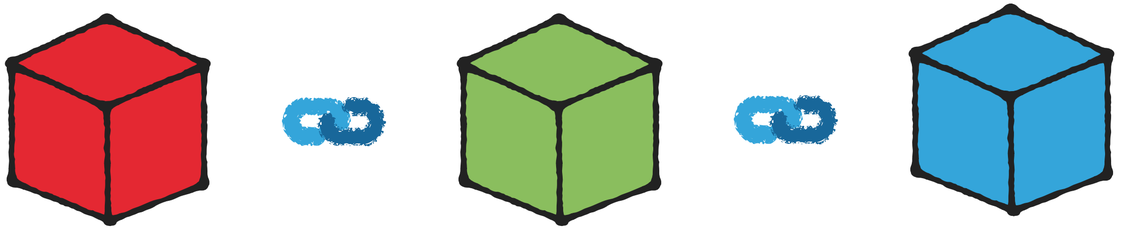


A Public Dataset for the ZKsync Rollup

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

Imperial College London

NOVA IMS Information Management School

Matter Labs

Johnnatan Messias, PhD

@johnnatan_me

CAAW–FC in Miyakojima, Japan

April 18th, 2025

Joint work with Maria Inês Silva, and Benjamin Livshits

johnnatan-messias.github.io

Blockchain Data Challenges 🤭

Key Challenges ⚙️

Data is not easily accessible

- Not everyone is able to deploy an Archive Node
- APIs rate limit might be too low.
- Many chains available!

High costs of running archive nodes

- Computer resources requirements might be too high for low-budget groups.
- Relying on external data providers leads to costs that might be hard to predict.

Limited research accessibility for non-technical users

- Public but not necessary accessible!
- Hard to gather and parse.
- Run analysis on a laptop!

Solution 💡

Dune Analytics nansen INFURA

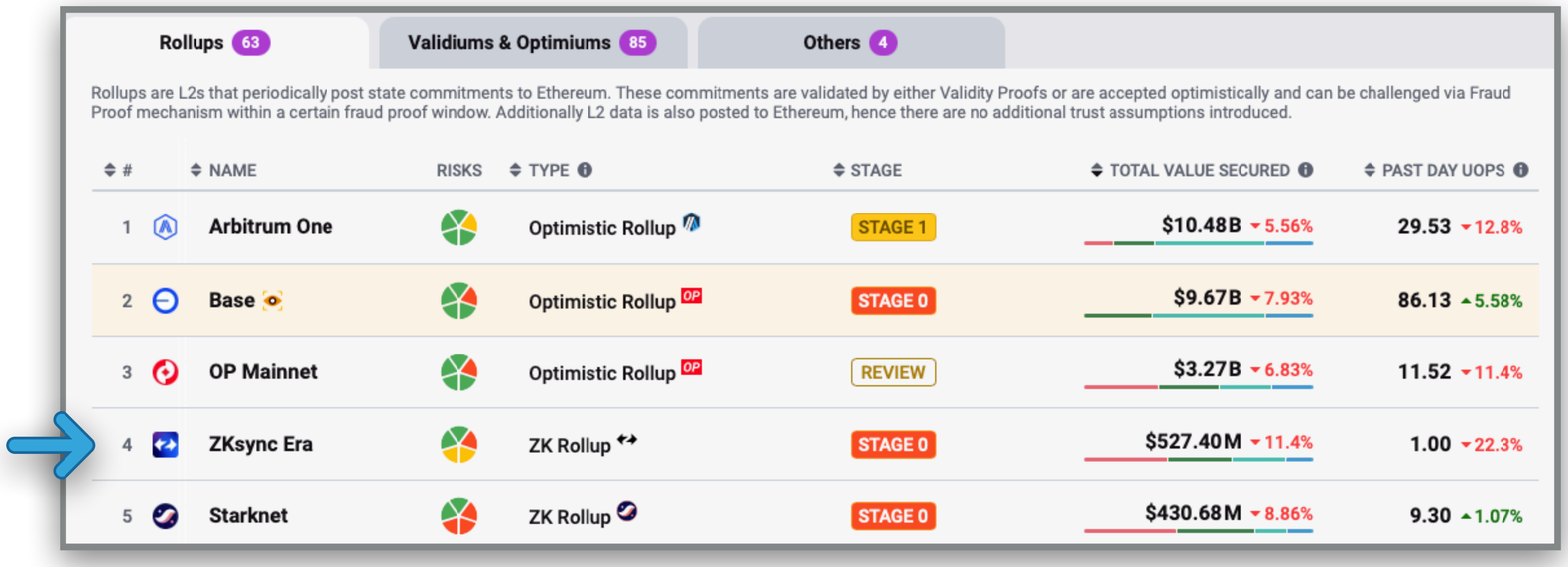
- **Curating and releasing a public dataset of ZKsync Era.**

