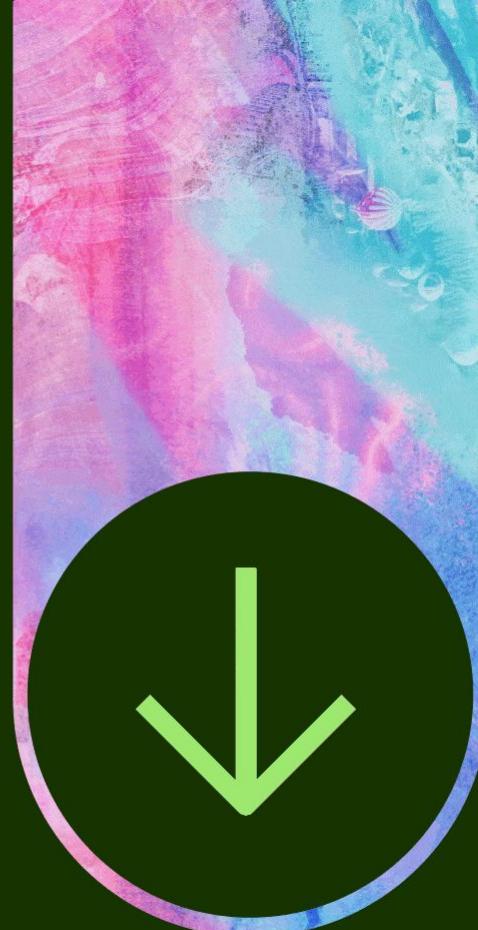


Graphs on Big Data

Why don't you use Neo4j?





Otto von Sperling

Data Scientist
Servicing Platform

Servicing customers

TALK

KNOW

PROTECT

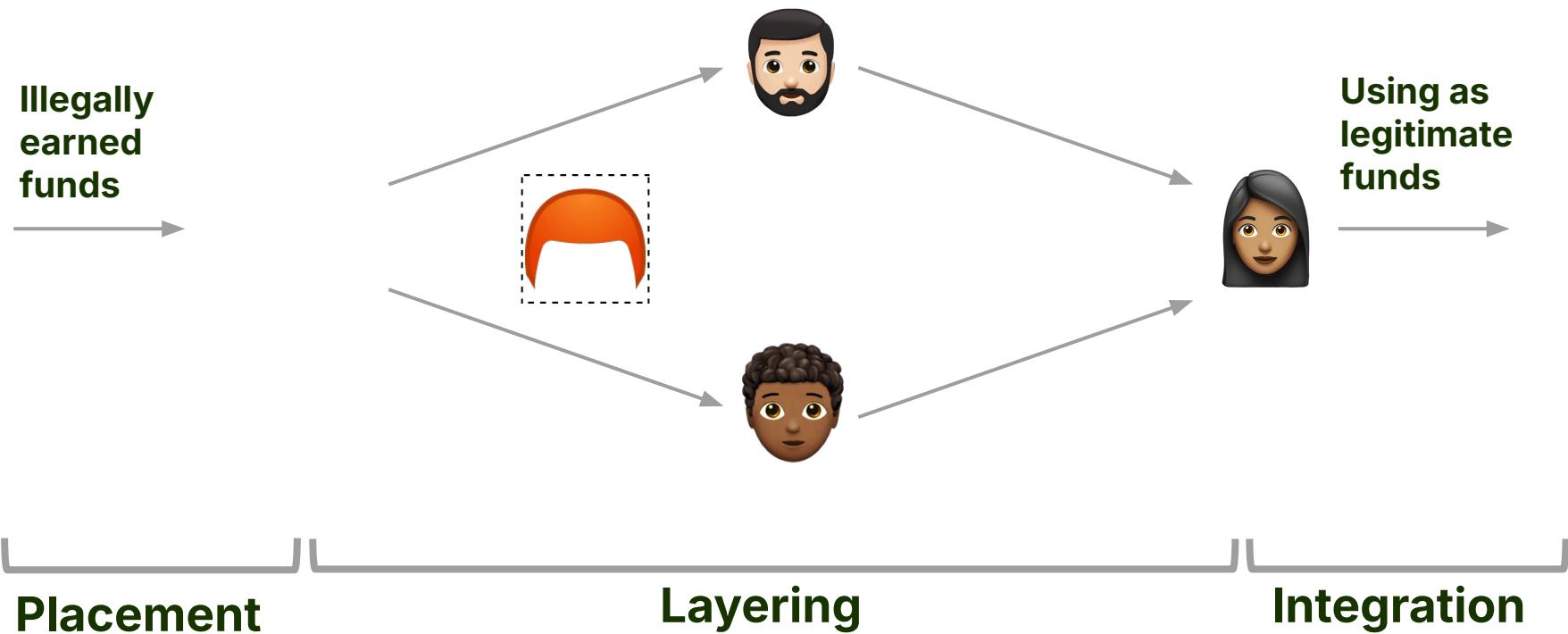


PLATFORM

PROBLEM

Finding accounts linked to criminals

Money Laundering - simple example



Networks without graphs

Feature Engineering

E.g. Ratio of identities marked as fraudulent in one's recipient list

Organic Art?

