Why is Git not considered a "block chain"?

Asked 7 years, 3 months ago   Modified 5 days ago   Viewed 53k times

336

Git's internal data structure is a tree of data objects, wherein each objects only points to its predecessor. Each data block is hashed. Modifying (bit error or attack) an intermediate block will be noticed when the saved hash and the actual hash deviate.

How is this concept different from block chain?
Git is not listed as an example of block chains, but at least in summaries, both data structure descriptions look alike: data block, single direction reverse linking, hashes, ...).

So where is the difference, that Git isn't called a block chain?

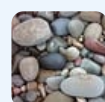git   hash   blockchain

Share
Improve this question
Follow

3   *Git is not listed as an example of block chains* When I first tried to learn what a blockchain was, I was referred to git as the most prominent example (I don't have the exact link now, but it was from the top of the list returned by Google search for "blockchain") – Leon Sep 13, 2017 at 8:58

9   Both Git and blockchain are using merkle trees as their fundamental underlying data structure. But that alone does not make Git a blockchain, or the other way around. – If you do know Git (and its internals), you do know merkle trees though, which can be a very helpful revelation to understand how blockchains work. – poke Sep 13, 2017 at 10:06 ✏

2   It's your opinion that "it is NOT considered..." bitcoin.stackexchange.com/a/43627/77469 – v.oddou Feb 22, 2018 at 8:37

10   @v.oddou Merkle trees exist since 1979. Just because two technologies are using Merkle trees prominently as part of their concept, that does not make them the same. It is incorrect to reduce either Git or block chains to just merkle trees as neither of them *are* merkle trees. They only use them. That makes the linked post completely irrelevant since it is actually talking about merkle trees, and not block chains. – poke Feb 22, 2018 at 9:26

3   Can you cite the statement: "git is not considered a block chain"? – Jannis Ioannou May 18, 2018 at 8:00 ✏

11 Answers

Sorted by: Highest score (default) ⇕

▲

184

The reason why Git and blockchains appear similar is because they are both using merkle trees as their underlying data structure. A merkle tree is a tree where

each node is labeled with the cryptographic hash value of their contents, which includes the labels of its children.

Git's directed acyclic graph is exactly that, a merkle tree where each node (tag, commit, tree, or blob object) is labeled with the hash of its content and the label of its "child". Note that for commits, the "child" term conflicts a bit with Git's understanding of parents: Parent commits *are* the children of commits, you just need to look at the graph as a tree that keeps growing by re-rooting it.

Blockchains are very similar to this, since they also keep growing that way, and they are also using its merkle tree property to ensure data integrity. But usually, blockchains are understood as way more than just merkle trees which is where they are separating from the ["stupid content tracker" Git](#). For example, blockchains usually also means having a highly decentralized system on a block level (not all blocks need to be in the same place).

Understanding blockchains is kind of difficult (personally, I'm still far away from understanding everything about it), but I consider understanding Git internals as a good way to understand merkle trees which definitely helps understanding a fundamental part about blockchains.

Share   Improve this answer

Follow

answered Sep 13, 2017 at 10:27

poke
387k ● 80 ● 587 ● 627

72   I'm sorry but nowhere blockchains bring anything more than git does. blockchains are exactly as stupid as git. If you don't

believe so, you are overhyped. The peer network and the consensus systems are a separate thing. – v.oddou Feb 22, 2018 at 8:23

12   private ledgers (blockchains) are conceptually identical to git – Munhitsu Mar 29, 2018 at 11:09

13   Git is a blockchain. Objections such as requiring "a highly decentralized system" are implementation details. – Lumi Apr 14, 2021 at 9:18

2   So, what this answer is trying to say is that Git is not a block chain because *blockchains usually also means having a highly decentralized system on a block level*.? This answer doesn't seem to describe why, except having this one "example" which many people considered false. What am I missing? – Yasushi Shoji Nov 11, 2021 at 4:52

5   "*blockchains usually also means having a highly decentralized system, and not all blocks need to be in the same place*" - so does Git? You can do shallow or partial clones. – Bergi Nov 14, 2021 at 1:00

104

The question reads: *Why is Git not considered a "block chain"?* So this is asserting that there is a wide-spread opinion that Git is not a blockchain (an assertion that is illustrated and corroborated by the answers preceding mine on this page) and asking for the reason of the prevalence of this opinion. This is a good question.

