# Bot and Anonymous Wikipedia Contributions

# Wikipedia Recent Changes Stream

https://stream.wikimedia.org/v2/stream/recentchange



src: https://meta.wikimedia.org/wiki/Help:Recent_changes

# Stream items

```
{ ⊟
    "bot":false,
    "comment":"",
    "id":1032731839,
    "length":{ ⊟
        "new":16642,
        "old":16642
    },
    "meta":{ ⊟
        "domain":"en.wikipedia.org",
        "dt":"2018-02-27T23:05:42+00:00",
        "id":"bc8de3e7-1c12-11e8-911f-1866da994faa",
        "request_id":"aa861e55-0cc6-4053-8243-5ef8b330a3c9",
        "schema_uri":"mediawiki/recentchange/1",
        "topic":"eqiad.mediawiki.recentchange",
        "uri":"https://en.wikipedia.org/wiki/Glenn_Hughes",
        "partition":0,
        "offset":741265433
    },
    "minor":false,
    "namespace":0,
    "parsedcomment":"",
    "revision":{ ⊟
        "new":827995515,
        "old":826644393
    },
    "server_name":"en.wikipedia.org",
    "server_script_path":"/w",
    "server_url":"https://en.wikipedia.org",
    "timestamp":1519772742,
    "title":"Glenn Hughes",
    "type":"edit",
    "user":"177.43.85.218",
    "wiki":"enwiki"
}
```

# Stream items

- Bot or not

```
{
  "bot":false,
  "comment":"",
  "id":1032731839,
  "length":{
    "new":16642,
    "old":16642
  },
  "meta":{
    "domain":"en.wikipedia.org",
    "dt":"2018-02-27T23:05:42+00:00",
    "id":"bc8de3e7-1c12-11e8-911f-1866da994faa",
    "request_id":"aa861e55-0cc6-4053-8243-5ef8b330a3c9",
    "schema_uri":"mediawiki/recentchange/1",
    "topic":"eqiad.mediawiki.recentchange",
    "uri":"https://en.wikipedia.org/wiki/Glenn_Hughes",
    "partition":0,
    "offset":741265433
  },
  "minor":false,
  "namespace":0,
  "parsedcomment":"",
  "revision":{
    "new":827995515,
    "old":826644393
  },
  "server_name":"en.wikipedia.org",
  "server_script_path":"/w",
  "server_url":"https://en.wikipedia.org",
  "timestamp":1519772742,
  "title":"Glenn Hughes",
  "type":"edit",
  "user":"177.43.85.218",
  "wiki":"enwiki"
}
```

# Stream items

- Bot or not
- Magnitude of change

```
{
  "bot":false,
  "comment":"",
  "id":1032731839,
  "length":{
    "new":16642,
    "old":16642
  },
  "meta":{
    "domain":"en.wikipedia.org",
    "dt":"2018-02-27T23:05:42+00:00",
    "id":"bc8de3e7-1c12-11e8-911f-1866da994faa",
    "request_id":"aa861e55-0cc6-4053-8243-5ef8b330a3c9",
    "schema_uri":"mediawiki/recentchange/1",
    "topic":"eqiad.mediawiki.recentchange",
    "uri":"https://en.wikipedia.org/wiki/Glenn_Hughes",
    "partition":0,
    "offset":741265433
  },
  "minor":false,
  "namespace":0,
  "parsedcomment":"",
  "revision":{
    "new":827995515,
    "old":826644393
  },
  "server_name":"en.wikipedia.org",
  "server_script_path":"/w",
  "server_url":"https://en.wikipedia.org",
  "timestamp":1519772742,
  "title":"Glenn Hughes",
  "type":"edit",
  "user":"177.43.85.218",
  "wiki":"enwiki"
}
```