That slow down investigations

A lot of data



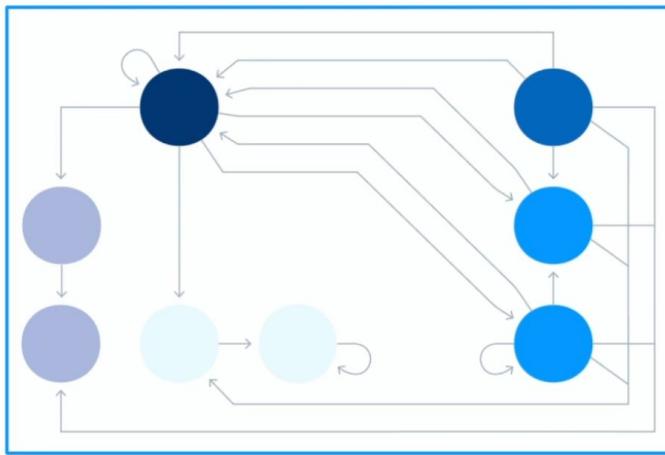
Neo4j

On big data?

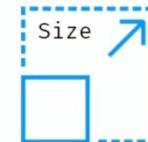
Approach #2: Graph Native Sharding



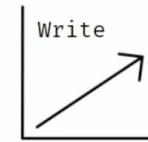
Invest in
Data Modeling



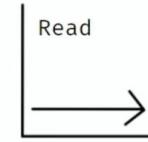
Get These Magical
Properties:



Unbounded size



Linear write throughput



Constant read performance

Hassle-free performance?



Vendors

Buy → Build → Bye-Bye

**ONE STEP
BACK**

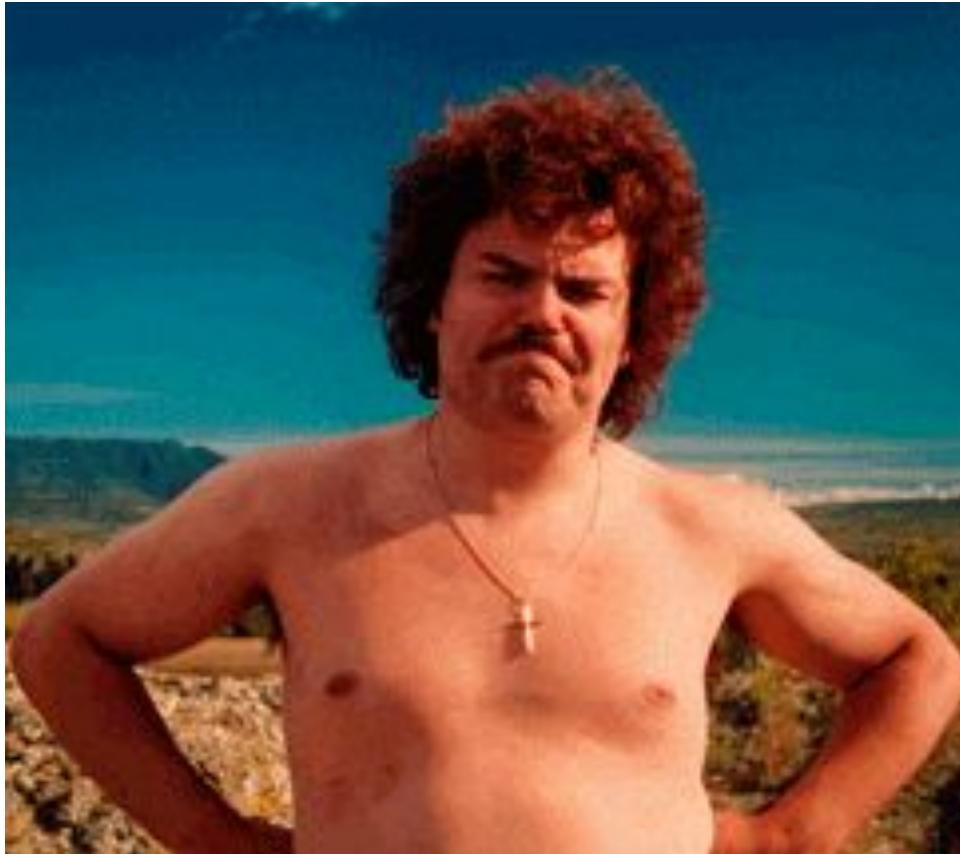
What are databases for?

