HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.

ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Hệ thống Thu thập, Lưu trữ, Xử lý và Phân tích Dữ liệu thời tiết dự báo thiên tai

Group: 9

ONE LOVE. ONE FUTURE.

Dataset

- 4 bo du lieu:

+ Historical Hourly Weather Data 2012-2017 NYC

+ Vehicle Collisions in NYC

+ Yellow Taxi NYC

+ 311 Service Requests NYC

Historical Hourly Weather Data

- Du lieu thoi tiet tinh theo gio o New York tu 2012-2017

```
--- Preview dữ liệu ---
              datetime  temperature  humidity  pressure weather_desc  \
0 2012-10-01 12:00:00        15.07      58.0    1012.0   few clouds
1 2012-10-01 13:00:00        15.07      58.0    1012.0   few clouds
2 2012-10-01 14:00:00        15.10      57.0    1012.0   few clouds
3 2012-10-01 15:00:00        15.18      57.0    1012.0   few clouds
4 2012-10-01 16:00:00        15.26      57.0    1012.0   few clouds

   wind_direction  wind_speed
0           260.0         7.0
1           260.0         7.0
2           260.0         7.0
3           260.0         7.0
4           260.0         7.0
```

Vehicle Collisions in NYC

- Du lieu cac vu tai nan tai thanh pho New York
- Cac truong thong tin quan trong: Datetime, So nguoi chet, so nguoi bi thuong, dia diem, nguyen nhan tai nan,...
- Tong so ban ghi: ~200.000 (2016)

Yellow Taxi NYC

- Du lieu taxi vang o thanh pho New York
- So ban ghi: ~(10.000.000) (2016)

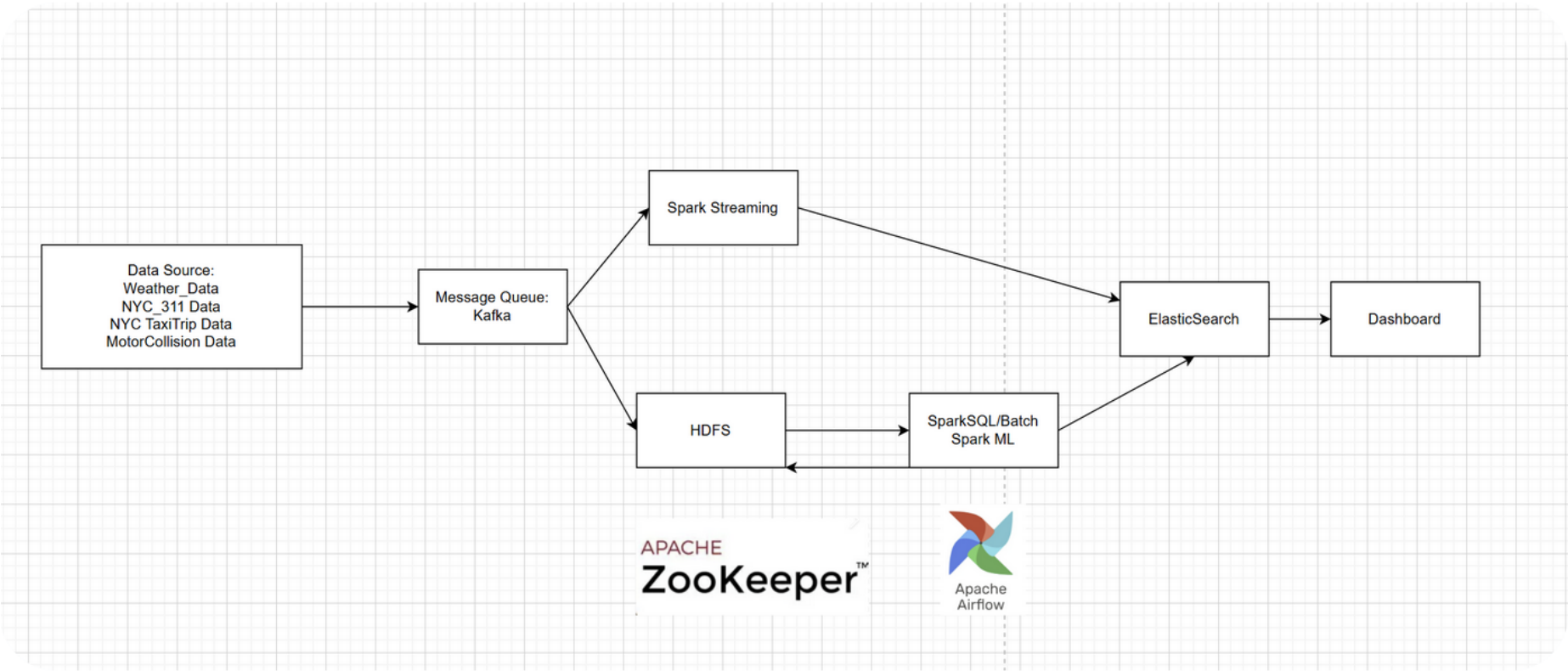
311 Request NewYork

- Du lieu phan nan cua dan cu New York gui ve
- So ban ghi (~2.200.000) (2016)

--- KẾT QUẢ KHAI PHÁ: THỜI TIẾT TÁC ĐỘNG GÌ ĐẾN KHIẾU NẠI? ---

	antecedents	consequents	support	confidence	lift
67	(Spike_DOOR, Weather_Cold)	(Spike_PLUMBING, Spike_HEAT)	0.051913	0.633333	9.658333
69	(Spike_PLUMBING, Weather_Cold)	(Spike_DOOR, Spike_HEAT)	0.051913	0.575758	8.104895
62	(Spike_DOOR, Spike_PLUMBING, Weather_Cold)	(Spike_HEAT)	0.051913	0.826087	7.752508
64	(Spike_DOOR, Spike_HEAT, Weather_Cold)	(Spike_PLUMBING)	0.051913	0.904762	6.758017
19	(Weather_Hot)	(Spike_WATER)	0.051913	0.513514	6.712355
44	(Spike_DOOR, Weather_Cold)	(Spike_HEAT)	0.057377	0.700000	6.569231
71	(Spike_HEAT, Weather_Cold)	(Spike_DOOR, Spike_PLUMBING)	0.051913	0.612903	6.409217
65	(Spike_PLUMBING, Weather_Cold, Spike_HEAT)	(Spike_DOOR)	0.051913	0.904762	6.020779
38	(Spike_PLUMBING, Weather_Cold)	(Spike_HEAT)	0.057377	0.636364	5.972028
57	(Spike_DOOR, Weather_Cold)	(Spike_PLUMBING)	0.062842	0.766667	5.726531

Pipeline trien khai



Kafka

- 4 producer, moi producer tuong ung voi 1 loai du lieu
- >=2 group consumer

Spark Streaming

SparkML

ElasticSearch

HUST

hust.edu.vn fb.com/dhbkhn

HUST

hust.edu.vn fb.com/dhbkhn

THANK YOU !