Money in the Big Data Age - Analyzing Blockchains

Frederik The

29 Januar 2018

Betreuer:

Dr. Julian Kunkel

Übersicht

1. Motivation
2. Technische Hintergründe
3. Datenvolumen & Transaktionsvolumen
4. Analytics Tools
5. Big Data Outlook
6. Zusammenfassung
7. Appendix

Motivation

Was ist Kryptogeld?
Versuch einer Definition

Kryptogeld ist digitales "Vermögen", das Kryptografie verwendet, um (1) seine Transaktionen zu sichern, (2) die Erstellung zusätzlicher Einheiten zu kontrollieren und (3) die Übertragung von "Vermögen" zu überprüfen.

Blockchain-Tokens sind digitale „Nutzungsrechte", die Kryptografie verwenden, um (1) Transaktionen zu sichern, (2) die Erstellung zusätzlicher Einheiten zu kontrollieren und (3) die Übertragung von „Nutzungsrechten“ zu überprüfen.

Zentralisiertes vs. Dezentralisiertes Kryptogeld

- Die Idee "kryptografischen Geldes" existiert seit 1980er Jahren (Chaum, 1983) [1]
- Bitcoin war das erste dezentralisierte Konzept [2]
- Heute gibt es 1000+ verschiedene "Krypto-Coins", wovon die meisten dezentrale Systeme verfolgen
- Viele dieser Blockchains erheben keinen Anspruch darauf ein Finanzinstrument zu sein, sondern sind eher mit Anteilen (sog. „Tokens“) an einer Firma oder einer speziellen Dienstleistung vergleichbar

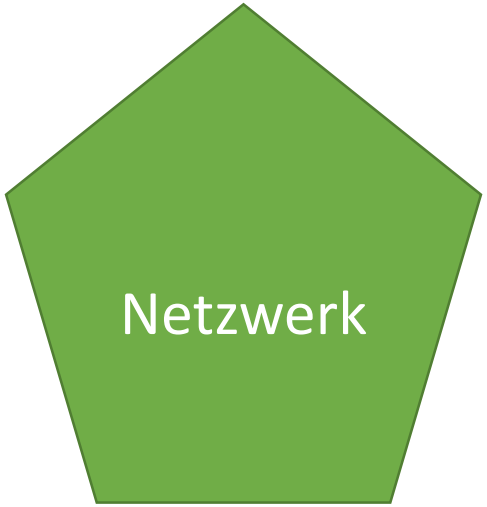
Zentralisiertes vs. Dezentralisiertes Kryptogeld

"*[T]he 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...What is needed is an electronic payment system based on cryptographic proof instead of trust...*" (S. Nakamoto, 2009) [2]

Zentralisierte Entscheidungen

“Vertrauen”

Dezentralisierte Entscheidungen



Kryptografischer Beweis

3 Arten von *Coins/Tokens*

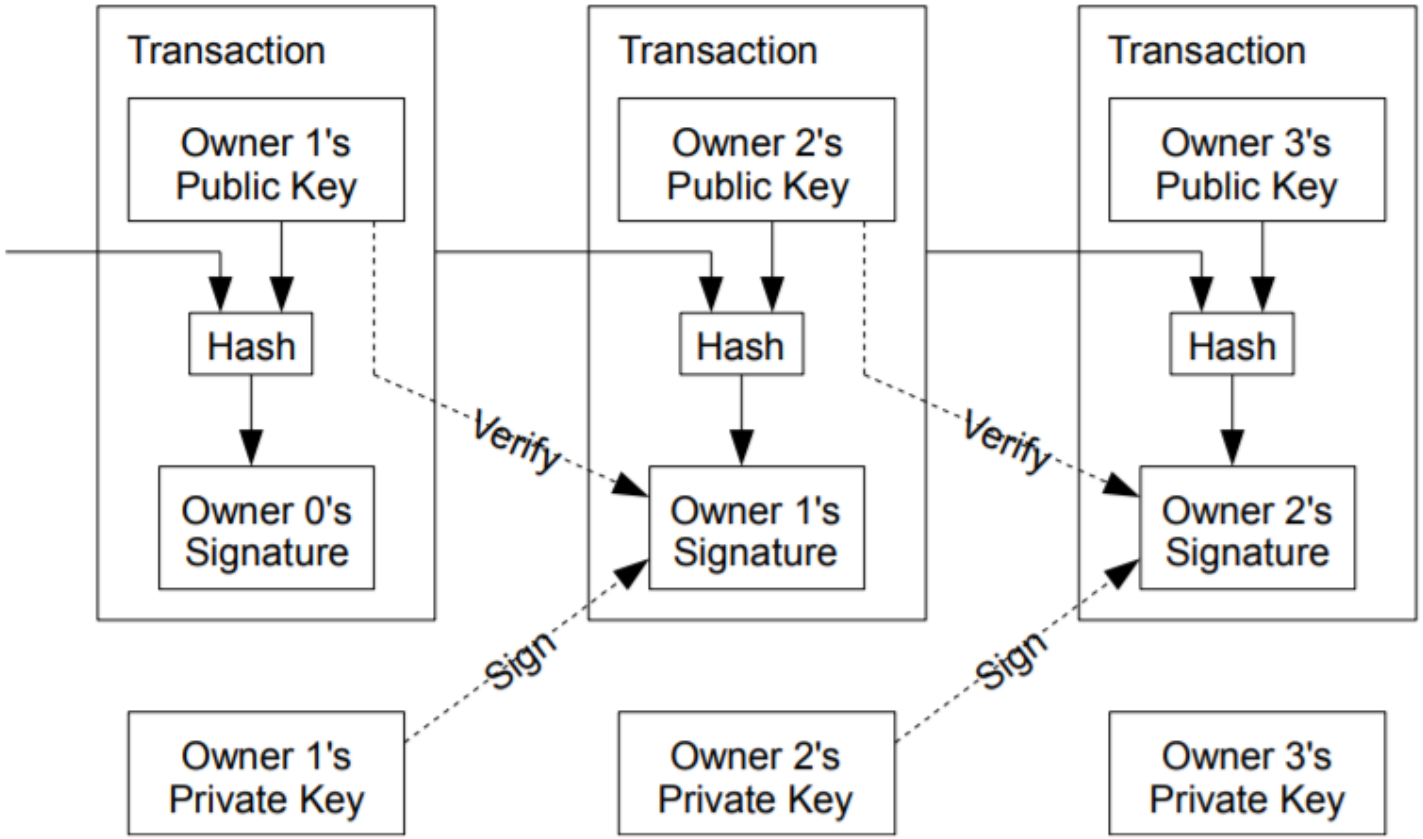
- Bitcoin
  - Die “Genesis Blockchain”, kreierte den Markt für *Distributed Ledger Technologie* (DLT)
  - Kein formaler Entstehungsprozess: Bitcoin existierte von heute auf morgen und Menschen haben angefangen es zu nutzen
  - Vergleichbar mit “digitalem Gold” als natürlicher Rohstoff/Ressource
- Security Tokens (ST)
  - STs repräsentieren Eigenkapitalanteil an einer Organisation/Projekt/Firma bzw. einen Anspruch auf zukünftig generierte Gewinne derselben (Investitionsvertrag)
  - Ether (Ethereum) ist nach Howey-Testkriterien der meistgenutzte *Security Token*
- Utility Tokens (UT)
  - UTs stellen das Recht zur Nutzung einer Dienstleistung/Technologie dar (meistens basierend auf einem Smart Contract). Diese Tokens sind mit *in-game currencies* oder *pay-per-use SaaS* Angeboten zu vergleichen.

