

COPENHAGEN, DENMARK & ONLINE

POLKADOT DECODED 2023

JUNE 28-29

Polkadot data for
beginners

Karim Jedda

Parity Data

Who am I

Karim Jedda: Data nerd

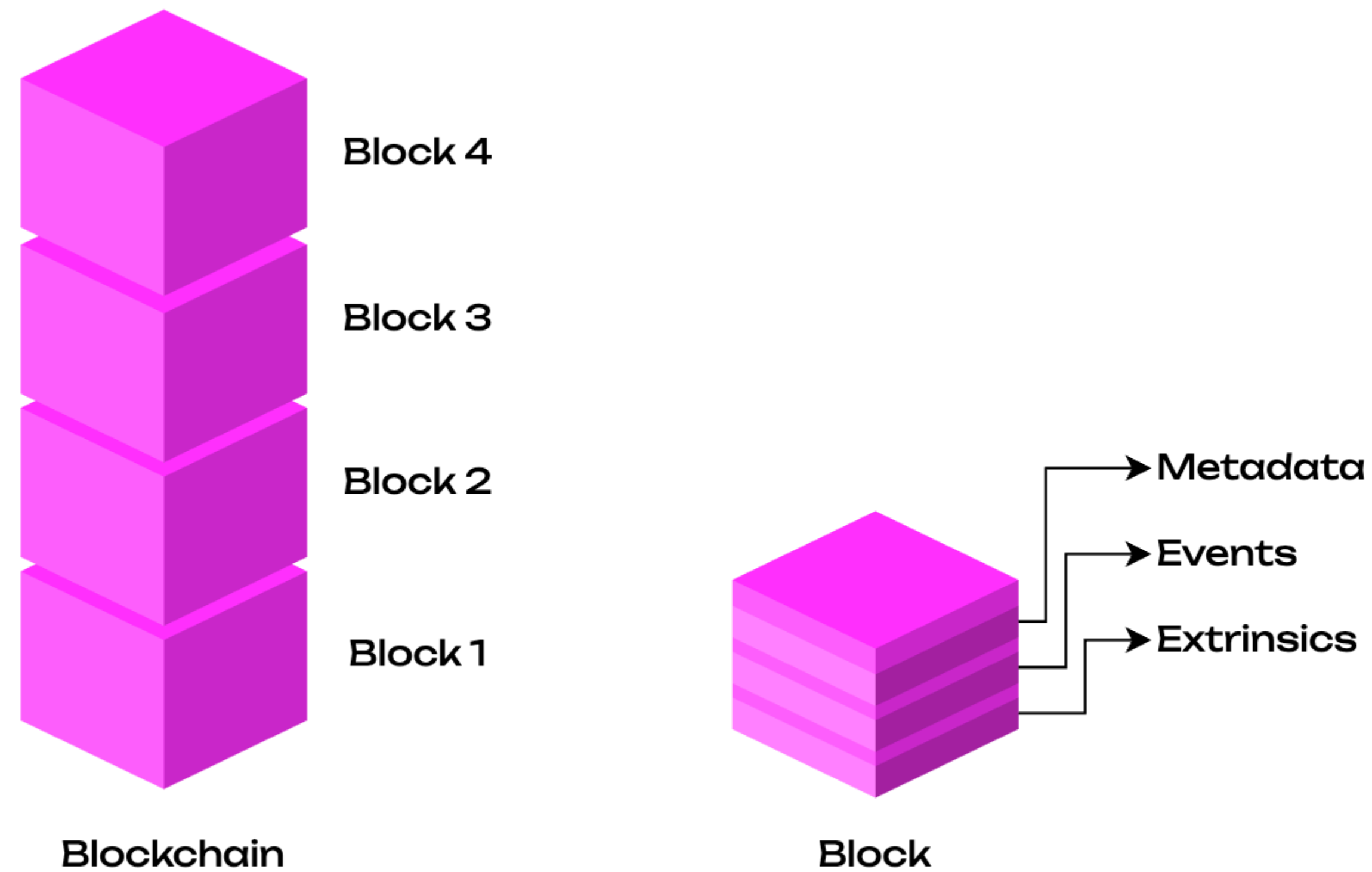
- Lead data engineer @ Parity
- <https://karimjedda.com/decoded>
 - Links
 - Socials



Me, 2022 - spot the differences

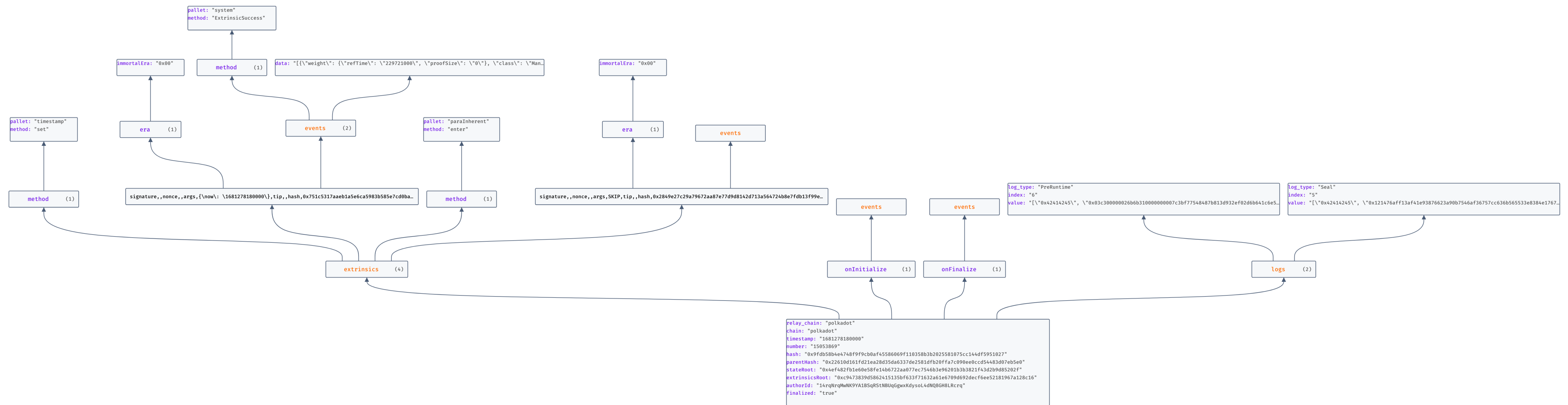
What is Polkadot data?

