# Mastering Git's index

Charles Bailey

Bloomberg LP

# What is Git?

# Attributes of Git

Git is considered to be hard to learn

Git has “the index”

# Gitless

- What’s Wrong with Git? A Conceptual Design Analysis (Onward! 2013)
- Purposes, Concepts, Misfits, and a Redesign of Git (OOPSLA 2016)

*S. Perez De Rosso and D. Jackson*

http://gitless.com

What is the index?

# Git’s object model

- blob
- tree
- commit
- tag

Why is it called the index?

and what are its other names?

# Alternative names

- index
- cache
- staging area

# A brief history of the index

## Git in 2005

```
cat-file        commit-tree     init-db
read-cache      read-tree       show-diff
update-cache    write-tree
```

The index contains the complete project tree but often we only consider what has changed

e.g. `git status`

# Index tools

# “Unstage”

```
git reset -- <path>
```

# List files in the index

```
git ls-files -s
git ls-files --debug
```

# Moving stuff around

# Why use index manipulation?

- speed
- part of a script, e.g. `filter-branch`

# Clear the index

or create a new one (`GIT_INDEX_FILE`)

```
git read-tree --empty
```

# Populate the index

```
git read-tree -m <tree-ish>
```

# Merge in more trees

```
git read-tree --prefix=<prefix>/
```

# What about moving things up?

`git read-tree` takes a “tree-ish”

# Ignoring stuff

# Assume unchanged (?)

```
git update-index --assume-unchanged
```

# What you probably meant

```
git update-index --skip-worktree
```

# Sparse checkout

`core.sparseCheckout` (boolean config)

`$GIT_DIR/info/sparse-checkout`

# Performance considerations

# `stat` data in the cache matters

```
~/git$ git read-tree HEAD
~/git$ time git diff

real    0m2.046s
user    0m0.219s
sys     0m1.813s

~/git$ time git diff

real    0m0.128s
user    0m0.000s
sys     0m0.125s
```

```
~/git$ git read-tree HEAD
~/git$ time git update-index --refresh

real    0m0.773s
user    0m0.063s
sys     0m0.719s

~/git$ time git diff

real    0m0.124s
user    0m0.000s
sys     0m0.125s
```

```
~/git$ git read-tree -m HEAD
~/git$ time git diff

real    0m0.121s
user    0m0.000s
sys     0m0.125s
```

# Restoring the stat cache

Happens automatically on many “porcelain” commands

```
git update-index --refresh
```

# What might invalidate state information?

- Copying the working tree
- Using `read-tree` instead of `reset`

# Git’s index

- is core to Git
- helps Git to perform
- can help you be productive