Analyzing Cryptocurrencies on HBase

For Finance, Forensics, and Fraud.

April 2017

Agenda

1. What is Ripple?
2. Difference from Bitcoin
3. HBase Infrastructure
4. Demo
5. Q + A

What is Ripple?

ripple

OUR VISION

The Internet of Value:
Enable the world to move money
like information moves today.

ripple

OUR MISSION

Make cross-border payments truly efficient for banks and their customers.

Bitcoin Blockchain
vs
Ripple Consensus Ledger

Bitcoin Blockchain vs Ripple Consensus Ledger

A Tale of Two Different Technologies

Bitcoin Blockchain (BTC)

- 8 years old
- Native Digital Asset - BTC
- $20+ Billion Market Cap
- Inflationary (increasing supply)
- Validation by Proof-of-Work (Mining)
- Confirmation by Chain of Blocks (10 mins)
- No Scripting
- No Decentralized Exchange
- Hash Tree Data Structure

Ripple Consensus Ledger (XRP)

- 4 years old
- Native Digital Asset - XRP
- $5+ Billion Market Cap
- Deflationary (fixed supply)
- Validation by Consensus (No-Mining)
- Confirmation by Last Ledger Closed (5 secs)
- Automated Scripting
- Decentralized Exchange
- Hash Tree Data Structure

Reporting Needs

ripple

Enable a real time transaction visualization showing trends and patterns of financial information.

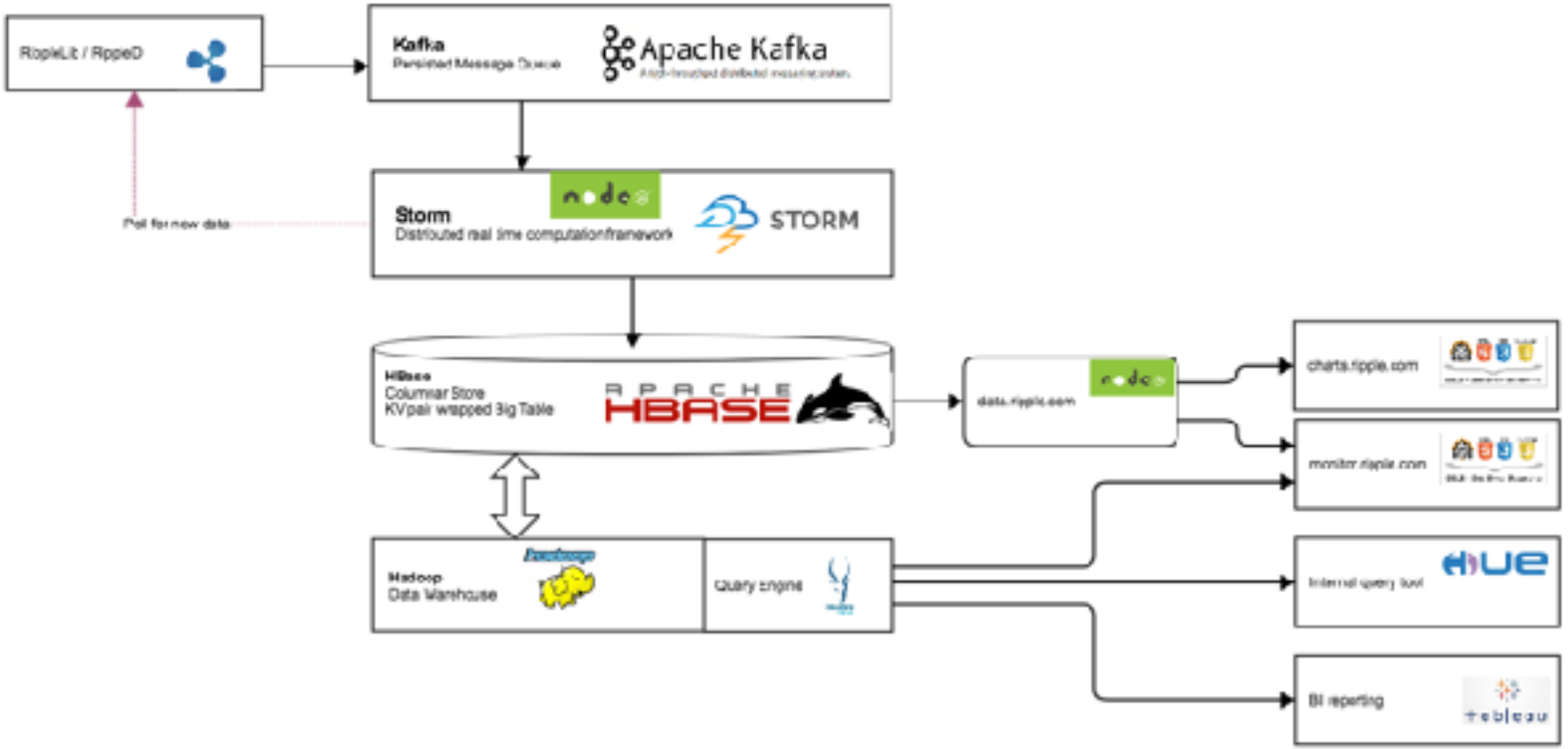
Relieve the application network the burden of expensive queries of static data