Technische Hintergründe

Elemente dezentralisierten Kryptogeldes

1. Zustandsübergangssystem
2. Konsessystem (PoS, PoW, etc.)
3. Timestampserver

Transaktionen (Zustandsübergangssystem)

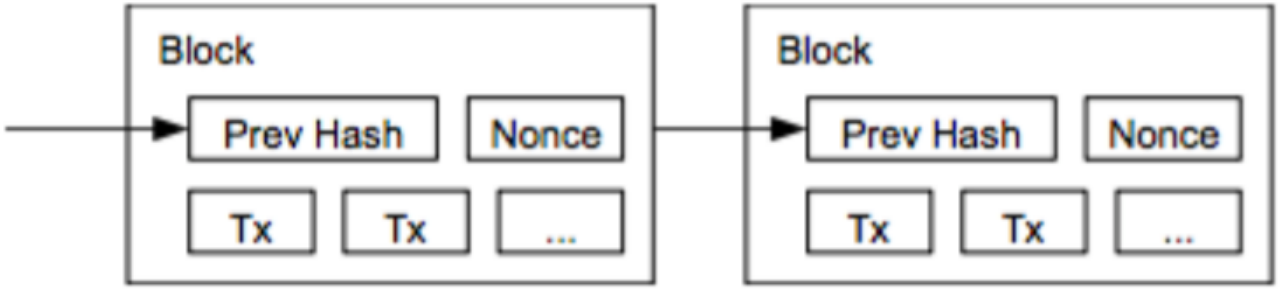


Quelle: “Bitcoin: Peer-to-Peer Electronic Cash System”, S. Nakamoto (2009) [2]

Konses durch Kryptografie - "Proof of Work"

"Proof-of-work is essentially one-CPU-one-vote" (S. Nakamoto, 2009)

- Finde eine kryptografische Nonce, die einen Hash (double-SHA-256) unterhalb eines Zielwerts Z liefert, dem Schwierigkeitsgrad des Systems zum Zeitpunkt t
- Schwierigkeitsgrad wird durch einen gleitenden Mittelwert anhand der durchschnittlichen Anzahl an Blöcken pro Stunde bestimmt
- Zeitlich lineares Aneinanderreihen von Blöcken ermöglicht eine einheitliche "Geschichtsschreibung"



Source: *"Bitcoin: A Peer-to-Peer Electronic Cash System", Nakamoto, 2009* [2]

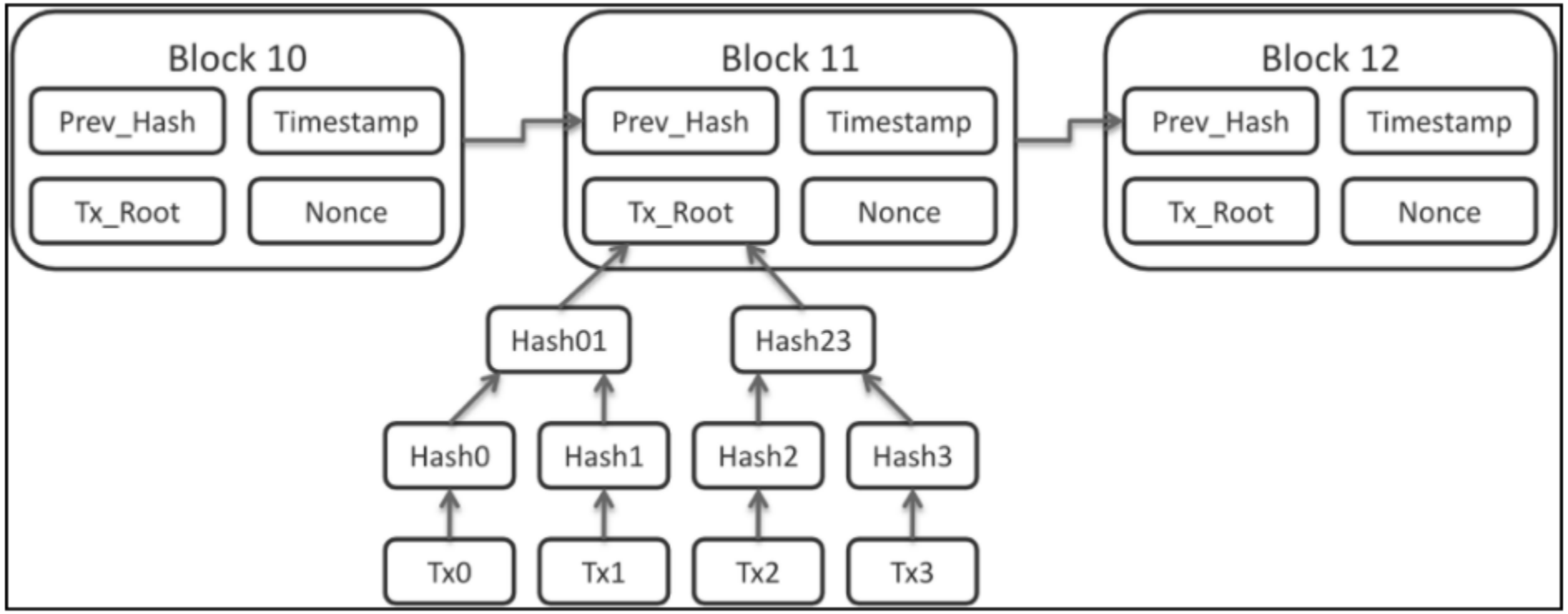
Blockvalidierungsalgorithmus

Algorithmus:

1. Prüfe ob der vorherige Block existiert und valide ist
2. Prüfe, dass Timestamp des Blocks zum Zeitpunkt t > t-1
3. Prüfe, dass proof of work für den Block valide ist
4. S[0] sei der Zustand am Ende des vorherigen Blocks
5. Angenommen TX ist die Liste aller n Transaktionen des Blocks. Für alle i in 0...n-1, prüfe dass, S[i+1] = APPLY(S[i],TX[i]) - Im Falle eines Fehlers, EXIT und return FALSE
6. Ansonsten TRUE und registriere S[n] als den Status am Ende des jetzigen Blocks

Quelle: Buterin, V. “A next-generation smart contract and decentralized application platform.” [3]

Vom Block zur Blockchain



Quelle: Rosic, A. “What Is Hashing? Under The Hood Of Blockchain.” Blockgeeks, Sep. 2017 [4]

Blockstruktur (Bitcoin)

Feld	Beschreibung	Größe
Fixer Wert	immer 0xD9B4BEF9	4 Bytes
Blockgröße	Zahl der Bytes bis zum Ende des Blocks	4 Bytes
Blockheader	besteht aus sechs Teilen	80 Bytes
Transaktionszähler	positive Ganzzahl	1 bis 9 Bytes
Transaktionen	die (nichtleere) Liste von Transaktionen	So viele Transaktionen, wie im Transaktionszähler genannt

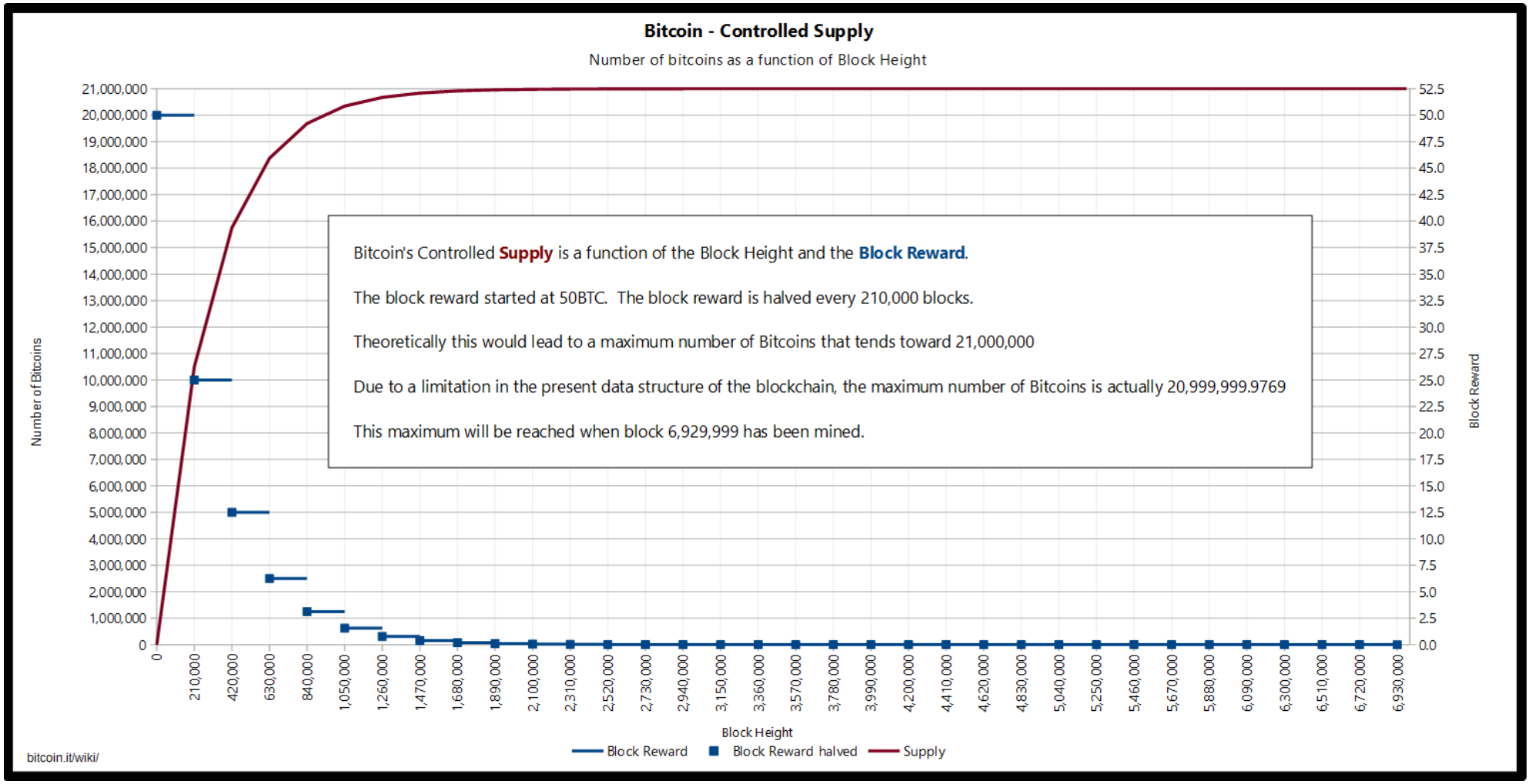
Quelle: “Block.” In Bitcoin Wiki. [5]

