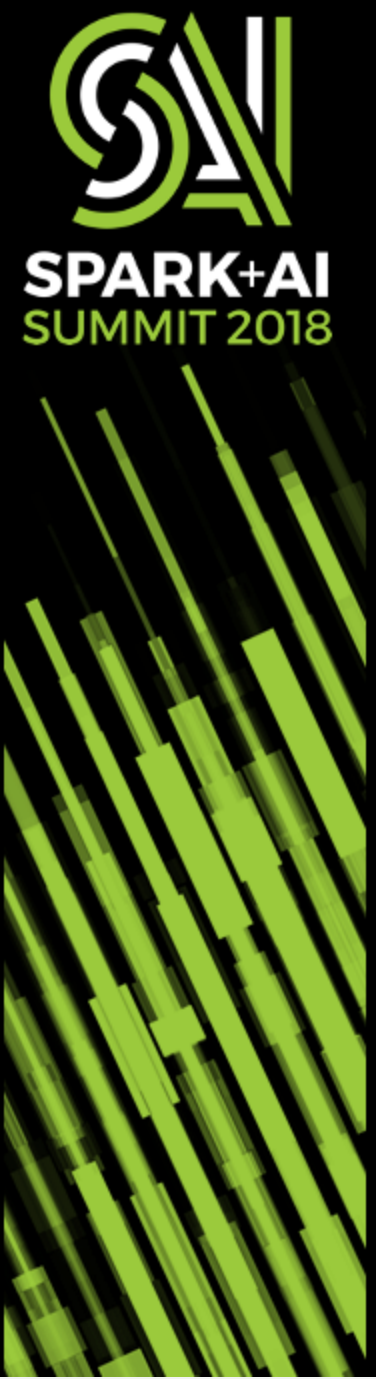


Analyzing Blockchain Transactions

Jiri Kremser, Red Hat, Inc.

