How to build a Blockchain and why you should (not)

Guest Lecture Advanced Data Architectures

22-04-2022

Stefan Driessen

JADS Jheronimus Academy of Data Science

The founders of JADS

TU/e Technische Universiteit Eindhoven University of Technology

Tilburg University

's-Hertogenbosch

Provincie Noord-Brabant

Before we begin

Who here has ever bought cryptocurrency?

Who here can give me a definition of blockchain technology?

My Definition

Blockchain is *Decentralized, Distributed Ledger* Technology (DDLT)

[3]

Outline

1. **What** do these properties really mean?
2. **Why** do we want these properties? (and why shouldn’t we want them?)
3. **How** do we ensure these properties?
4. **When** do we want to use Blockchain?

DDLT: Ledger

- Ordered list of mutations of balances.
- From, To, Amount, Value after transaction, Description, etc.
- Let’s look at a Bitcoin Transaction
- Blocks are ordered sets of transactions → Blockchain
- A blockchain as a database.

Source: Investopedia

[5]

DDLT: Distributed

What does it mean?

- *All* the data in the ledger/blockchain is stored in *multiple locations*.

Why would we do this?

- Transparency
- Redundancy (immutability)
- Easily accessible

Why doesn't everyone do this?

- All of the above!
- High(er) cost
- Necessary communication delay

[6]

DDLT: Decentralized

What does it mean?

- No one entity is in charge of the ledger
  - Adding new data and removing / changing old data
  - We need consensus

Why?

  - Autonomous – Example
  - Trust(less) / Immutable – No one can change your data.
  - Democratic

Why not?

  - Hard to correct mistakes
  - Necessary Delay

The How

Cryptographic encryption for accounts

(Even though everything public, only I can use my account)

The consensus protocol

(How do we ensure that everyone’s copy is the same?)

The challenges of consensus: A thought experiment

- We are starting the new JADS blockchain with JADS-coin.
- You can earn JADS-coin by attending lectures.
- You can use JADS-coin to buy beer with Patterns.
- You can trade JADS-coin with your peers.
- We **all** keep a ledger of **every** transaction of JADS-coin!
- We can instantly broadcast our transactions to another (e.g. groupchat).

What happens when a transaction is missing from someone's ledger?

[9]

Why everything needs to be in everyone’s ledger

If a transaction to you is missing

- You can’t spend your JADS-coin with that person.

If a transaction from you is missing

- You can re-spend your JADS coin with that person (doublespending).

What about consensus

Why is it hard to achieve consensus?

- We don’t know each other, we don’t trust each other.
- We check each other, to keep each other honest.
- Naïve solution: For every new block, we vote: if we can convince a majority of the nodes, we are happy.

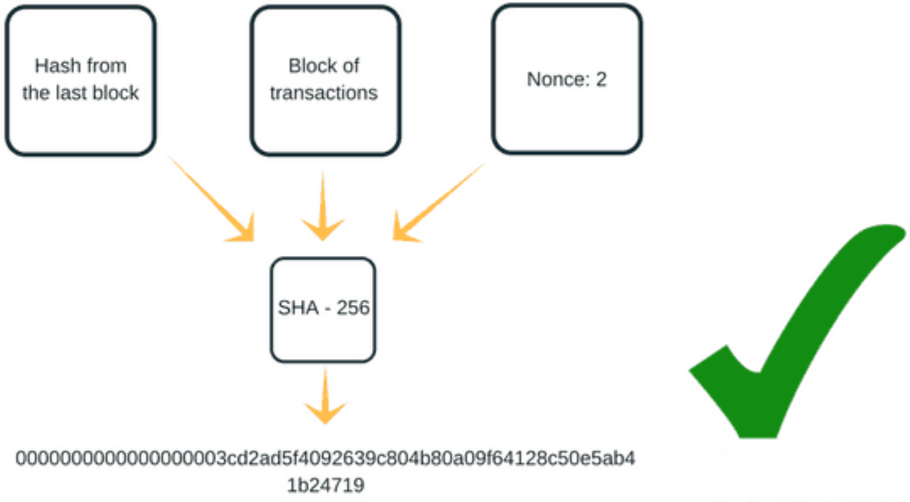
Solution: Proof of Something

- Adding a block costs **time** & **money**!
- Checking a block is **quick** & **free**.
- We all follow the **longest chain** (i.e. the one with the most blocks). So you want to add blocks to the longest chain only.
- Because you invest time & money, you are rewarded: coinbase + fees if other people verify your block.
- Does this work?