# Stream items

- Bot or not
- Magnitude of change
- Wikipedia Page URI

```
{
  "bot":false,
  "comment":"",
  "id":1032731839,
  "length":{
    "new":16642,
    "old":16642
  },
  "meta":{
    "domain":"en.wikipedia.org",
    "dt":"2018-02-27T23:05:42+00:00",
    "id":"bc8de3e7-1c12-11e8-911f-1866da994faa",
    "request_id":"aa861e55-0cc6-4053-8243-5ef8b330a3c9",
    "schema_uri":"mediawiki/recentchange/1",
    "topic":"eqiad.mediawiki.recentchange",
    "uri":"https://en.wikipedia.org/wiki/Glenn_Hughes",
    "partition":0,
    "offset":741265433
  },
  "minor":false,
  "namespace":0,
  "parsedcomment":"",
  "revision":{
    "new":827995515,
    "old":826644393
  },
  "server_name":"en.wikipedia.org",
  "server_script_path":"/w",
  "server_url":"https://en.wikipedia.org",
  "timestamp":1519772742,
  "title":"Glenn Hughes",
  "type":"edit",
  "user":"177.43.85.218",
  "wiki":"enwiki"
}
```

# Stream items

- Bot or not
- Magnitude of change
- Wikipedia Page URI
- IDs for revision

```
{
  "bot":false,
  "comment":"",
  "id":1032731839,
  "length":{
    "new":16642,
    "old":16642
  },
  "meta":{
    "domain":"en.wikipedia.org",
    "dt":"2018-02-27T23:05:42+00:00",
    "id":"bc8de3e7-1c12-11e8-911f-1866da994faa",
    "request_id":"aa861e55-0cc6-4053-8243-5ef8b330a3c9",
    "schema_uri":"mediawiki/recentchange/1",
    "topic":"eqiad.mediawiki.recentchange",
    "uri":"https://en.wikipedia.org/wiki/Glenn_Hughes",
    "partition":0,
    "offset":741265433
  },
  "minor":false,
  "namespace":0,
  "parsedcomment":"",
  "revision":{
    "new":827995515,
    "old":826644393
  },
  "server_name":"en.wikipedia.org",
  "server_script_path":"/w",
  "server_url":"https://en.wikipedia.org",
  "timestamp":1519772742,
  "title":"Glenn Hughes",
  "type":"edit",
  "user":"177.43.85.218",
  "wiki":"enwiki"
}
```

# Stream items

- Bot or not
- Magnitude of change
- Wikipedia Page URI
- IDs for revision
- Wiki domain

```
{
  "bot":false,
  "comment":"",
  "id":1032731839,
  "length":{
    "new":16642,
    "old":16642
  },
  "meta":{
    "domain":"en.wikipedia.org",
    "dt":"2018-02-27T23:05:42+00:00",
    "id":"bc8de3e7-1c12-11e8-911f-1866da994faa",
    "request_id":"aa861e55-0cc6-4053-8243-5ef8b330a3c9",
    "schema_uri":"mediawiki/recentchange/1",
    "topic":"eqiad.mediawiki.recentchange",
    "uri":"https://en.wikipedia.org/wiki/Glenn_Hughes",
    "partition":0,
    "offset":741265433
  },
  "minor":false,
  "namespace":0,
  "parsedcomment":"",
  "revision":{
    "new":827995515,
    "old":826644393
  },
  "server_name":"en.wikipedia.org",
  "server_script_path":"/w",
  "server_url":"https://en.wikipedia.org",
  "timestamp":1519772742,
  "title":"Glenn Hughes",
  "type":"edit",
  "user":"177.43.85.218",
  "wiki":"enwiki"
}
```