Blockheader (Bitcoin)

Field	Purpose	Updated when...	Size (Bytes)
Version	Block version number	You upgrade the software and it specifies a new version	4
hashPrevBlock	256-bit hash of the previous block header	A new block comes in	32
hashMerkleRoot	256-bit hash based on all of the transactions in the block	A transaction is accepted	32
Time	Current timestamp as seconds since 1970-01-01T00:00 UTC	Every few seconds	4
Bits	Current target in compact format	The difficulty is adjusted	4
Nonce	32-bit number (starts at 0)	A hash is tried (increments)	4

Quelle: “Block hashing algorithm.” In Bitcoin Wiki. [6]

Angebot (Bitcoin)



Quelle: “Controlled supply.” In Bitcoin Wiki. [7]

Datenvolumen & Transaktionsvolumen

Größe der Blockchain

Big Data Potential!

No Big Data yet!

Bitcoin

- Blockchain Größe: ~ 140 GB (Jan 2018)
- Total Tx: ~ 300 million (seit 2009)
- Tx/sec: 3-7
- Market Cap: ~ 228 billion USD
- Umlauf: ~ 16 million BTC

Quelle: Block Explorer, Retrieved Jan 2018 from: https://blockexplorer.com/

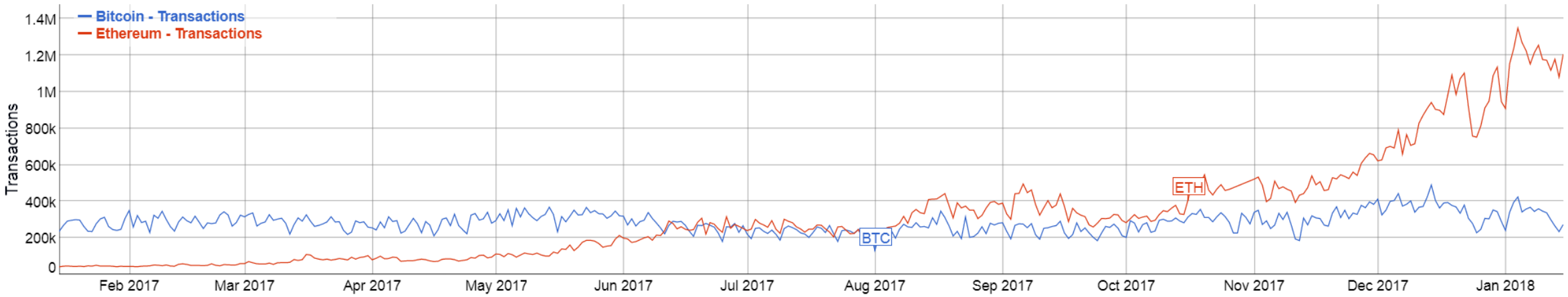
Ethereum

- Blockchain Größe: ~ 45 GB (Jan 2018)
- Total Tx: ~ 135 million (seit 2015)
- Tx/sec: 7-15 (25 max)
- Market Cap: ~ 123 billion USD
- Umlauf: ~ 97 million ETH

Quelle: Etherscan, Retrieved Jan 2018 from: https://etherscan.io/.

Vergleich: Tx/sec VISA ~ 2 thousand +

Bitcoin vs. Ethereum Transaktionen



Quelle: Bitinfocharts, Retrieved Jan 2018 from: https://bitinfocharts.com/comparison/ethereum-transactions.html

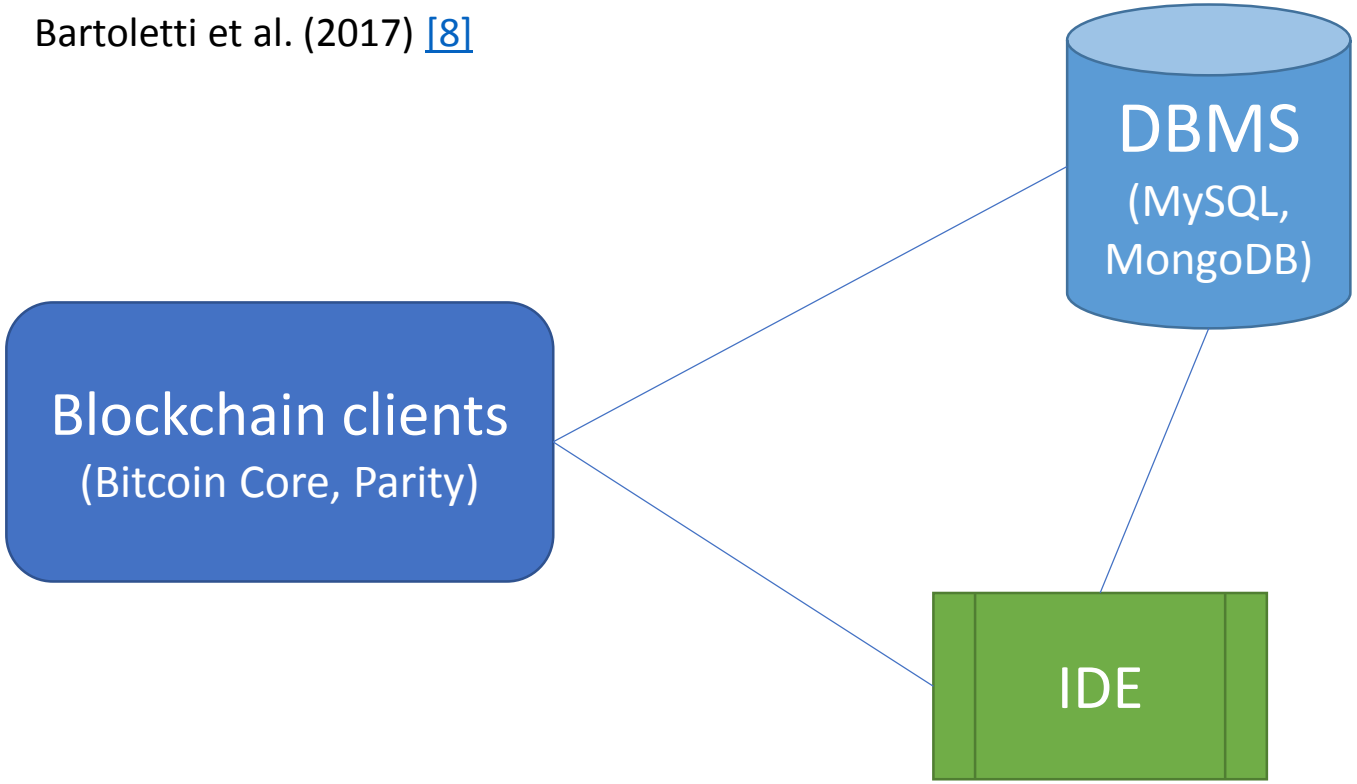
Analytics Tools

Bestehende Tools

- Blockchain Explorer APIs
  - Blockexplorer
  - Etherscan
  - blockchain.info
  - Coindesk
- Trading Platform APIs
  - Kraken
  - Coinbase
- Custom APIs
  - Project Scala-MongoDB
  - Andere Github repos

Custom API

- Als All-zweck Blockchain Analytics Tool gedacht
- Bartoletti et al. (2017) [8]



Kombiniere Daten innerhalb der Blockchain mit externen Daten (Wechselkurse, Tags, etc.)