Why ZKsync Era?

Key reasons 🤔

- Ethereum Layer-2 scaling solution using Zero-Knowledge Proofs.
- EVM compatibility and scalable architecture.
- Significant role in scaling Ethereum with low fees and high transaction throughput.
- Research potential in a growing ecosystem.
- ZKsync Era is among the top 5 rollups based on Total Value Secured (TVS).

Source L2BEAT



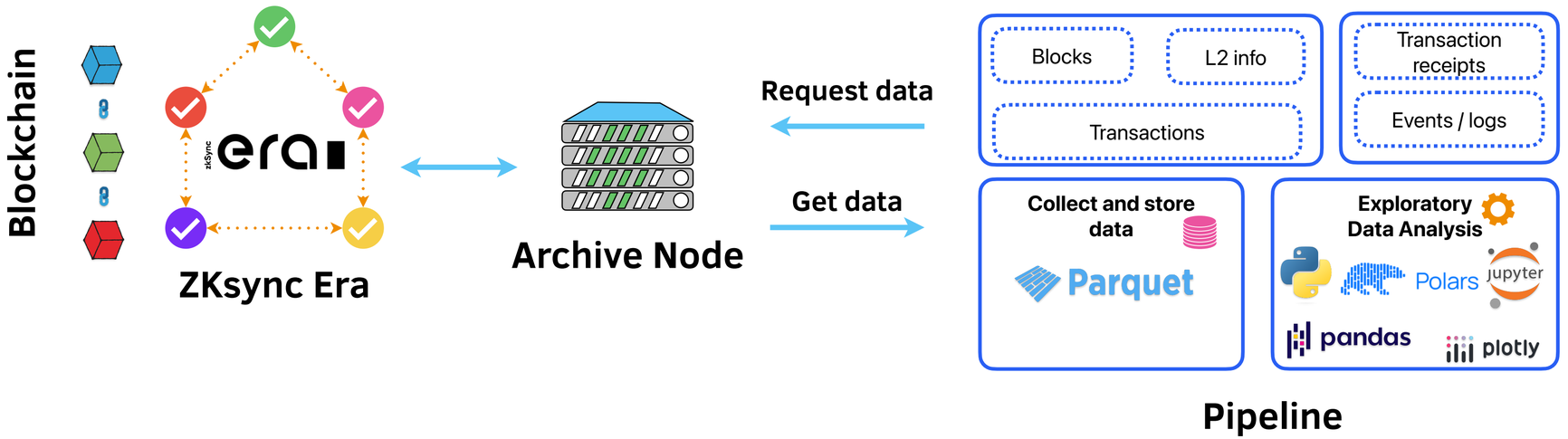
Rollups 63 | Validiums & Optimiums 85 | Others 4

Rollups are L2s that periodically post state commitments to Ethereum. These commitments are validated by either Validity Proofs or are accepted optimistically and can be challenged via Fraud Proof mechanism within a certain fraud proof window. Additionally L2 data is also posted to Ethereum, hence there are no additional trust assumptions introduced.