Proof of Work

- First one, all major blockchains use this.
- The real-world investment is computational power.
- Uses a hash function: Which is a function that takes a long number as input and maps it to a shorter output number.
- The output has to be **small** (this is what takes up the computational power)

Proof of Work: Calculation



Source: Async Labs

[14]

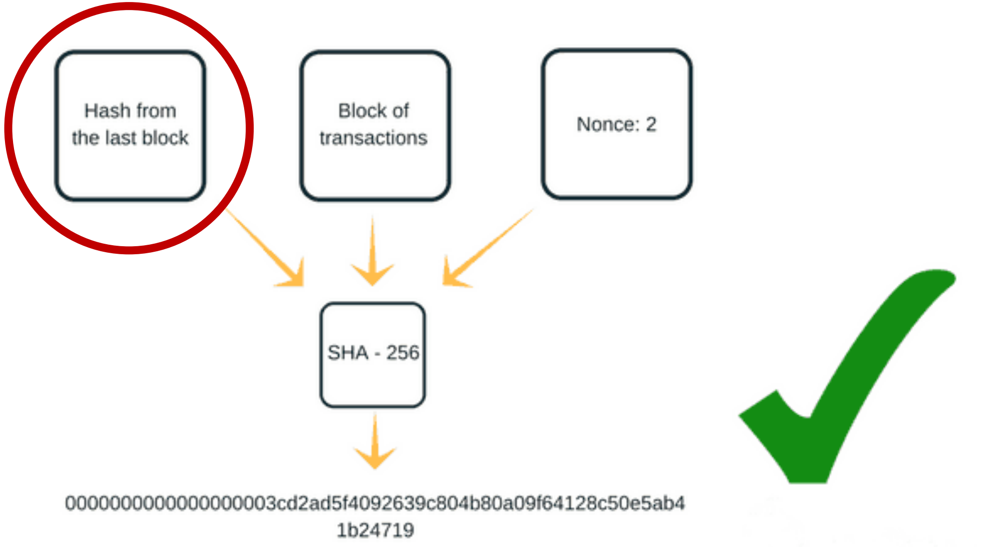
Proof of Work: continued

- Two properties of hashes that make this work:
  1. Hash function is non-surjective (i.e. it’s easy to check a solution, but hard to find one).
  2. Hash function is random (i.e. the most efficient method to get a low hash value is brute-force.

Disadvantages?

- Super (electricity) expensive
- Bad for decentralization: mining pools.

