MarineWatch

Find best navigable routes

Chloe Prouvost - Nathan Balblanc - Quentin Viozelange - Valerian Cazanave - Gautier Delorme

Summary

General Overview

What is it?

San Francisco --> Caracas

4

Technical Overview
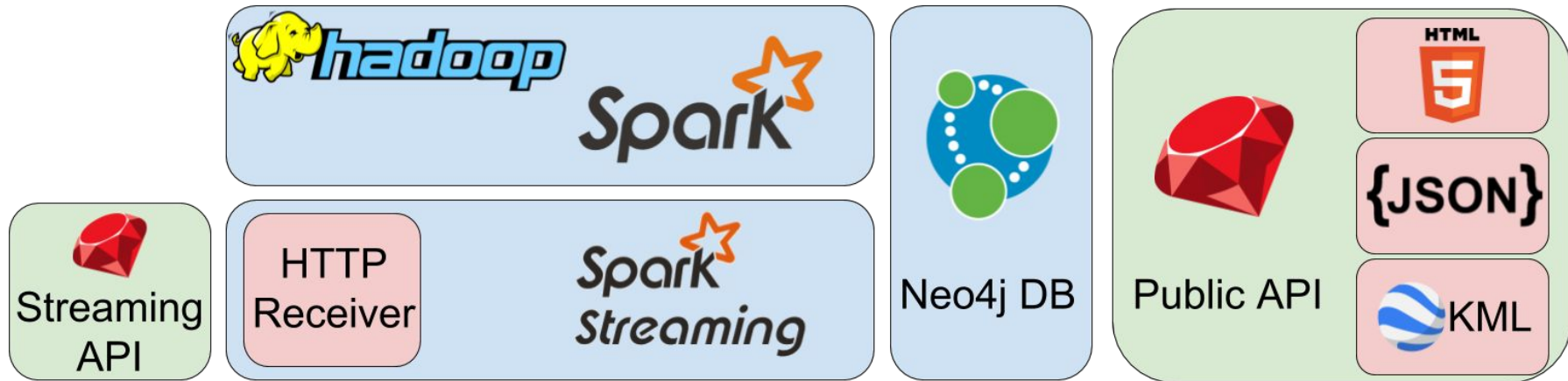
Big Data Environment

Batch Processing

CSV

0.2	0.5	0.3
0.1	0.7	0.9
0.1	0.4	0.2

RDD

(0, [(0, 0.2), (1, 0.5), (2, 0.3)]),
(1, [(0, 0.1), (1, 0.7), (2, 0.9)]),
(2, [(0, 0.1), (1, 0.4), (2, 0.2)])