Consistency

Availability

Partitions

Easy to get a new DB?

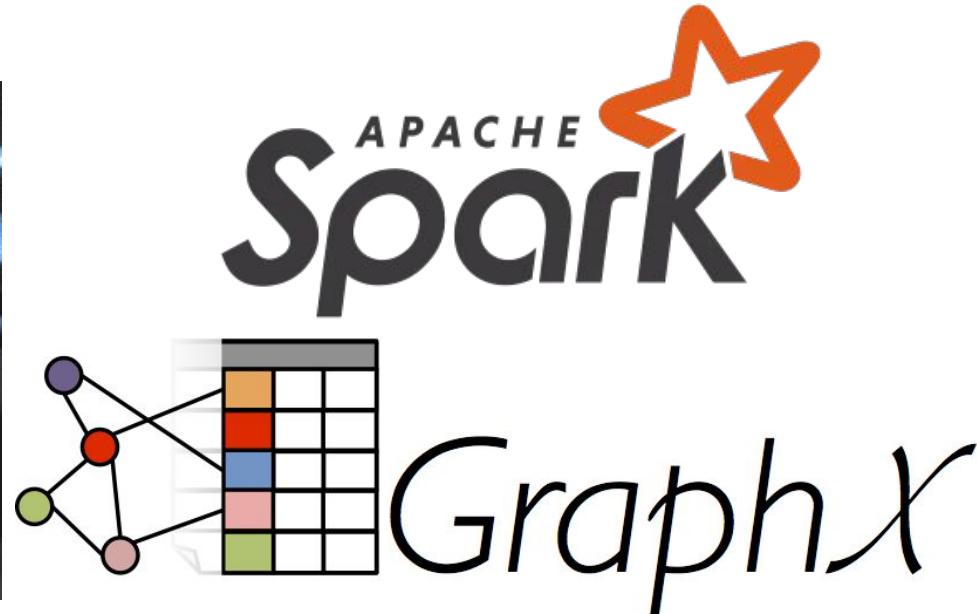


**ANOTHER
STEP BACK**

Alert investigation SLA?

one day, up to a week

Big data + longer SLAs...



GraphX

Long live...

[DISCUSS] Deprecation of GraphX

September 29, 2024

GraphX Deprecated

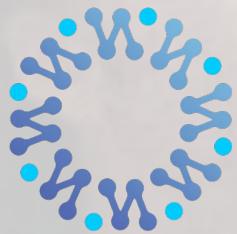
[ASF] Board Report: GraphX Deprecated

November 20, 2024

Find new maintainers... ?

