

DIFFERENT SHAPES OF DATA

SEMINAR 4

LASZLO SALLO & SZILVESZTER MOLNAR

NOVEMBER, 2017

TOOLS WE TRY TODAY

INFLUXDB

SOLR

NEO4J

SESSION RULES

WE ARE GOING TO USE SOCRATIVE ...

SERVER LOCATION AND ACCESS

ALL 3 TOOLS ARE INSTALLED ON AMAZON CLOUD

AND CAN BE ACCESSED AT:

<http://ceudsd.net/>

GENERAL RULE ABOUT THE ACCESS

DATA SOLUTION CAN BE GENERALLY ACCESSED:

PROGRAMING LANGUAGES (LIKE R, PYTHON)

NOTEBOOKS (ZEPPELIN, JUPYTER)

WEB INTERFACE ✓

HTTP SERVICE ✓



INFLUXDB - INSTRUCTIONS

BROWSER: <http://ceudsd.net/> > Influx DB



The screenshot shows the InfluxDB web interface. The top navigation bar is dark grey and contains the InfluxDB logo, links for 'Write Data' and 'Documentation', a dropdown menu for the database name 'Database: NOAA_water_database', and a settings gear icon. A red circle highlights the database dropdown menu. Below the navigation bar is a large white query input field with a 'Query:' label on the left. To the right of the input field are two buttons: 'Generate Query URL' and 'Query Templates'. A green circle highlights the 'Query Templates' button.

InfluxDB Write Data Documentation Database: NOAA_water_database ⚙

Query:

Generate Query URL Query Templates ▾

INFLUXDB - INSTRUCTIONS

THE DATA IS ALREADY PRE-LOADED FROM

https://s3-us-west-1.amazonaws.com/noaa.water.database.0.9/NOAA_data.txt

LINKS TO HELP YOU:

https://docs.influxdata.com/influxdb/v1.0/query_language/data_exploration/

https://docs.influxdata.com/influxdb/v1.0/query_language/math_operators/

https://docs.influxdata.com/influxdb/v1.0/query_language/functions/

NOAA_WATER_DATABASE SCHEMA

h2o_feet

time
level description
water_level
location

h2o_temperature

time
degrees
location

~ **80.000 ROWS**

average_temperature

time
degrees
location

h2o_ph

time
pH
location

h2o_quality

index
index
location
randtag



FACETED SEARCH



Shop by category ▼

Search...

Cell Phones & Smartphone: ▼

Related: [lg g3](#) [lg g4](#) [lg g2](#) [htc one m8](#) [note4](#) [htc one](#) [note 3](#) [htc](#) [samsung](#) [blackberry](#) [oneplus one](#) [htc one m7](#)

Categories

Cell Phones & Accessories

Cell Phones & Smartphones

Smart Watches

Smart Watch Accessories

Cell Phone Accessories

Display Phones

Phone Cards & SIM Cards

More ▼

Features

[see all](#)

- ☐ 3G Data Capable (31,735)
- ☐ 4G Data Capable (20,628)
- ☐ Bluetooth Enabled (34,541)
- ☐ GPS (30,787)
- ☐ Internet Browser (27,341)
- ☐ Music Player (30,791)
- ☐ Touchscreen (28,508)
- ☐ Wi-Fi Capable (28,179)

Brand

[see all](#)

- ☐ Apple (30,278)
- ☐ BlackBerry (4,672)
- ☐ HTC (6,151)

All Listings

Auction

Buy It Now

Sort:

Best Match ▼

View:



[Cell Phones & Accessories](#) > [Cell Phones & Smartphones](#)

Cell Phones and Smartphones 296,477 listings [+ Follow this search](#)

Shop by Color



Black



White



Gold



Gray



Factory Unlocked Apple iPhone 5S A1533 16GB 4G LTE Touch ID Mobile Smartphone

68,264.75 HUF
or Best Offer

From Hong Kong

FACETED SEARCH

Bundled Items

[see all](#)