Proof of Work: Immutability



Hash from the last block

Block of transactions

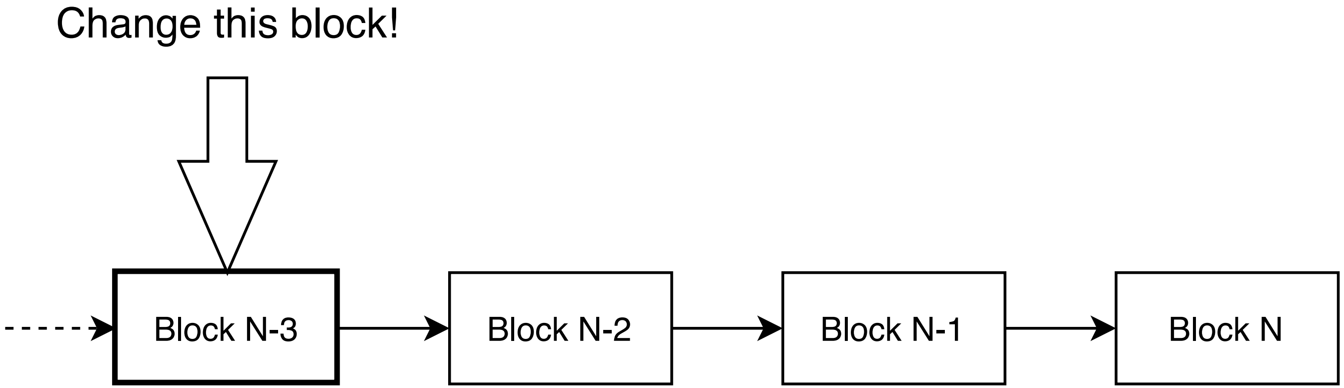
Nonce: 2

SHA - 256

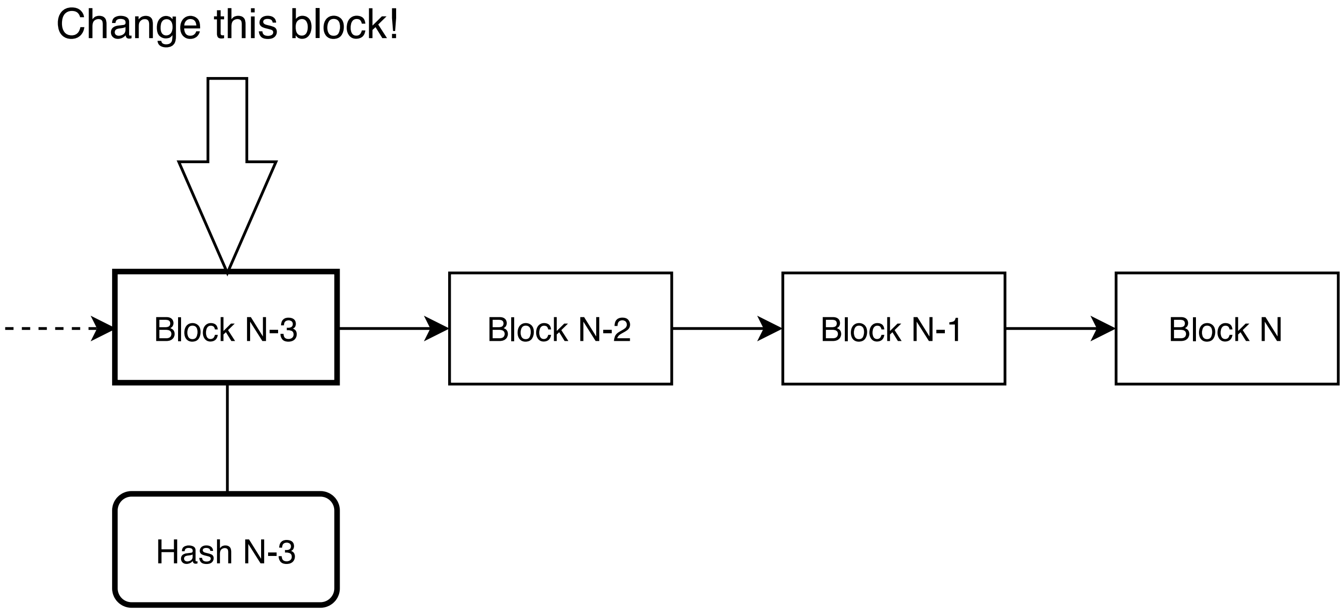
0000000000000000003cd2ad5f4092639c804b80a09f64128c50e5ab4
1b24719