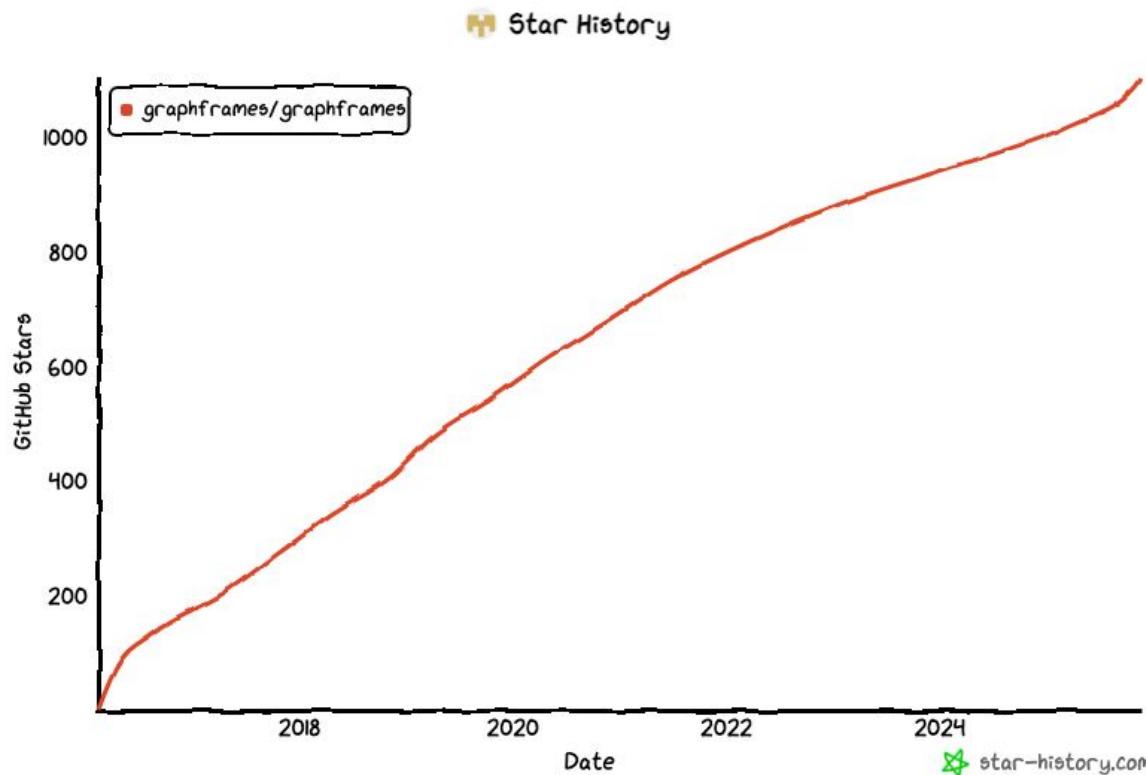






GraphFrames

GraphFrames is collaborative effort among UC Berkeley, MIT, and Databricks



"After a six year gap in additions, GraphFrames is back with Spark Connect support!"

MVP

AML

Tasks

1. Link identities

User → profile → recipients

2. Connect participants

User 1 → User 2

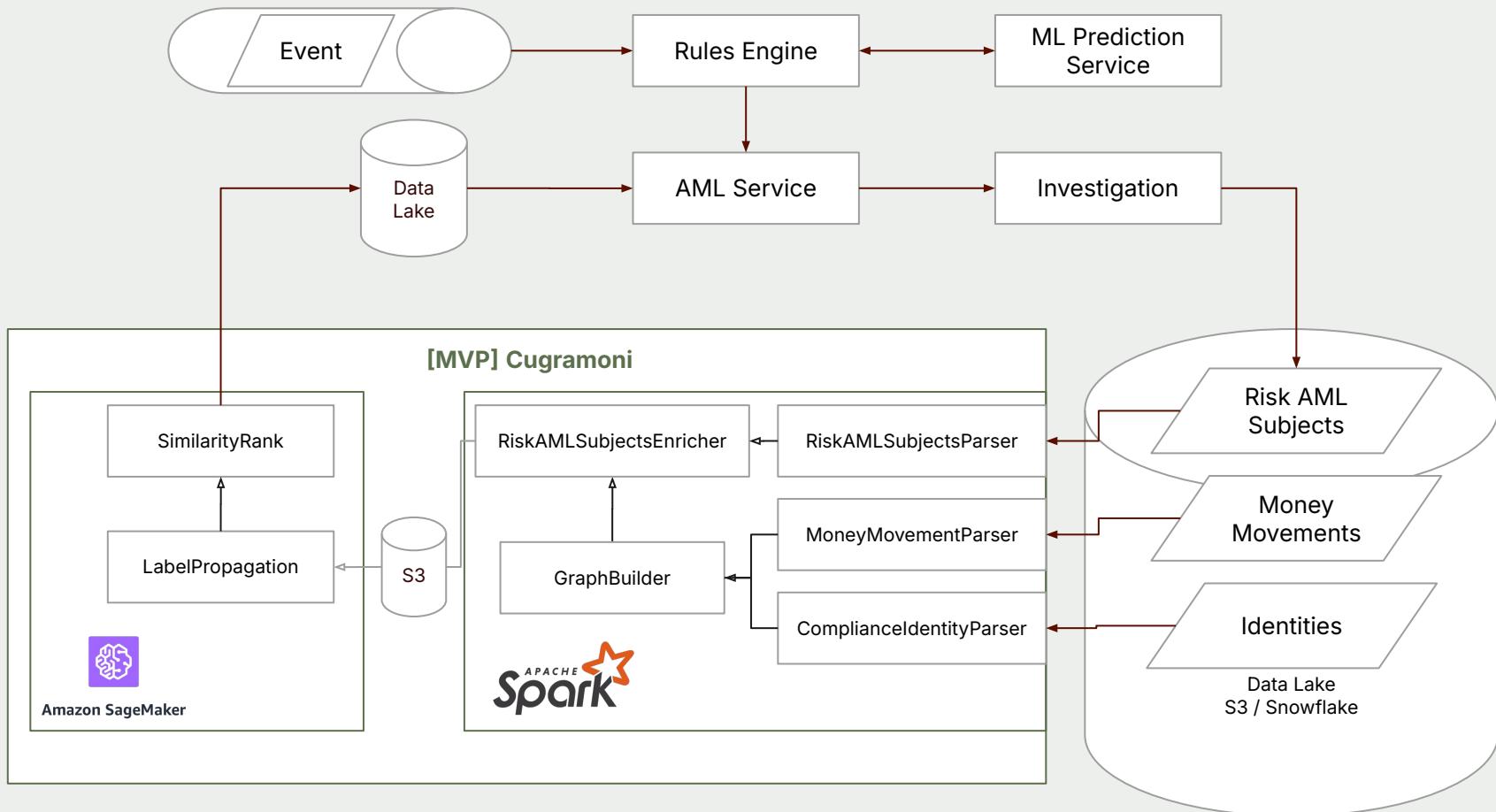
3. Add connected participants to AML Case

