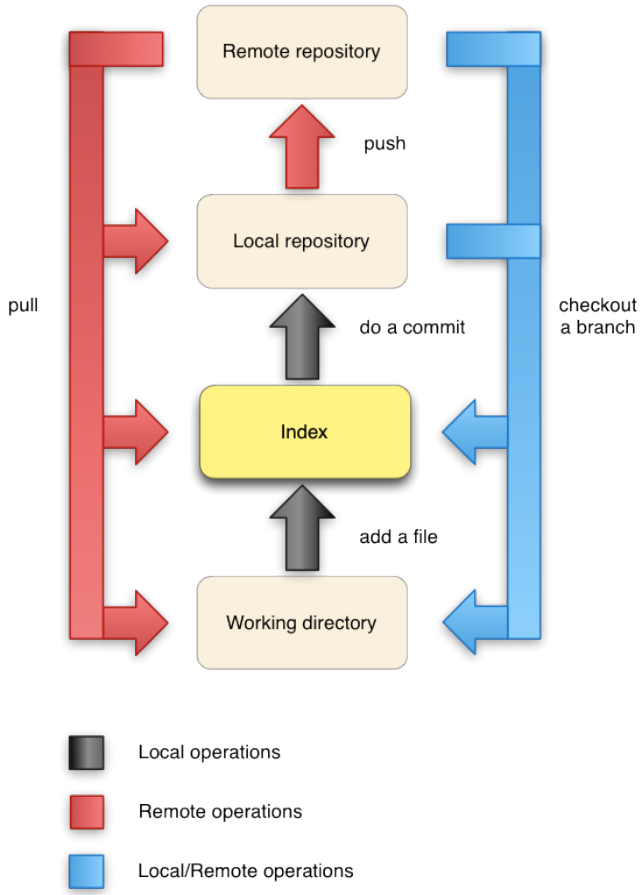
Working Directory

- Subset of a file system
- Has the content of a project
- Files can be the current files or files retrieved from the repository

```
sig Path {
    pathparent: lone Path,
    name: Name,
    unmerge: set State
}

one sig Root extends Path{}
```

Repository


Remote repository
push
Local repository
pull
do a commit
checkout
a branch
Index
add a file
Working directory
Local operations
Remote operations
Local/Remote operations