Workflow:

1. Konstruiere eine Ansicht der Blockchain und speichere sie im DBMS

2. Analysiere die Ansicht mit Hilfe der Query Sprache der Datenbank

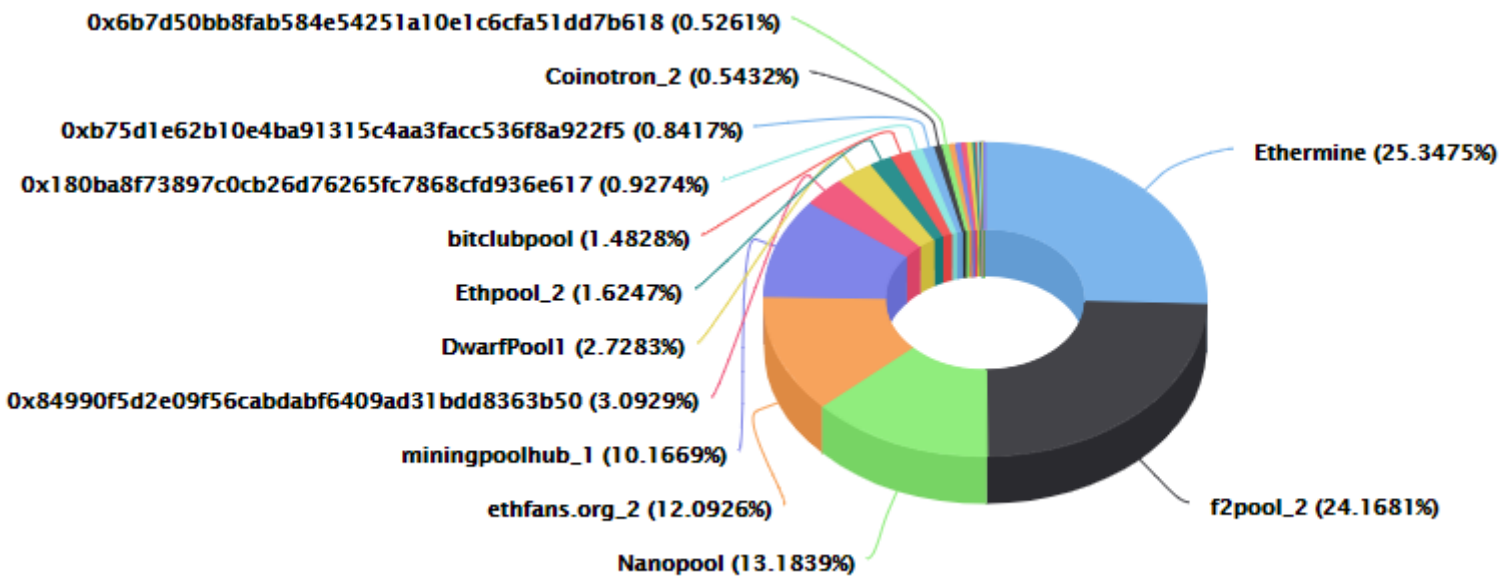
3. Nutze R für komplexere ML Algorithmen

Mongolite package

Analytics Beispiele

1. Top Miners
2. Adressen
3. Wer besitzt was?
4. Meta Daten / Smart Contracts
5. Transaktionsgebühren
6. Preisvorhersagen
7. Clustering

Ethereum Top 25 Miners by Blocks (letzten 7 Tage)



Quelle: EtherScan.io, Retrieved Jan 2018 from: https://etherscan.io/stat/miner?range=7&blocktype=blocks

Ethereum Addressenwachstum

Quelle: EtherScan.io, Retrieved Jan 2018 from: https://etherscan.io/chart/address

10 Größten Bitcoin Adressen in BTC

	Address	Balance Δ1w	% of coins	First In	Last In	Number Of Ins Δ1w	First Out	Last Out	Number Of Outs Δ1w
1	16rCmCmbuWDhPjWTrpQGaU3EPdZF7MTdUk	179,203 BTC ($2,056,842,937 USD) +0.0001 BTC	1.07%	2016-02-27 19:00:09	2018-01-27 03:51:51	93 +1	2016-11-16 21:50:07	2017-10-27 22:25:49	32
2	3D2oetdNuZUqQHPJmcMDDHYoqkyNVsFk9r wallet: Bitfinex-coldwallet	138,685 BTC ($1,591,787,529 USD) -12083 BTC	0.8247%	2017-01-05 13:34:15	2018-01-26 17:26:45	4255 +36	2017-01-06 11:29:06	2018-01-26 16:21:14	3884 +57
3	16ftSEQ4ctQFDtVZiUBusQUjRrGhM3JYwe	117,170 BTC ($1,344,845,418 USD)	0.6967%	2017-12-08 08:51:10	2018-01-14 03:29:12	81	2017-12-10 17:55:29	2018-01-08 04:24:45	39
4	3Nxwenay9Z8Lc9JBiywExpnEFiLp6Afp8v	83,348 BTC ($956,649,214 USD) +5000 BTC	0.4956%	2015-10-16 16:43:06	2018-01-26 14:16:01	154 +1	2015-10-29 11:44:26	2017-12-29 14:21:16	51
5	1FeexV6bAHb8ybZjqQMjJrcCrHGW9sb6uF	79,957 BTC ($917,726,393 USD) +0.0002 BTC	0.4755%	2011-03-01 11:26:19	2018-01-21 07:47:25	119 +1			
6	18rnfoQgGo1HqvVQaAN4QnxjYE7Sez9eca wallet: 29043297	69,600 BTC ($798,849,694 USD)	0.4139%	2014-10-24 12:40:08	2017-12-17 10:00:08	251	2014-10-27 07:55:11	2018-01-06 19:12:34	101
7	1HQ3Go3ggs8pFnXuHVHRytPCq5fGG8Hbhx	69,370 BTC ($796,211,041 USD)	0.4125%	2013-04-09 23:03:36	2017-12-15 04:48:03	82	2015-04-23 16:10:25	2015-04-23 16:10:25	1
8	1PnMfRF2enSZnR6JSexxBHuQnxG8Vo5FVK	66,452 BTC ($762,718,585 USD)	0.3952%	2013-11-22 20:06:31	2017-12-09 08:47:57	125			
9	1AhTjUMztCihiTyA4K6E3QEpobjWLwKhkR	66,379 BTC ($761,877,674 USD)	0.3947%	2014-02-25 06:33:06	2017-12-31 06:12:49	192			
10	1DiHDQMPFu4p84rkLn6Majj2LCZZZRQUaa	66,236 BTC ($760,236,531 USD)	0.3939%	2013-11-23 01:08:37	2017-12-29 04:20:37	138			

Quelle: Bitinfocharts.com, Retrieved Jan 2018 from: https://bitinfocharts.com/top-100-richest-bitcoin-addresses.html

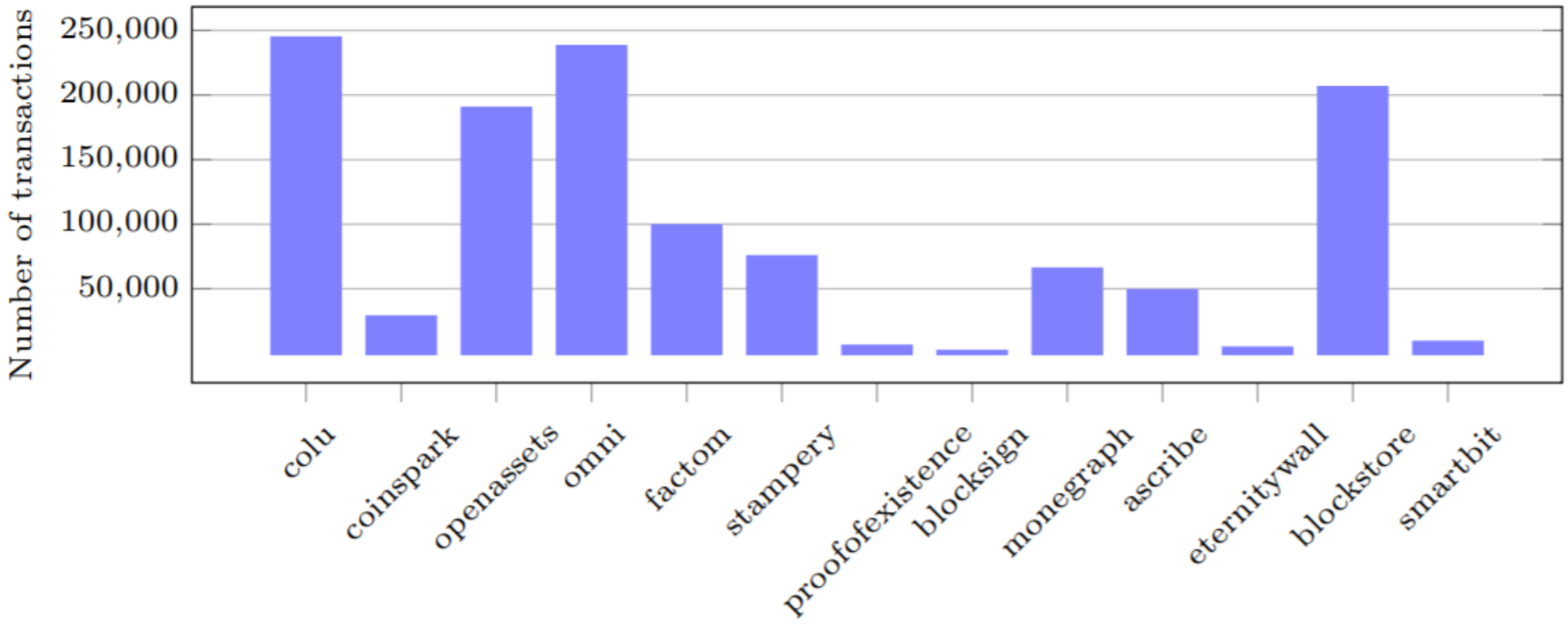
Extrahieren von Bitcoin Metadata

```
1  val opReturnOutputs = new Collection("opReturn", mongo)
2
3  blockchain.start(290000).end(473100).foreach(block => {
4      block.bitcoinTxs.foreach(tx => {
5          tx.outputs.foreach(out => {
6          if(out.isOpreturn()) {
7            opReturnOutputs.append(List(
8            ("txHash", tx.hash),
9            ("date", block.date),
10           ("protocol", OpReturn.getApplication(out.outScript.toString)),
11           ("metadata", out.getMetadata())
12           ))
13         }
14         })
15     })
16 })
```