Taking the question literally, the answer could be that the blockchain term and concept gained popularity as part of the digital currency operation called "Bitcoin", and hence came to be associated with how Bitcoin does things: which is by using a lot of computing power to calculate a

specific hash including a nonce to meet certain arbitrary requirements, which is by allegedly having no central authority, which is by being "independent", maybe even "democratic", and the rest of the kool aid; and as these things are not seen in Git, well, Git cannot be a blockchain, right? And so the question would be answered literally.

Hidden behind this prima facie question is another question: What is a block chain? Now you could look up a *definition* somehwere and copy it over here, but I didn't do that as I have *made up my mind* years ago, when listening to a podcast about Bitcoin that strove to explain the new concept of a blockchain, that *a blockchain works like Git* and I don't intend to let my precious understanding be misled by random claims on the internet.

So what is a blockchain? What's in the word?

Nothing in the term "blockchain" presupposes the requirement to include a nonce in the content so as to come up with a hash of so and so many leading zeros. (This requirement is only there to be able to control the blockchain by computing power and so, ultimately, by money.)

Nothing in the term "blockchain" presupposes the existence of a network, let alone a decentralized one.

Nothing in the term "blockchain" presupposes any "independence" from "central authority".

The term "block chain" only presupposes blocks (of data) chained together. Now what is a chain? Is it just a link? No, it is a *strong* link designed to hold things together by force.

A simple linked list doesn't qualify as a blockchain because the contents of the chunks of data in the list could be altered while the list would continue to link back and forth just fine. This is not how a chain works.

To make a *link* of blocks of data into a *chain* of blocks of data, the contents of the blocks need to be checksummed (digested) in one way or another and this checksum (digest) must be part of the link, making it a strong link protecting the content, preventing it from being altered. This is a blockchain.

And this is what Git does, and hence Git is a blockchain, or works as one, if you prefer.

To close the circle, let's ask again: *Why is Git not considered a "block chain"?* It could be because many people, perhaps even a large majority, do not focus on the essence of a concept but on blinking accidents.

Share Improve this answer

Follow

edited Jun 11, 2021 at 5:57

M Matiboux
112 ● 8

answered Apr 14, 2021 at 11:23

Lumi
15.2k ● 9 ● 61 ● 92

3   I agree with you (or at least I do not disagree) that Git should be considered a blockchain. But for corroborating evidence (beyond the other answers here) that many others think differently, see wikipedia: en.wikipedia.org/wiki/Blockchain – iconoclast Apr 26, 2021 at 20:35

Blockchain is immutable (Git allows rebase), and also relies on consensus via redundant network mining and security is increased by the size of the network. – lacostenycoder Jan 31, 2023 at 15:16

4   You can fork a blockchain and rewrite blocks just as you can fork a git repo and rewrite history. Both are immutable in a sense and mutable in another. – DylanYoung Feb 17, 2023 at 17:57

Blockchain is **not** just any chain of any blocks.

Blockchain is when *there is a way of determining the main chain when two or more are diverted*, and when no central authority is needed for that determination.

31

Share   Improve this answer

Follow

edited Jul 12, 2022 at 9:35

Lii
12.1k ● 9 ● 68 ● 89

answered Dec 23, 2017 at 7:42

Daniel Vartanov
3,267 ● 1 ● 21 ● 27

3   Using this definition, "permissioned blockchain" makes no sense since they do in fact have a central authority. So your definition is contradicting actual usage of the word. See e.g.

semi-centralised (federated) blockchains like Liquid.
– Janus Troelsen Nov 9, 2020 at 20:13

7   @JanusTroelsen "permissioned blockchain" (or "private blockchain") is an oxymoron and indeed makes absolutely zero sense, so called centralised blockchains don't have a single difference from a functionality of regular servers or regular P2P networks. These terms (private/permissioned/centralised blockchain) are used only outside of the professional community. – Daniel Vartanov Nov 11, 2020 at 18:01

1   R3 had been ridiculed in the community exactly for abusing the word 'blockchain' to market their software which could easily exist even if the initial paper was not published back in 2008. Reference to the authority of R3 is not valid I'm afraid.
– Daniel Vartanov Nov 20, 2020 at 15:24

