Event Stream Processing

UNIVERSITY OF
COPENHAGEN

Casper Lisager Frandsen

January 22, 2020

Outline

What is Event Stream Processing?

- The basics
- Comparison to conventional databases

The Assginment

- Number of tagged users

What is Event Stream Processing?

The basics

- Unbounded amounts of realtime data
- Continuous evaluation of queries
- In and out of order
- Windows

Comparison to conventional databases

- Queries
  - Continuous
  - One-time
- Scalability for unbounded inputs
- Precision
- Storage

Number of tagged users

- Filter for the ones we want
  - ”Man-in-the-middle” streams
- Normal SQL Query

```
CREATE TABLE close_number_tags AS
        SELECT
                t.user_id AS user_id,
                count(t.user_id) AS tagCount
        FROM taggedUsers t
                INNER JOIN GPS g
                WITHIN 15 SECONDS
                ON g.user_id = t.user_id
                WINDOW HOPPING (SIZE 15 SECONDS, ADVANCE BY 5 SECONDS)
                WHERE ABS(t.lat - g.lat) <= 5.0
                AND ABS(t.long - g.long) <= 5.0
        GROUP BY t.user_id;
```

Our query


```
-- Counts users that are far away from where picture was taken when tagged.
CREATE TABLE far_tag_count AS
        SELECT
                T.photo_id,
                COUNT(T.USER_ID)
        FROM TAGGEDUSERS T
                INNER JOIN GPS G
                WITHIN 1 HOURS
                ON G.USER_ID = T.USER_ID
                WINDOW HOPPING ( SIZE 15 SECONDS , ADVANCE BY 5 SECONDS )
                WHERE ABS(T.LAT - G.LAT) >= 5.0
                AND ABS(T.LONG - G.LONG) >= 5.0
        GROUP BY T.photo_id;
```