Source: Async Labs

Proof of Work: Immutability



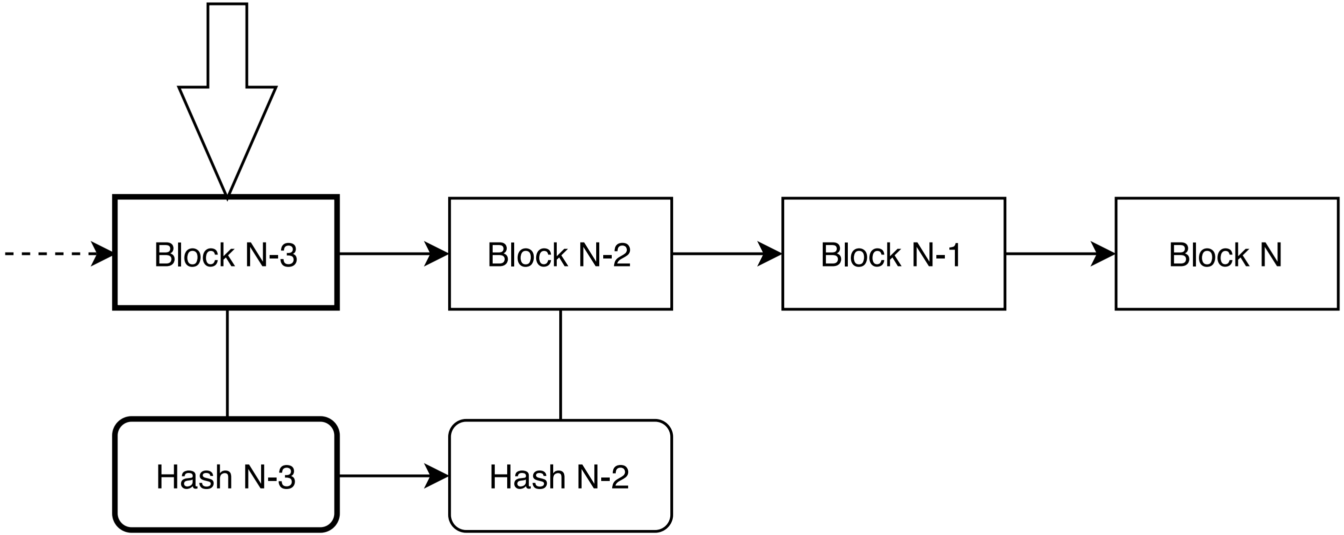
Proof of Work: Immutability



[18]