3   The central authority in a blockchain such as Bitcoin which requires unnecessary and irrational computing is simply the party assembling the most computing power. – Lumi Apr 14, 2021 at 9:31

1   @ocodo what do you mean by *"a way of determining the main chain when two or more are diverted, and when no central authority is needed for that determination"* in git?
– Daniel Vartanov Aug 12, 2022 at 17:20

▲
30
▼

🔖

Cyber Currencies like Bitcoin, use a distributed consensuses cryptographic chain of blocks (merkle tree). Common usage has shortened this to 'blockchain'

While git uses a chain of blocks (merkle tree), it lacks the distributed consensuses cryptographic components that common usage of the term 'BlockChain' imply.

5   Without specifying what "distributed consensus" exactly requires, this distinction is irrelevant. If the PoW threshold is low, anybody can overwrite your blockchain.
– Janus Troelsen Nov 9, 2020 at 20:11 ✎

1   The irrational and unnecessary cipher requirements in Bitcoin are only there so that block chaining requires computing power and thus can be dominated by purchasing computing power, and thus by financial power. – Lumi Apr 14, 2021 at 9:25

@JanusTroelsen I think what the author means is that there is an algorithm by which there is no central authority do decide on the main chain. In git there technically is no main chain, anyone can declare their own as such. Since there is no scarcity of some object (asset) then for git it doesn't matter. – Cigarette Smoking Man Jul 6, 2023 at 8:06

Unlike *cryptocurrency blockchains*; git doesn't have a p2p trustless consensus mechanism.

21

edited Jul 2, 2018 at 11:12

10   Why do you consider a trustless consensus system as part of a block chain? There are many ways to create trust in a block chain, for git it is just that you know that everything in your local copy cannot be removed by the next pull and you specifiy that you want the changes in the remote copy. You only need trustless consensus when it would otherwise be unclear what's right. In git multiple branches can be "right" and get eventuell merged together. – allo May 18, 2018 at 14:45

1   @allo GitHub is typically used as the central source of truth but what's stopping an admin from force pushing and overriding history? If there was no GitHub and you pulled from your peers then how do you handle merge conflicts? How do you determine whose right? – Miguel May 18, 2018 at 17:02

7   Nothing stops you from force pushing. But like a blockchain guarantees me, I can detect it because my chain cannot verify these commits as being based on it. *That's* the point with a blockchain, not the decentral consent. And in git I explicitely *do not* want to have a consent protocol for what I merge (development is not a democracy), but I actually read the new commits when merging them into my chain. So my copy *is* right, because it consists of stuff I already have and thus can verify (i.e. by seeing merge conflicts) and stuff I review and then accept into it. – allo May 19, 2018 at 18:08
🖉

2   @allo you're correct in that regard, however I stated in the answer "cryptocurrency blockchains", not blockchains in general, but I now that I think about it my answer doesn't really seem to fit the question being asked because I was thinking about the system as a whole rather than the underlying data structures – Miguel May 24, 2018 at 22:25

You're completely right about the difference of the block chains used in git and cryptocurrencies. It is just not an

answer to the question why (or if) git is not considered a block chain, when using the term rigorouly. Even the currently accepted answer is similar to your answer. I still prefer the answer which got the bounty. – allo May 25, 2018 at 10:26

To sum it up (for me):

While Git offers you complete full freedom of choice, Blockchains are a highly political system, where you are forced to trust in others:

- Git is a Merkle Tree without a predefined consensus algorithm.

- Blockchains are Merkle Trees with a predefined consensus algorithm.

Hence if you are all alone, there is no difference between Git and a Blockchain. As you trust Git and yourself, you already have that predefined consensus.

But things start to become different, when you are in a Network.

Notes:

- For Blockchains there is absolutely no requirement for the hash to be difficult to calculate or to define something like "Mining" or have some specific software which ensures you take part of a certain Network.

This all might be a requirement for something like Bitcoin (which usually is referred to as Cryptocurrency, which I cannot fully agree to), but neither is BitCoin defining what a Blockchain is, nor does a Blockchain need to be something like BitCoin.

- The consensus algorithm does not necessarily be something which is based on some cryptographic protocol. For example it would be enough to publish your TIP in a local newspaper each day to (ab)use Git as some Blockchain.