Quelle: Bartoletti et al. (2017), Fig.4 [8]

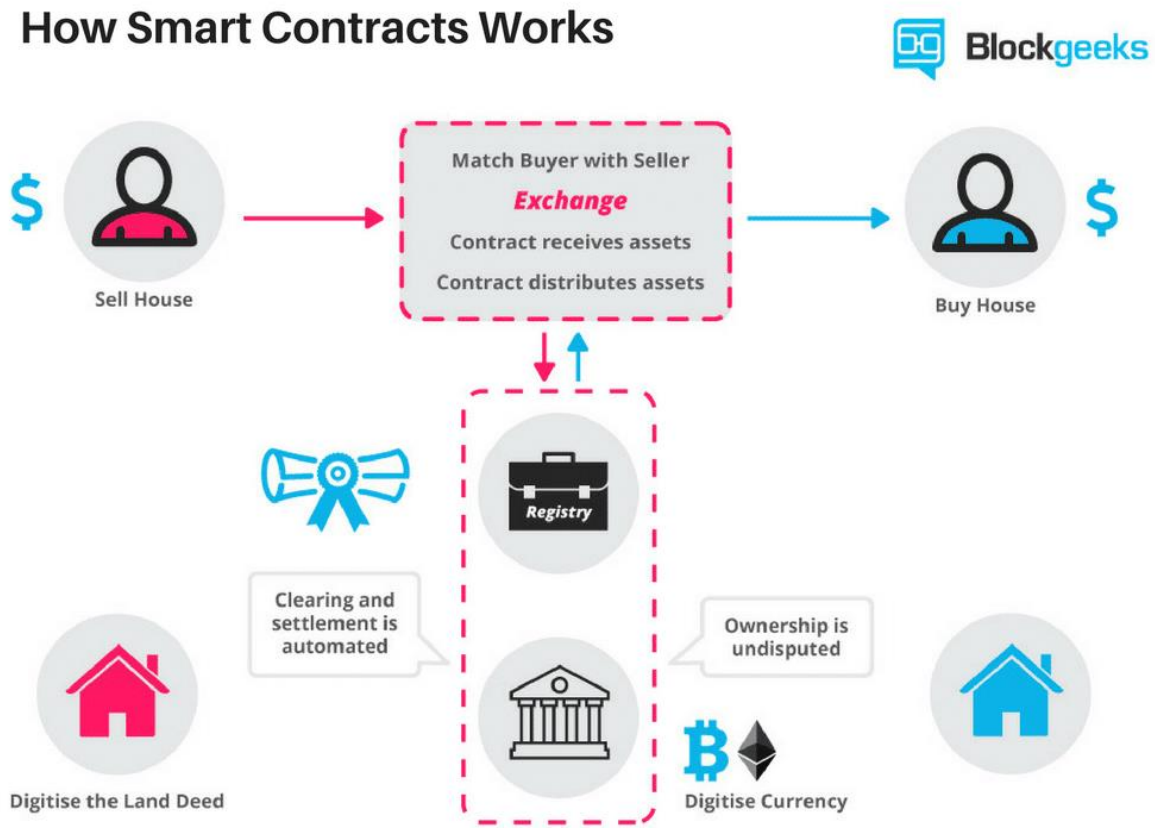
Anzahl der Transaktionen pro Protokoll

(Protokolle mit > 1000 tx)



Quelle: Bartoletti et al. (2017), Fig.5 [8]

Smart Contracts ("DAPPs" – Decentralized Applications)

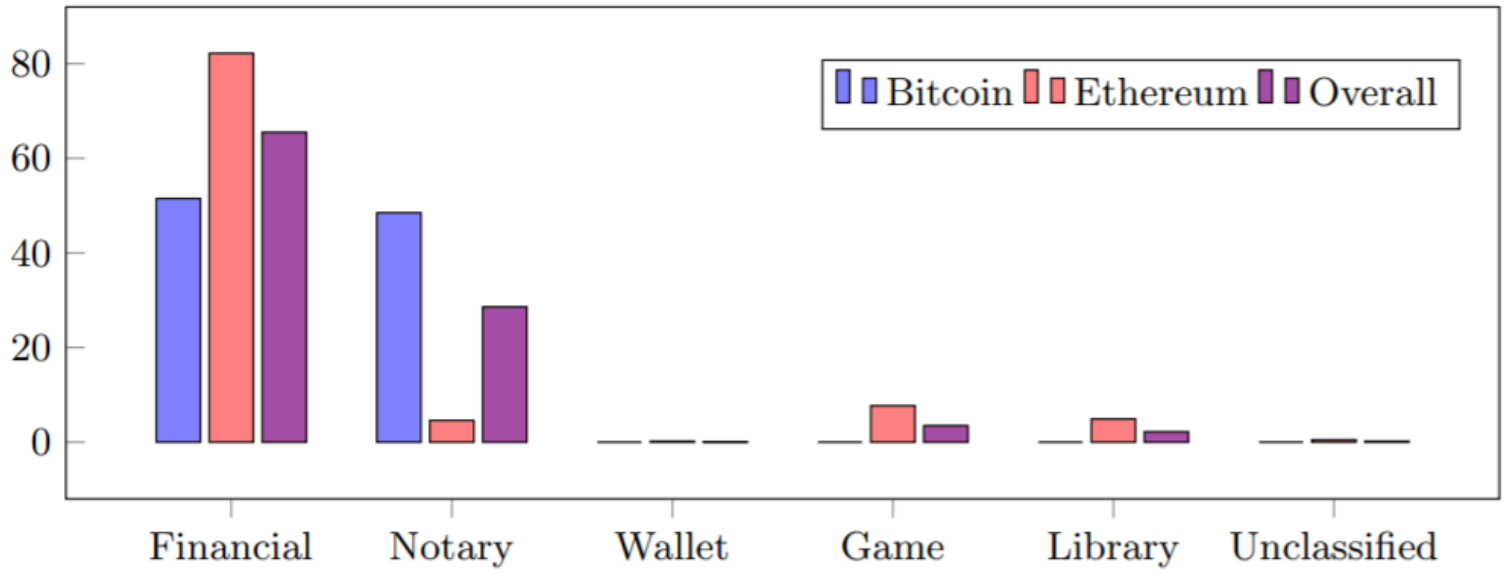


Quelle: Rosic, A. "Smart Contracts: The Blockchain Technology That Will Replace Lawyers." Blockgeeks, Feb. 2017 [9]

Smart Contract Kategorien

Category	Platform	Contracts	Transactions
Financial	Bitcoin	6	470,391
	Ethereum	373	624,046
Notary	Bitcoin	17	443,269
	Ethereum	79	35,253
Game	Bitcoin	0	0
	Ethereum	158	58,257
Wallet	Bitcoin	0	0
	Ethereum	17	1,342
Library	Bitcoin	0	0
	Ethereum	29	37,034
Unclassified	Bitcoin	0	0
	Ethereum	155	3,679
Total	Bitcoin	23	913,660
	Ethereum	811	759,611
	Overall	834	1,673,271