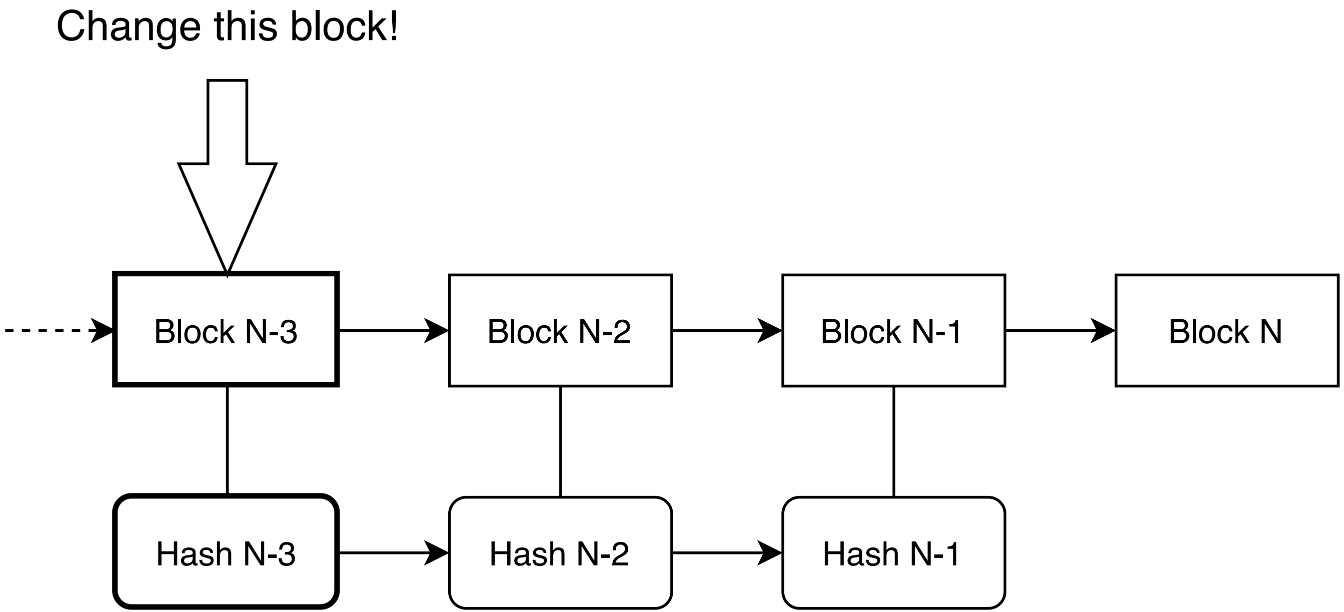
Proof of Work: Immutability

Change this block!



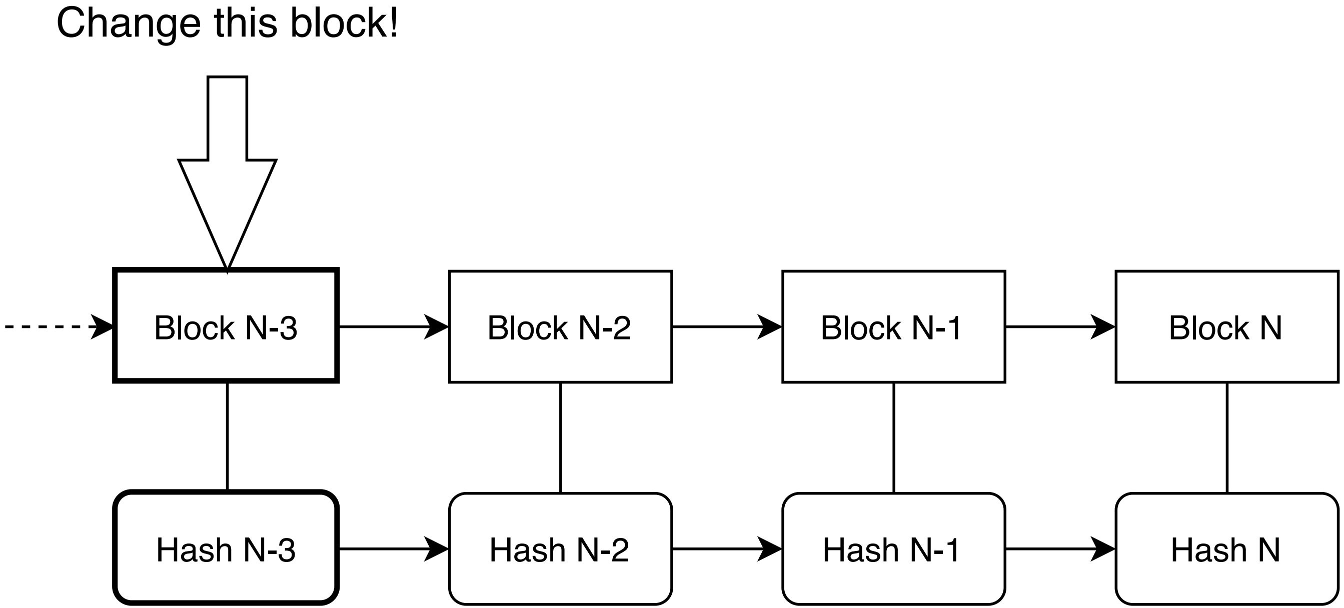
[19]

Proof of Work: Immutability



[20]

Proof of Work: Immutability



Proof-of-Work: Consequences

- Changing something in one block, means having to re-mine **all subsequent blocks.**
- This is practically impossible, **unless?**