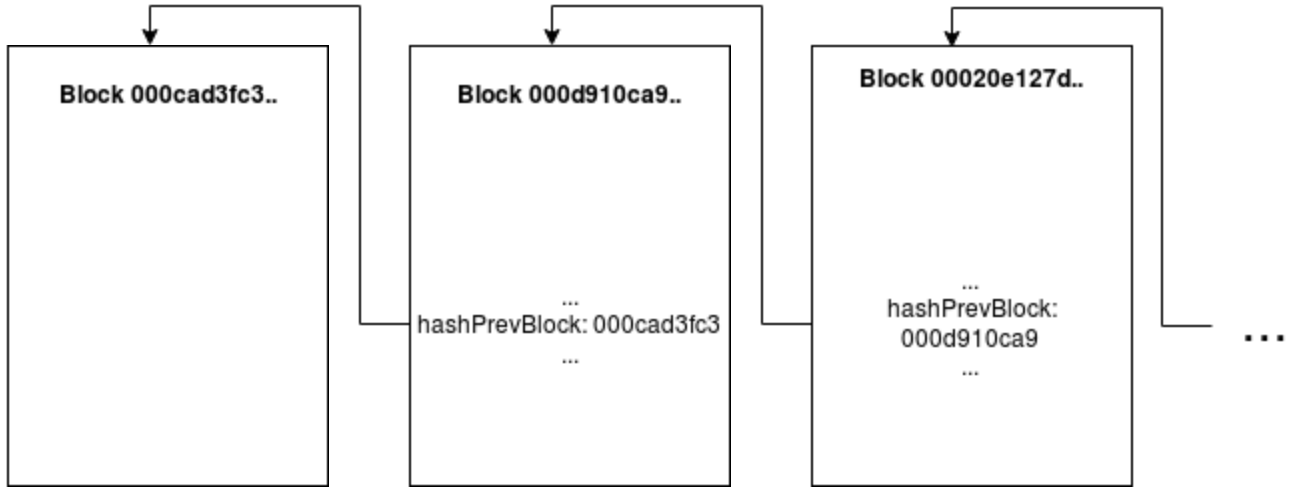
#Exp6SAIS

Outline

- Blockchain 101
- Blockchain to graph
- Graph data in Spark
- DEMO

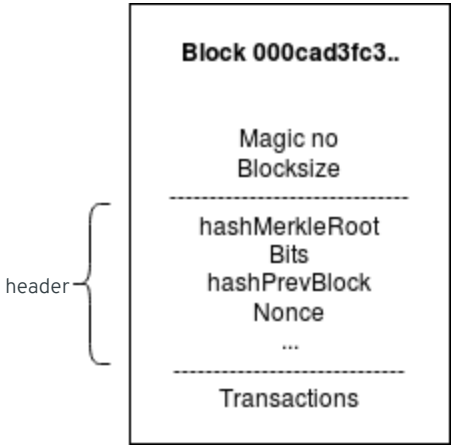
What is Blockchain

- Distributed ledger
- Linked list of blocks
- Trust stems from Merkle trees and proof of work (aka mining)
- Cryptography



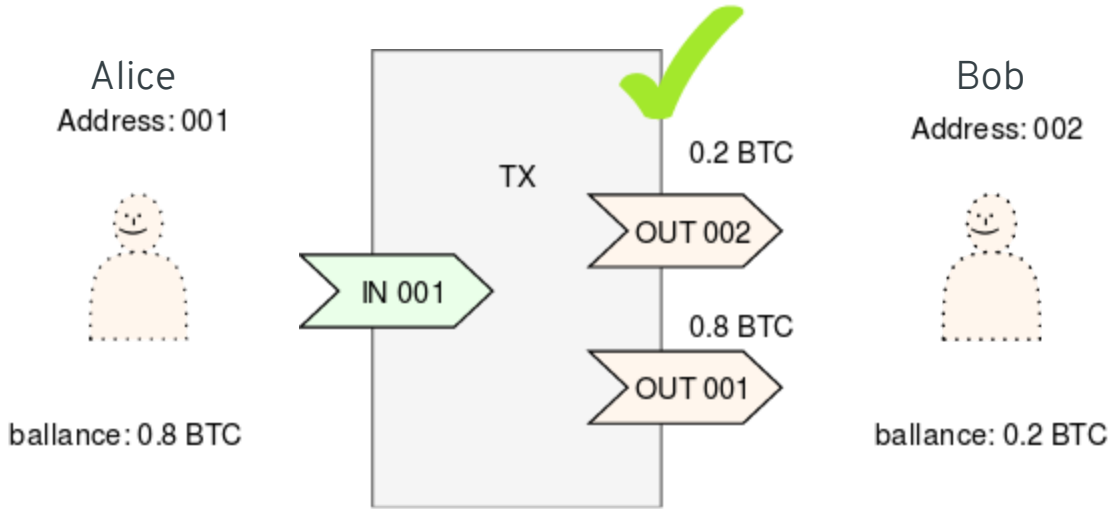
What is Block

- Set of transactions approved at once
- Metadata
- Hard limit 1 MB (*)

