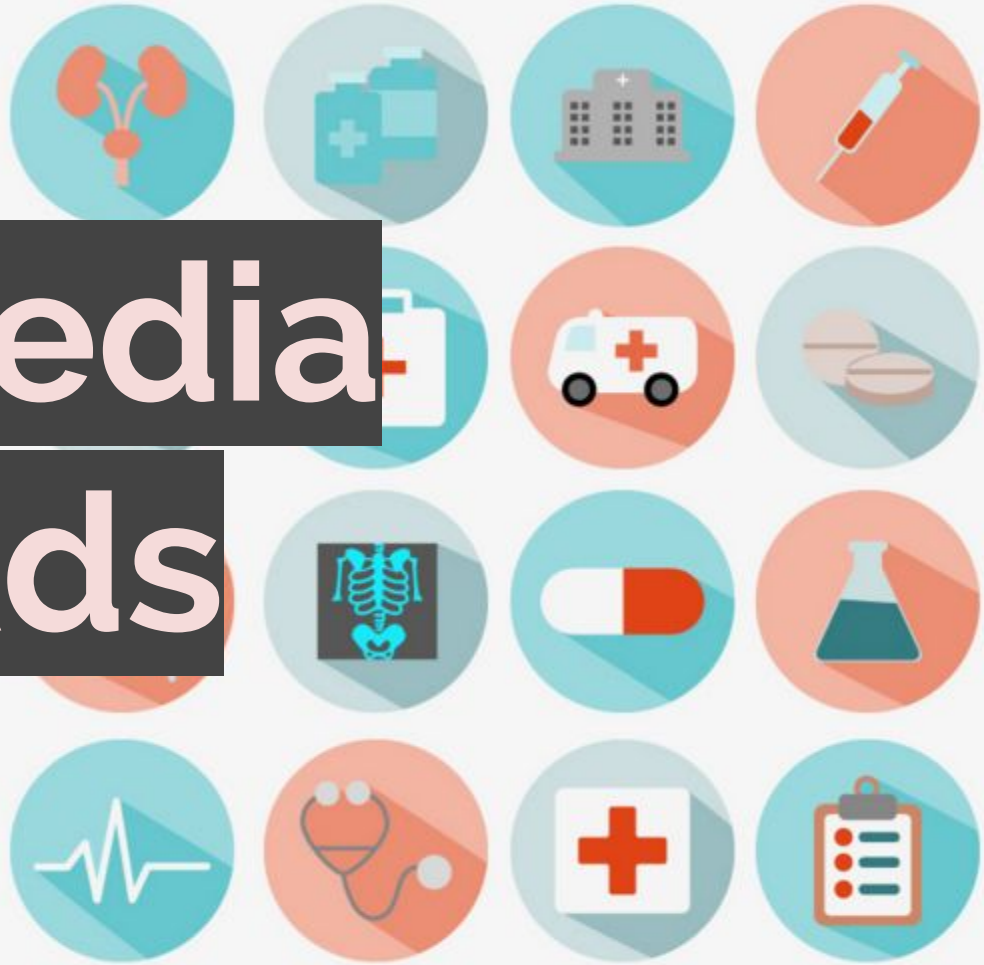
