TUM I 13
Chair of Application and Middleware Systems
Prof. Dr. H.- A. Jacobsen
Author: Doblander

Tutorial
Event Processing
SS 2020

Issued on: May 2020

Course Project

All exercises can be implemented in your prefered event processing system. We recommend either Apache Flink or Kafka + KSQL. Summarize the findings in a report, around 0.75 pages per teammember. Maximum teamsize 3. Use the ACM Master Article Template using the ACM SIGPLAN style (two-column). Needless to say use good scientific practice which means no plagirism and so on. In the final lecture slot your team should present the outcomes.

You can put snippets of your code, queries, drawings of topologies, visualizations of results and so on in your report. Discuss the guarantees, how it can be distributed and what happens in case of faults. Try to measure latency and throughput and put it into a graph and discuss the outcomes. Also try to get the memory consumption of the operator. Either use the integrated methods of the systems or add timestamp to the event which denotes when it was sent.

Exercise 1 and 2 are mandatory, Exercise 3 only if your team has 2-3 members, each member should implement a query.

Submission: A single ZIP-File containing the code and the report (pdf) and presentation (pdf).

Exercise 1 Send Json Messages

Each message is in a single line. You can use a simple lineIterator such as available in Apache Commons to read the file line per line. Of course you can use Python or any other programming language. Some systems may provide integrated methods.

```
LineIterator it = FileUtils.lineIterator(file, "UTF-8");
 try {
   while (it.hasNext()) {
     publishMessage(it.nextLine()); //You can insert System.nanoTime() in the message to
         measure the time a message takes through the system
     Thread.sleep(100); //depends on your system
   }
 } finally {
   it.close();
 }
```

Example event from the stream, indented for readability.

```
{"utc_offset":3600000,
 "venue":{"country":"de",
        "city":"Munchen",
        "address_1":"Tal 8",
        "name":"Tegernseer Tal - Brauhaus",
        "lon":11.578209,"lat":48.13599},
 "rsvp_limit":0,
 "venue_visibility":"public",
 "visibility":"public",
```

```
"maybe_rsvp_count":0,
"description":"<p>Hi everyone,<\/p>\n<p>Lets get .....snip....",
"mtime":1479116040446,
"event_url":"https:\/\/www.meetup.com\/The-Munich-iOS-Developers-Meetup\/events\/235150101\/",
"yes_rsvp_count":21,
"payment_required":"0",
"name":"Social drinks meetup",
"id":"235150101",
"time":1479232800000,
"group":{"join_mode":"open",
       "country":"de",
       "city":"Munchen",
       "name":"The Munich iOS Developers Meetup",
       "group_lon":11.58,
       "id":10979402,
       "urlname":"The-Munich-iOS-Developers-Meetup",
       "category":{"name":"tech","id":34,"shortname":"tech"},
                "group_photo":{"highres_link":"http:\/\/photos1.meetupstatic.com\/photos\/event\/8\/0\/0\/a\/highres_438692778.jpeg",
                "photo_link":"http:\/\/photos1.meetupstatic.com\/photos\/event\/8\/0\/0\/a\/600_438692778.jpeg",
                "photo_id":438692778,
                "thumb_link":"http:\/\/photos1.meetupstatic.com\/photos\/event\/8\/0\/0\/a\/thumb_438692778.jpeg"},
                "group_lat":48.14},
                "status":"upcoming"}
```


Exercise 2 Query on top of Stream

Create a topology of two event processors. The first one filters the output stream down to Germany. The second further filters down the output to just Munich.

Below is an example for KSQL. You need to define the schema before being able to express a query on top.

```
CREATE STREAM meetup_events_stream(
 utc_offset BIGINT,
 venue STRUCT<country VARCHAR,
           city VARCHAR,
           ...>
 ...) WITH (KAFKA_TOPIC='meetup-events', VALUE_FORMAT='JSON');
```

Exercise 3 Optional queries

- Heatmap - Split Munich into 1km x 1km squares. Where are the Meetup Hubs?
- Visualization - Show the heatmap on top of a map.
- TopK - What are the Top 10 Meetup cities in Europe
- Your own query - Should show some interesting insight, e.g., Android vs. iOS who has more meetups?