DataFrames

Cell_ID	Lat	Long
0	-19.7	67.9
1	34.4	-145.2

Rel_ID	Cost	From	To
0	34.54	0	1
1	234.6	3	2

Neo4j

cells.csv
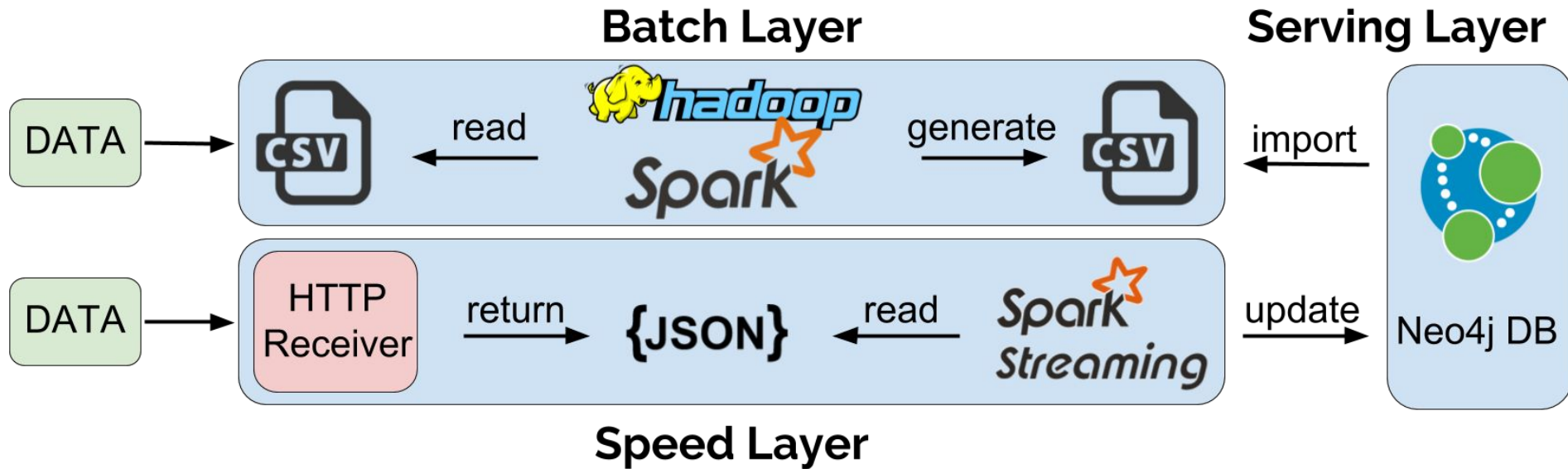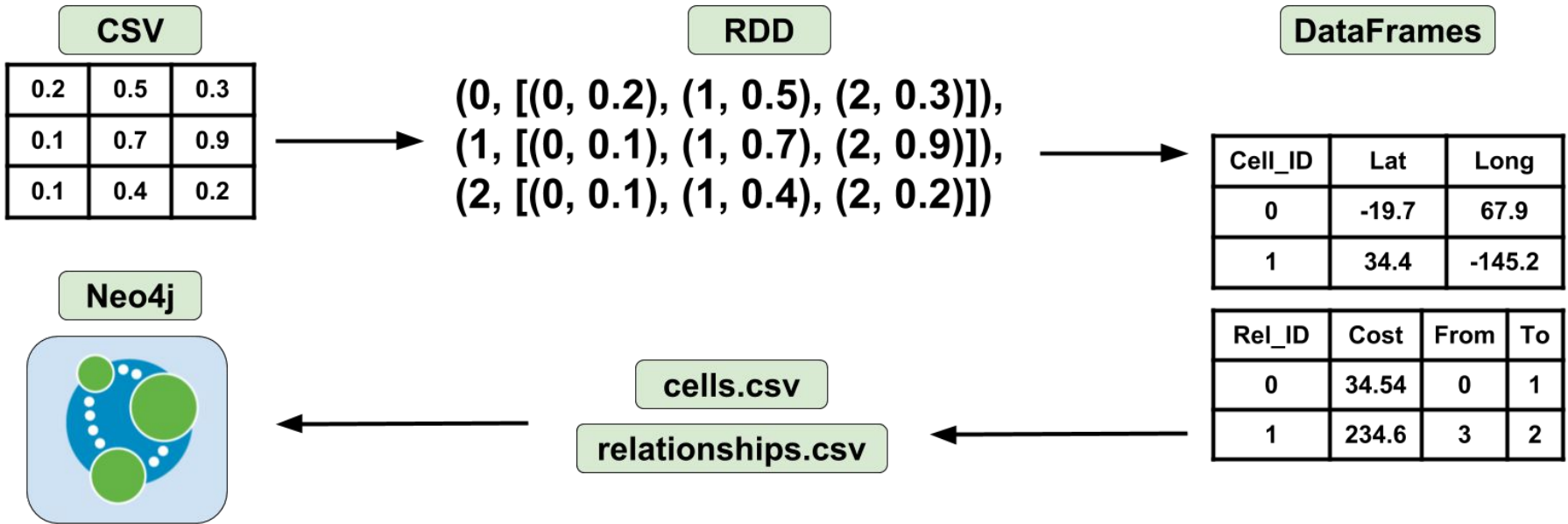
relationships.csv