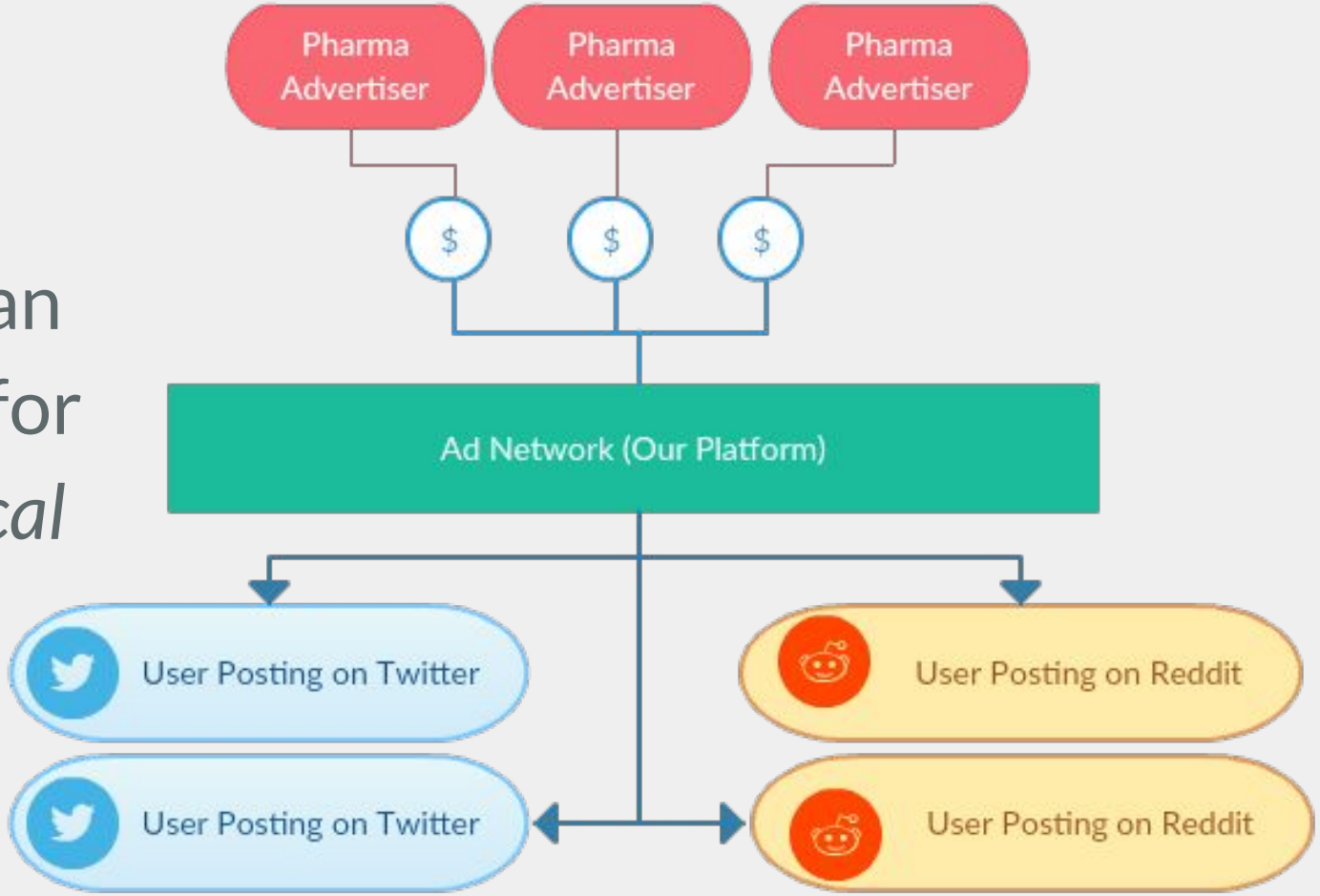
Social Media Health Ads