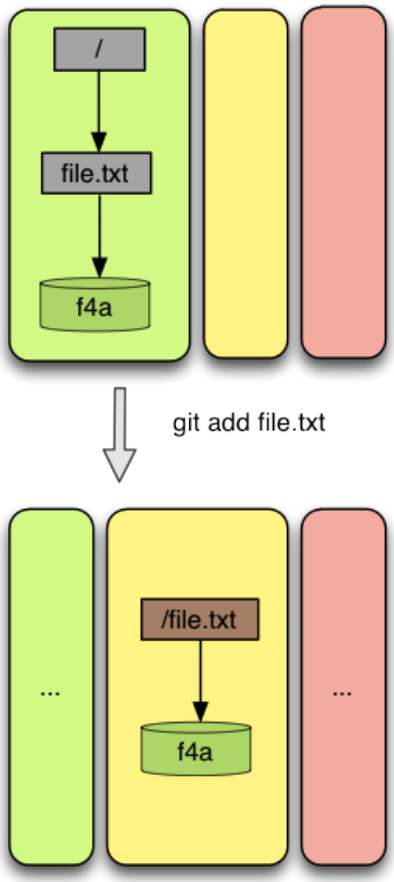

CSAIL

Index

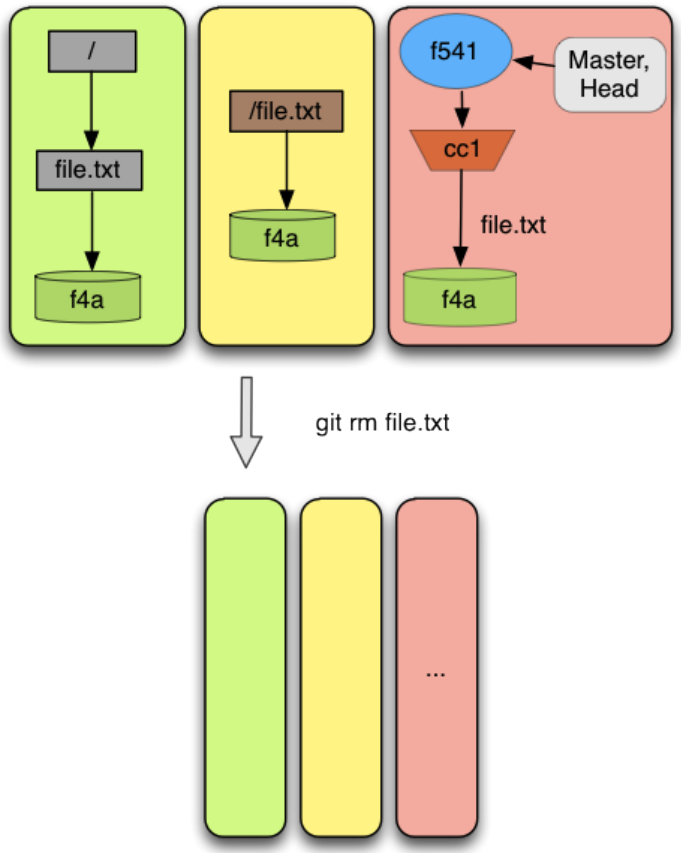
- Contains all files that are going to be committed on the next commit
- The index is not necessarily equal to the working directory
- If an user wants to commit a new file or a newly modified file, first he must add it to the index

```
sig File{
    path: Path,
    blob: Blob,
    index: set State
}
```