Give traders and market makers aggregated information as it happens

Of course, make it scalable

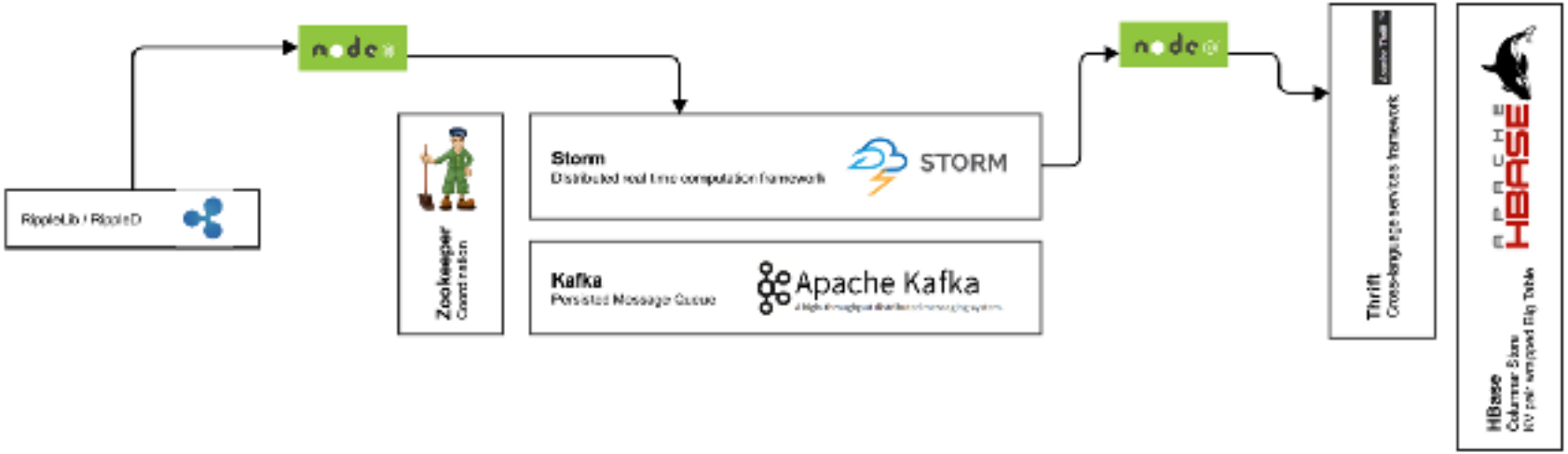
Our Stack

How we designed a high speed reporting system



Our Data Pipeline

How we get data in hbase



ripple

Some key stats

Our Cluster

- 5 node hbase cluster
  - Active replication to 3 node backup cluster
- 2x 2.4GHz Intel Xeon-Haswell (E5-2620-V3-HexCore)
- 8x 16GB (128 GB RAM)
- 6 4 TB HDD