Basics



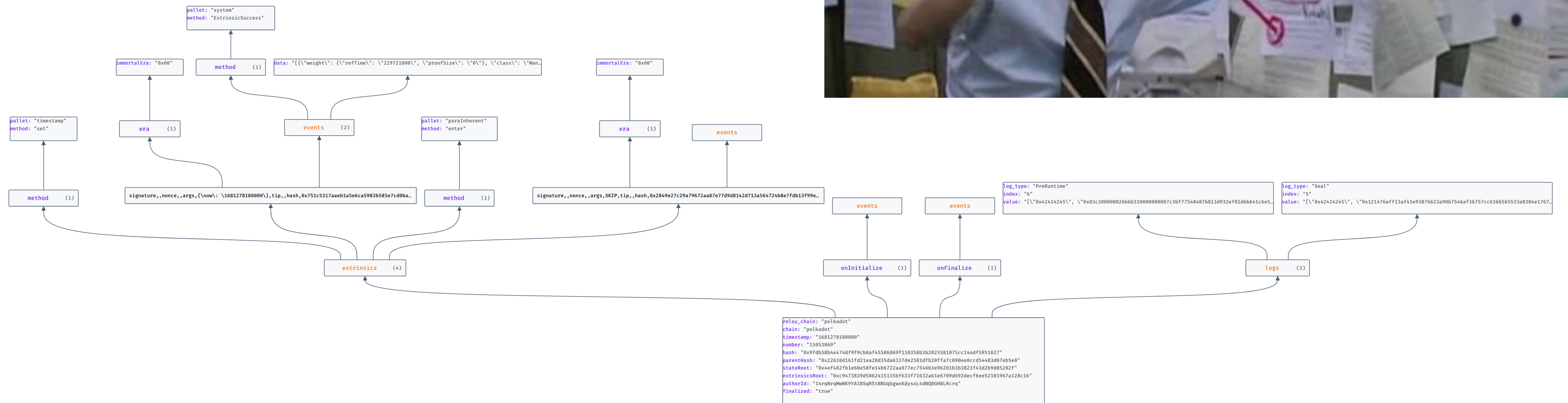
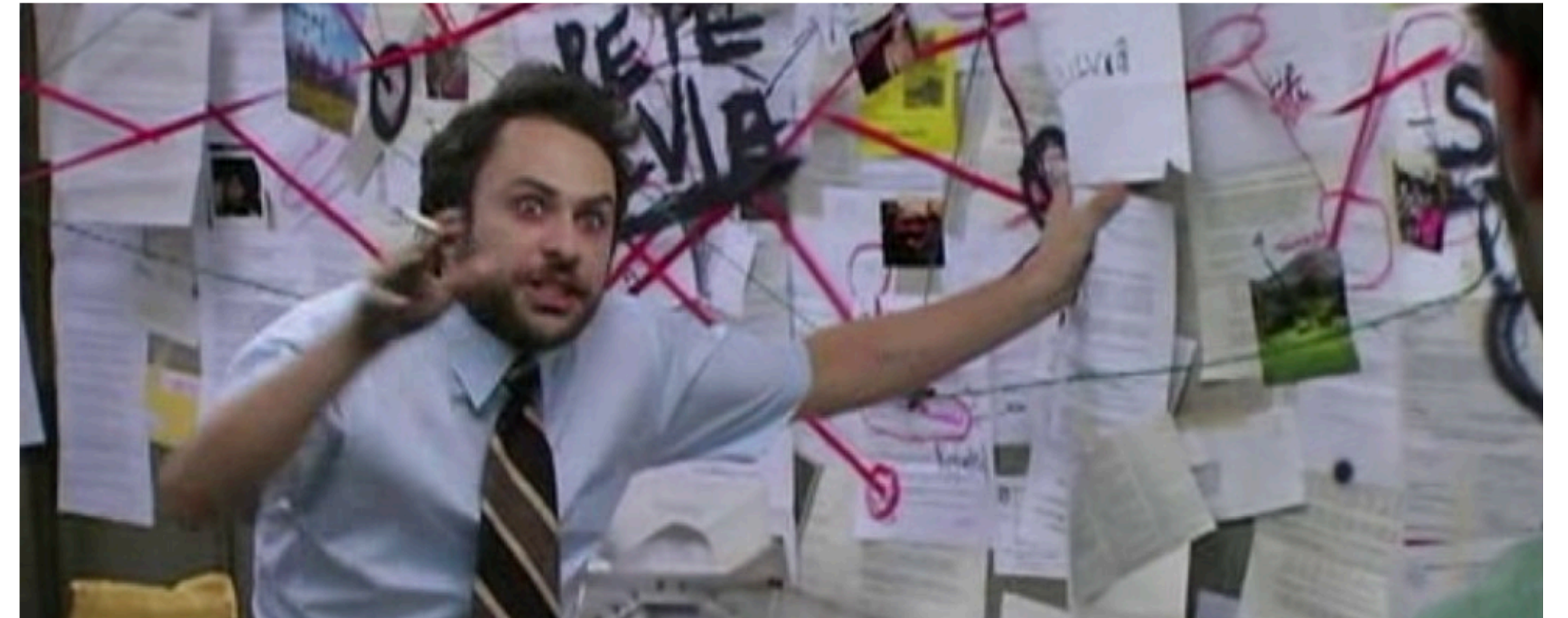
What is Polkadot data?

Show me a block



What is Polkadot data?

Show me a block



What is Polkadot data?

