Bitcoin Gain/Loss Analysis

Authors: Haifa Alquwaiee, Amble (Xu) Xing, Jesse Campbell

Motivation

- Bitcoin is a new type of payment system

- It was recently classified by the IRS as an asset, like stock
    - Income tax and capital gains tax are now applied retroactively on U.S. accounts

- How much profit was made? What amount is taxable? What is the total tax?

Design

- Bitcoin operates as a P2P network
  - All transactions are saved on each peer

- Data from the Bitcoin client was copied from the Bitcoin client's binary data files

- The data was mined into MySQL tables using "Bitcoin Abe", a python script.

Data Implementation

- Input Data
  - Bitcoin client (32.5 GB, 1 day)
    - https://bitcoin.org/bin/blockchain/bootstrap.dat.torrent was imported to speed up sync speed
  - MySQL (90 GB, 1 week)
    - on a 500MBps SSD
  - CSV (13 GB, 2 days)
    - 150 million transactions
    - 34 million accounts
- Output Data
  - CSV
    - bitcoin addr, profit, short term tax, long term tax

Server Implementation

- Cloud Computing Cluster
    - Amazon Web Services
        - EC2
            - Ubuntu Server
                - 32-bit
                - m1.small
            - Hadoop 2.2.0
        - S3
            - File storage

Algorithm Implementation

- Accounting for Taxable capital gain

    FIFO method ➡️ Most widely used

    Sell by Specific ID method

    Average Cost (Mutual funds only)

- Taxable profits calculation

    35% for selling those kept less than 1 year

    15% for selling those kept more than 1 year

Evaluation

Data nodes	Data size	Time needed
1	150 million transactions	?
2	150 million transactions	?
4	150 million transactions	?
8	150 million transactions	?
16	150 million transactions	?

Evaluation with Failures

Data nodes	Failed data nodes	Data size	Time needed
2	1	150 million transactions	?
4	2	150 million transactions	?
8	3	150 million transactions	?
16	4	150 million transactions	?

Parallel Processing

- Each bitcoin address can be calculated individually
  - ~34 million addresses
  - ~150 million transactions
  - About 4 transactions per address

Speed-up Graph

Conclusion

- Performance
  - Faster with Hadoop?
  - Bitcoin transactions are difficult to link to individuals
- Estimated profit & tax
  - Total profit: ?
  - Total tax: ?
- Question/Comments?

- Thanks

References

- Bitcoin mining process
  - https://docs.google.com/document/d/1bEKBQM8pfDkthvMpEpMXzJGB3vcHewxlCSfDx_F2sLs/edit?usp=sharing
- Cluster configuration
  -
- Hadoop MapReduce programming
- Bitcoin Abe
  - https://github.com/bitcoin-abe/bitcoin-abe
- Blockchain.info
  - https://blockchain.info/
- BitcoinCharts.com
  - http://bitcoincharts.com/
  - bitcoin/USD pricing data
- Blockr
  - http://blockr.io/
- GainsKeeper
  - tax rules
  - https://www.gainskeeper.com/siebert/siebert_fund.htm