#	NAME	RISKS	TYPE	STAGE	TOTAL VALUE SECURED	PAST DAY UOPS
1	Arbitrum One		Optimistic Rollup	STAGE 1	$10.48B ▼5.56%	29.53 ▼12.8%
2	Base		Optimistic Rollup	STAGE 0	$9.67B ▼7.93%	86.13 ▲5.58%
3	OP Mainnet		Optimistic Rollup	REVIEW	$3.27B ▼6.83%	11.52 ▼11.4%
4	ZKsync Era		ZK Rollup	STAGE 0	$527.40M ▼11.4%	1.00 ▼22.3%
5	Starknet		ZK Rollup	STAGE 0	$430.68M ▼8.86%	9.30 ▲1.07%

Data Pipeline and Overview

1 data source



- **Data Format:** Parquet for efficient storage and loading.
- **Processing Tool:** Polars (for better memory management).

Polars Parquet

Data Pipeline and Overview

1 data source

zkSync era

Notebook file	Description
01-zksync-data.ipynb	It computes the basic statistics of the dataset. We use four main sources of data: blocks, transactions, transaction receipts, and logs.
02-data-exploration-fees.ipynb	It analyses gas usage and transaction fees for ZKsync. We use two main sources of data: blocks and transaction receipts.
03-data-exploration-contracts.ipynb	It analyzes the contract deployment and events triggered on ZKsync. We use one main source of data: transaction logs but also load blocks data to extract timestamps information.
04-data-exploration-swaps.ipynb	It analyzes the swap events on ZKsync. We use one main source of data: transaction logs but also load blocks data to extract timestamps.

These notebook files are available in our GitHub repository and show how to interact and process our ZKsync dataset.

Time period

- February 14th, 2023 - March 24th, 2024 (1 year of data).

Data Volume

- 29.7 thousand blocks, 327 million transactions, 1.6 million contracts, 2 billion events or logs, and 7.3 million accounts.
- 150.97 GB

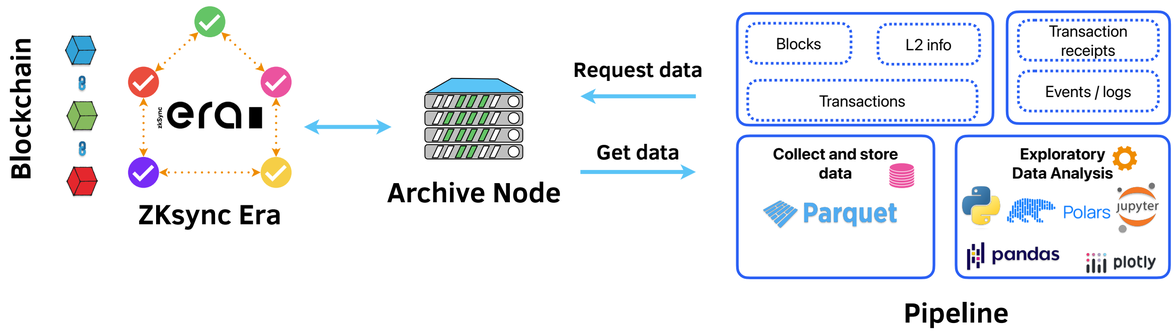
Dataset

Data Access

- Publicly available via GitHub.

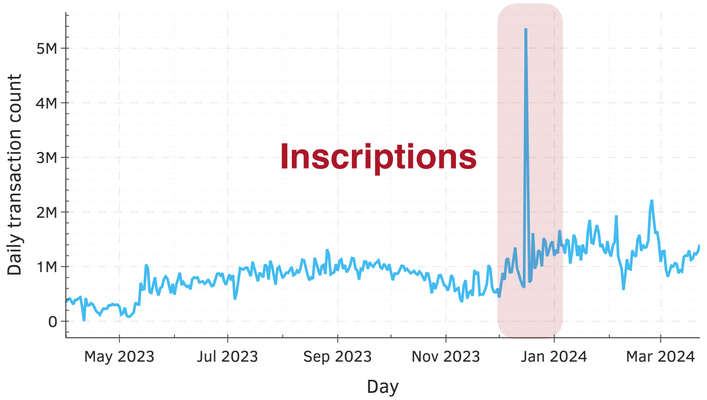
Key Components:

- Block and transaction attributes, detailed receipts and logs.

