



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?





OUR VISION

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





OUR MISSION

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





Bitcoin Blockchain
vs
Ripple Consensus Ledger



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



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

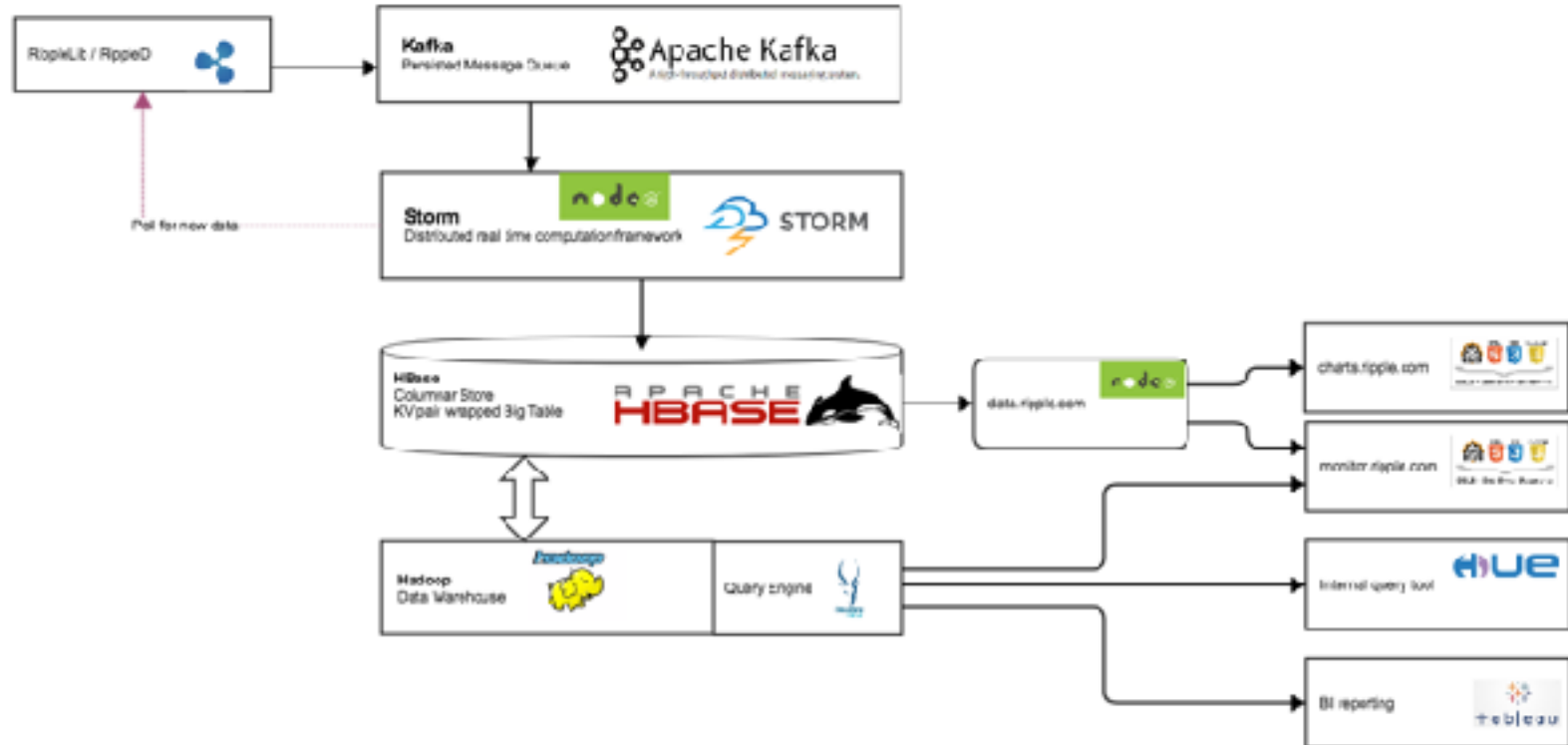
Of course, make it scalable



Our Stack



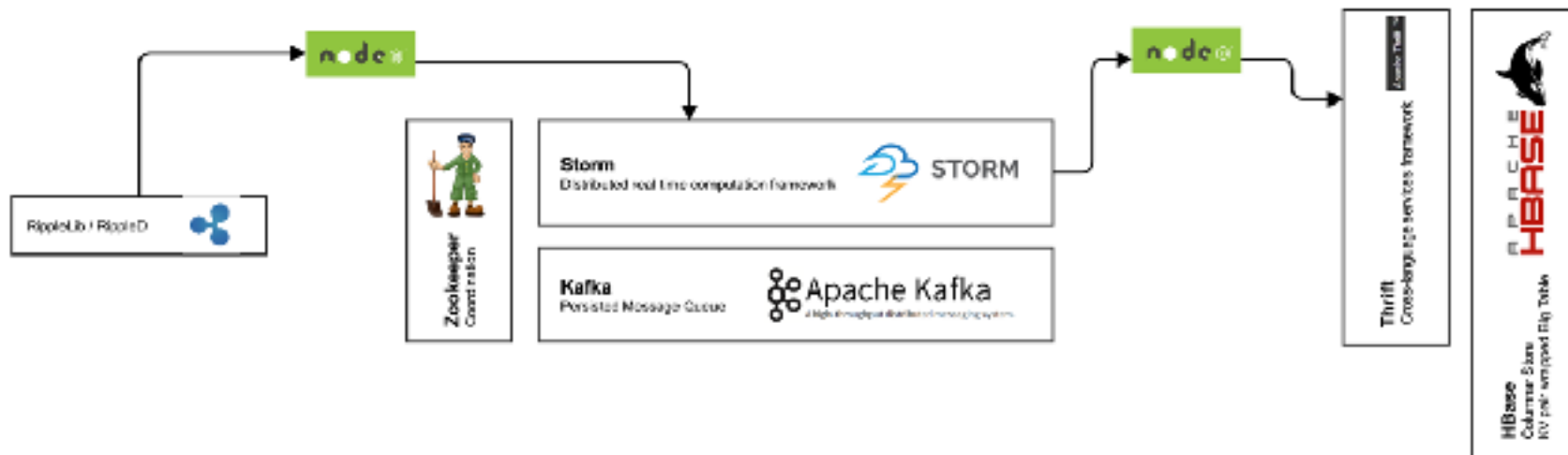
How we designed a high speed reporting system



Our Data Pipeline



How we get data in hbase



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

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

<input type="checkbox"/> prod_exchanges	<input type="checkbox"/> prod_exchanges
<input type="checkbox"/> prod_fee_stats	<input type="checkbox"/> prod_fee_stats
<input type="checkbox"/> prod_gateways	<input type="checkbox"/> prod_gateways
<input type="checkbox"/> prod_issuer_balance_snapshot	<input type="checkbox"/> prod_issuer_balance_snapshot
<input type="checkbox"/> prod_ledgers	<input type="checkbox"/> prod_ledgers
<input type="checkbox"/> prod_li_account_memos	<input type="checkbox"/> prod_li_account_memos
<input type="checkbox"/> prod_li_account_offers_by_sequence	<input type="checkbox"/> prod_li_account_offers_by_sequence
<input type="checkbox"/> prod_li_account_transactions	<input type="checkbox"/> prod_li_account_transactions