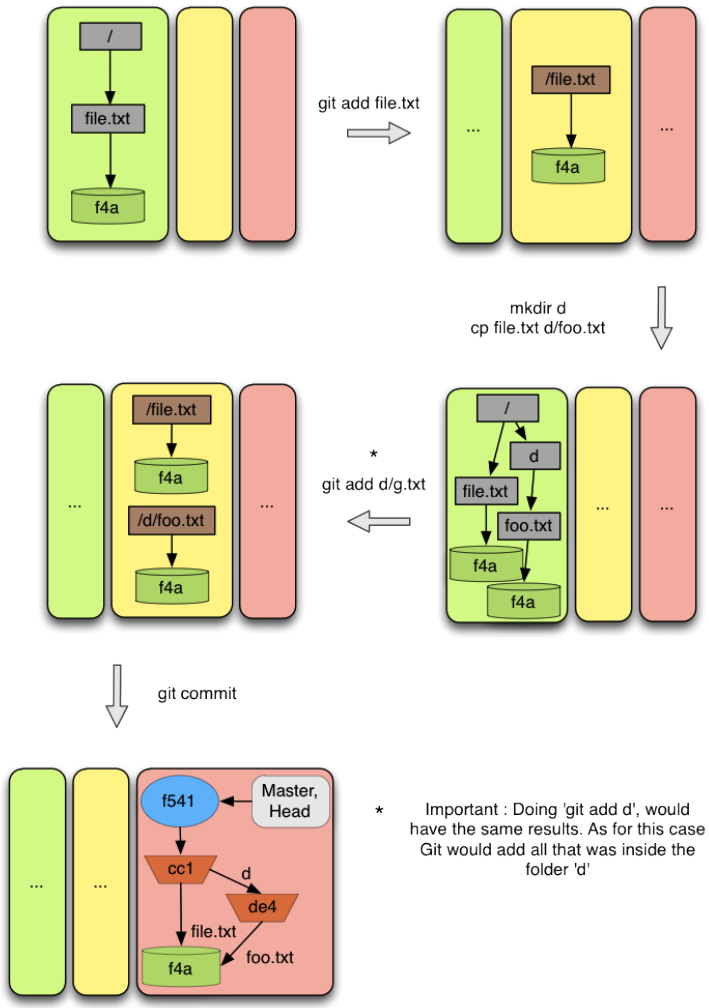
Add



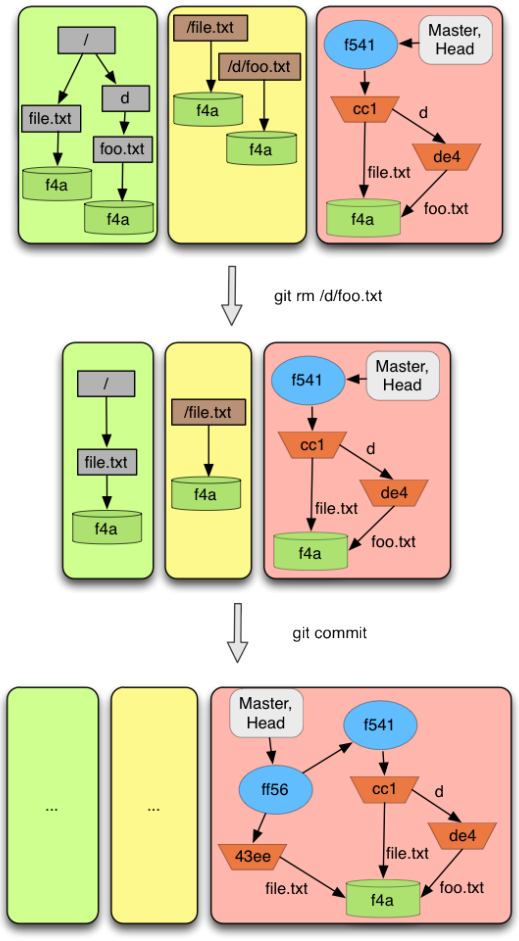
Remove


/
file.txt
f4a
/file.txt
f4a
f541
Master,
Head
cc1
file.txt
f4a
git rm file.txt
...

Commit


/
file.txt
f4a
git add file.txt
/file.txt
f4a
mkdir d
cp file.txt d/foo.txt
/file.txt
f4a
/d/foo.txt
f4a
* git add d/g.txt
/
d
file.txt
foo.txt
f4a
f4a
git commit
f541
Master, Head
cc1
d
de4
file.txt
foo.txt
f4a
* Important : Doing 'git add d', would have the same results. As for this case Git would add all that was inside the folder 'd'

