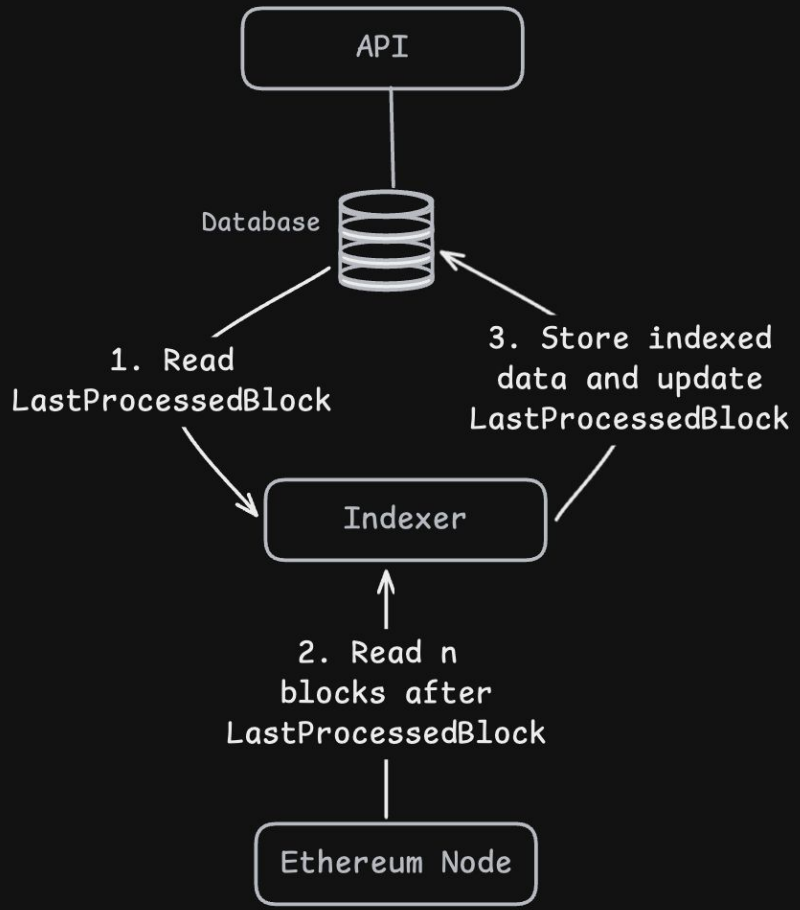
Indexing Entire 2.4 Billion Transactions on Ethereum in 10 Hours

Panjamapong **“PanJ”** Sermsawatsri

CTO, CLEVERSE



Conventional Ethereum Indexer

A typical Ethereum indexer would execute the indexing program **sequentially** while mutating the indexed data in the database.

AlphaTrace Beta

Type Wallet Address

0 P

Donate NEW

Wallet#0x11

New tab

0x74c1...4675

0x74c1...4675

Net Worth: $262,667 | ETH Balance: 0.19 ETH | Volume: 18.99M | Txs: 6,693 | Realized PnL: +$2M | Unrealized PnL: +$167K

Trade Performance | Trading History

Updated: 2 mins ago

Ethereum

Win Rate

7D | 15D | 30D | All

30D
48.8%
20/41

Top profit tokens

Token	Realized PnL	ROI
ETHEREUM	+$10,103	+117%
FTP	+$5,473	+148%
QF	+$1,325	+5%
oGPU	+$605	+3%
KAGE	+$592	+20%

Top Holding Tokens

Token	Avg Cost/Market	Total USD/PnL
SBR STRATEGIC BITCO…	$0.012068 $0.83725	$215,606 +$212,499 (+6,838%)
BOOE Book of Ethereum	$0.65455 $0.56394	$20,584 -$3,308 (-14%)
PIPI Pipi	$0.0₆79765 $0.0₅10358	$20,454 +$4,704 (+30%)
FBB4 Faith	$0.0042512 $0.0021236	$1,519 -$1,522 (-50%)
BLEPE Blepe the Blue	$0.0₄54318 $0.00010126	$1,475 +$684 (+86%)

Active Trades

1D | 7D | 30D | All

Last Tx	Token	Avg Buy Price	Initial Value	Current Value	% Change	Action
18h 56m 44s	SBR	$0.011948	$3,076.69	$215,606.32	+6,907.74%	Uniswap Scan ···
1d 4m 20s	BOOE	$0.27962	$10,206.34	$20,584.42	+101.68%	Uniswap Scan ···
10 Nov 15:42	PIPI	$0.0₆93029	$18,369.97	$20,454.37	+11.35%	Uniswap Scan ···

The Requirements

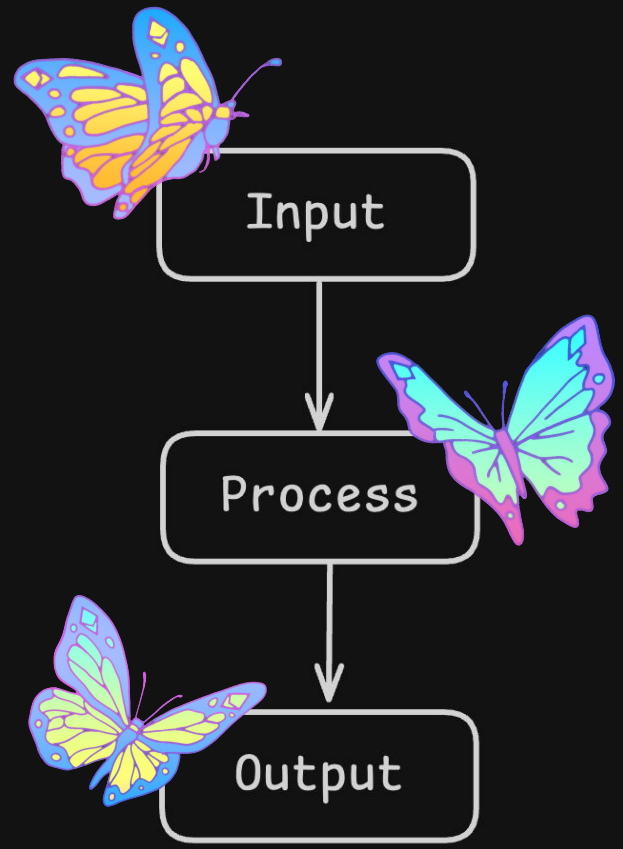
- To be able to view any wallet address on Ethereum
- Calculate and display the wallet's PnL (Profit and Loss)
- Data needed to be indexed to meet the product requirements
  - ERC-20 token transfers with price
  - Ether spent for gas
  - Ether transfers in the transaction
  - Index the data since genesis (otherwise PnL will be inaccurate)
  - Index every single transaction

A conventional indexer would not work, we need to rethink how to implement an indexer at large scale

Back to the basics

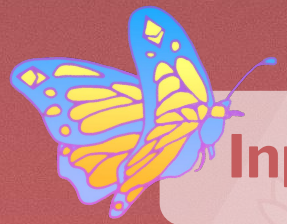
Let’s look back and see the blockchain indexing as
input, process, and output

In each component, we have to identify the factors that are required to build a large scale blockchain indexer.



We need a compact and structured data format

We need a compact and structured data format



Input

ETH Node

paradigmxyz/**cryo**

cryo is the easiest way to extract blockchain data to parquet, csv, json, or python dataframes