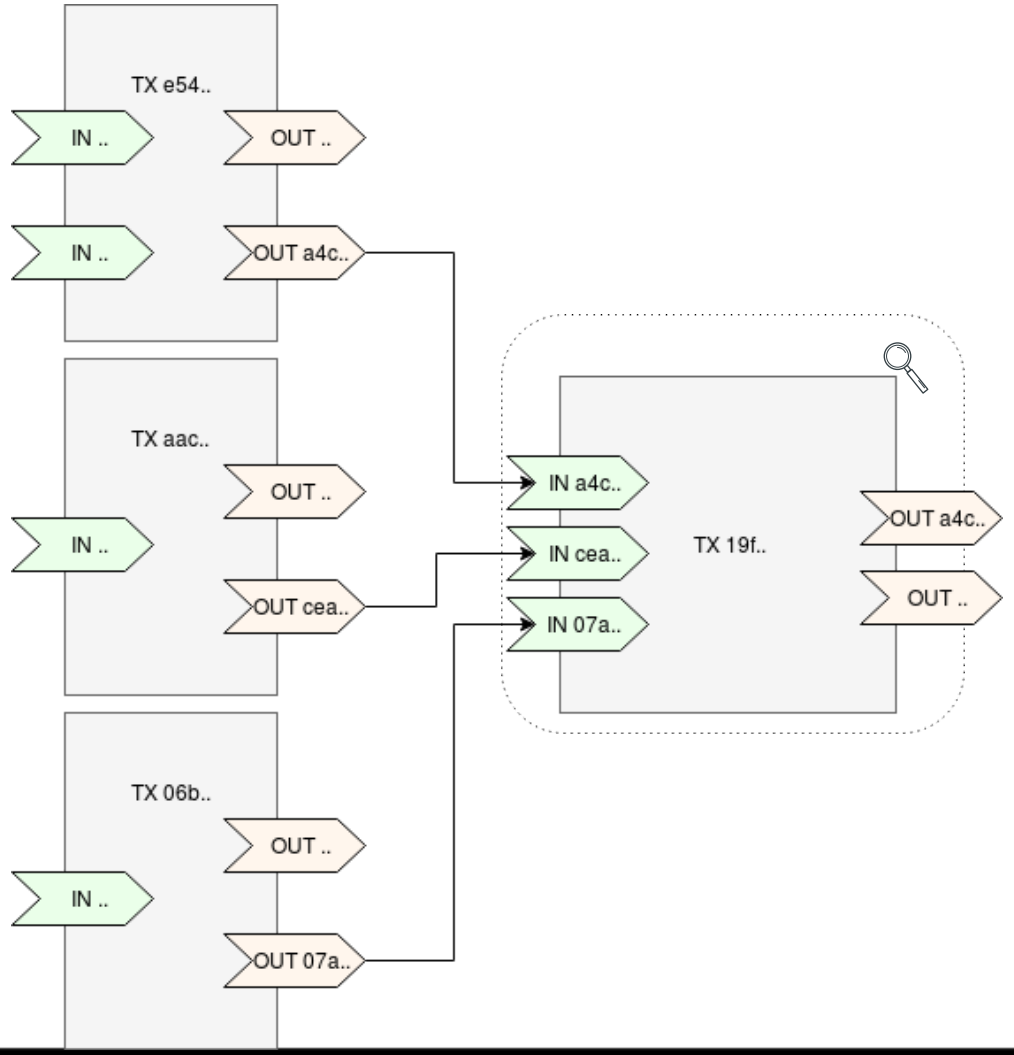


What is Transaction

- (INs, OUTs)
- sum of INs ≥ sum of OUTs
- confirming ~ including it to a new block and finding the "nonce"