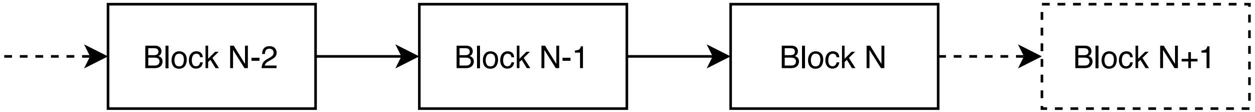
If you can mine faster than all other miners combined, you can change the past in a blockchain

This is known as a [51% attack]

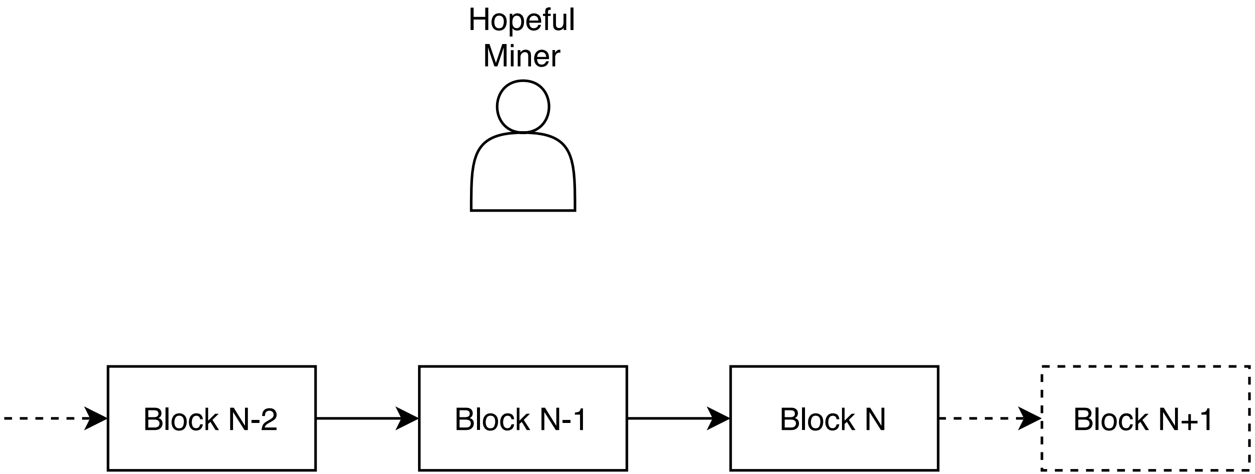
Proof of Stake

- You lock cryptocurrency, your chance to add a block is proportional to your stake compared to total stake.
- The “real world” investment is cryptocurrency from the chain. Is this a real-world investment?
- Every time a block is mined the miners randomly assign a next person or group of persons.
- Solves the scalability and energy problems!
- Not really being used in practice yet.