- ☐ Case or Bag (841)
- ☐ Extra Battery (630)
- ☐ Extra Battery Charger (331)
- ☐ Flash (389)
- ☐ Lens (3,319)
- ☐ Memory Card (926)
- ☐ Strap (Neck or Wrist) (1,266)
- ☐ Tripod (337)

Series

[see all](#)

- ☐ Canon EOS (2,643)
- ☐ Canon EOS Rebel (426)
- ☐ Canon PowerShot (44)
- ☐ Fujifilm X Series (77)
- ☐ Nikon COOLPIX (56)
- ☐ Nikon D (2,951)
- ☐ Olympus OM-D (89)
- ☐ Sony Alpha (463)

Color

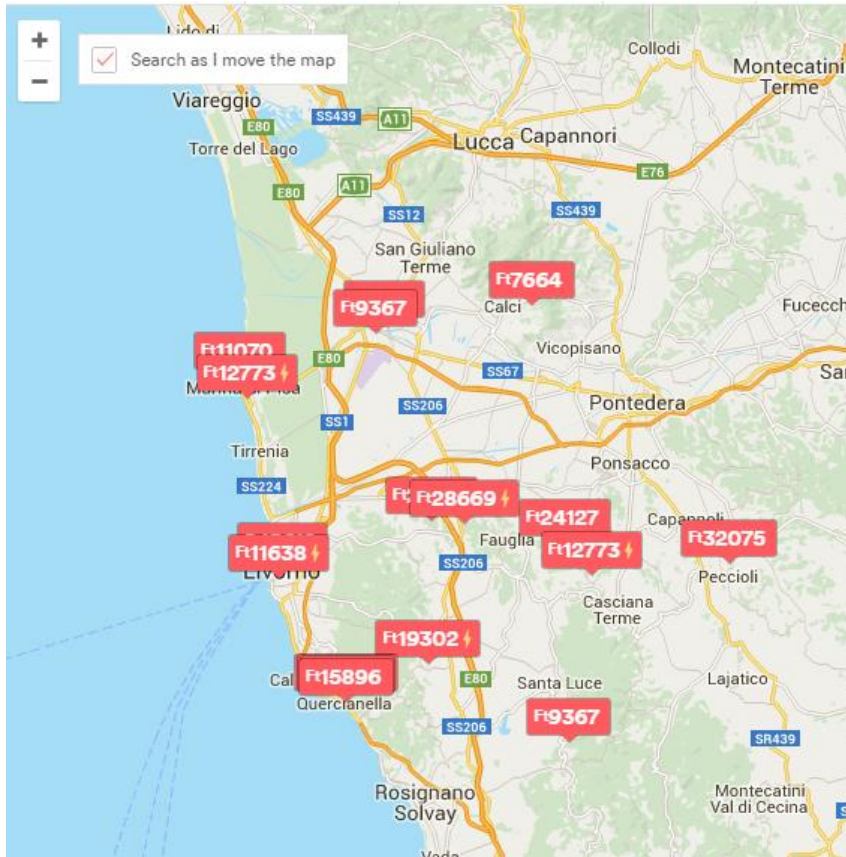
[see all](#)

- ☐ Black (3,014)
- ☐ Blue (40)
- ☐ Gray (60)
- ☐ Pink (14)
- ☐ Red (162)
- ☐ Silver (265)
- ☐ White (172)

GUIDED QUERIES

E-COMMERCE: GUIDE YOUR CUSTOMERS
ANALYTICS: SLICING AND DICING

GEOSPATIAL SEARCH



COORDINATE TO ITEMS

AREA BY DIFFERENT SHAPES
PROXIMITY

SOLR - INSTRUCTIONS

BROWSER: <http://ceudsd.net/> > Influx DB

THE DATA IS ALREADY PRE-LOADED FROM

<https://cran.r-project.org/web/packages/nycflights13/nycflights13.pdf>

LINKS WHICH MIGHT HELP YOU:

<https://cwiki.apache.org/confluence/display/solr/The+Standard+Query+Parser>

<https://cran.r-project.org/web/packages/solr/solr.pdf>

NYCFLIGHTS SCHEMA

year,month,day Date of departure

dep_time,arr_time Actual departure and arrival times, local tz.

sched_dep_time,sched_arr_time Scheduled departure and arrival times, local tz.

dep_delay,arr_delay Departure and arrival delays, in minutes. Negative times represent early departures/arrivals.

hour,minute Time of scheduled departure broken into hour and minutes.

carrier Two letter carrier abbreviation. See airlines to get name

tailnum Plane tail number

flight Flight number

origin,dest Origin and destination. See airports for additional metadata.

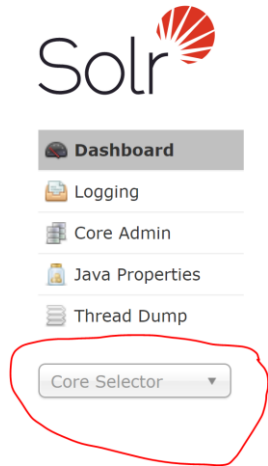
air_time Amount of time spent in the air

distance Distance flown

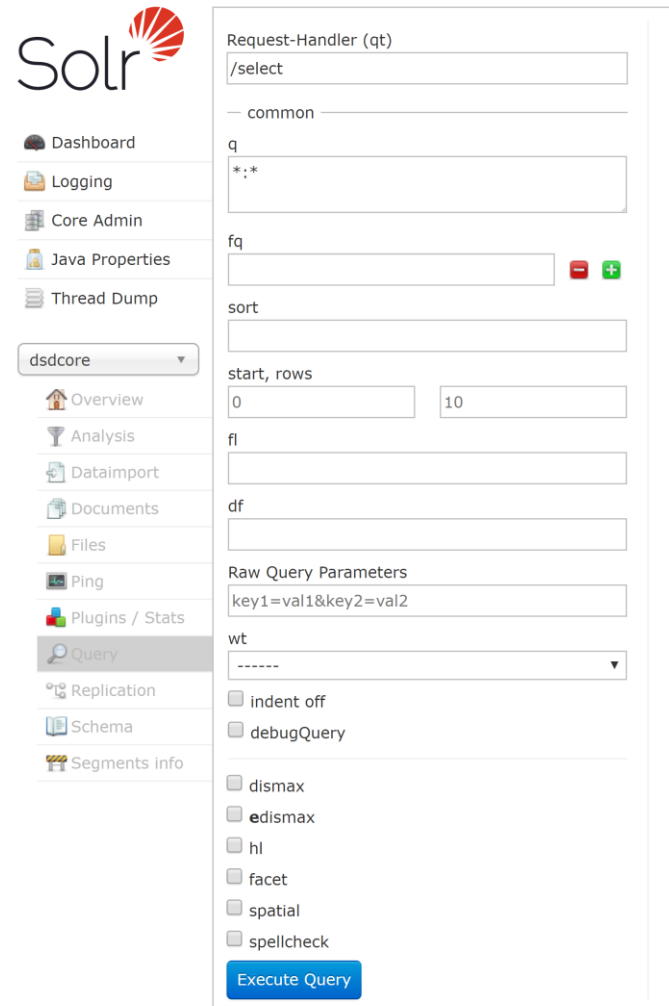
time_hour Scheduled date and hour of the flight as a POSIXct date. Along with origin, can be used to join flights data to weather data.