Joyce Chan



Goal

To simulate an **ad network** for *pharmaceutical* agencies in social media networks



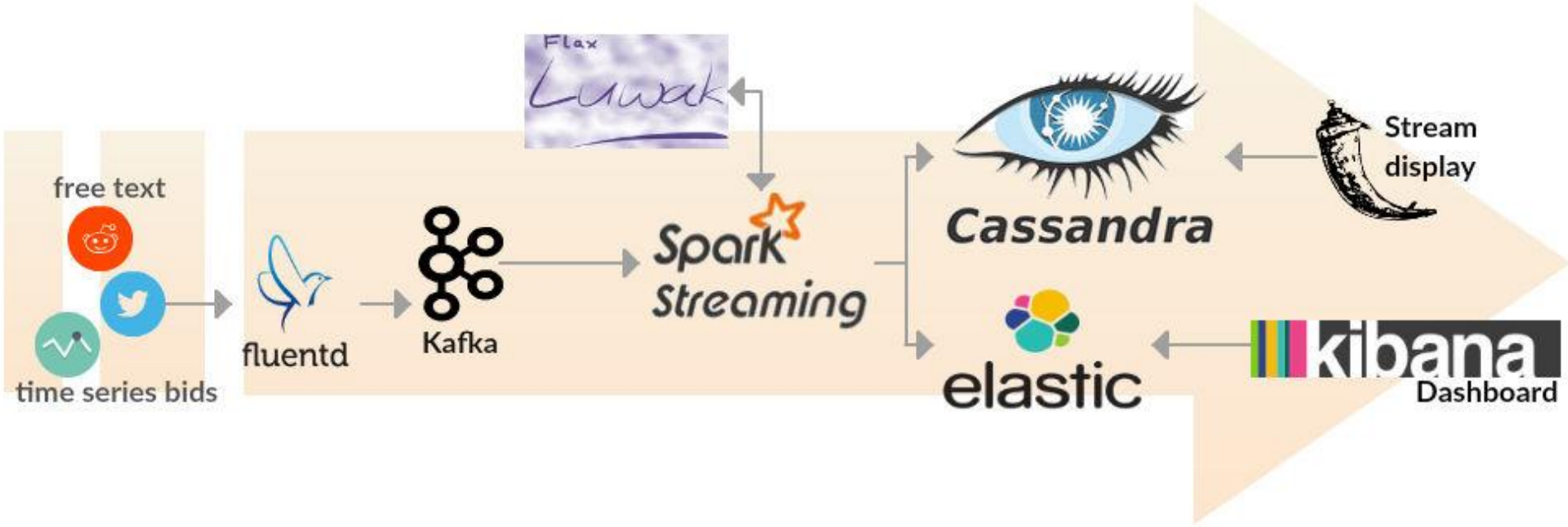
Motivations

1. **Search** on the steam efficiently
   - ES percolate vs Luwak

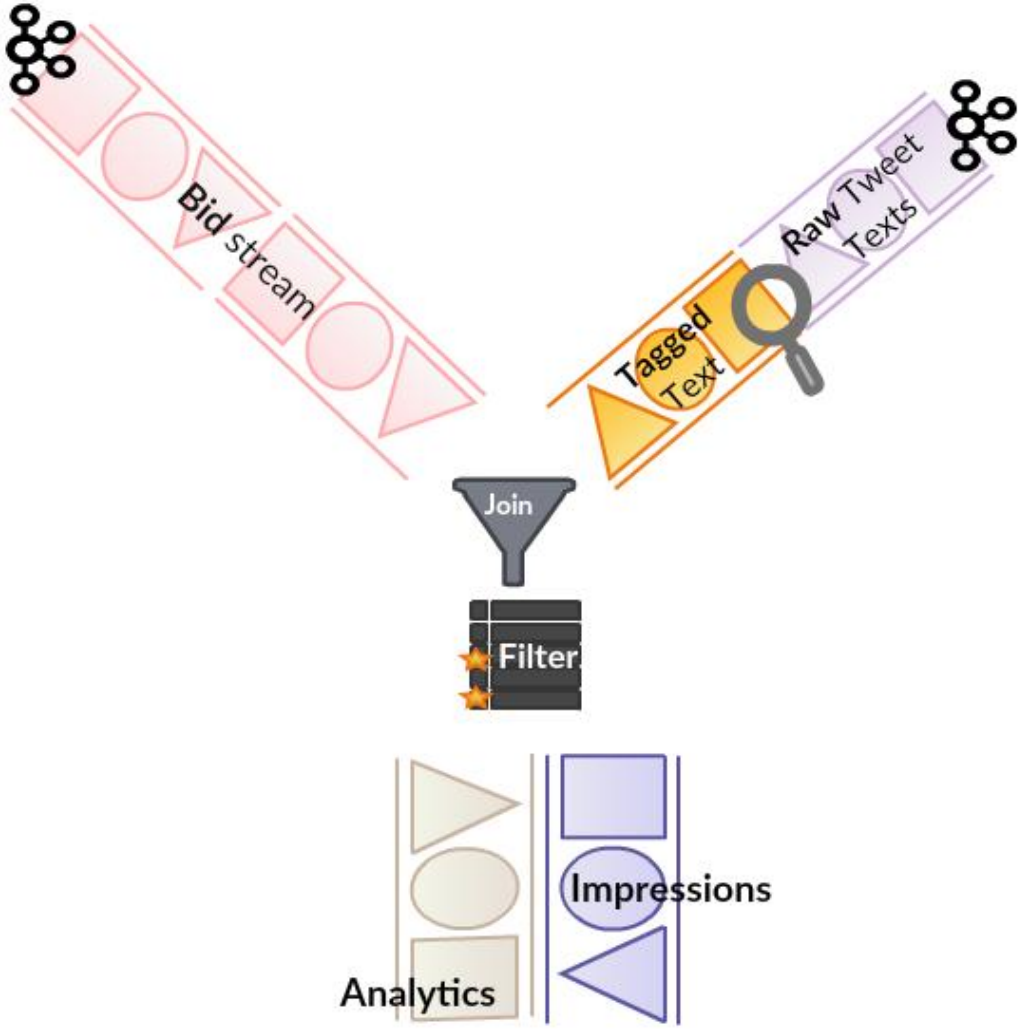
2. Learn how to **join** streams

Demo

Pipeline

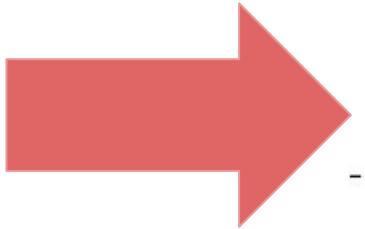

free text
time series bids
fluentd
Kafka
Flax
Luwak
Spark
Streaming
Cassandra
elastic
Stream
display
kibana
Dashboard

Spark Streaming



Luwak search


```
{
    id: 693898976458903600,
    id_str: "693898976458903552",
    body: "When I started this
    medication, I really felt very calm.
    It is like when the rain stops, you
    realize how noisy it was few seconds
    ago. But I got the ugly side effect
    of coughing which did not go away
    until 1 month after I stopped
    taking.",
    created_at: 1454273382000,
    lang: "EN",
    created_utc: 1454273382,
    user_screen_name: "YOUTUBEAR1",
    idAsStr: "693898976458903552"
}
```


```
{
    id: 693898976458903600,
    id_str: "693898976458903552",
    body: "When I started this medication, I really
    felt very calm. It is like when the rain stops,
    you realize how noisy it was few seconds ago.
    But I got the ugly side effect of coughing which
    did not go away until 1 month after I stopped
    taking.",
    created_at: 1454273382000,
    lang: "EN",
  - pharmatags: [
        "abbvie_inc",
        "eli_lilly_and_company",
        "lannett_company_inc",
        "lundbeck_inc",
        "lupin_pharmaceuticals_inc",
        "novartis_pharmaceuticals_corporation"
    ],
  - symptomtags: [
        "Coughing up blood or mucus",
        "Coughing, especially in the early morning or
        at night"
    ],
  - conditiontags: [
        "Pneumonia",
        "Malaria",
        "Asthma",
        "Pulmonary embolism",
        "Tuberculosis"
    ],
    created_utc: 1454273382,
    user_screen_name: "YOUTUBEAR1",
    idAsStr: "693898976458903552"
}
```