Proof of Stake: Example

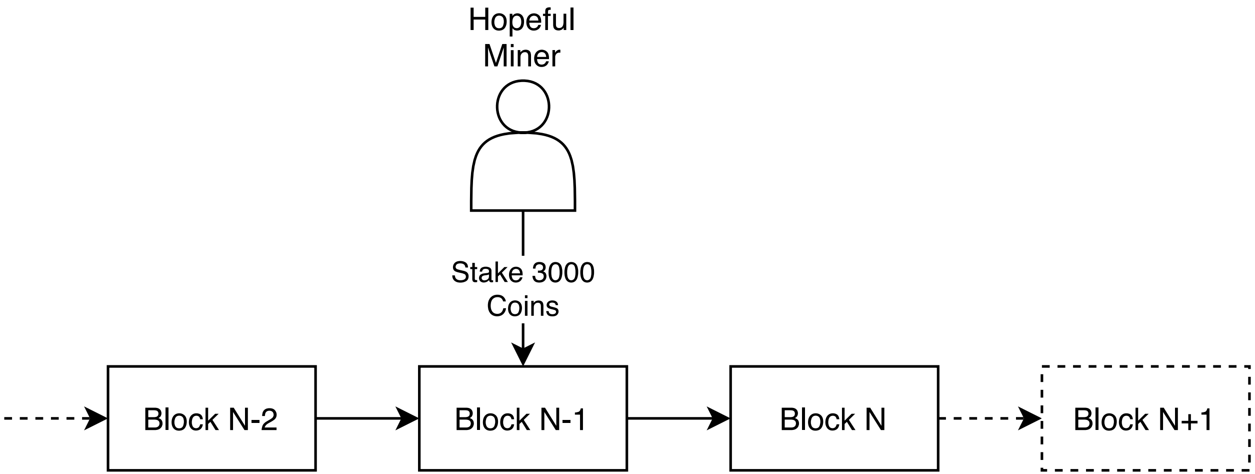


Proof of Stake: Example



[25]

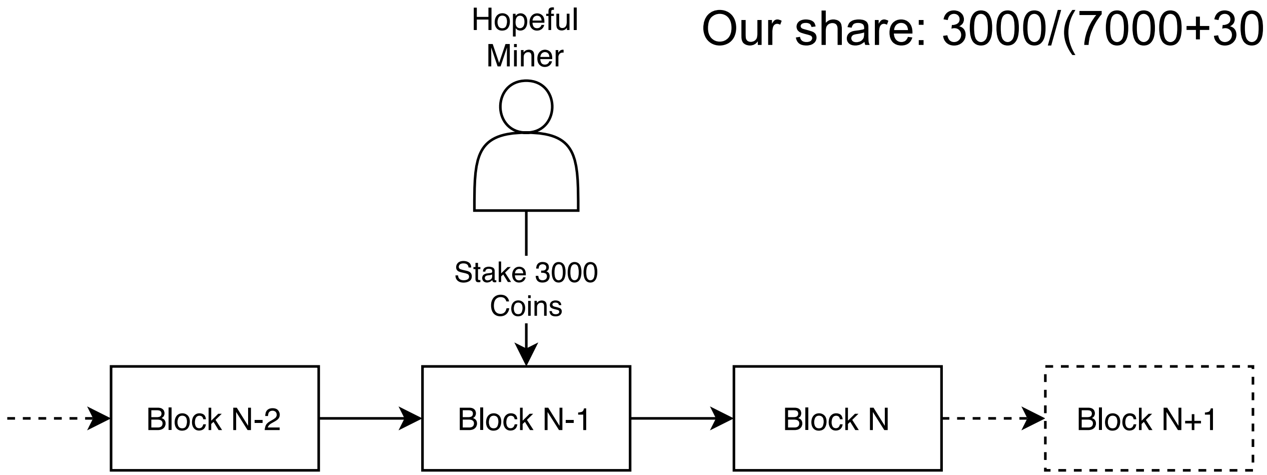
Proof of Stake: Example



[26]

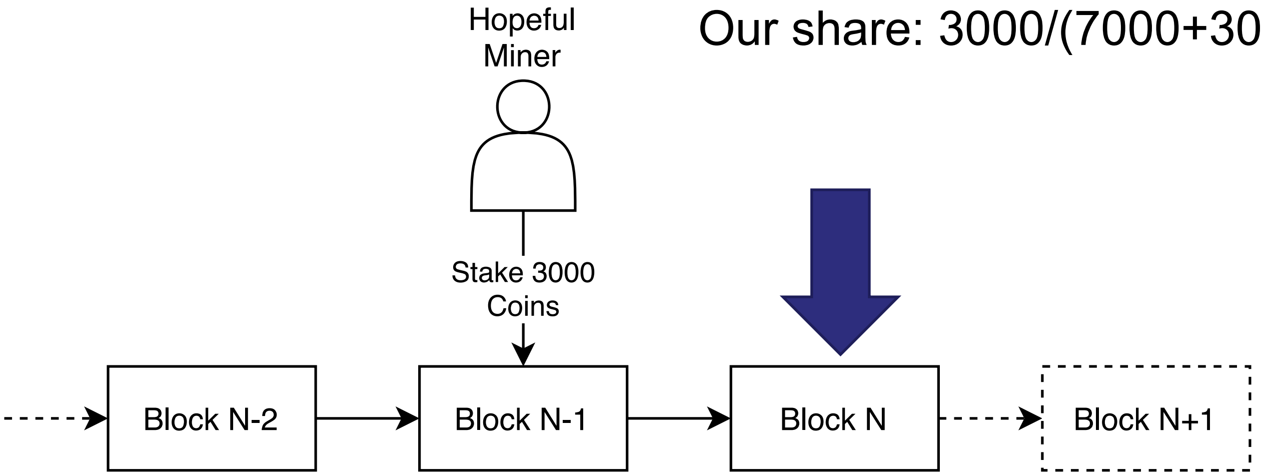
Proof of Stake: Example

Already staked: 7000 coins
Our share: 3000/(7000+300) = 30%



Proof of Stake: Example

Already staked: 7000 coins
Our share: 3000/(7000+300) = 30%



[28]

Smart Contracts

- Remember: *any* data can be on the blockchain!
- We can put *code* on the blockchain.

- Only need two things:
  1. Execution environment that comes with the blockchain
  2. Miners keep track of code state in addition to transactions.

Smart Contracts

Why do we want smart contracts?

- Decentralized
- Distributed

Transparency

Easy Access

Democratic

Redundancy

Cost reduction

Immutable

Trust(less)

Autonomous

[32]

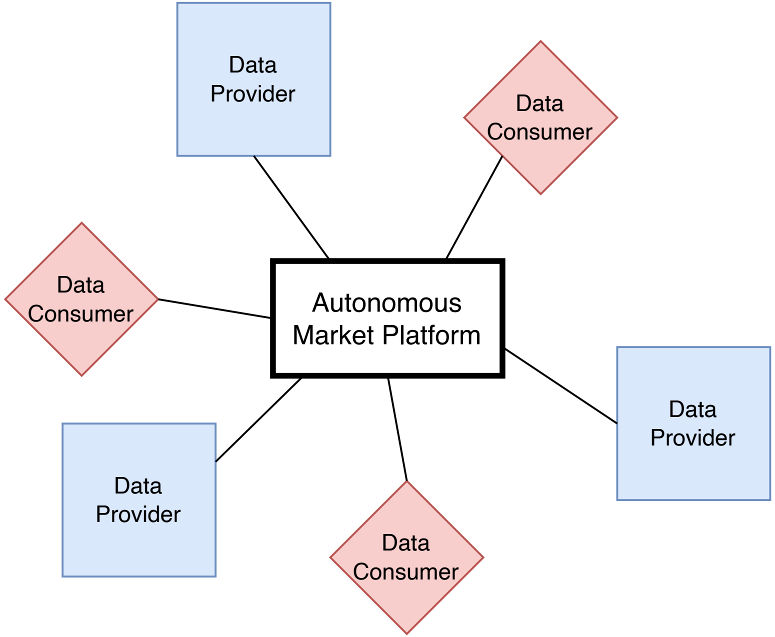
Application: Non Fungible Tokens

- Smart contract to keep track of who claims (digital) ownership of what.
- Advantages?
- Disadvantages?
  - No legal recognition!
  - Hard to enforce ownership

©rv

Application: Data Market

- Data Providers and Data Consumers exchange data.
- The exchange is facilitated by a fully autonomous, smart-contract based data market.
- Advantages?
- Disadvantages?



[34]

Application Domains

Disclaimer: I cannot predict the future!

Blockchain makes sense when:

- You benefit from *Decentral, Distributed* data architectures.
- Multiple parties share the same infrastructure, but they don't want to have to trust each other.
- Agreements can be automated
- The people involved understand the code / smart contracts

In general Blockchain technology can be used to eliminate the need for trusted third parties

[35]

Master Thesis Proposal

Automate Data Product Monitoring
with Smart Contracts

[36]

Thank you for listening