Code versioning principles (GIT)

What is Git?

DVCS(Distributed Version Control System)

Made-by Linus Torvalds For Linux

Why Git?

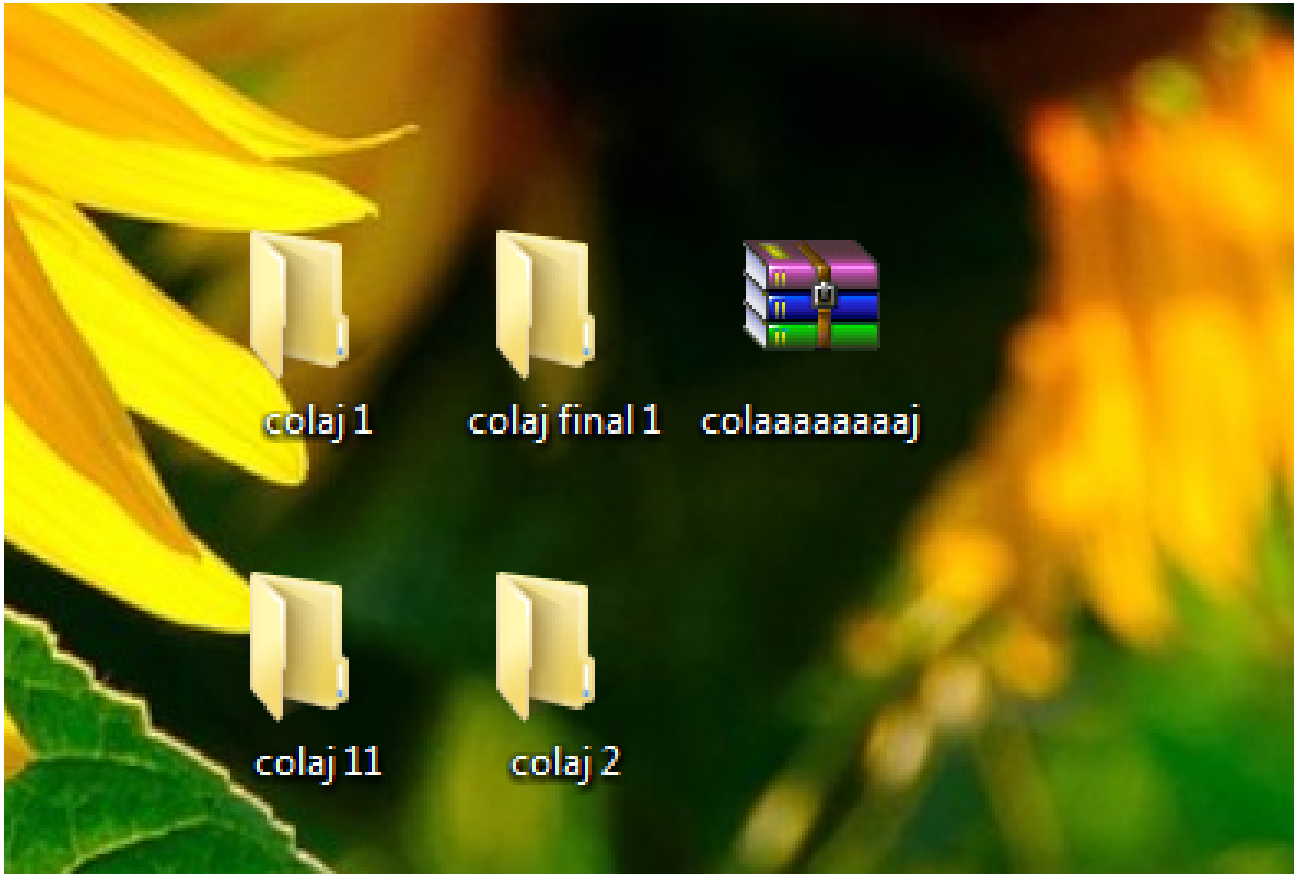
Alternatives?

- Subversion.
- Mercurial.
- CVS.
- Helix VCS.
- Microsoft Team Foundation Server.
- Rational ClearCase.
- AccuRev.
- Kiln.

Git internals

how do you think the git works?

In a parallel life :)



colaj final

Папка с файлами

ие: C:\Users\home\Desktop

245 МБ (257,732,710 байт)

246 МБ (258,564,096 байт)

Файлов: 407; папок: 1

Usual Life Of File

We Need Version Control System
VCS Would...

- Record Every Changes Safely, Efficiently
- Able To Check Out Any Version
- Easy To Read History

.git folder

```
$ ls
HEAD
branches/
config
description
hooks/
index
info/
objects/
refs/
```

.git folder

config contains your project-specific configuration options
info contains info from .gitignore file

objects stores all the content for your database

refs stores pointers into commit objects in that data (branches, tags ...)

HEAD points to the branch you currently have checked out

index where Git stores your staging area information

One thing to understand about git is that git doesn’t store diff of the contents of your files !!!

.git/objects folder

Git blob object.

A blob object is used for storing the contents of a single file.