Streaming Processing

JSON

{"lat":19.7,"long":36.4,"cell":0.7}
…
{"lat":56.2,"long":134.9,"cell":0.2}

RDD

(19.7, 36.4, 0.7),
…
(56.2, 134.9, 0.2)

Neo4j

Cypher

```
MATCH (c:Cell{lat:19.7,long:36.4})<-[r]-()
SET r.cost = 1.43
```

Neo4j Database
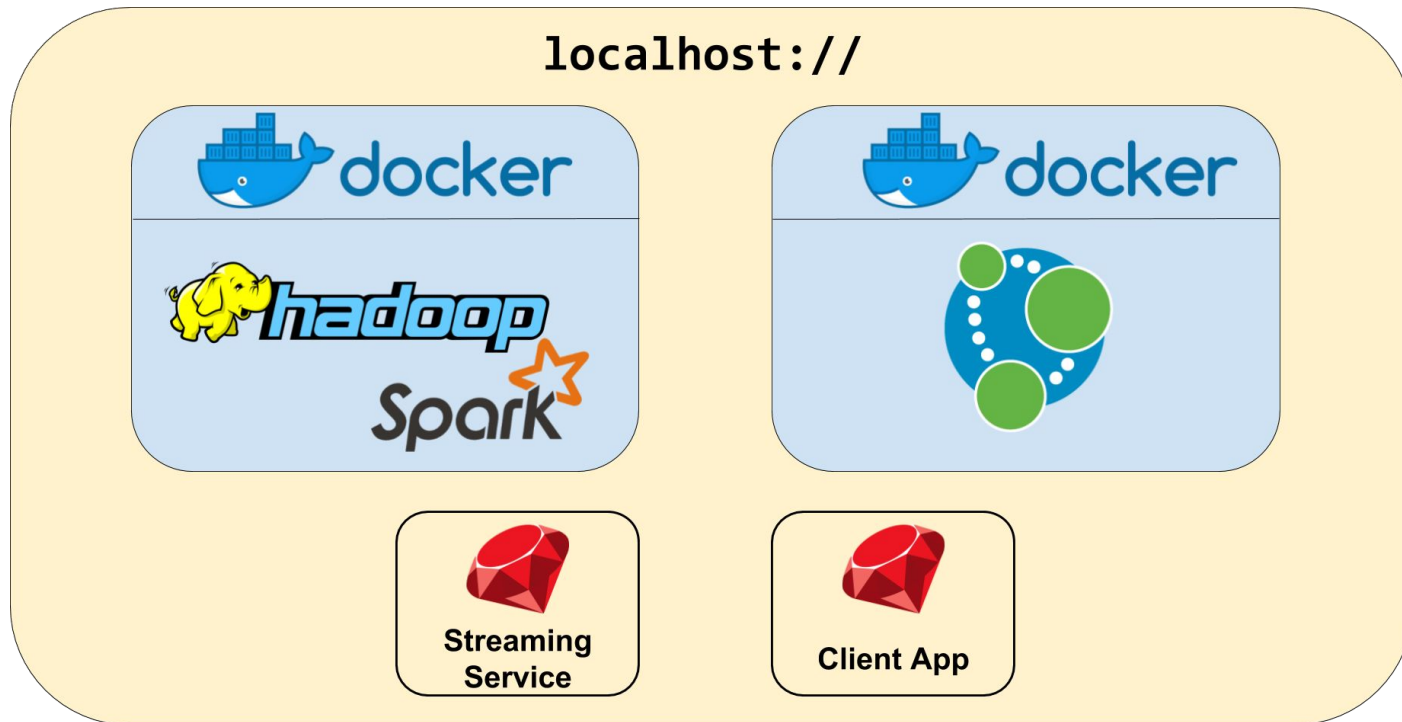
32 868 925 nodes

225 213 632 relationships

A* to find the "best" path
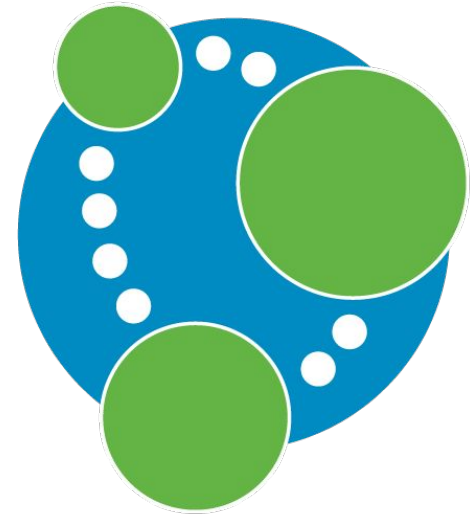
Development Environment

CLI Tool

```
$ ./marinewatch-cli -h
Usage: marinewatch-cli [-h] [-b <int>] [-c <string>] [-u <string>] [-d <string>] [-s] [-t]

  -h  Help. Display this message and quit.
  -b  <int>  Run batch process with specified accuracy.
  -c  <string>  Run streaming process listening on specified address.
  -u  <string>  Create new database with name.
  -d  <string>  Start database with name.
  -s  Start web server.
  -t  Start streaming server.
```

11

Test #1

San Francisco --> Caracas computed in 200ms
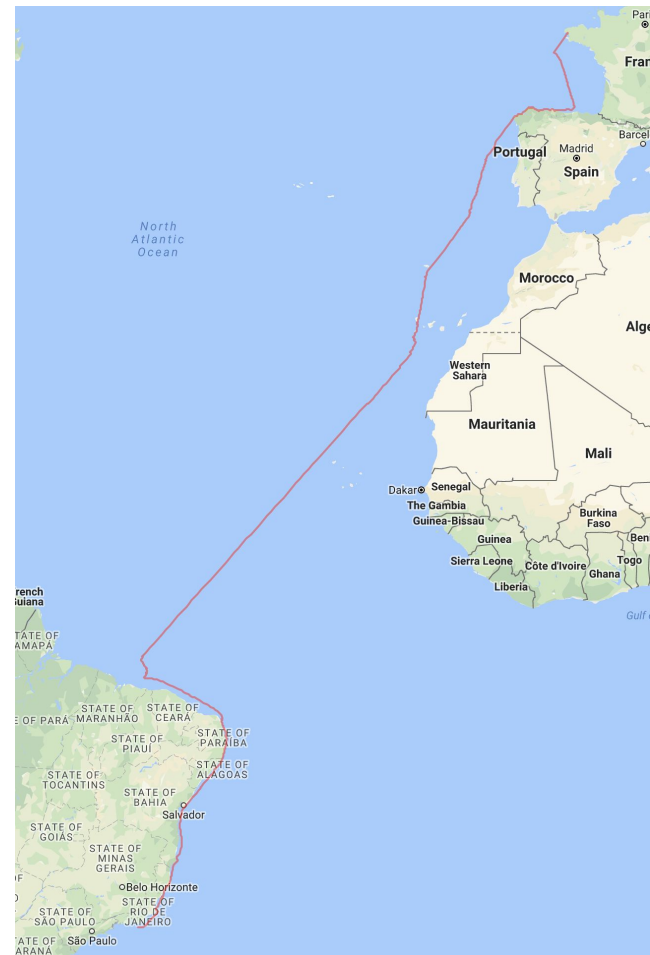
Using the Panama Canal

Test #2

Rotterdam --> Kuala Lumpur computed in 2s
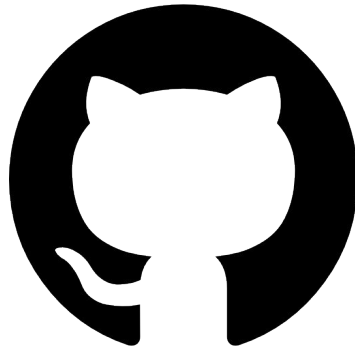
Using the Main, the Danube and the Suez Canal

Test #3

Brest --> Rio de Janeiro

computed in 250ms

14

Thanks for your attention!

github.com/fogglyorg/marinewatch