Quelle: Bartoletti et al. (2017), Table 2 [10]



Quelle: Bartoletti et al. (2017), Fig. 2 [10]

Transaktionsgebühren

```
1  val blockchain = BlockchainLib.getBitcoinBlockchain(new BitcoinSettings("user","
       password","8332",MainNet,true))
2  val mongo = new DatabaseSettings("myDatabase", MongoDB, "user", "password")
3  val txWithFees = new Collection("txWithFees", mongo)
4
5  blockchain.end(473100).foreach(block => {
6      block.bitcoinTxs.foreach(tx => {
7          txWithFees.append(List(
8              ("blockHash", block.hash),
9              ("txHash", tx.hash),
10             ("fee", tx.getInputsSum() - tx.getOutputsSum()),
11             ("date", block.date),
12             ("rate", Exchange.getRate(block.date))
13         ))
14     })
15 })
```

Quelle: Bartoletti et al. (2017) [8]

Top 5 *"Whale Transactions"*

Average: $\bar{x} = 0.41$ Standard Deviation: $\sigma = 12.09$

Fee (USD)	Date	Transaction hash
136243.37	2016-04-26 14:15:22	cc455ae816e6cdafdb58d54e35d4f46d860047458eacf1c7405dc634631c570d
56493.50	2017-01-04 20:01:28	d38bd67153d774a7dab80a055cb52571aa85f6cac8f35f936c4349ca308e6380
39502.15	2017-05-31 14:28:51	cb95ab3aef378c14bc59d0db682d96202b981c1f8fad7d66e23e0be06f2a00c4
25095.71	2017-05-31 14:28:51	8e12a1aba87e4657f5fabec1121ed57f706805ad6d4ffe88c6fce78596bd9b75
23518.00	2013-08-28 10:45:17	4ed20e0768124bc67dc684d57941be1482ccdaa45dadb64be12afba8c8554537

Quelle: Bartoletti et al. (2017) [8]

Features für Bitcoin Preis Vorhersagen

FEATURE	DEFINITION
Average Confirmation Time	Ave. time to accept transaction in block
Block Size	Average block size in MB
Cost per transaction percent	Miners revenue divided by the number of transactions
Difficulty	How difficult it is to find a new block
Estimated Transaction Volume	Total output volume without change from value
Hash Rate	Bitcoin network giga hashes per second
Market Capitalization	Number of Bitcoins in circulation * the market price
Miners Revenue	(number of BTC mined/day * market price) + transaction fees
Number of Orphaned Blocks	Number of blocks mined / day not off blockchain
Number of TXN per block	Average number of transactions per block
Number of TXN	Total number of unique Bitcoin transactions per day
Number of unique addresses	Number of unique Bitcoin addresses used per day
Total Bitcoins	Historical total Number of Bitcoins mined
TXN Fees Total	BTC value of transaction fees miners earn/day
Trade Volume	USD trade volume from the top exchanges
Transaction to trade ratio	Relationship of BTC transaction volume and USD volume

Quelle: Madan, I. et al. (2015) [11]

Binomiale Klassifizierung

STATISTIC	BINOMIAL GLM	SVM	RANDOM FOREST
Sensitivity (TPR)	0.9790	0.0348	1.0
Specificity (TNR)	0.9939	0.5514	0.9392
Precision (PPV)	0.9790	0.0839	0.7762
Accuracy (ACC)	0.9879	0.2716	0.9498

Daily Interval

STATISTIC	10 SECOND GLM	10 MINUTE GLM	10 MINUTE RANDOM FOREST
Sensitivity (TPR)	0.5429	0.524	0.540
Specificity (TNR)	0.577	0.576	0.619
Precision (PPV)	0.574	0.551	0.581
Accuracy (ACC)	0.085	0.539	0.574

Dies ist vergleichbar mit einem Münzwurf, d.h. Raten!

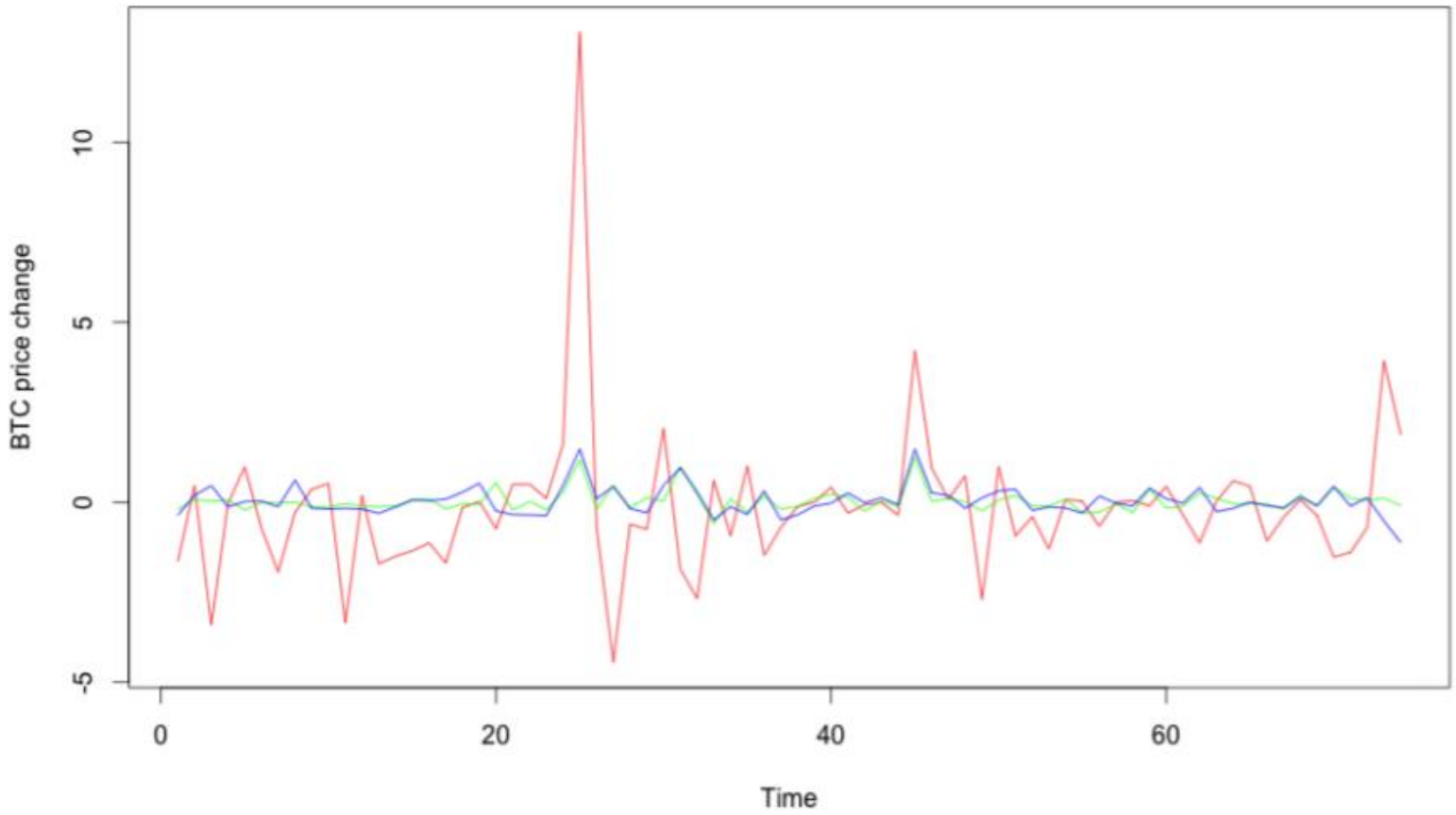
Quelle: Madan, I. et al. (2015) [11]