.git/objects folder

Git tree object.

A tree object contains references to other blobs or subtrees.

.git/objects folder

Git commit object.

A commit object contains the reference to another tree object and some other information(author, committer etc.)

.git/objects folder

Git tag object.

A tag or a tag object is just another reference to a commit object and just makes for easier referencing.

Let the Magic Begin

git cat-file - Provide content or type and size information for repository objects

```
find .git/objects/ -type f
```

Content-addressable storage system

Hash map

Content-addressable storage system

Key: sha-1 Hash of object content
Value: Compressed content

Hash map algorithm

Content-addressable storage system

Same content never saved twice ?

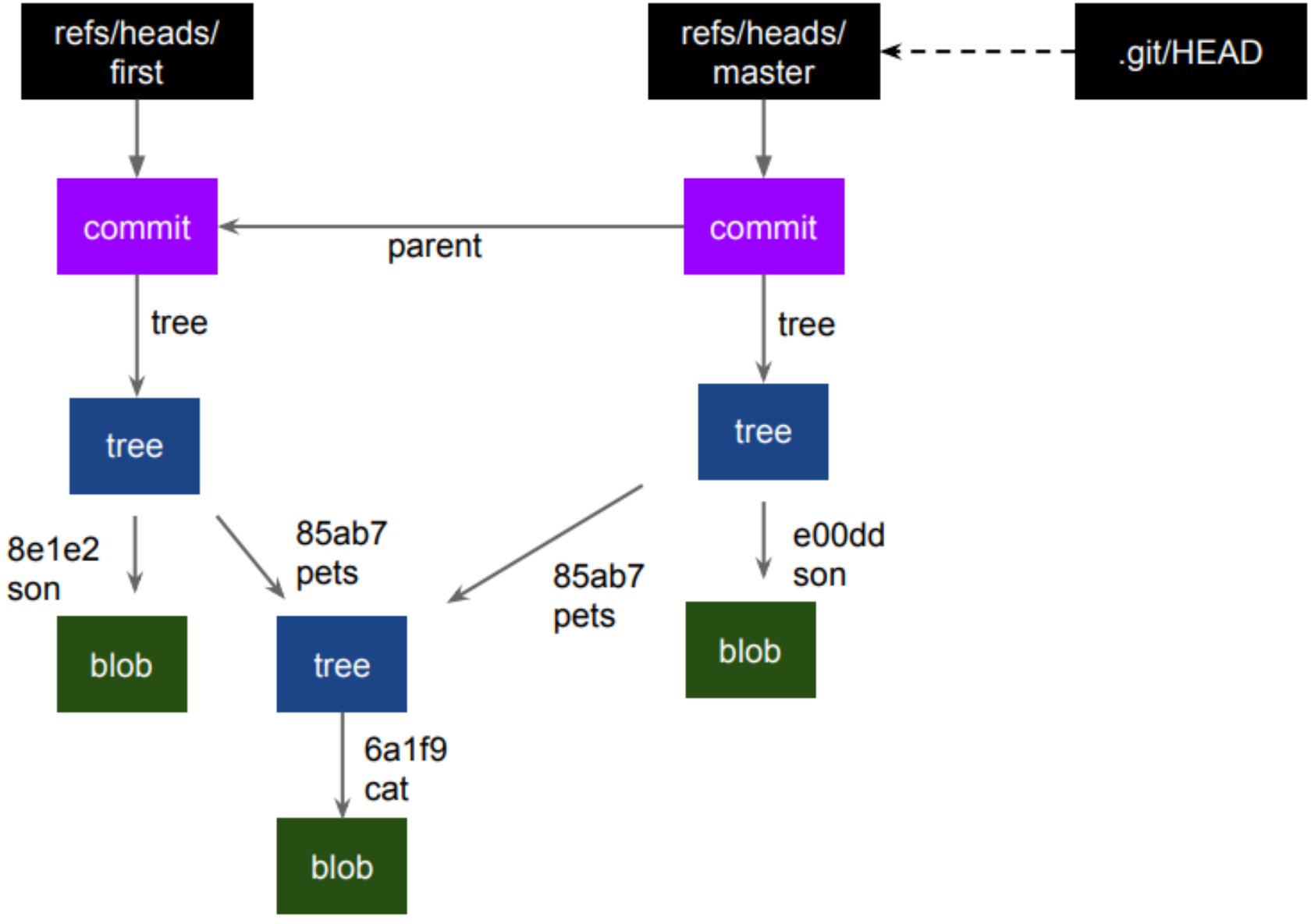
Content-addressable storage system

git hash-object

hash_object_w

```
# Save compressed header + content at sha1 path
def hash_object_w(content):
    header = 'blob %d/0' % len(content)
    store = header + content
    sha1 = sha.new(store).hexdigest()
    dir = '.git/objects/' + sha1[0:2] + '/'
        filename = sha1[2:]
    open(dir + filename, 'w').write(
        zlib.compress(store))
```

Internal Data Structure



How git knows current commit?

HEAD

Let's play with git

What If Small Changes Inside A Big File?