SOLR ADMIN

select dsd core



explore menu items, especially “query”

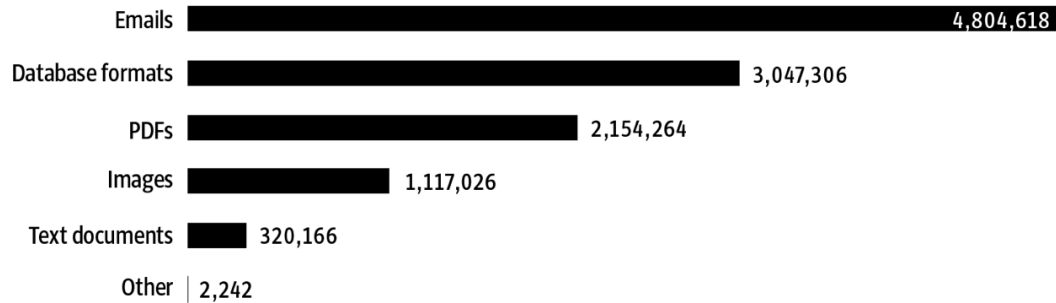




USE CASE – PANAMA PAPERS

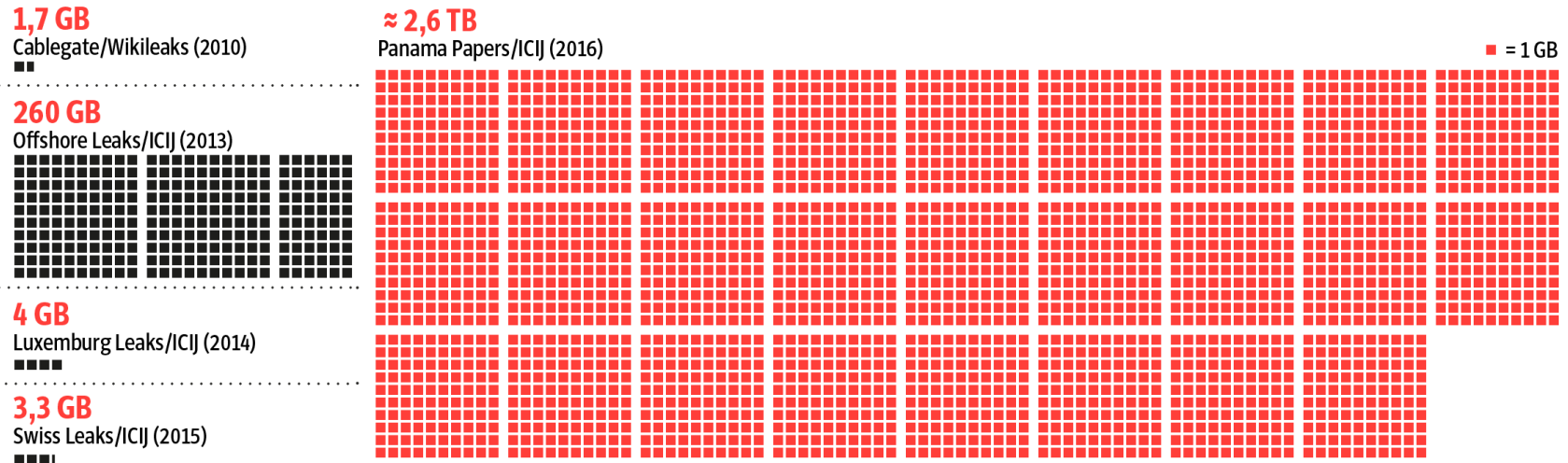
The structure of the leak

The 11,5 millionen contain the following file types