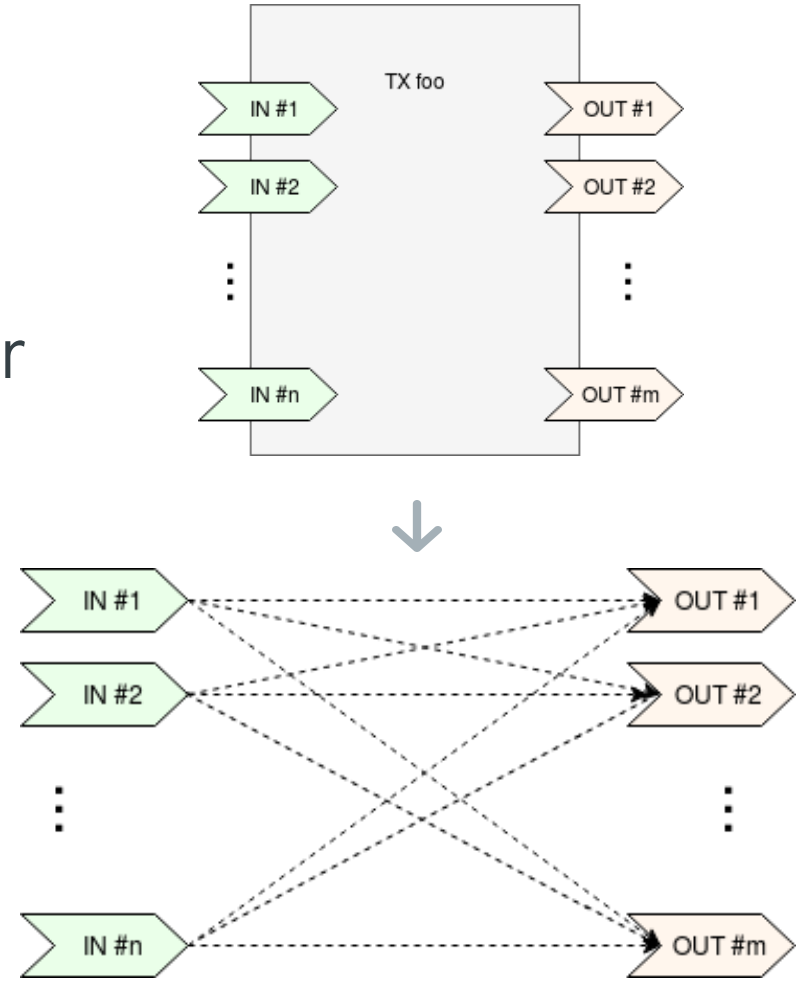


What is Transaction (more general case)



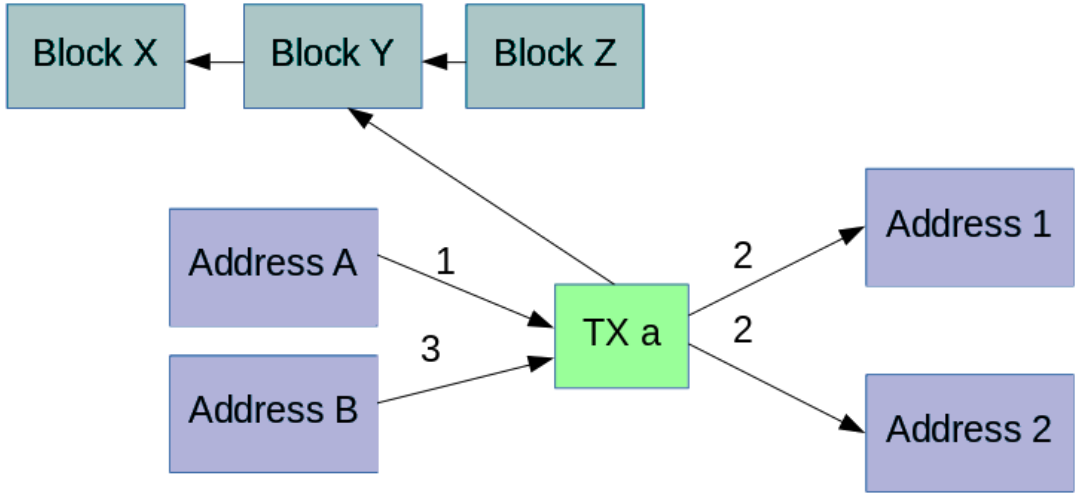
Transaction to Graph

- M:N transactions produces a lot of edges
- Apache Parquet
- blockchain binary data -> parquet converter



Transaction to Graph

- # of Satoshis sent on the edges of type [T→A] and [A→T]
- timestamp on the block nodes
- more suitable for querying the graph