Git readily offers multiple possible consensus algorithms you can chose from:

- Publishing the SHA in a Newspaper or similar (something which is distributed and hard to fake)

- If you are in the rare situation that you are already part of some GnuPG Web Of Trust, you readily can use Signed Commits (or Signed Tags) to agree to the consensus.

- The "Signed off:" variant does not offer cryptographically secure consensus, but in combination with [something like Gerrit](#) and Fast-Forward-Only pushes it is some pretty well defined consensus algorithm.

Hence to make Git a Blockchain, [all you need is to add some air](#).

Some different view:

Git is no Blockchain on itself. In contrast, it is far less than a Blockchain (lacking the predefined consensus algorithm) and much more than a Blockchain (allows a plethora of consensus algorithm to chose from, is meant as an SCM etc.).

Some other observations:

- Git branches are the same as Blockchain splits. While Blockchain splits happen rarely, most Git repositories have less branches (`master` + `HEAD`) than [BitCoin had splits](#).

- Git always has an explicit consensus done by you, that is, the TIP you push to. However this only applies to you and nobody else.
  Pushing the Git repository to some shared Git Service can also be seen as a consensus. There is no requirement for such a consensus to be based on Democratic principles.

Very personal thoughts:

While Blockchain is some [overhyped buzzword](#), something you can happily live without, Git is an inevitable fundamental tool for getting your work done, one of the basic must-haves you cannot live without, something as important as air and water. This is probably, why people like me do not refer to Git as a Blockchain ..

YMMV

answered Sep 11, 2021 at 9:04

Tino
10.4k ● 5 ● 57 ● 62

---

2   Fantastic summary. – stevegt Jan 6, 2022 at 22:59 ✏

> This is probably my favourite answer, including the nod at the
> end to the most important part - `git` is more useful than
> almost any product, platform, or tool that's billed as a
> "blockchain". – Matthew Strasiotto May 27 at 2:07

17

There is no reason to not consider Git as a blockchain.
Git is focused in a very particular (and important) set of
assets: source code. The consensus in this case is
manual, and we can consider that a transaction (commit)
is accepted when it is merged into the release branch.
Actually, considering the number of transactions
(commits), Git is by far the most successful blockchain.

Extracted from: https://arxiv.org/pdf/1803.00892.pdf "...
...We define"blockchain" and "blockchain network", and
then discuss two very different, well known classes of
blockchain networks: cryptocurrencies and Git
repositories..."

See also next paper that explain why Google use a single
monorepo as single source of truth (basically, as a
blockchain). https://research.google/pubs/pub45424/

Update 2024: Newest version of Git have also integrated
"best-pattern" from major cryptocurrency public networks.

In particular non "secure" md5sum has been replaced by the safer sha256. Bloom filters have been integrated to provide "fast" search over the commit chain. Also, Git has always allowed to use signed transactions (commits in git parlance) with the help of PGP signatures, long way before Bitcoin allowed to sign "payments".

Share   Improve this answer

Follow

edited Dec 14 at 13:43

answered Nov 23, 2020 at 9:12

earizon
2,208 ● 20 ● 32

8

As poke said:

Git and Blockchains appear similar because they are both using Merkle Trees to store ordered timestamped transactions. A merkle tree is a tree data structure where each node is labeled with the cryptographic hash value of their contents, which includes the labels of its children.

The first difference is the Hash function: Blockchain has a very expensive hash function so that each block has to be mined, wheras a Git "block" can be created with a simple commit message.

The purpose of Bitcoin is to add trust to the order of transactions. The focus is on the longest chain, since

that is most expensive to compute and thus most likely to be the truth.

Bitcoin accomplishes this by requiring that the hash meets certain parameters (begins with a specific number of 0s), by incrementing a value ("nonce") in the message until a satisfactory hash is found. This takes effort to find, but only 1 calculation to verify for a nonce; and if multiple nonces produce a satisfactory hash, then one will be lower and taken as the truth. Other authentication schemes make the hash trustworthy by centralizing the issuing of the hash to an authority, perhaps voted by network agreement, or some other method.

Blockchain data is limited to transactions, which must must conform to validation. Transaction must be valid to be included in the next block. **A Bitcoin transaction corresponds to something important in the real world that justifies using an expensive block to record this transfer, like exchange of money value.** We don't actually care about the final ledger, it's a metaphor for something in the real world.