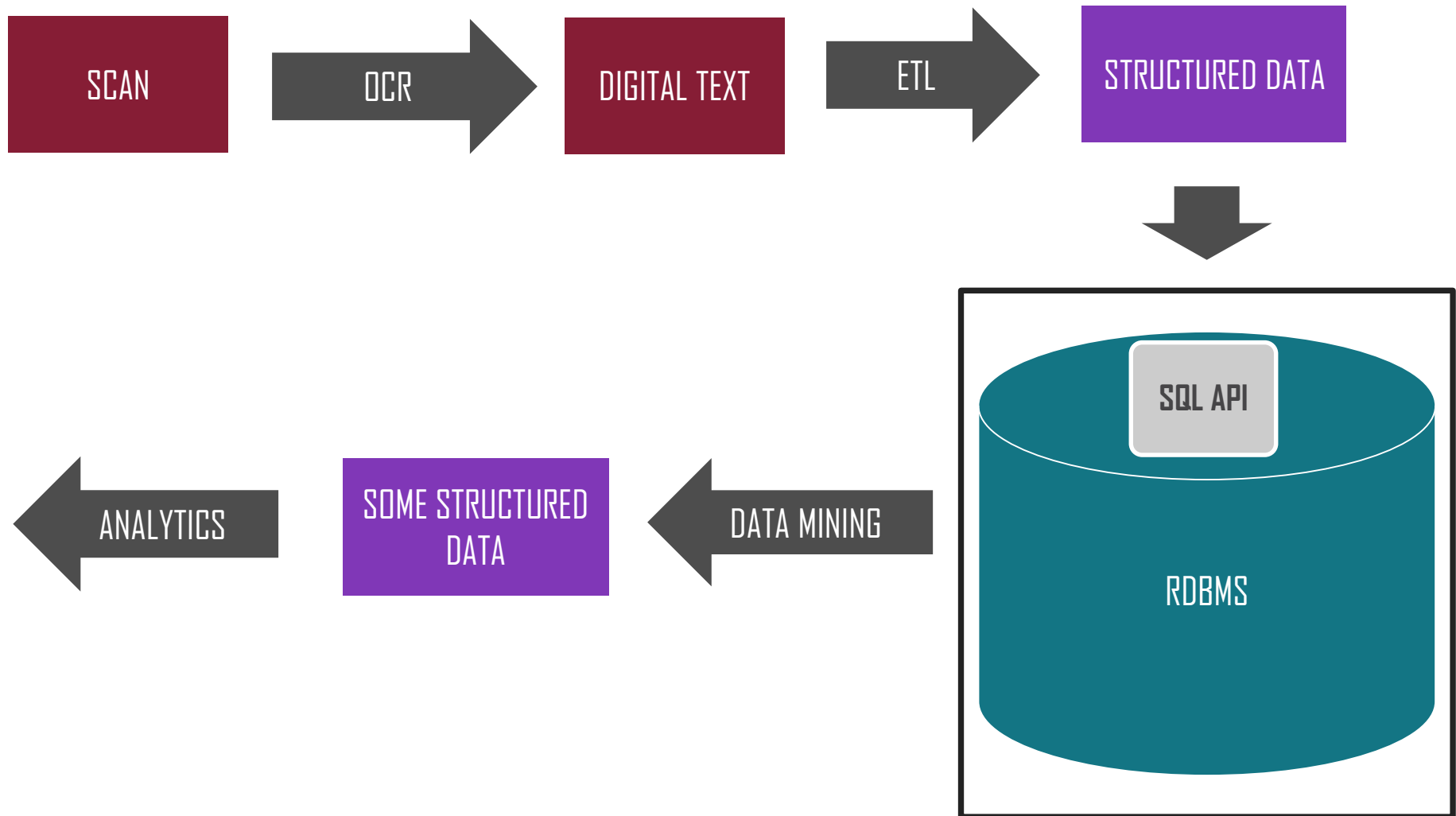


The scale of the leak

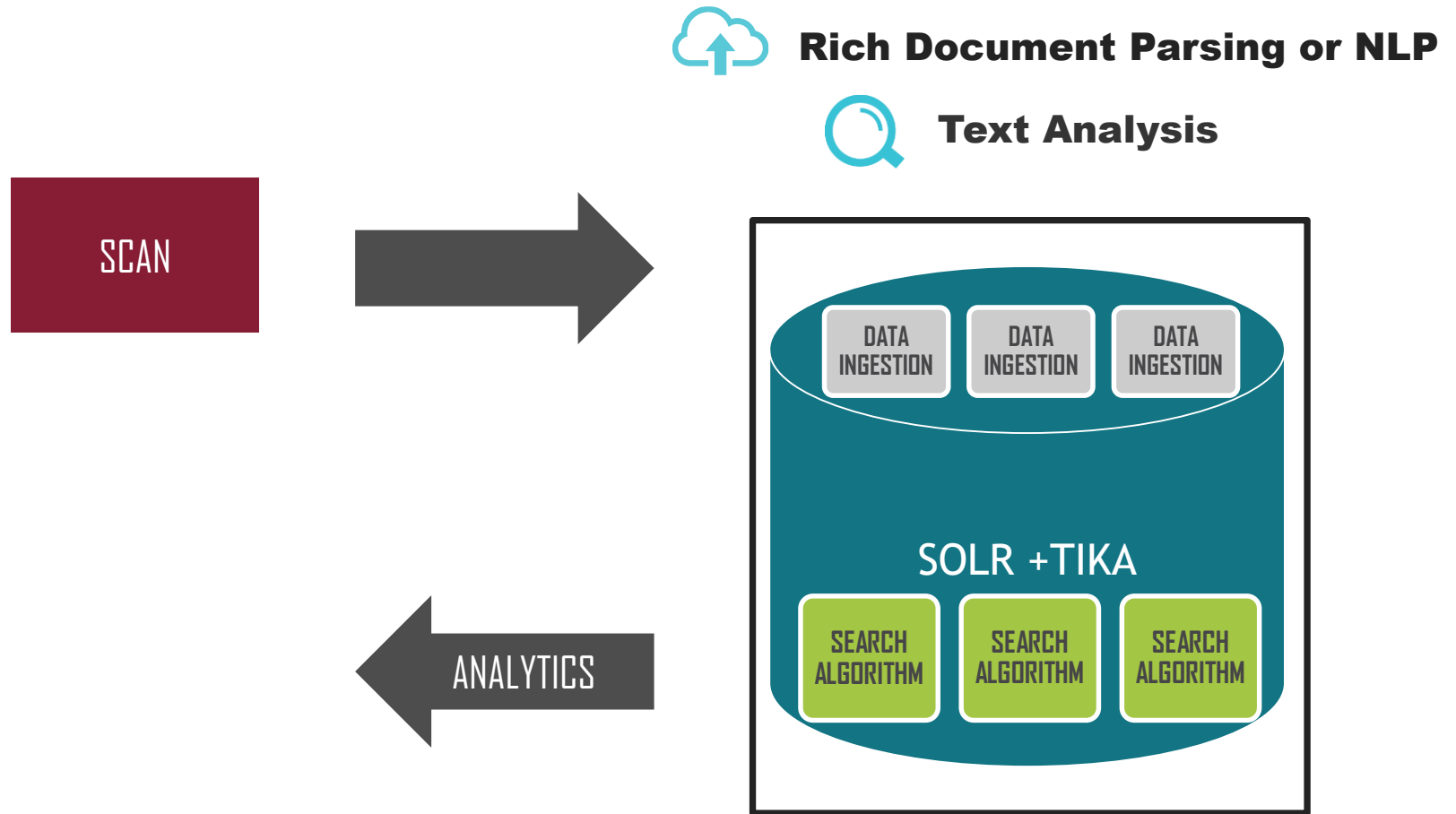
Volume of data compared to previous leaks