Tatsächliche Preise vs. Vorhersage

Actual

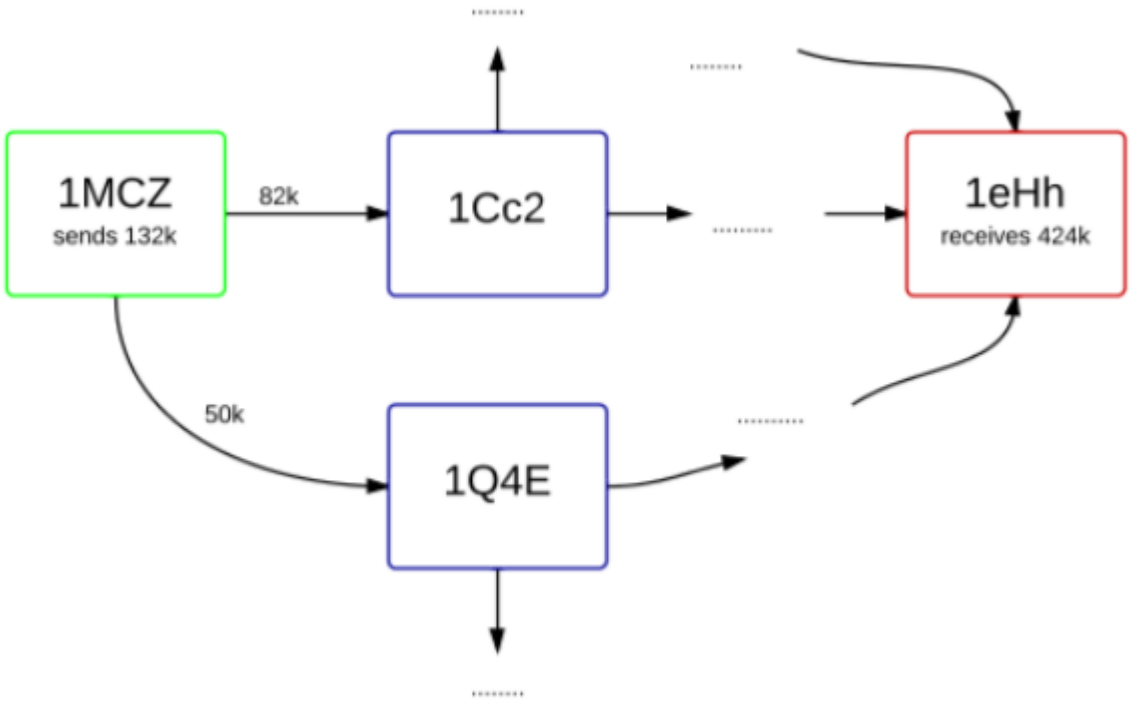
GLM

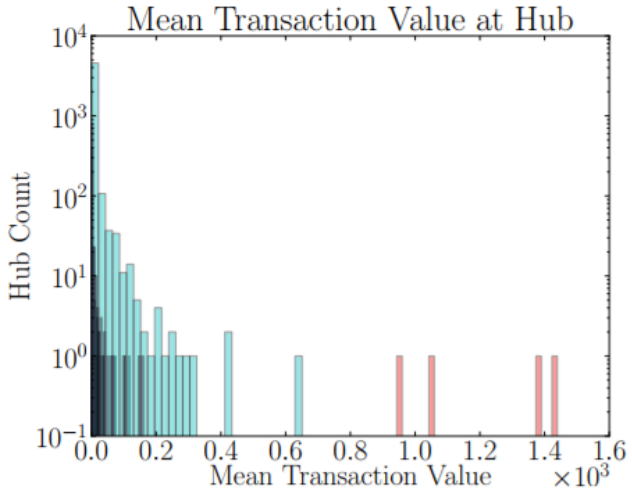
Random Forest

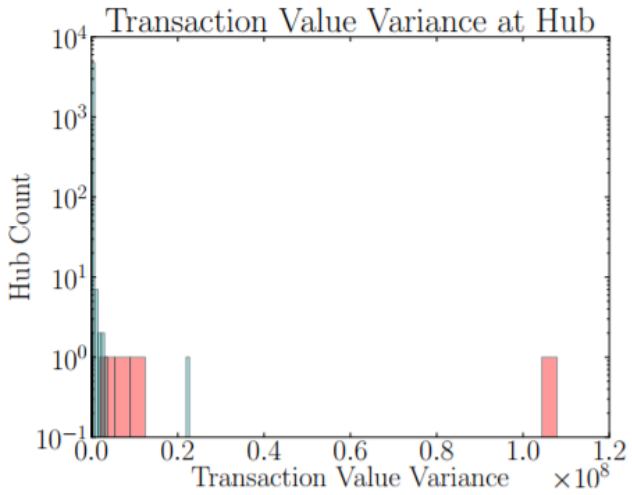


Quelle: Madan, I. et al. (2015) [11]

Clustering Addressen/Transaktionen







Quelle: Hirshman et al. (2013), Fig. 3 und Fig. 6 [12]

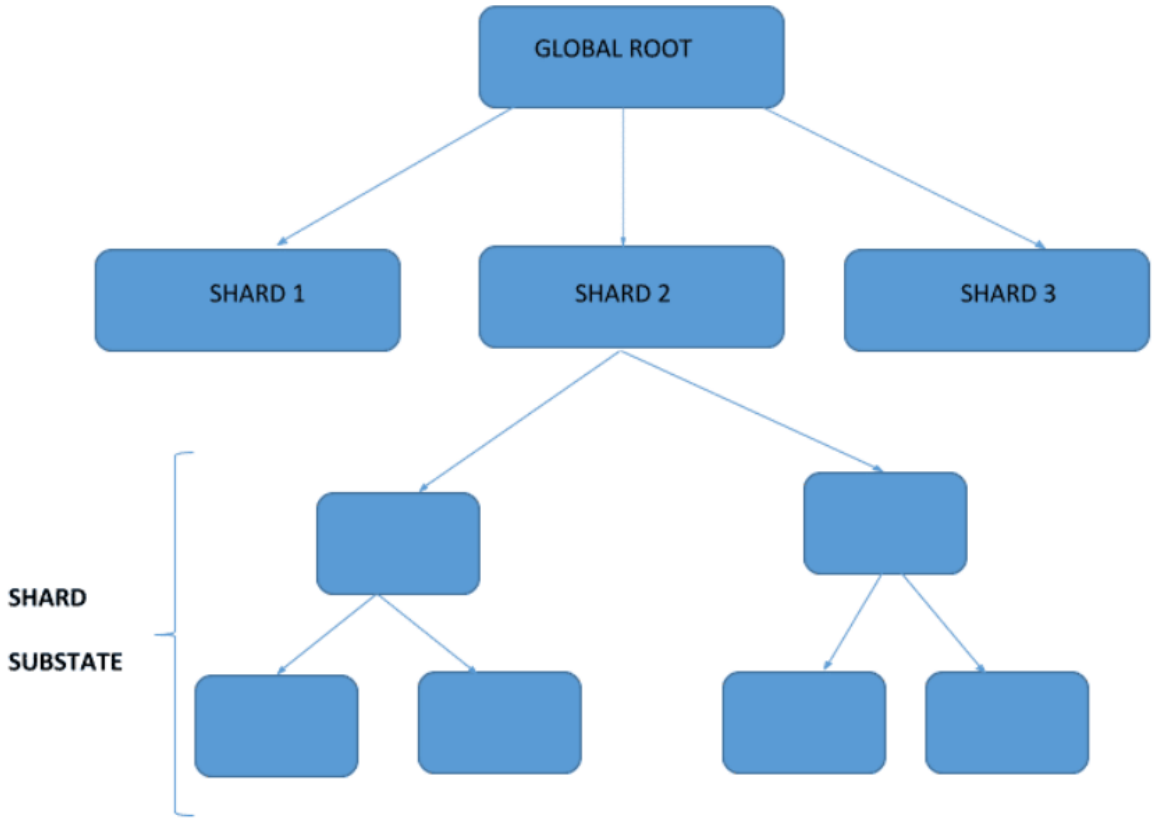
Big Data Outlook

- Zukünftiges Transaktionsvolumen: 2000 tx/s
- Zukünftige Marktkapitalisierung: 1 Billion USD
- Zukünftige Größe einer Blockchain: 100 TB

Skalierbarkeit

1. “Scale-Up”
   - Increasing Block Size → Bitcoin Cash
   - Merge Mining
2. “Scale-Out” (parallelization)
   1. **On-Chain**
      - Sharding
        - Existiert seit den 90er Jahren
        - “Vertical Partitioning”
   2. **Off-Chain**
      1. Channel-based technologies
         - Raiden (Ethereum)
         - Lightning Network (Bitcoin)
      2. Subchains
         - Plasma (Ethereum)

Sharding



Shard

Shard ID:	43	<sig #1284>	<sig #2543>
Pre state:	a138b3ff	<sig #7821>	<sig #6118>
Post state:	835680cc	<sig #9053>	<sig #4337>
Receipt root:	fa3819d4	<sig #1662>	<sig #4785>