# Stream items

- Bot or not
- Magnitude of change
- Wikipedia Page URI
- IDs for revision
- Wiki domain
- User

```
{
  "bot":false,
  "comment":"",
  "id":1032731839,
  "length":{
    "new":16642,
    "old":16642
  },
  "meta":{
    "domain":"en.wikipedia.org",
    "dt":"2018-02-27T23:05:42+00:00",
    "id":"bc8de3e7-1c12-11e8-911f-1866da994faa",
    "request_id":"aa861e55-0cc6-4053-8243-5ef8b330a3c9",
    "schema_uri":"mediawiki/recentchange/1",
    "topic":"eqiad.mediawiki.recentchange",
    "uri":"https://en.wikipedia.org/wiki/Glenn_Hughes",
    "partition":0,
    "offset":741265433
  },
  "minor":false,
  "namespace":0,
  "parsedcomment":"",
  "revision":{
    "new":827995515,
    "old":826644393
  },
  "server_name":"en.wikipedia.org",
  "server_script_path":"/w",
  "server_url":"https://en.wikipedia.org",
  "timestamp":1519772742,
  "title":"Glenn Hughes",
  "type":"edit",
  "user":"177.43.85.218",
  "wiki":"enwiki"
}
```

# Strange Behavior (among non-bots)

**NonBot**

| USER | COUNT |
|---|---|
| Queryzo | 503 |
| MisterSynergy | 38 |
| Danrok | 19 |
| Vinkje83 | 15 |
| Laukatu | 13 |
| PKM | 11 |
| Adert | 9 |
| Димон мп | 7 |
| Asqueladd | 7 |
| Bamyers99 | 6 |
| Blythwood | 5 |
| 80.12.43.156 | 5 |
| Simon Villeneuve | 5 |
| Nono314 | 5 |

# Strange Behavior (among non-bots)

**NonBot**

| USER | COUNT |
|------|-------|
| Queryzo | 503 |
| MisterSynergy | 38 |
| Danrok | 19 |
| Vinkje83 | 15 |
| Laukatu | 13 |
| PKM | 11 |
| Adert | 9 |
| Димон мп | 7 |
| Asqueladd | 7 |
| Bamyers99 | 6 |
| Blythwood | 5 |
| 80.12.43.156 | 5 |
| Simon Villeneuve | 5 |
| Nono314 | 5 |

Queryzo 503
MisterSynergy 38

# Strange Behavior

Reverting possible vandalism by 76.233.6.154

# Strange Behavior



**1828 in the United States: Difference between revisions**

From Wikipedia, the free encyclopedia

| Browse history | ⌄ |
|---|---|

**Revision as of 23:08, 27 February 2018 (edit)**
76.233.6.154 (talk)
(←Replaced content with ' hi')
(Tags: blanking, Replaced)
← Previous edit

**Latest revision as of 23:08, 27 February 2018 (edit) (undo)**
CitueBot NG (talk | contribs)
m (Reverting possible vandalism by 76.233.6.154 to version by 47.21.17.246.
Report False Positive? Thanks, ClueBot NG. (3300421) (Bot))
(Tag: Rollback)

**Line 1:**

− hi

**Line 1:**

+ {{Yearbox US|1828}}

+ Events from the year '''1828 in the United States'''.

+

+ == Incumbents ==

+ === [[Federal government of the United States|Federal Government]] ===

+ * [[President of the United States|President]]: [[John Quincy Adams]] ([[Democratic-Republican Party|DR/NR]]-[[Massachusetts]])

+ * [[Vice President of the United States|Vice President]]: [[John C. Calhoun]] ([[Democratic Party (United States)|D]]-[[South Carolina]])

# Objectives

1.  Create aggregate windowed counts of non-bot users and send to a kafka stream for notifications downstream

2.  Send anonymous changes for English Wikipedia to their own Kafka stream so that the contents of the changes can be scheduled for collection and the enriched data can be stored in HDFS for later analysis