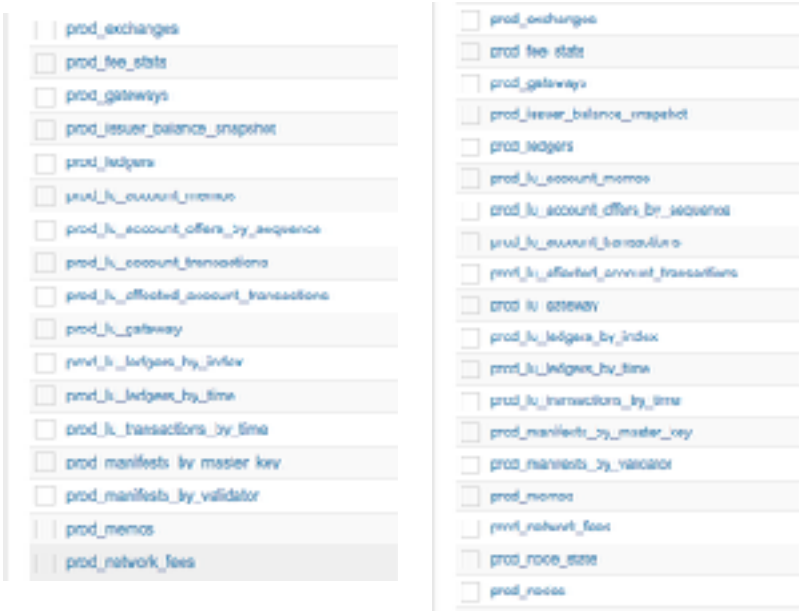
Metrics

- Peak so far 1.2 million operations per second
- Approximately 60 purpose built tables
- Co-Processors (none yet)
- Data ingestion averaging 100ms from application commit to population of all tables
- Data retrieval under 30ms

Our Tables

How we designed a high speed reporting system

- Purpose built tables based on known query patterns and paths
- Data is pre-aggregated to certain levels to allow for fast retrieval
- Multiple “Look Up” tables to help translate keys



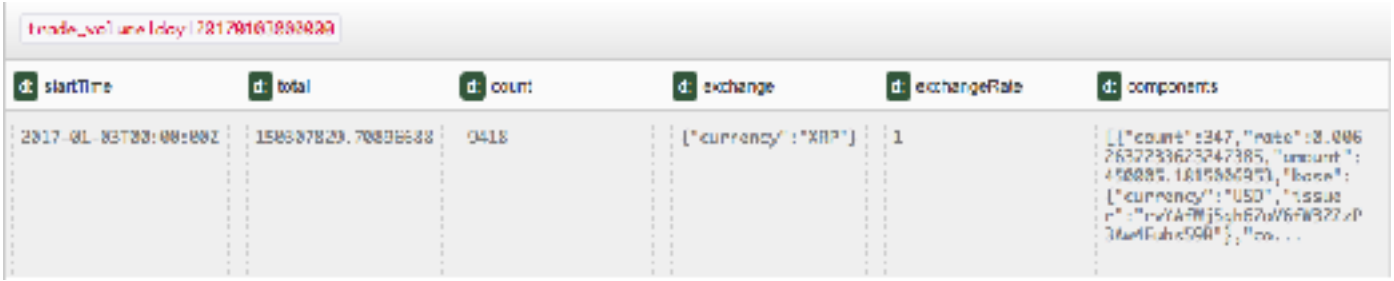
How we designed a high speed reporting system

Query Based Design

- Queries go after a subject and have a strict condition set
- Keys are designed around the condition set
- Mostly date driven before grain

Data

- 1 ledger can hold over 1000 transactions
- Each transaction serves a purpose to the ledger
- Data Consistency
- Bonus - I get a data warehouse with Impala



Demo

charts.ripple.com

Q + A

URL
https://charts.ripple.com
https://data.ripple.com

Our stuff is open sourced
Github Repo
Storm ETL ingestion - https://github.com/ripple/rippled-historical-database
Ripple Charts - https://github.com/ripple/ripplecharts-frontend