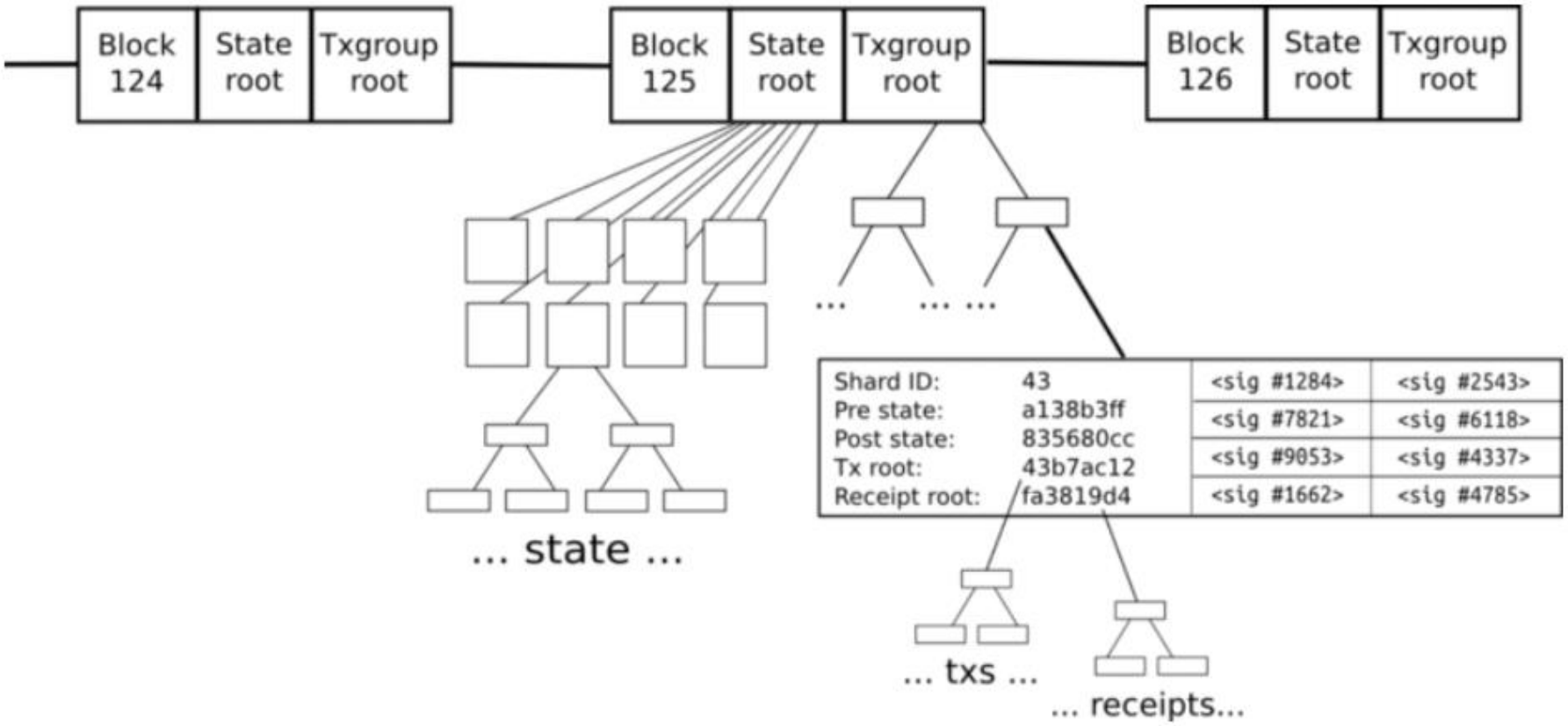
Transaction group header

Tx a142	Tx a558	Tx eca6
Tx a35f	Tx e25a	Tx 34ac
Tx 2308	Tx 6987	Tx f260
Tx 9f14	Tx ec30	Tx 5fc3

Transaction group body

Quelle: Rosic, A. “What are Ethereum Nodes And Sharding.” Blockgeeks, Oct. 2017 [13]

Sharding



Quelle: Rosic, A. “What are Ethereum Nodes And Sharding.” Blockgeeks, Oct. 2017 [13]

Zusammenfassung

- 1st (Bitcoin) → 2nd (Ethereum) → 3rd Generation Blockchains (Cardano, Neo, Ethereum 2.0)
- Smart Contracts (meta data & Tokens) führen zu Anstieg der…
  - Transaktionen/s
  - Network bandwidth
  - Data
- Noch nicht “Big Data” (aber bald?)
  - Mehr als genug Daten für ein Analytics Projekt
  - Zahlreiche Repos auf Github
- Die Idee von dezentralisiertem “Trust” ist nicht wirklich neu (siehe Appendix). DLT beschreitet allerdings neue Wege bestehende Probleme zu lösen.
  - Tradeoff zwischen *Consistency* und *Availability* im Falle von Network Failures (CAP Theorem)
  - Das Web 1.0 und 2.0 waren u.A. erfolgreich aufgrund des 5-Layer Modells, in dem nicht jeder alle Informationen brauchte (Widerspruch zu DLT)
- DLT steht vor großen Hürden, hat aber das Potential einen Paradigmenwechsel im Bereich Informationsinfrastruktur sowie Finanzinfrastruktur einzuleiten

Literatur

- [1] Chaum, D. “Blind signatures for untraceable payments.” Advances in Cryptology, pp. 199-203, 1983.
- [2] Nakamoto, S. “Bitcoin: A Peer-to-Peer Electronic Cash System.” 2008 [Online]. Available: https://bitcoin.org/bitcoin.pdf
- [3] Buterin, V. “A next-generation smart contract and decentralized application platform.” Ethereum white paper. 2014 [Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper
- [4] Rosic, A. “What Is Hashing? Under The Hood Of Blockchain.” Blockgeeks, Sep. 2017. Retrieved Jan 2018 from: https://blockgeeks.com/guides/what-is-hashing/
- [5] “Block.” In Bitcoin Wiki, n.d. Retrieved Jan 2018 from: https://de.bitcoin.it/wiki/Block
- [6] “Block hashing algorithm.” In Bitcoin Wiki, n.d. Retrieved Jan 2018 from: https://en.bitcoin.it/wiki/Block_hashing_algorithm
- [7] “Controlled supply.” In Bitcoin Wiki, n.d. Retrieved Jan 2018 from: https://en.bitcoin.it/wiki/File:Controlled_supply-supply_over_block_height.png
- [8] Bartoletti, M., Bracciali, A., Lande, S., & Pompianu, L. “A general framework for blockchain analytics.” 2017. arXiv preprint arXiv:1707.01021. [Online]. Available: https://arxiv.org/abs/1707.01021v2. Github Repo [Online]. Available: https://github.com/bitbart/blockchain-analytics-api
- [9] Rosic, A. “Smart Contracts: The Blockchain Technology That Will Replace Lawyers.” Blockgeeks, Feb. 2017. Retrieved Jan 2018 from: https://blockgeeks.com/guides/smart-contracts/
- [10] Bartoletti, M., Pompianu, L. “An empirical analysis of smart contracts: platforms, applications, and design patterns.” In International Conference on Financial Cryptography and Data Security, pp. 494-509, April, 2017. Springer, Cham. [Online]. Available: https://arxiv.org/abs/1703.06322v1
- [11] Madan, I., Saluja, S., and Zhao, A. “Automated Bitcoin Trading via Machine Learning Algorithms.” Technical report, Stanford University, 2015 [Online]. Available: http://cs229.stanford.edu/proj2014/Isaac%20Madan,%20Shaurya%20Saluja,%20Aojia%20Zhao,Automated%20Bitcoin%20Trading%20via%20Machine%20Learning%20Algorithms.pdf
- [12] Hirshman, J., Huang, Y., and Macke, S. "Unsupervised approaches to detecting anomalous behavior in the bitcoin transaction network.” Technical report, Stanford University, 2013 [Online]. Available: http://cs229.stanford.edu/proj2013/HirshmanHuangMacke-UnsupervisedApproachesToDetectingAnomalousBehaviorInTheBitcoinTransactionNetwork.pdf
- [13] Rosic, A. “What are Ethereum Nodes And Sharding.” Blockgeeks, Oct. 2017. Retrieved Jan 2018 from: https://blockgeeks.com/guides/what-are-ethereum-nodes-and-sharding/

Appendix

Ausblick

“The whole human memory can be, and probably in a short time will be, made accessible to every individual. [T]his new all-human cerebrum…can have at once the concentration of a craniate animal and the diffused vitality of an amoeba…”

H. G. Wells, World Brain, 1938

Ausblick

- H. G. Wells, World Brain, 1938
- Ted Nelson, Project Xanadu, 1960
- Google “BackRub”, 1997
- Leslie Lamport, “Time, Clocks, and the Ordering of Events in a Distributed System”, 1978