Join + Filter for Winning bid


```
{"pharmatags":"merck_sharp_dohme_corp","price":0.15}
{"pharmatags":"mutual_pharmaceutical_company_inc","price":0.06}
{"pharmatags":"mylan_pharmaceuticals_inc","price":0.14}
{"pharmatags":"nostrum_laboratories_inc","price":0.06}
{"pharmatags":"novartis_pharmaceuticals_corporation","price":0.13}
{"pharmatags":"otsuka_america_pharmaceuticals_inc","price":0.03}
{"pharmatags":"par_pharmaceutical","price":0.06}
{"pharmatags":"mcneil_consumer_healthcare_div_mcneilppc_inc","price":0
{"pharmatags":"msd_consumer_care_inc","price":0.09}
{"pharmatags":"mylan_institutional_inc","price":0.02}
{"pharmatags":"northstar_rxllc","price":0.15}
{"pharmatags":"novartis_consumer_health_inc","price":0.06}
{"pharmatags":"organon_usa_inc","price":0.09}
{"pharmatags":"pack_pharmaceuticals_llc","price":0.06}
{"pharmatags":"parkedavis_div_of_pfizer_inc","price":0.07}
```

```
{
    id: 6949537138955182OO,
    pharmatag:
    "novartis_pharmaceuticals_corporati
  - symptomtags: [
        "Coughing up blood or mucus",
        "Coughing, especially in the ea
        morning or at night"
    ],
  - conditiontags: [
        "Pneumonia",
        "Malaria",
        "Asthma",
        "Pulmonary embolism",
        "Tuberculosis"
    ],
    price: "0.03",
    created_utc: 1454273382
}
```

Challenges

- for stream joins, while stream mismatch can be avoided with more thoughtful tuning of batch duration, it's not ***handled well by spark streaming***

- what's the best way to join tables - order by most popular bidder

Joyce Chan

University of Toronto

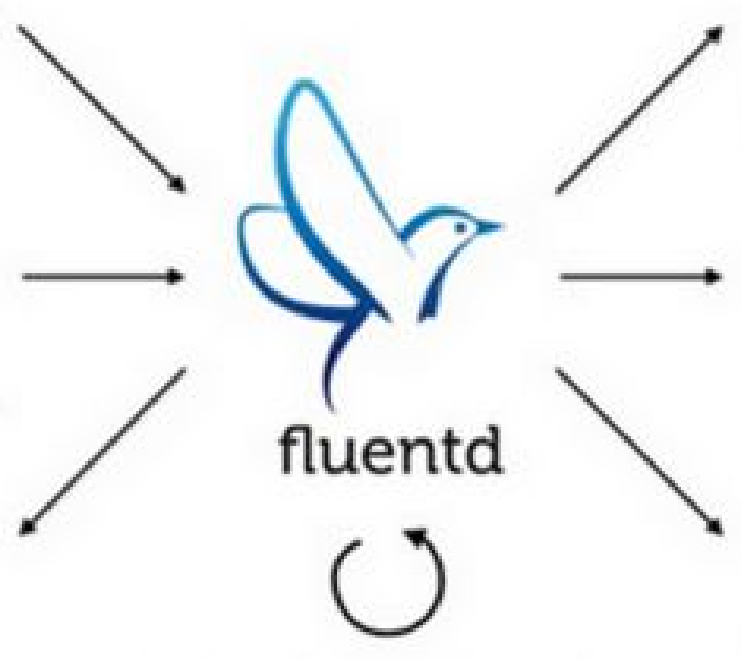
B.Sc Computer Science & Math

- analytics and data engineer at health care startup
- full stack engineer for Sysomos
- full stack engineer for Granata Decisions Systems
- search & api engineer Mercatus

Other Slides

Ingestion

1. fluentd tails files downloaded from S3 & filters reddit by desired subreddit
2. fluentd acts as twitter collector, filter by desired languages
3. fluentd writes to kafka topics



Social Media Health Ads

http://healthads.instapage.com

https://github.com/holajoyce/twitter.rx

Joyce Chan

http://linkedin.com/in/joyschan