By contrast, Git blocks are arbitrary, as a commit can contain any amount of data. The value lies in the changes of data being organized into the git tree because we care about the final product, it's validated by the existence of the git repository.

The purpose of Git is to allow cheap "ledgers" to track multiple product alternatives. The "ledger" in Git is what we care about, it's our final product; the

transactions data just record how the product was built. We want to make it very cheap to make multiple versions of final products, just enough overhead to require the creator to record how they built this product. No explicit validation is done on the data, you maintain the end-product if it looks good, and that existence makes it useful to have the chain of this product's creation. If the end-product is bad or the order of commits is invalid, this "ledger" gets deleted during garbage collection.

The second difference is that Blockchain transactions must come from a prior valid source. In Git, we don't care what data you use to extend the tree. In Blockchain, the transactions must come from a prior valid source. In that sense, **Git tracks the extension of our environment, whereas Blockchain tracks the exchange of value within a closed environment.**

Share   Improve this answer

Follow

The Goals are different for blockchain and git although both use merkle trees as data structure.

A `blockchain` is typically managed by a peer-to-peer network adhering to a protocol for inter-node communication and validating new blocks. Once recorded, the data in any given block cannot be altered retroactively without alteration of all subsequent blocks, which requires consensus of the network majority.

As According to Bitcoin whitepaper :

> A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network. The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain not only serves as proof of the sequence of events witnessed, but proof that it came from the largest pool of CPU power. As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they'll generate the longest chain and

> outpace attackers. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone

While `Git` is a distributed version-control system for tracking changes in source code during software development.It is designed for coordinating work among programmers, but it can be used to track changes in any set of files. Its goals include speed, data integrity, and support for distributed, non-linear workflows.

As according to Linus Torvalds:

> In many ways you can just see git as a filesystem – it's content-addressable, and it has a notion of versioning, but I really designed it coming at the problem from the viewpoint of a filesystem person (hey, kernels is what I do), and I actually have absolutely zero interest in creating a traditional SCM system.

0

Git is a block chain in which successive data entries include the hash of the predecessor in the data such that the entire data set can be verified against data corruption by recomputing the hash codes of each successive entry and comparing the final hash code with the separately recorded latest hash code for the data set.

Other sequential data sets use this efficient consistency check, for example the Kafka log files do this.

In contrast, cryptocurrencies for example Bitcoin, use block chains whose hash codes are crypographically (slowly) calculated so that it is practically impossible to corrupt the data to arrive at a given hash code.

Git and Kafka use very efficient hash codes that do not have this anti-hacker feature. They only detect ordinary corruption from example missing, duplicated or garbled data as opposed to malicious data falsification.

"Blockchain" in common speech has the meaning of cryptographic block chain and that is why Git is not ordinarily considered to be a "Blockchain" despite it

having data blocks whose verification is performed by chained (efficient non-cryptographic) hash functions.

Share   Improve this answer

Follow

Stephen Reed
1 ● 1

▲

-1

▼

🔖

🕑

A great way to understand any given technology is to ask, **"what problem does it solve"?** Git's use case is quite simple in that it's intended use is for version control / source code control.

What is Git?

"Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency." See https://www.git-scm.com/

So it is clear that the intended use case problem to solve is "distributed version control". That is all, nothing more, nothing less. Many proofs of this are readily available.

"Version control — also known as source control or revision control — is an important software development practice for tracking and managing changes made to code and other files. It is closely related to source code management." gitlab source

What is Blockchain?

"Blockchain is a peer-to-peer decentralized distributed ledger technology that makes the records of any digital asset transparent and unchangeable and works without involving any third-party intermediary. It is an emerging and revolutionary technology that is attracting a lot of public attention due to its capability to reduce risks and fraud in a scalable manner." blockchain-council.org

Without repeating the technical details of blockchain already outlined in previous answers(i.e. mining, distributed networking), simply put, blockchain is a solution to an entirely different problem than those solved by Git.

Share Improve this answer

Follow

edited Jan 31, 2023 at 19:43

answered Jan 31, 2023 at 15:34

lacostenycoder

11.1k ● 4 ● 36 ● 50