<input type="checkbox"/> prod_li_effective_account_transactions	<input type="checkbox"/> prod_li_effective_account_transactions
<input type="checkbox"/> prod_li_gateway	<input type="checkbox"/> prod_li_gateway
<input type="checkbox"/> prod_li_ledgers_by_index	<input type="checkbox"/> prod_li_ledgers_by_index
<input type="checkbox"/> prod_li_ledgers_by_time	<input type="checkbox"/> prod_li_ledgers_by_time
<input type="checkbox"/> prod_li_transactions_by_time	<input type="checkbox"/> prod_li_transactions_by_time
<input type="checkbox"/> prod_manifests_by_master_key	<input type="checkbox"/> prod_manifests_by_master_key
<input type="checkbox"/> prod_manifests_by_validator	<input type="checkbox"/> prod_manifests_by_validator
<input type="checkbox"/> prod_memos	<input type="checkbox"/> prod_memos
<input type="checkbox"/> prod_network_fees	<input type="checkbox"/> prod_network_fees
	<input type="checkbox"/> prod_offers
	<input type="checkbox"/> prod_race_state
	<input type="checkbox"/> prod_races

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

trade_volume day 20170107000000					
d: startTime	d: total	d: count	d: exchange	d: exchangeRate	d: components
2017-01-07T00:00:00Z	150337829.70836533	9418	[{"currency": "XRP"}]	1	[{"count": 347, "rate": 0.0062637233623247385, "amount": 450000.101558295}, {"currency": "USD", "issuer": "rN7YA7Wj5sh6Zu6fW377zP3ha4Ehuc59A"}, {"co...



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>