Commit



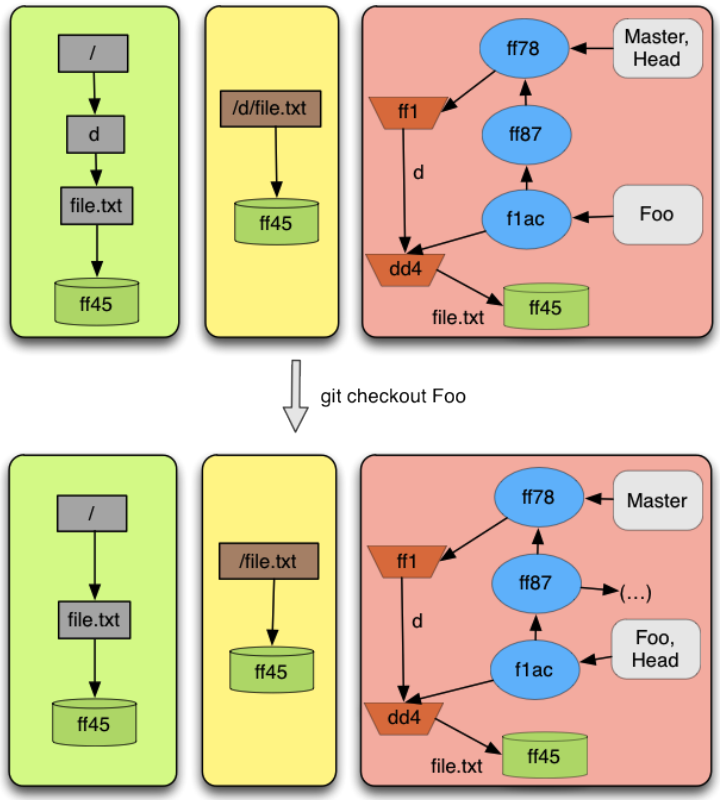
CSAIL

Checkout

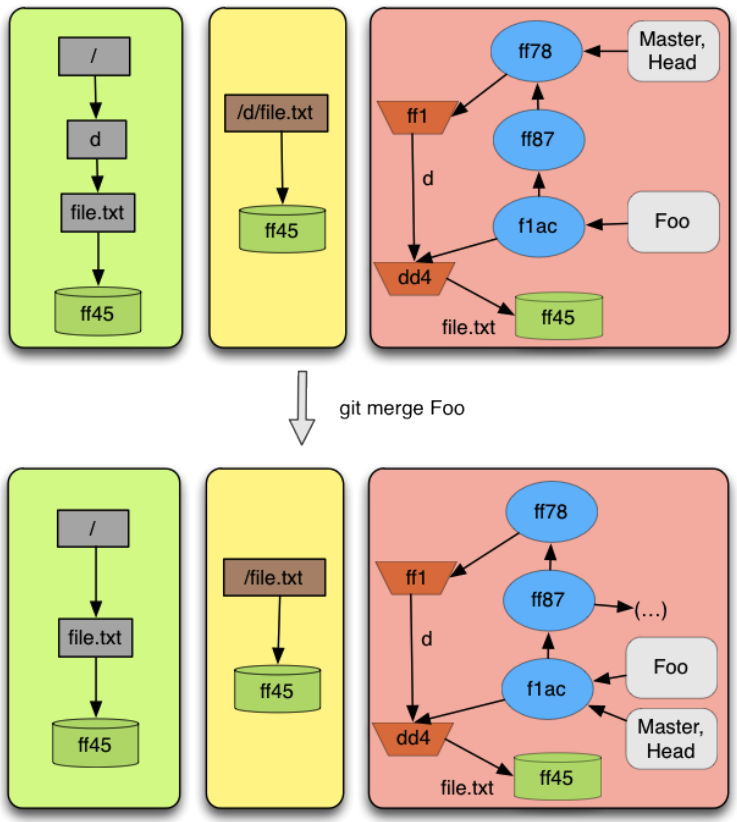
The most difficult operation to specify

- Expected pre-conditions were not found
- Instead we found weaker pre-conditions
- Strange behaviour caught. Discussed about it with the Git mailing list. Concluded that was a bug

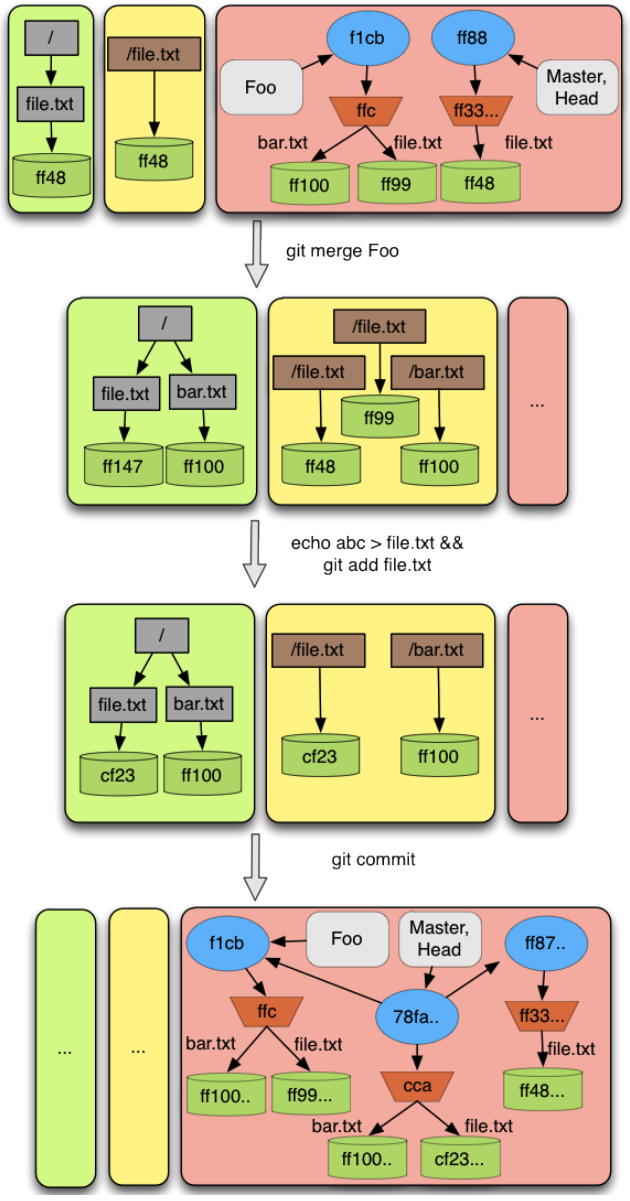
Checkout


/
d
file.txt
ff45
/d/file.txt
ff45
ff78
Master, Head
ff1
ff87
d
f1ac
Foo
dd4
file.txt
ff45
git checkout Foo
/
file.txt
ff45
/file.txt
ff45
ff78
Master
ff1
ff87
(...)
d
Foo, Head
f1ac
dd4
file.txt
ff45

A fast-forward Merge



A 2-way Merge



CSAIL

Modeled Operations

- Add and Remove
- Commit
- Branch and Branch Remove
- Checkout
- Merge (2-way and fast-forward)

Website

- Website created based on the manual
- http://nevrenato.github.com/CSAIL_Git

Future Work

- Model more operations (rebase, fetch, 3-way merge...)
- Specify more properties that the model does (not) guarantee
- Build interactive diagrams of concrete examples of operations

Conclusions