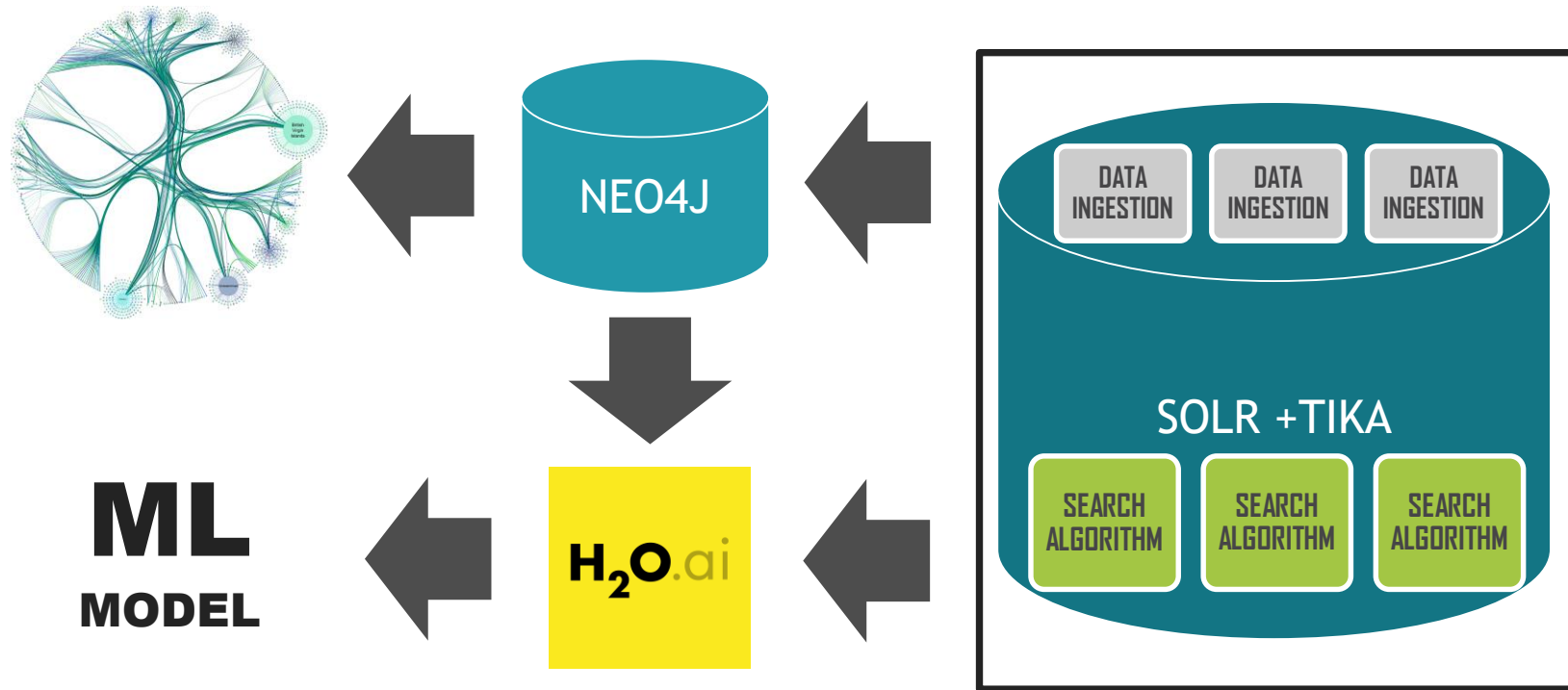
PANAMA PAPERS – CLASSICAL FLOW



WITH SEARCH ENGINE



ADVANCED ANALYTICS



NEO4J - INSTRUCTIONS

BROWSER: <http://ceudsd.net/> > Neo4j

PASSWORD: [ceudsd357](#)

THE DATA IS ALREADY PRE-LOADED FROM

<https://offshoreleaks.icij.org/pages/database>

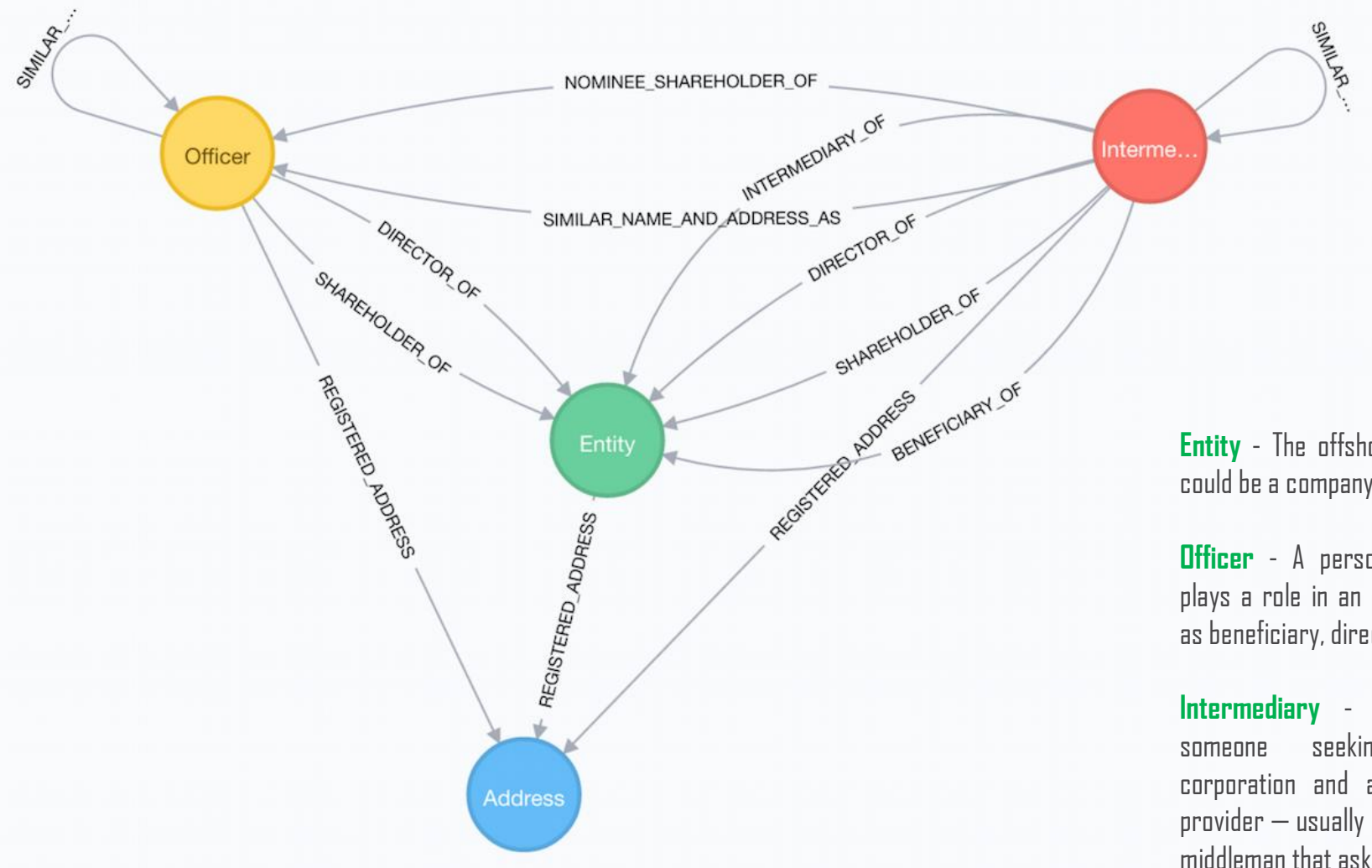
LINKS WHICH MIGHT HELP YOU:

<https://neo4j.com/developer/cypher-query-language/>

<http://neo4j.com/docs/developer-manual/current/cypher/>

<https://cloudfront-files-1.publicintegrity.org/offshoreleaks/neo4j/guide/index.html>

PANAMA PAPERS SCHEMA



Entity - The offshore legal entity. This could be a company, trust, foundation ...

Officer - A person or company who plays a role in an offshore entity, such as beneficiary, director, or shareholder.

Intermediary - A go-between for someone seeking an offshore corporation and an offshore service provider — usually a law-firm or a middleman that asks an offshore service provider to create an offshore firm.

Address - The registered address