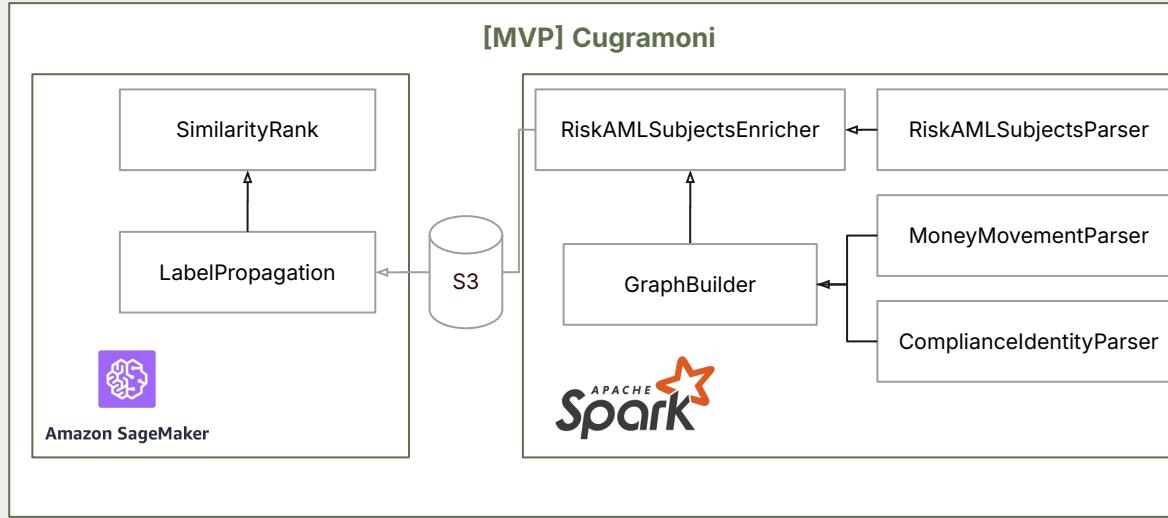
Reported User 1 → New User 2

4. Filter top-k participants

5. Create new cases for investigation

ARCHITECTURE





DEMO

Why don't you use Neo4j for that?

Misses the point

Tools



Solutions

NEXT STEPS

LIVE CODE

github.com/ottobricks/demo-graphs-spark

INTERESTED?

JOB RESULTS

39 results

**Mohammed Topiwala**

Lead Data
Scientist - Fraud
Prevention



London



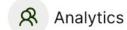
Analytics

**Luca Traverso**

Director of Data
Science -
Servicing



London



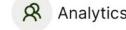
Analytics

**Pavlo Tertychnyi**

Lead Data
Scientist - Spend



Tallinn



Analytics





Rain Vagel

**Senior Data
Scientist- Risk**



Tallinn



Analytics



Alexandr Severinov

**Senior Data
Science Manager
- Financial...**



London



Analytics



**Lead Data
Scientist -
Treasury Liquidity**



London



Analytics





Pere Morell

Marketing Science
Manager

London

Analytics

Senior Data
Scientist -
Marketing

London

Analytics

Lead Marketing
Scientist

London

Analytics





It's a me!

Senior ML Engineer
Senior Data Scientist



Tallinn



Analytics



Σ