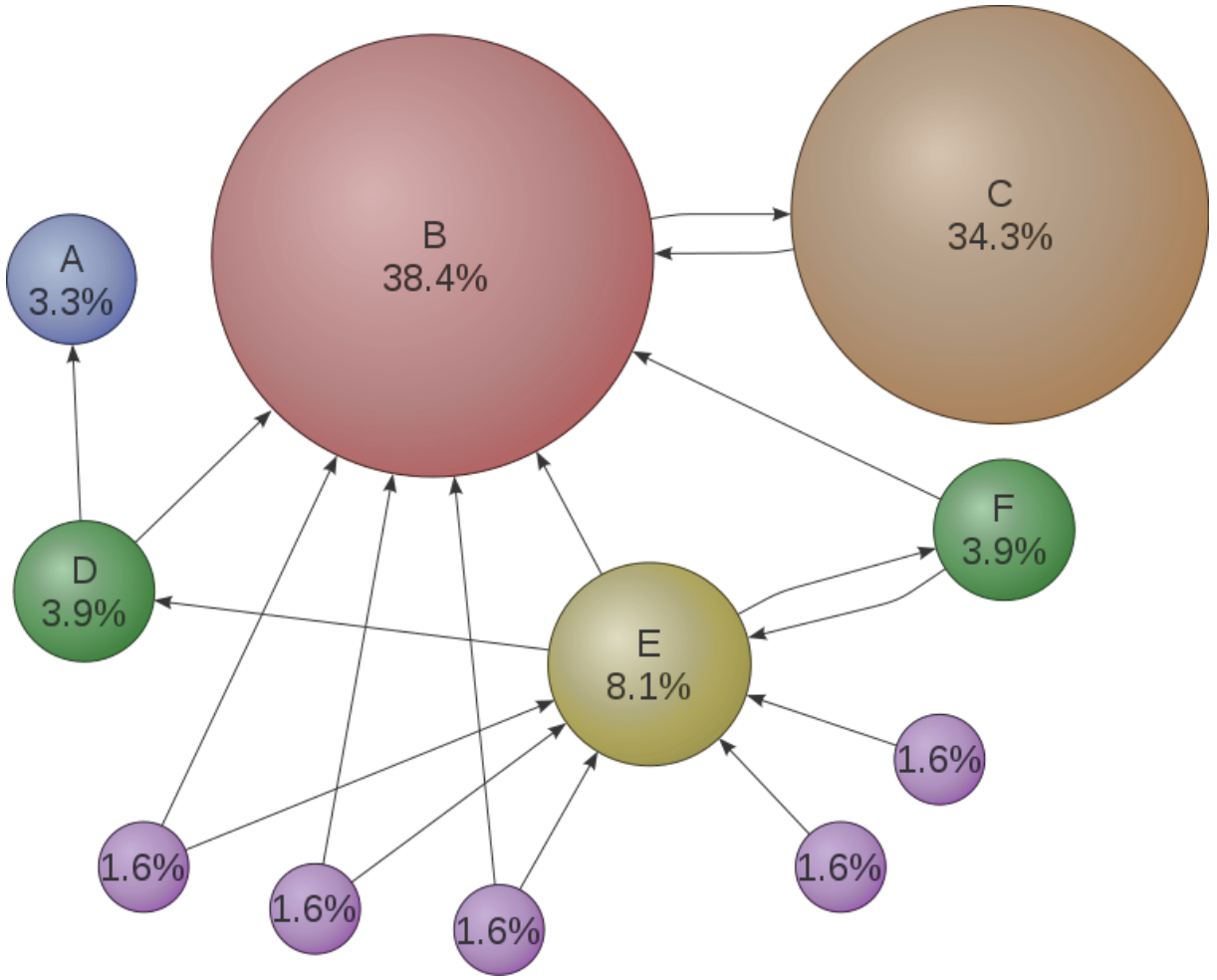
Graphs and Spark

- GraphX
- GraphFrames
- built-ins (label propagation, pagerank, triangles, bfs, etc.)
- motif ~ cypher
- Pregel

Talk is Cheap

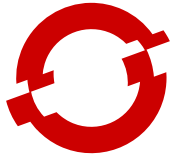
Demo time

Page Rank



Takeways

- Blockchain is out there
- GraphFrames vs GraphX
- Reproducible experiments with notebooks and containers



How to get started

More projects, tutorials and examples can be found at

radanalytics.io

Thank You!

This presentation

http://bit.ly/sais18

Jiri-Kremser/bitcoin-insights

@JirkaKremser