The important bits: substrate is key

- Block can be stored like a file
- Can be in JSON format (human readable)
- ~ Looks the same for all of the ecosystem
 - On block level (not pallet)
 - Optimisations possible
- polkadot-block-150293839.json
- kusama-block-18567368.json

```
[{
  "relay_chain": "polkadot",
  "chain": "polkadot",
  "timestamp": "1681278180000",
  "number": "15053869",
  "hash": "0x9fdb58b4e4748f9f9cb0af45586069f110358b3b2025581075cc144df5951027",
  "parentHash": "0x22610d161fd21ea28d35da6337de2581dfb20ffa7c090ee0ccd54483d07eb5e0",
  "stateRoot": "0x4ef482fb1e60e58fe14b6722aa077ec7546b3e96201b3b3821f43d2b9d85202f",
  "extrinsicsRoot": "0xc9473839d5862415135bf633f71632a61e6709d692decf6ee52181967a128c16",
  "authorId": "14rqNrQmWNK9YA1BSqRStNBUqGgwxKdysol4dNQ8GH8LRcrq",
  "finalized": "true",
  "extrinsics": [{
    "method": {
      "pallet": "timestamp",
      "method": "set"
    },
    "signature": null,
    "nonce": null,
    "args": "{ \"now\": \"1681278180000\" }",
    "tip": null,
    "hash": "0x751c5317aaeb1a5e6ca5983b585e7cd0baca054d9bcd412e6c3b2b0792b351c",
    "info": "{}",
    "era": {
      "immortalEra": "0x00"
    }
  }
}
```


What is Polkadot data?

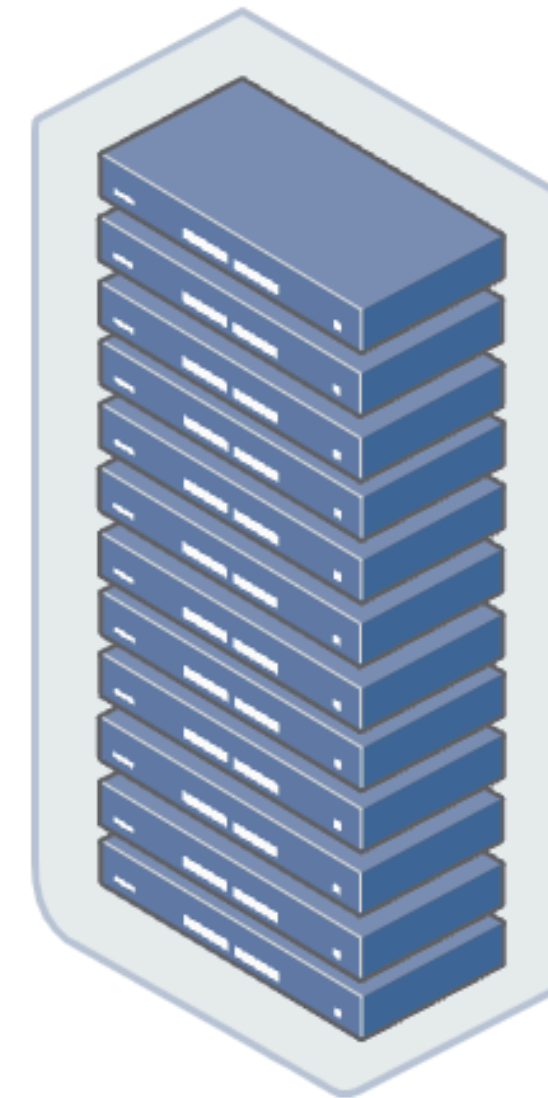
Where does it come from?