# Approach...

1.  Ingest data

2.  Filter out the fields we want from raw json

3.  Clean and enhance data:

    a.   Convert Unix time to datetime (for window aggregations)

    b.   Identify whether anonymous or not

4.  Run window of counts per user

5.  Write data to Kafka topics for downstream notification, analysis, and data enrichment by other processes.
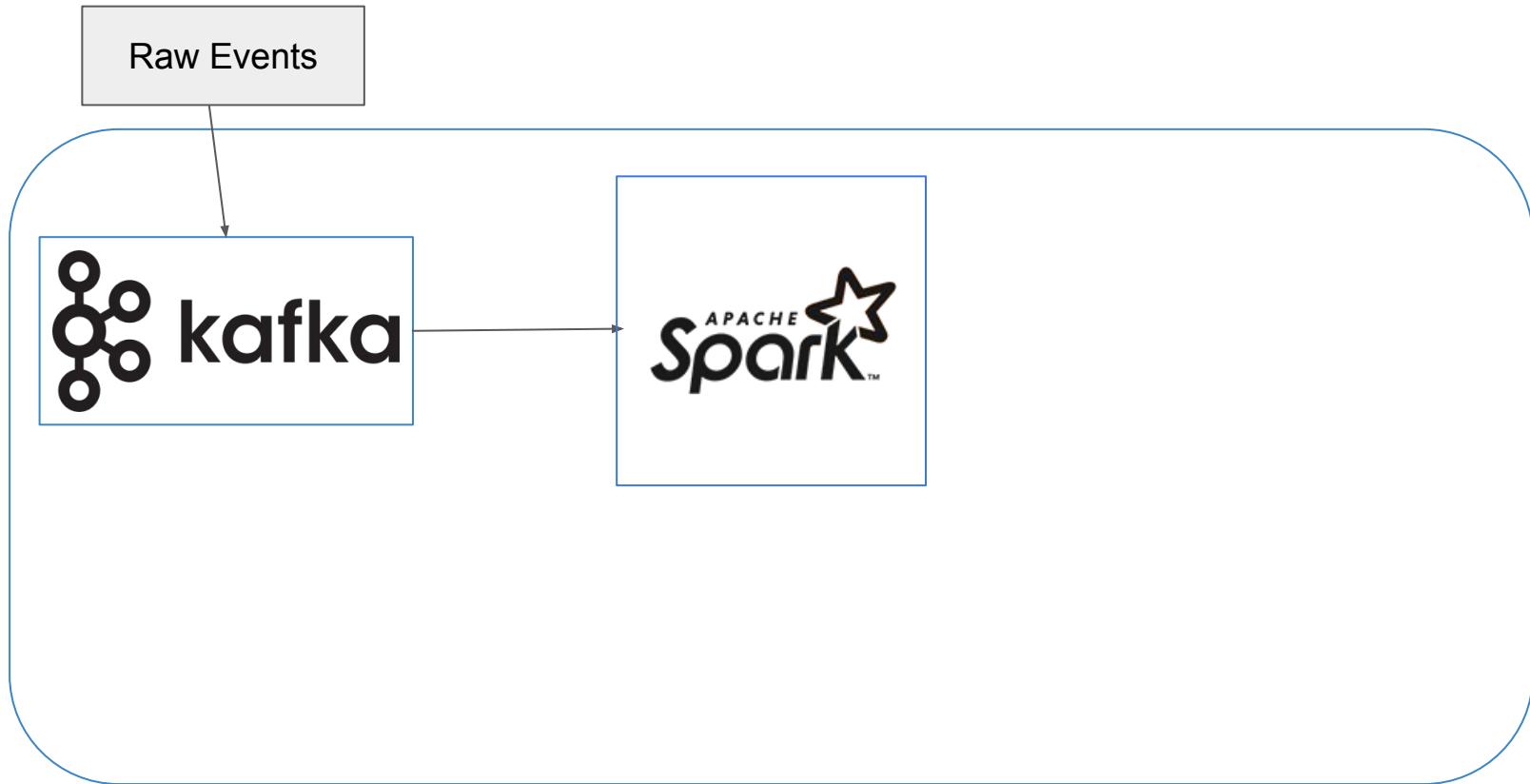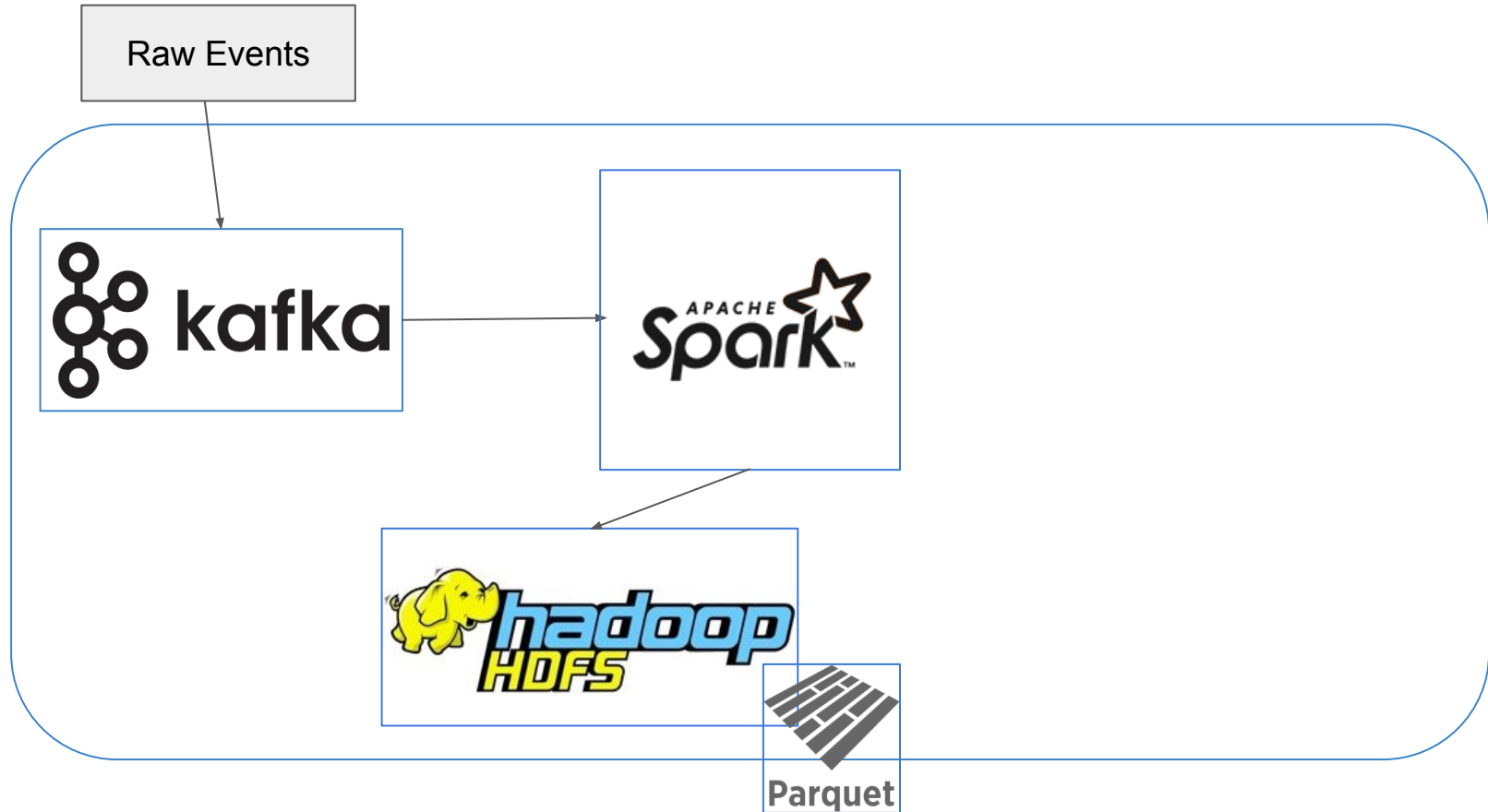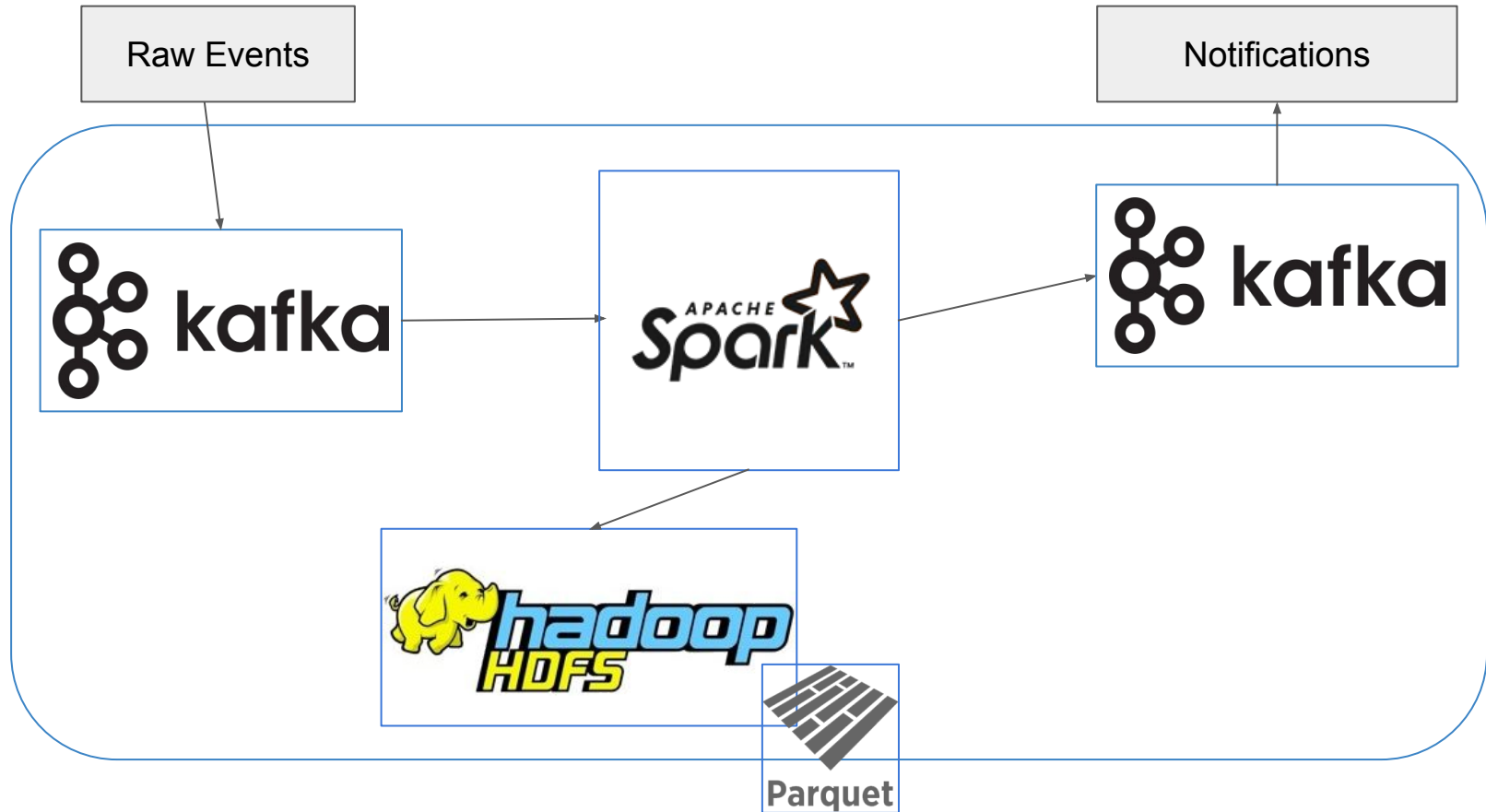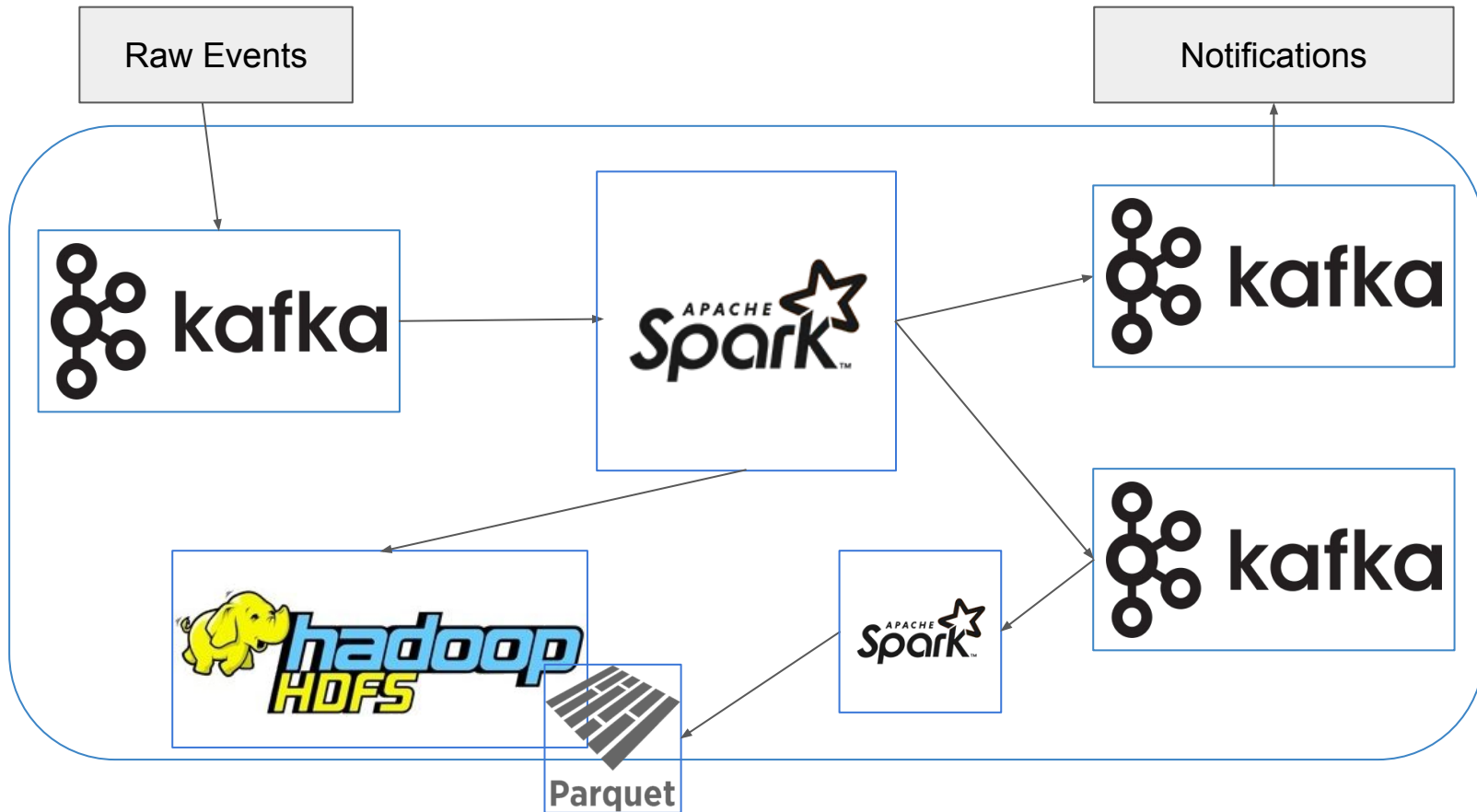
# Approach...

Raw Events

# Approach...

# Approach...

Raw Events



Parquet

# Approach...

# Approach...

# Conclusion

## Comments

- Done in notebooks, not in a more production-like fashion
- Is a bit contrived, reduce steps
- Collecting the raw text for anonymous events might get behind stream, how to address

## Take it further

- Set an retention policy on Kafka topics based on their use
- Better way to handle events that have been dealt with
- Better concept for persistence (database rather than HDFS flat file storage for querying and fault tolerance)