17 Contributors

9 Used by

1k Stars

106 Forks

Parquet

GCS

Data processing tools that support parallel execution

Data processing tools that support parallel execution

Data processing tools that support parallel execution

Google Dataflow

beam

Output

Database engine that can scale reads & writes horizontally

Output

Database engine that can scale reads & writes horizontally

Output

Database engine that can scale reads & writes horizontally

Now piece everything together

Google BigTable

beam

Google Dataflow

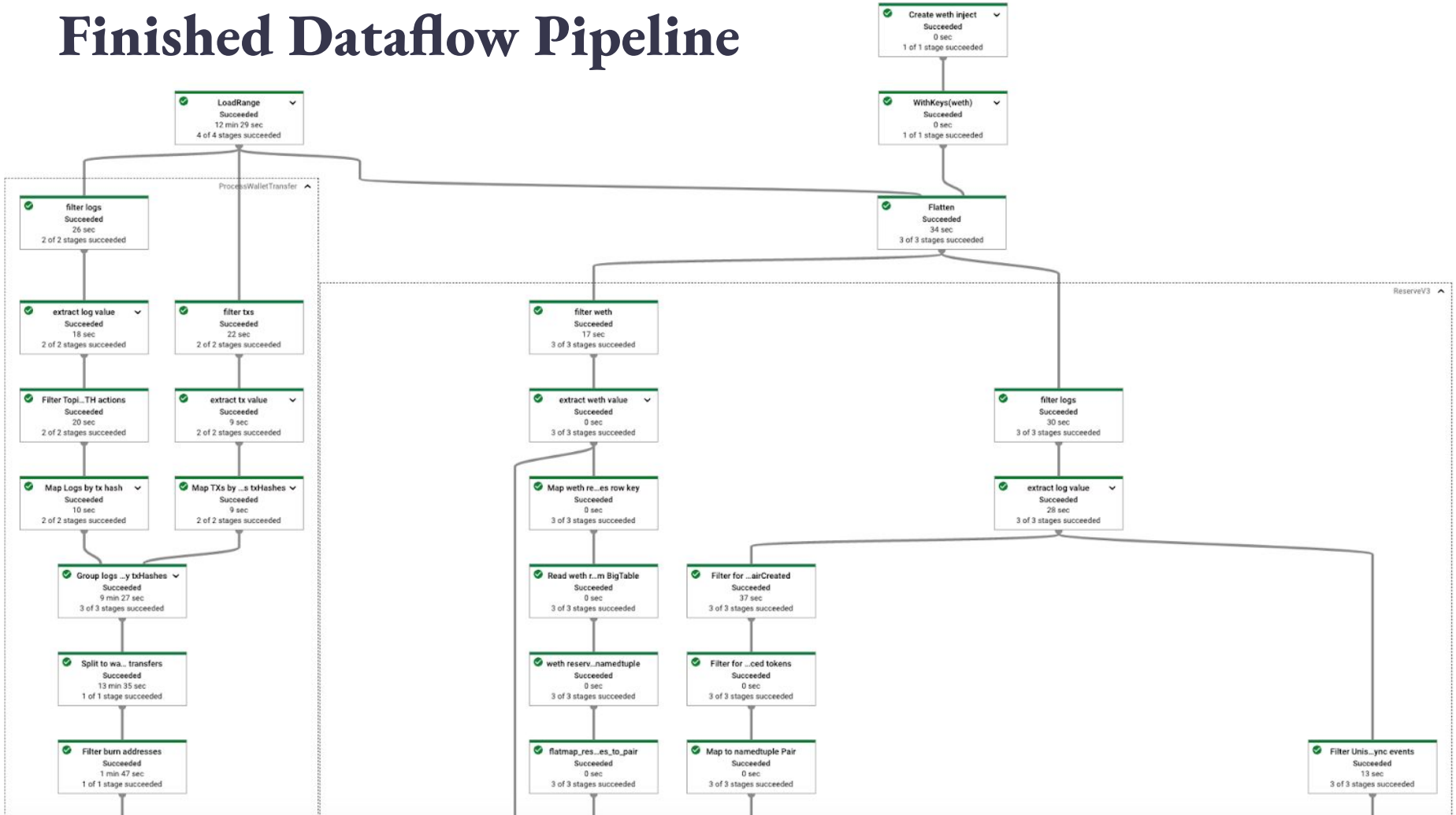
paradigmxyz/**cryo**

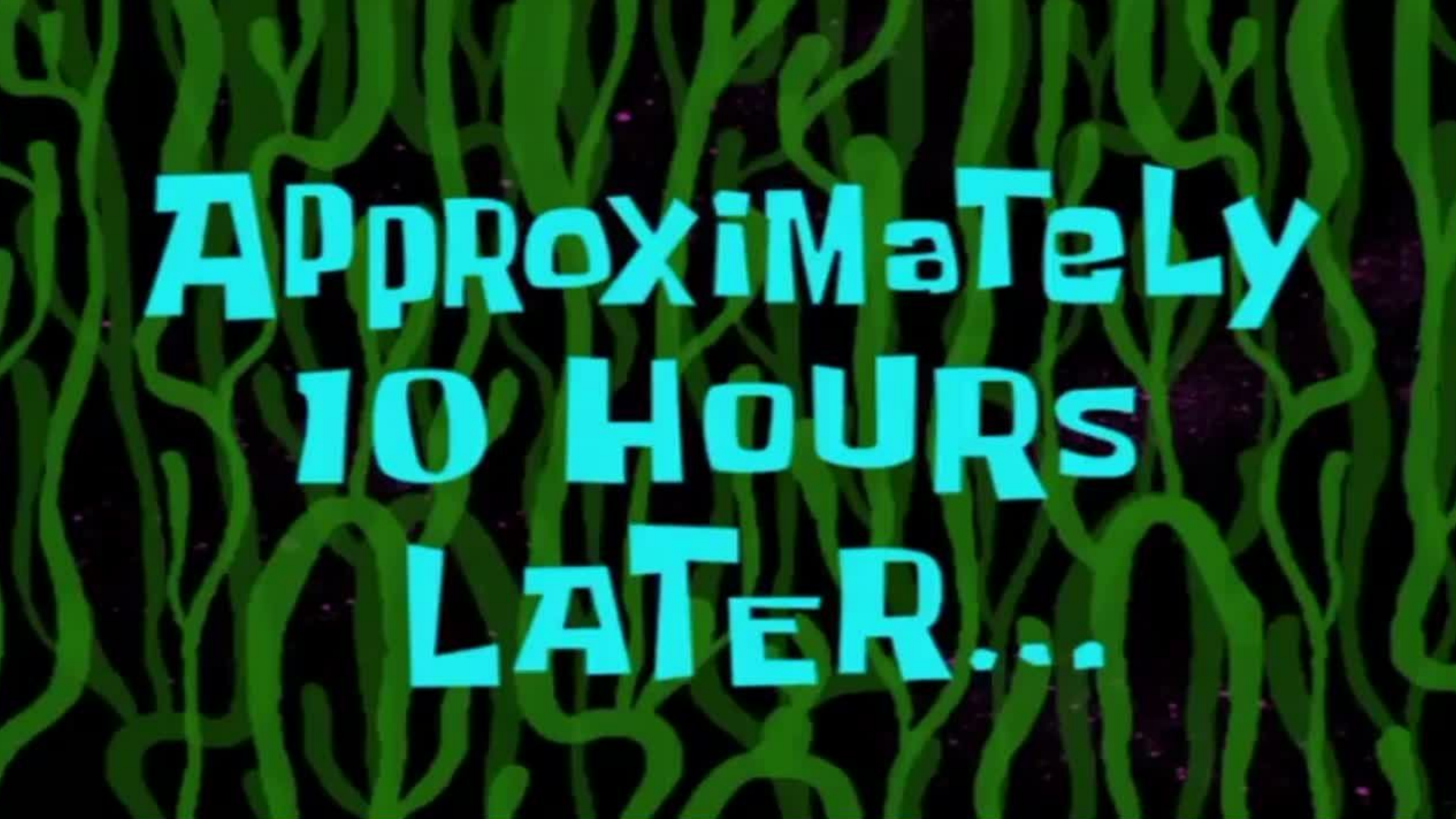
cryo is the easiest way to extract blockchain data to parquet, csv, json, or python dataframes

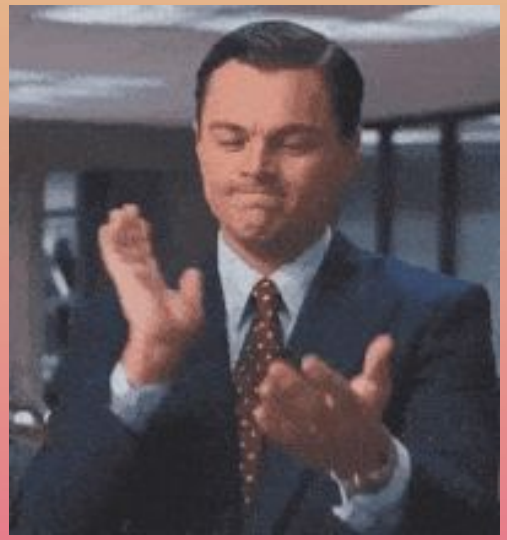
17 Contributors · 9 Used by · 1k Stars · 106 Forks

Parquet

Finished Dataflow Pipeline




APPROXIMATELY
10 HOURS
LATER...



Indexing complete!!!

But what's next?

- That was one big batch processing, but we need to keep our data updated
- One option is to do incremental batch processing, a batch that only process the new transactions
- To make indexed data real-time, you will need to implement streaming data processing that process data one block at a time when a new block arrives
- We implemented streaming data processing in a conventional way using Golang to make our data real-time

176

vCPUs

352

GB Memory

20

Bigtable Instances

7.1B

Records Written

$202

Dataflow Cost

$157

Bigtable Cost


Thank you!
Panjamapong “PanJ” Sermsawatsri
CTO, CLEVERSE
@PanJ