Light node



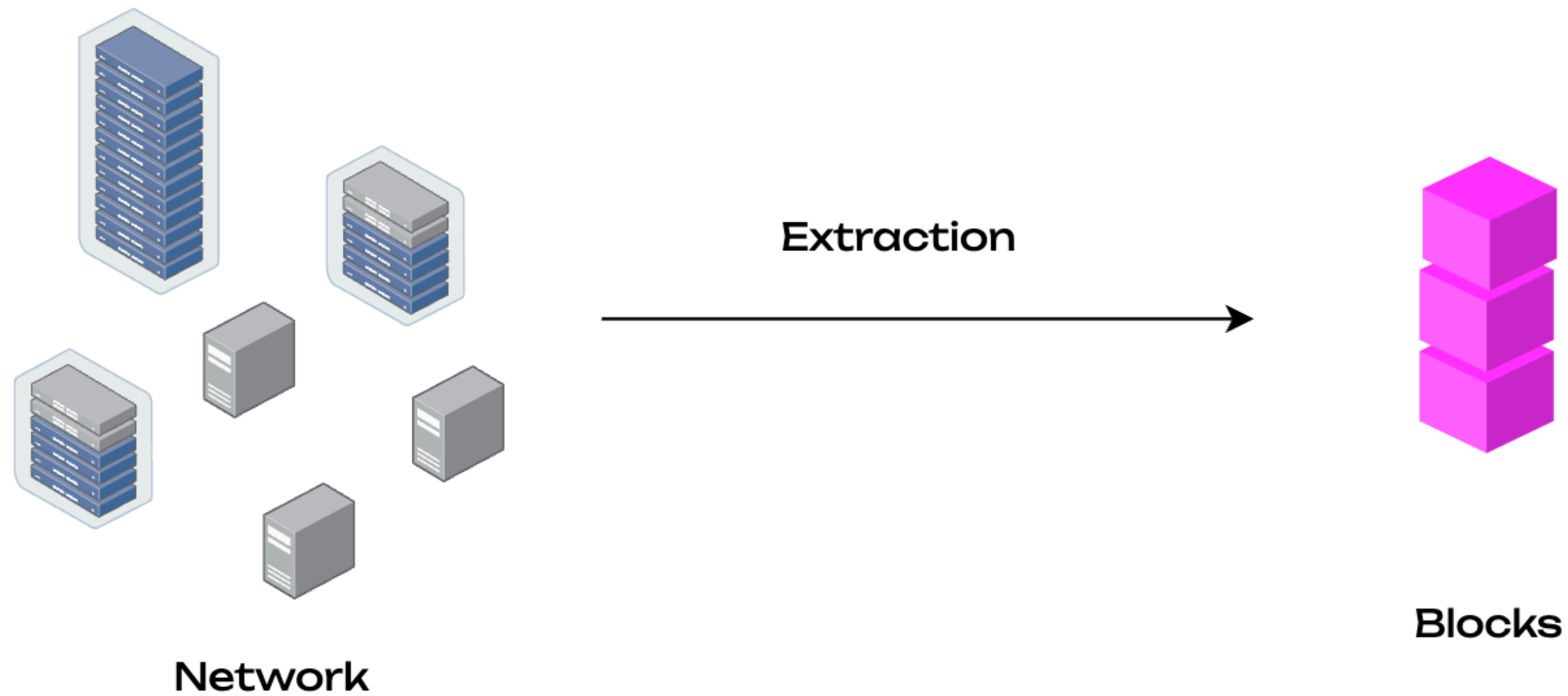
Full node



Archive node

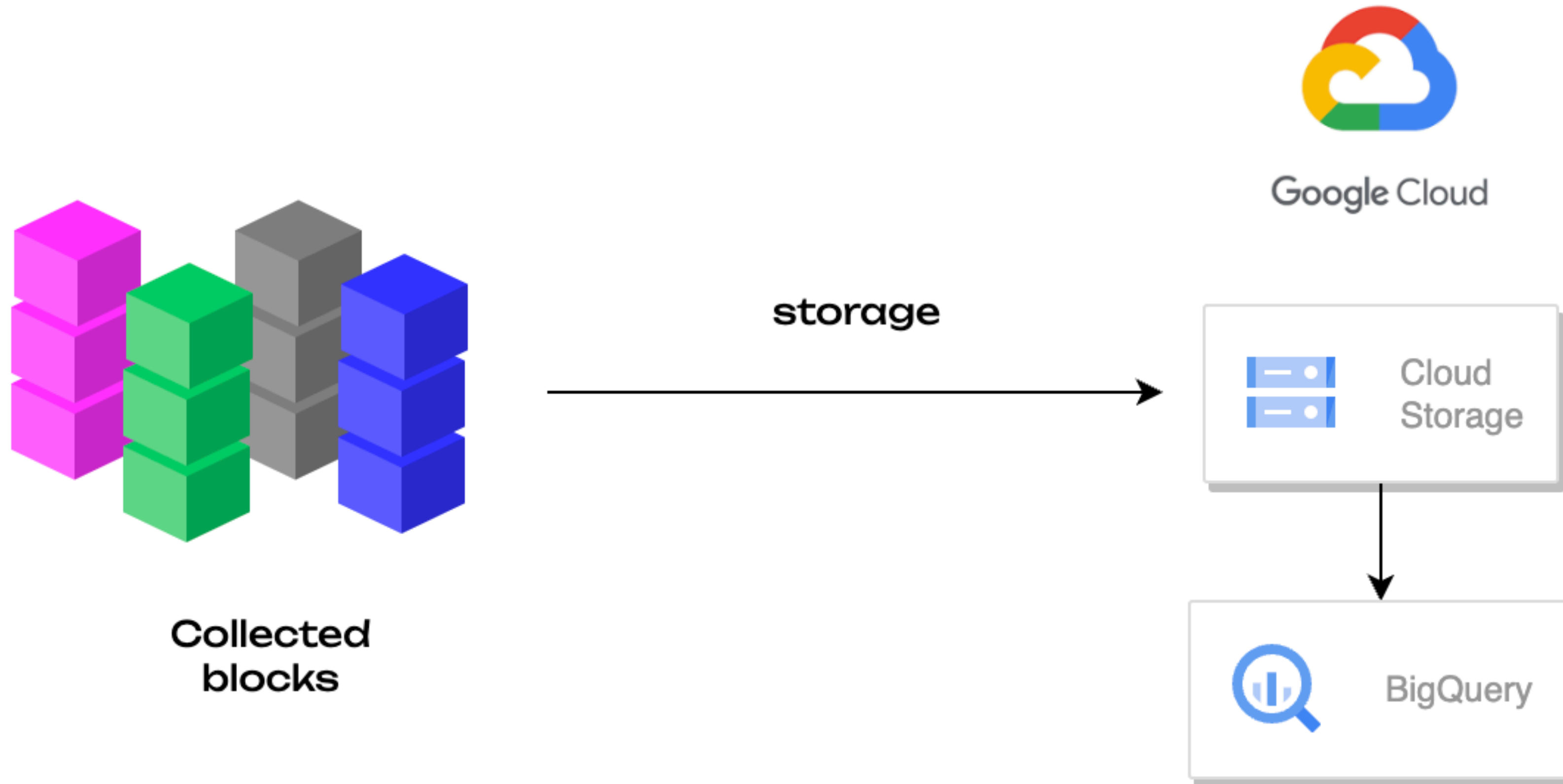
What is Polkadot data?

Simplified overview: one chain



What is Polkadot data?

Our setup moving forward



What can we do with it?

Data driven decision making for web3 using SQL

- Reads “naturally”, without fetching every block manually
- `SELECT * FROM polkadot;`
- `SELECT * FROM polkadot WHERE event = “transfer” AND account_balance > 10 DOT ORDER BY account_balance;`
- `SELECT block_id FROM kusama WHERE event = “transfer” AND account = “XXXXX” ORDER BY timestamp;`
- All of this, accross all of the chains

What can we do with it?

Data driven decision making for web3 using SQL

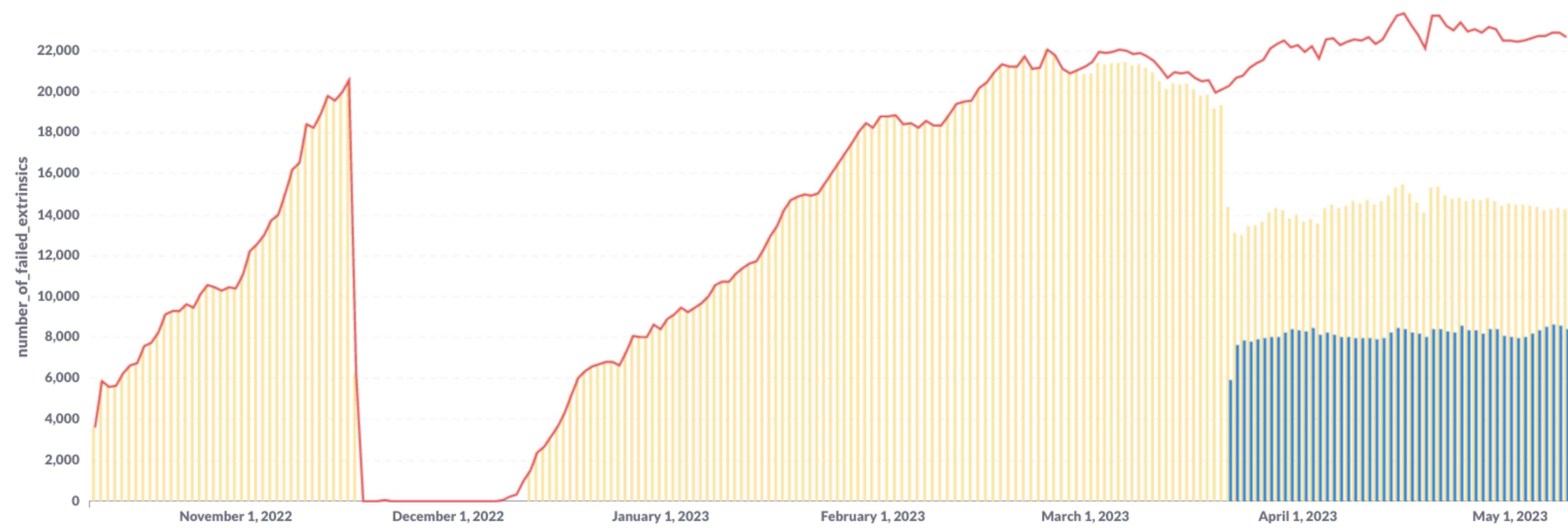
- Answer ecosystem wide queries
 - Eg: “Where’s my DOT?!”
- Verify & investigate
 - Eg: “What happened last month on X?”
- Dashboards, analyses & more
 - Interactive queries

Some examples of data in action

Failed extrinsics on Kusama

Kusama: Failed Transfer Extrinsics by Receiving Address

● Gk2FdnhLNnumXWrxjyk7yQstq3CL1Ni4QNCPuTeSjCdDoMh ● CdjNHo2d5h6o6fSWZBp2QPnJBcKqWq7EbtpG4sgt6BdAtvN ● total_failed_extrinsics



Some examples of data in action

Search for function usage in all deployed contracts

WASM Contract Search Tool

Parachain to search

astar

Search term

terminate

Toggle menu bar

astar contracts that contain `terminate` (Note: A blank page indicates nothing found.)

Show 50 entries

Search:

	chain	code_hash	segment
1	astar	0x5049a866128e9ddaf09a037296a63bb8aca536095679b692d0ea1555a45ae726	"seal1" "terminate" (func (;
2	astar	0xb03992f14a651e374bb7b0c9840a1264636d868296740ed87ce0f7ee7f138b0e	"seal1" "terminate" (func (;

Previous

1

Next

Some examples of data in action

Ecosystem Dashboards


















Polkadot: Number of Validators

















BigQuery - Evaluation




















```
1 SELECT
2   date,
3   max(staking_validatorCount) as `number_of_validators`
4 FROM `metabase_views.staking_nominators_polkadot_view`
5 WHERE {{timestamp}}
6 GROUP BY date
7 ORDER BY date desc
```


















Some examples of data in action

And more!

	Relay Chain: Number of Voters 
	Relay Chain: Number of Voters (Gov1) 
	Relay Chain: Number of Voters (Monthly) 
	Relay Chain: Number of Votes by Durat... 
	Relay Chain: Referendum
	Relay Chain: Sum of Tokens Voted 
	Relay Chain : Sum of Tokens Voted by R... 
	Relay Chain: Sum of Tokens Voted (Gov1) 

Type	Name ^
	Polkadot: Average Number of Members in ...
	Polkadot: Minimum Active Stake for Nomin...
	Polkadot: Number of Addresses Staked
	Polkadot: Number of Addresses Unbon... 
	Polkadot: Number of Blocked Nomination ...
	Polkadot: Number of Destroying Nominati...
	Polkadot: Number of Inactive Nominat... 
	Polkadot: Number of Inactive Nominat... 
	Polkadot: Number of Members in Nomi... 
	Polkadot: Number of Nomination Pools
	Polkadot: Number of Nominators 

	Answered Questions 
	Avg Time Question and Answer (Days)
	Avg Time Question and Answer (Hours)
	Questions 
	Stack Exchange: Answers by Score 
	Stack Exchange: Answers Last 7 Days
	Stack Exchange: Answers Last 7 Days 
	Stack Exchange: Answers Over Time
	Stack Exchange: Answers User id 
	Stack Exchange: Questions Answered 
	Stack Exchange: Questions by Tag Over Time
	Stack Exchange: Questions by Tag (Top ... 

	EGv2: No. of events 
	EGv2: No. of events - Daily
	EGv2: No. of transactions (Signed - Fee paying) 
	EGv2: No. of transactions (Signed- includes non-fee paying) - Daily 
	EGv2: No. of transactions (Total - includes non-fee paying) 
	EGv2: No. of transaction (Total - including non-fee paying) - Daily
	No. of transactions (Signed - Fee paying) 
	No. of transactions (Signed - Fee paying)
	No. of unique accounts
	Number of Events for Polkadot ecosystem 

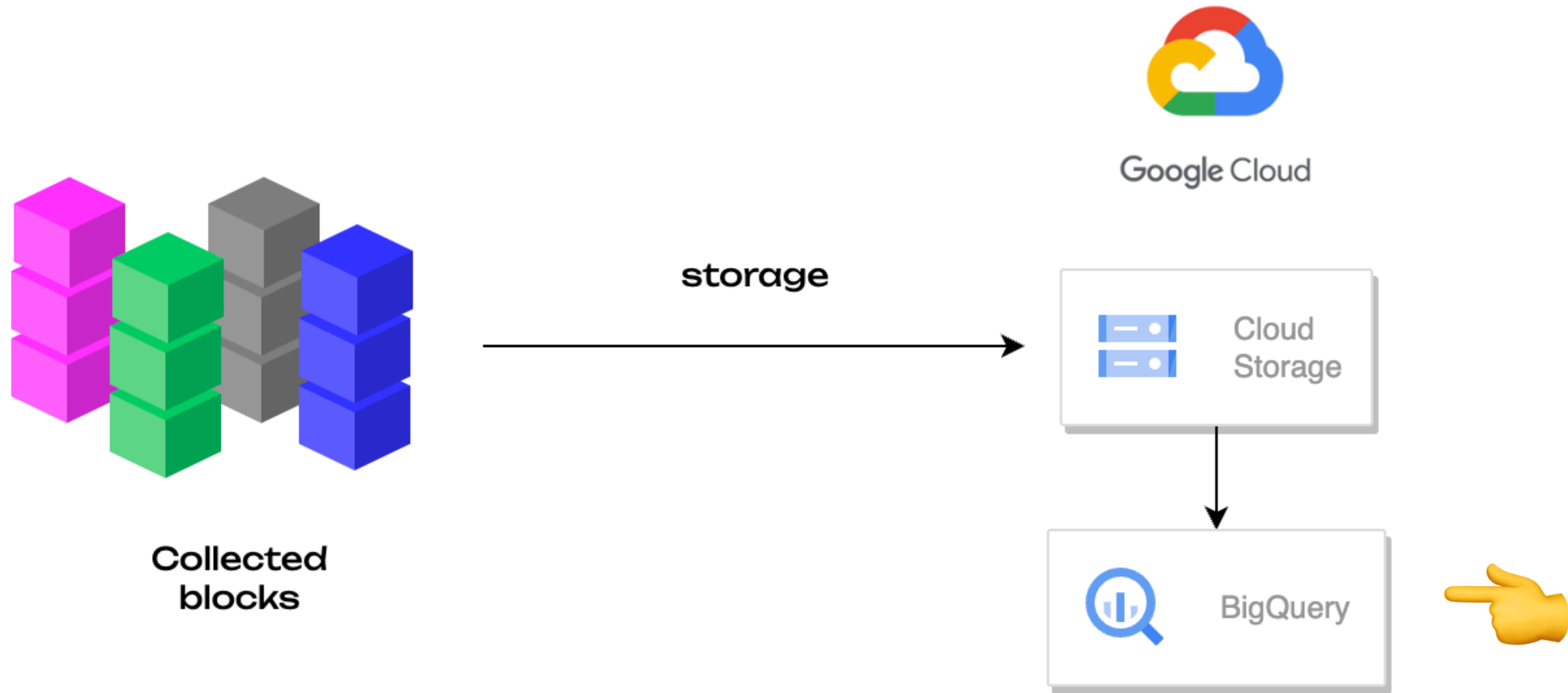
DEMO

How to use this: BigQuery



Cool stuff, but it's centralised

Saving costs, at a cost



Cool stuff, but it's centralised

Some data gymnastics

- JSON files are okay
- Optimising storage & querying
 - Range queries
 - Columns
- Other performant data formats
 - parquet

polkadot-289392.json
polkadot-382392.json

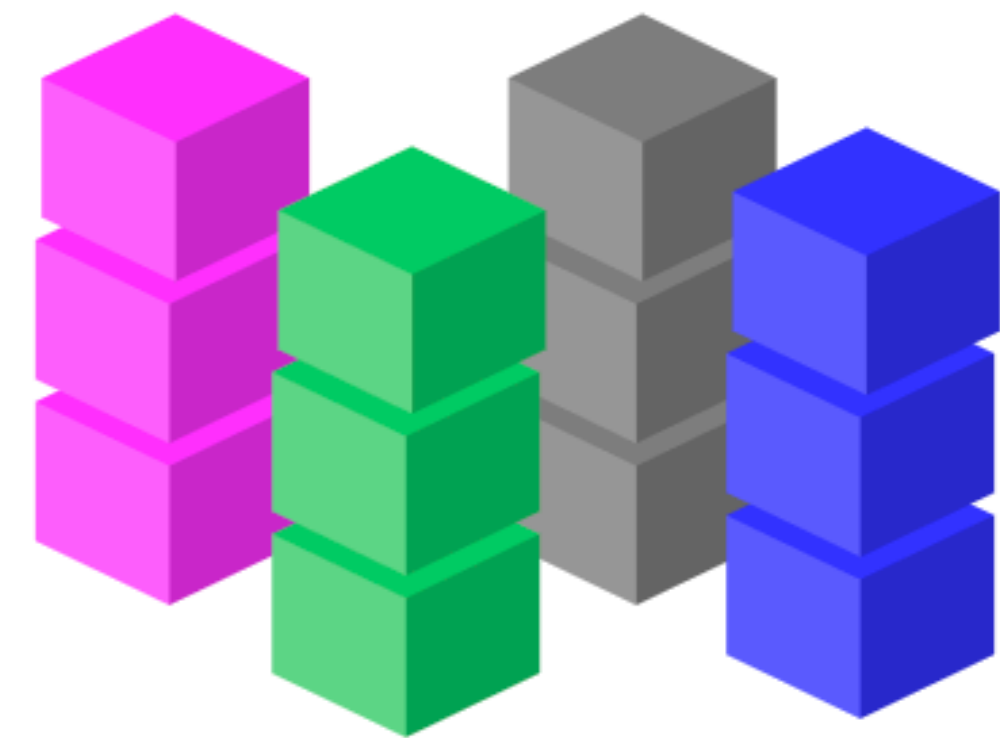
...



polkadot-blocks.parquet

Cool stuff, but it's centralised

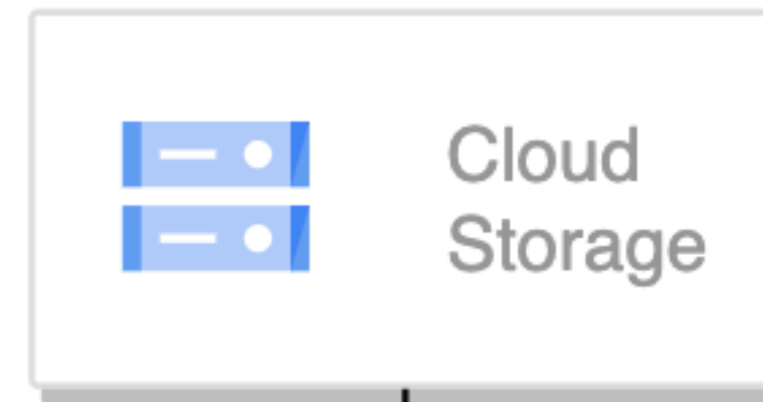
DuckDB & WASM to the rescue



Collected blocks



Google Cloud



Anywhere



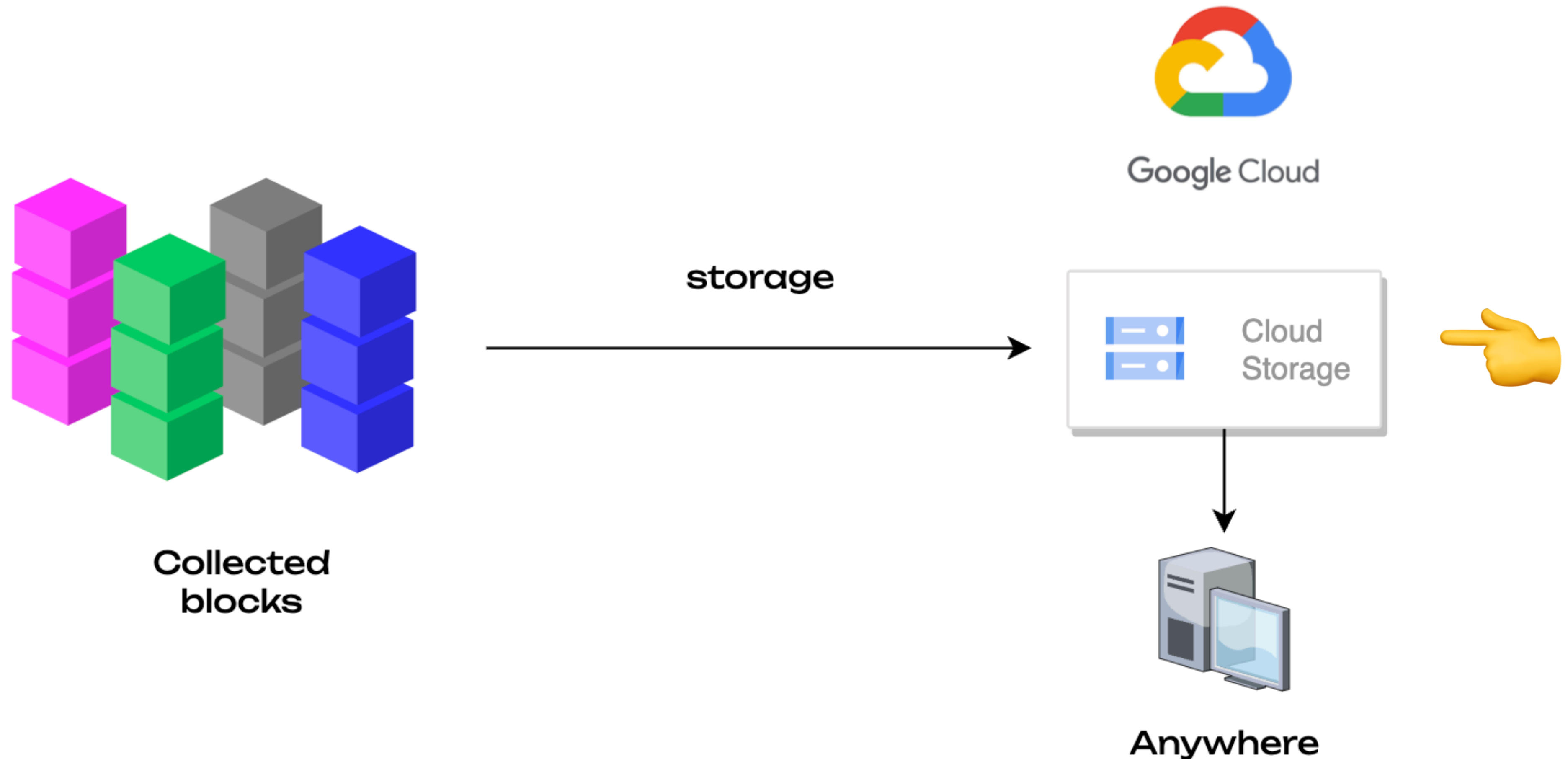
DEMO

Query Polkadot data from your browser



Still too centralised

Big potential egress costs at query layer, billing



Still too centralised

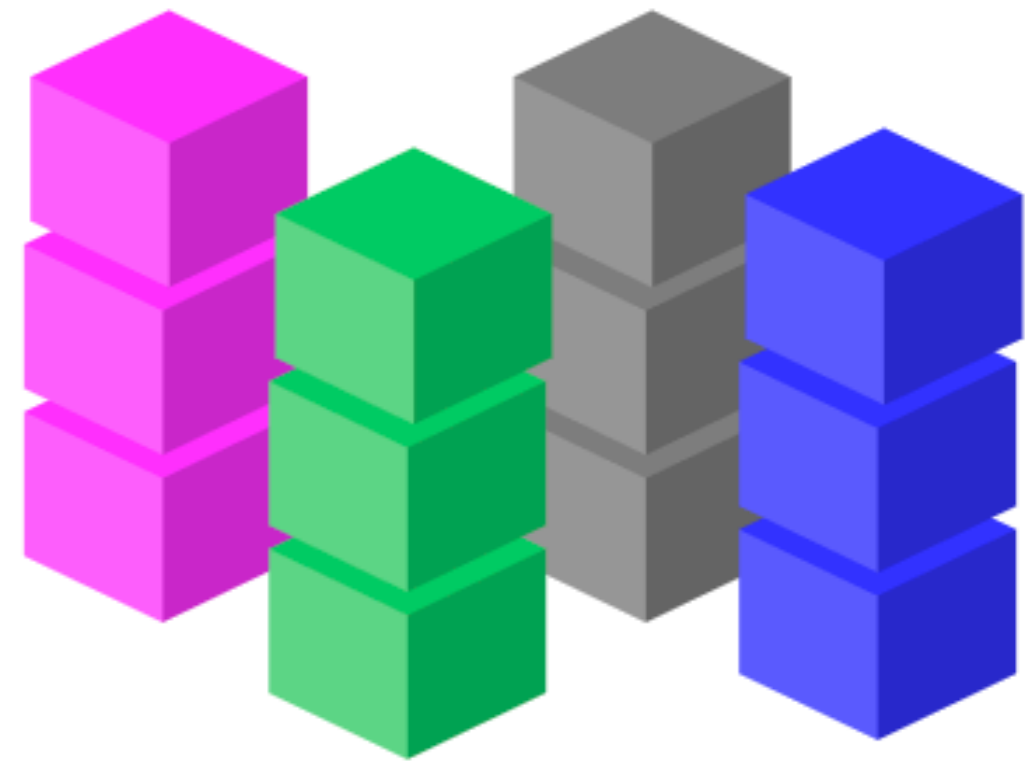
But wait!

- Cloud Storage
 - Stores files (objects)
 - Under a certain path
 - Unique identifier
 - Accessible over the internet
 - R2/S3/Snowflake etc



Still too centralised

Let's change one little thing



Collected
blocks

storage



Anywhere



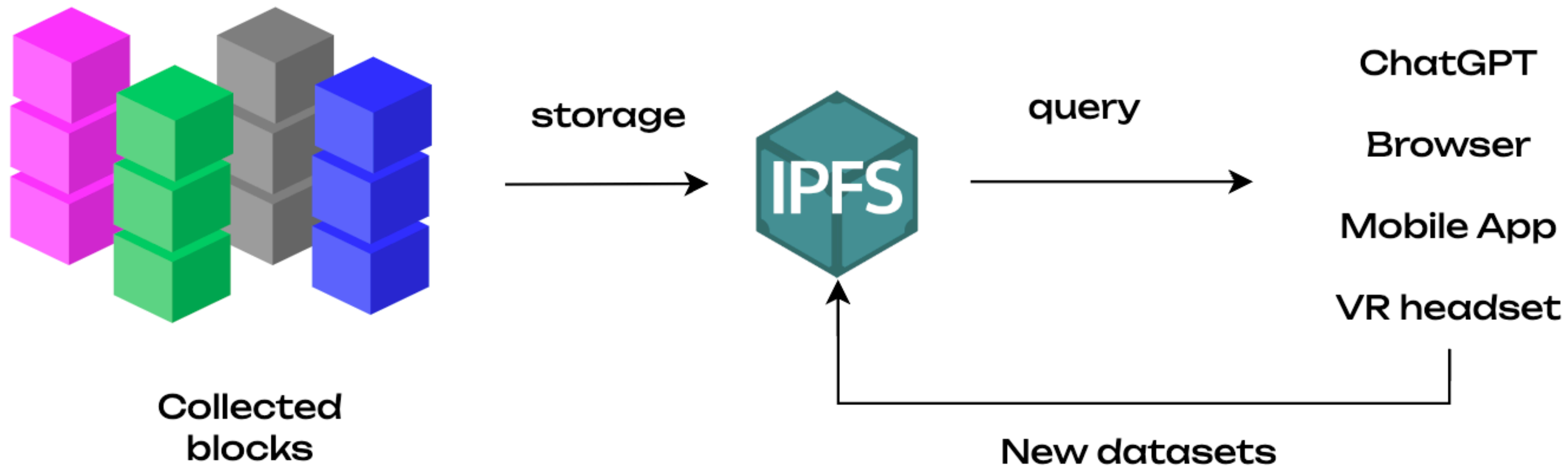
DEMO

Query Polkadot data from your browser over Crust IPFS Gateway



MVP of a Decentralised Dune

If it works in your browser, it works in your Dapp











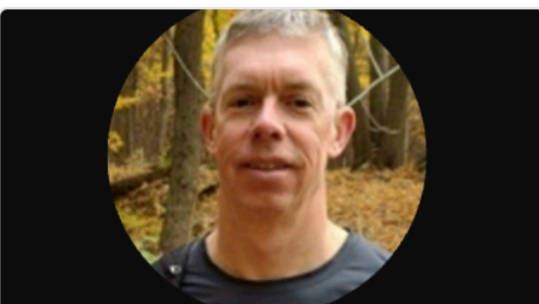
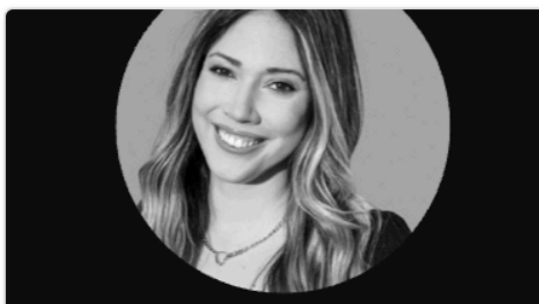

With Web3 primitives

“But will it scale tho?” Let’s find out.

Parity Data Team

We're not keeping this to ourselves

- Path to open sourcing
 - Tech
 - Data
 - Insights
- Get in touch 
- Support
- Questions & ideas

 <p>Karim - Lead Data Engineer Data Engineering Central team</p>	 <p>Pranay - Data Engineer Data Engineering Central team</p>	 <p>Marek - Data Engineer Data Engineering Central team</p>
 <p>Ollie - Ecosystem Success Engineer Data Analytics Central team</p>	 <p>William - Data Analyst Data Analytics Central team</p>	 <p>Rui Ferreira - Data Analyst Data Analytics Central team</p>
 <p>Pavla - Data Platform Lead Data Platform Central team</p>	 <p>Roger Bos - Data Analyst (Growth, Eco Ops) Data Analytics Extended team</p>	 <p>Alexa Berard - People Operations Data Analyst Data Analytics Extended team</p>
 <p>Marina Ariamnova - Data Analyst, Marketing & Communications Data Analytics Extended team</p>	+ New	

Get started today

Dotlake, substrate-etl and more

- Dotlake
 - Reach out to us
- Substrate-ETL (public BigQuery)
 - Colorful Notion
- Polkadot Data Alliance
 - Next steps

More data