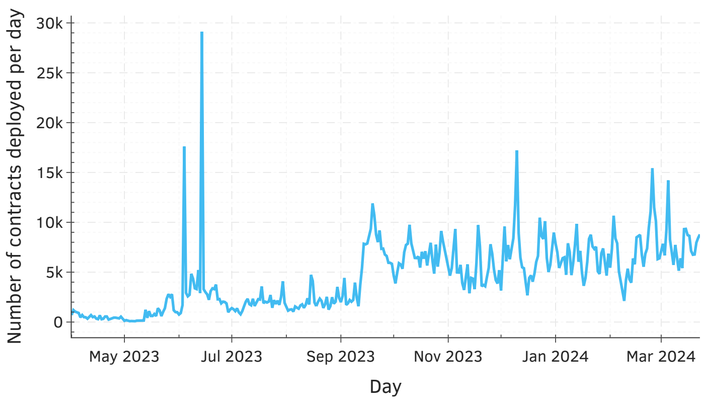


ZKsync Data Characterization

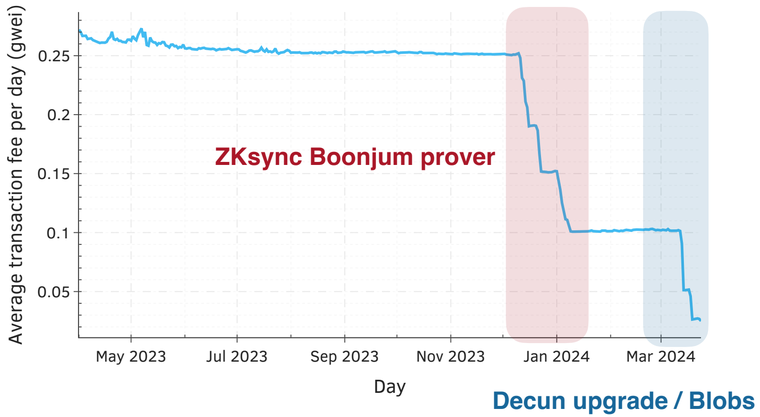
Daily number of transactions



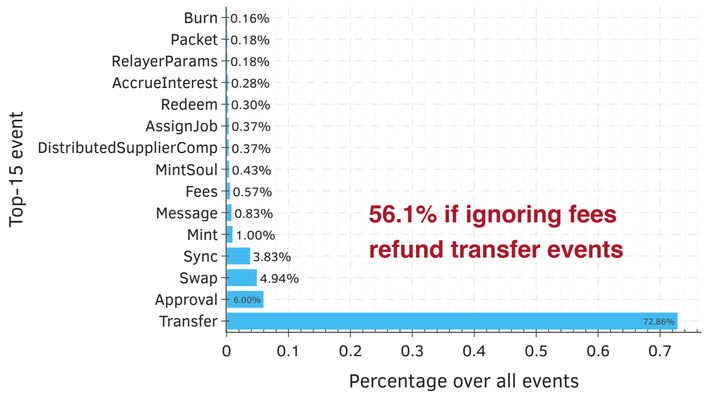
Daily number of contracts deployed



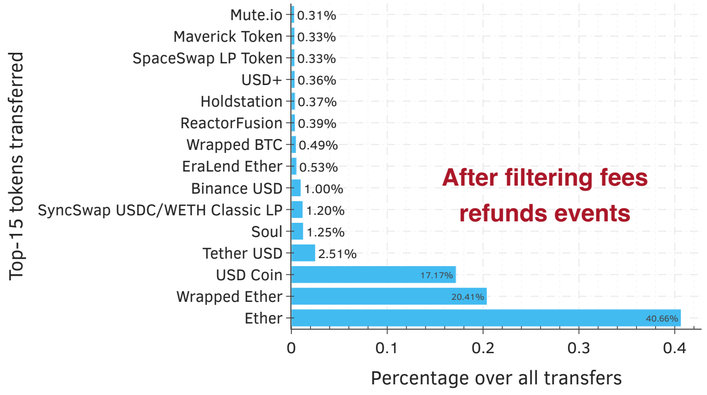
Daily average transaction fees



Top events emitted

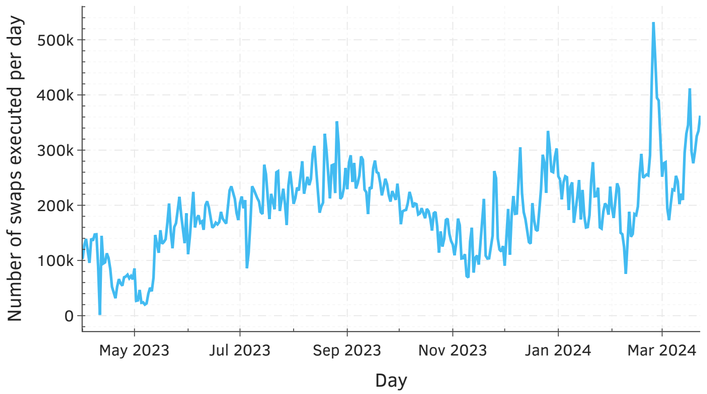


Most frequently transfers

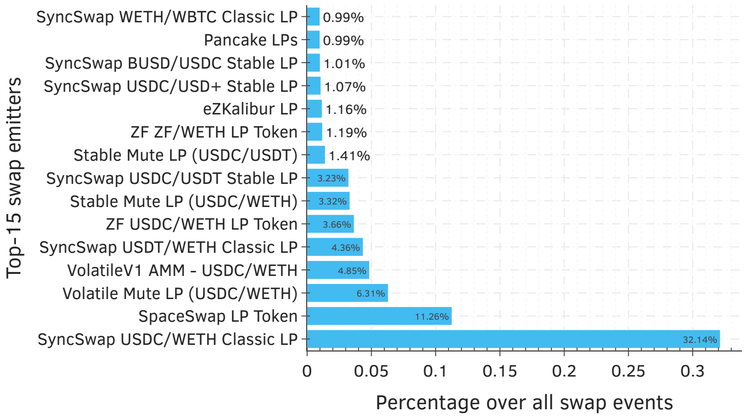


Focus on Swap Events

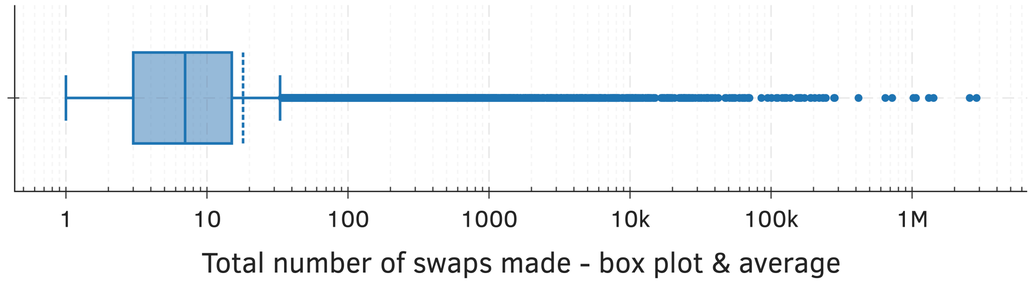
Daily swap event distribution



Key pools and contracts driving swap activity



Distribution of swaps per unique wallet addresses



Concluding Discussion

Practical Applications

- **For Researchers:** Explore **Layer-2 blockchain mechanics**.
- **For Data Scientists:** Learn **blockchain-related data science skills** using our dataset.
- **Integration:** Easy integration with **Python libraries** like **Pandas** and **Polars**.

Future Research Directions

- **MEV and Arbitrage:** Research opportunities for **ZKsync MEV**.
- **User Behavior Analysis: Sybil detection**, **airdrop farming**, and others.
- **Cross-Rollup Research:** Investigate **cross-rollup MEV** and arbitrage.
- **Governance Research:** Study **decentralized governance** and **voting** on ZKsync.

Impact of ZKsync Dataset on Blockchain Research

- A **public, accessible ZKsync dataset** to support **blockchain research**.
- **Fosters innovation** in **Layer-2 ecosystems**.
- **Available on GitHub** for **reproducibility** and further research.

Contact

johnme@mpi-sws.org
johnnatan-messias.github.io

Johnnatan Messias, PhD
Research Scientist

@johnnatan_me

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS