

WAZE TRAFFIC ANALYSIS

PRASHIL TUMBADE



What is Waze?

Waze is the world's largest community-based traffic and navigation app where drivers in your area share real-time traffic and road info, saving everyone time and gas money on their daily commute.

Goal

To analyze traffic notification alerts in San Francisco using waze data.

Data

- Real time traffic data in San Francisco city
- No official API for Waze
 - So data is very messy
- Collected data every hour from Waze LIVE map



Demo

"time_stamp": 1453696077.374517,

"alerts": [

```
{
  "country": "US",
  "latitude": "41.846912",
  "longitude": "-87.643182",
  "numOfThumbsUp": 19,
  "placeNearBy": null,
  "subType": "POLICE_HIDING",
  "type": "POLICE"
},
```

```
{
  "country": "US",
  "latitude": "41.756379",
  "longitude": "-87.552768",
  "numOfThumbsUp": 5,
  "placeNearBy": null,
  "subType": "",
  "type": "ACCIDENT"
},
```

```
...
{
  "country": "US",
  "latitude": "41.756379",
  "longitude": "-87.552768",
  "numOfThumbsUp": 3,
  "placeNearBy": null,
  "subType": "",
  "type": "JAM"
},
```

```
{
  "country": "US",
  "latitude": "41.824708",
  "longitude": "-87.719495",
  "numOfThumbsUp": 2,
  "placeNearBy": null,
  "subType": "HAZARD_ON_SHOULDER_CAR_STOPPED",
  "type": "HAZARD"
},
```

"jams": [

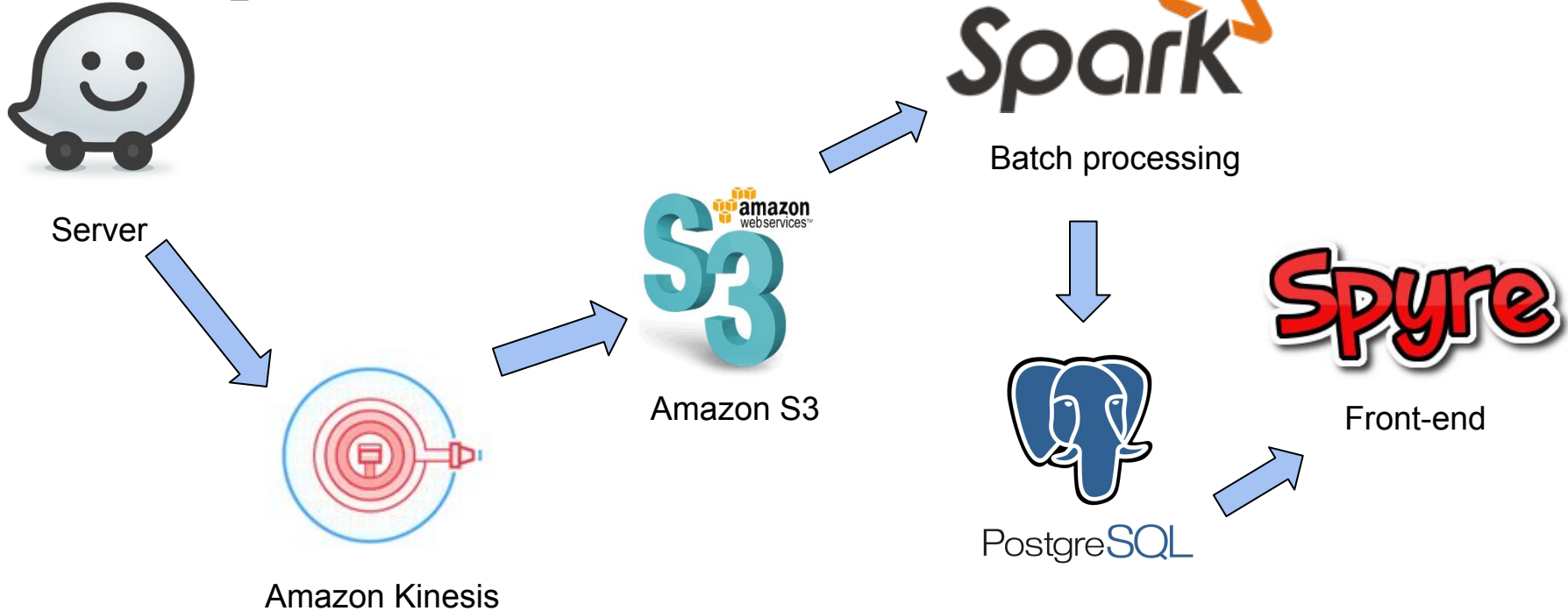
```
{
  "country": "US",
  "delayInSec": -1,
  "end": "W Congress Pkwy",
  "endLatitude": "41.875453",
  "endLongitude": "-87.642937",
  "severity": 3,
  "start": null,

```

```
{
  "country": "US",
  "latitude": "41.846912",
  "longitude": "-87.643182",
  "numOfThumbsUp": 19,
  "placeNearBy": null,
  "subType": "POLICE_HIDING",
  "type": "POLICE"
},
```

```
{
  "country": "US",
  "latitude": "41.824708",
  "longitude": "-87.719495",
  "numOfThumbsUp": 2,
  "placeNearBy": null,
  "subType": "HAZARD_ON_SHOULDER_CAR_STOPPED",
  "type": "HAZARD"
}
```

Pipeline Architecture



Cluster setup

- Server - 1 EC2 instance
- Kinesis - 1 EC2 instance
- Spark - 1 Master and 2 slaves
- Postgres - 1 RDS instance
- Spyre/Frontend - 1 EC2 instance

Challenges faced

- Data is very messy -- difficult to wrestle into a usable format.
 - Each GET Response from Waze's servers has changing nested json fields.
- Bringing nodes back online after failure.



Thank you