InfluxDB on Apache Flink

Open Source Data Processing – Data Engineering Systems Group

Ramin Gharib, Felix Seidel, and Leon Papke

Winter Semester 2020/2021

03.03.2021

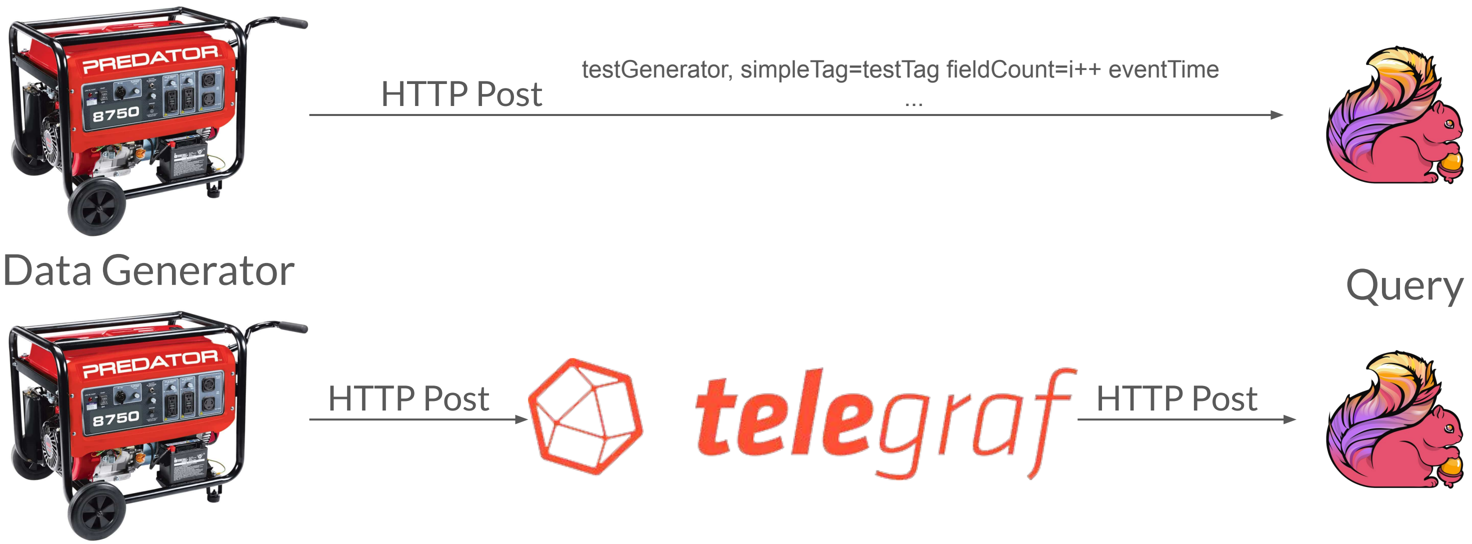
10 20 30 50 70 80 100 130 150
0 10 20 30 40 50 60 70 80 90 100
92259
1441
R N DRIVE L

Benchmarks

Experiment Design

- 4-node NUMA machine: 9 x 1GHz CPU (+9 virtual) per node
  - Execution on one node via *numactl*
- JVM heap limit: 10GB
- Flink settings:
  - Parallelism = 1
  - Object reuse enabled
  - No watermarking & checkpointing

Source Benchmarks



Throughput + Latency:

Data Generator

HTTP Post

testGenerator, simpleTag=testTag fieldCount=i++ eventTime

...

Query

Latency:

Data Generator

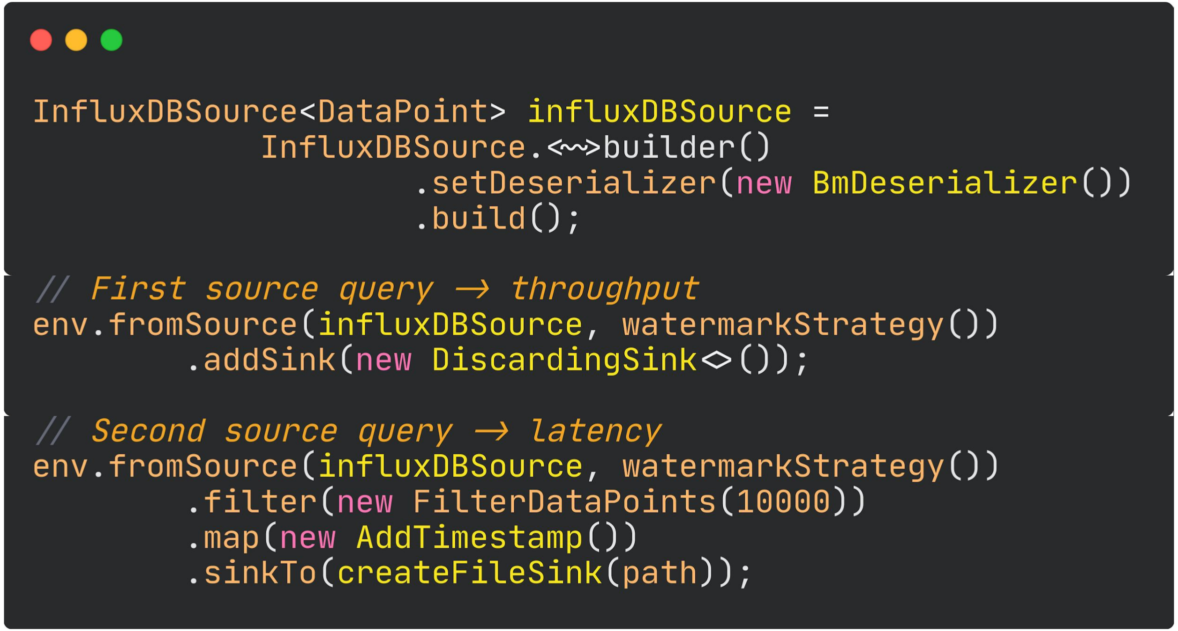
HTTP Post

telegraf

HTTP Post

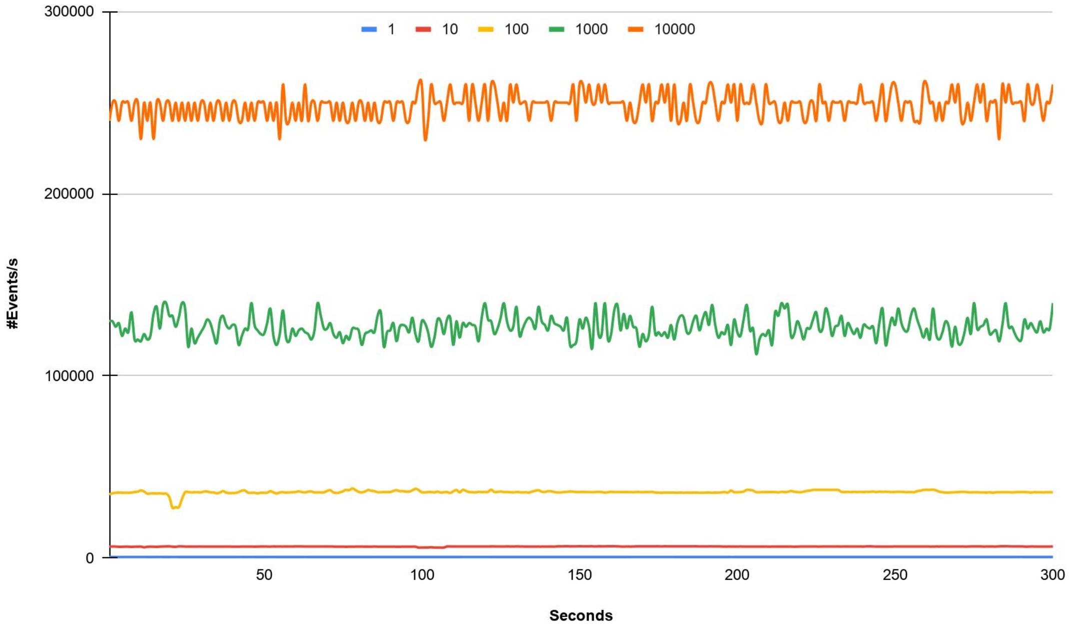
Query

Source Queries

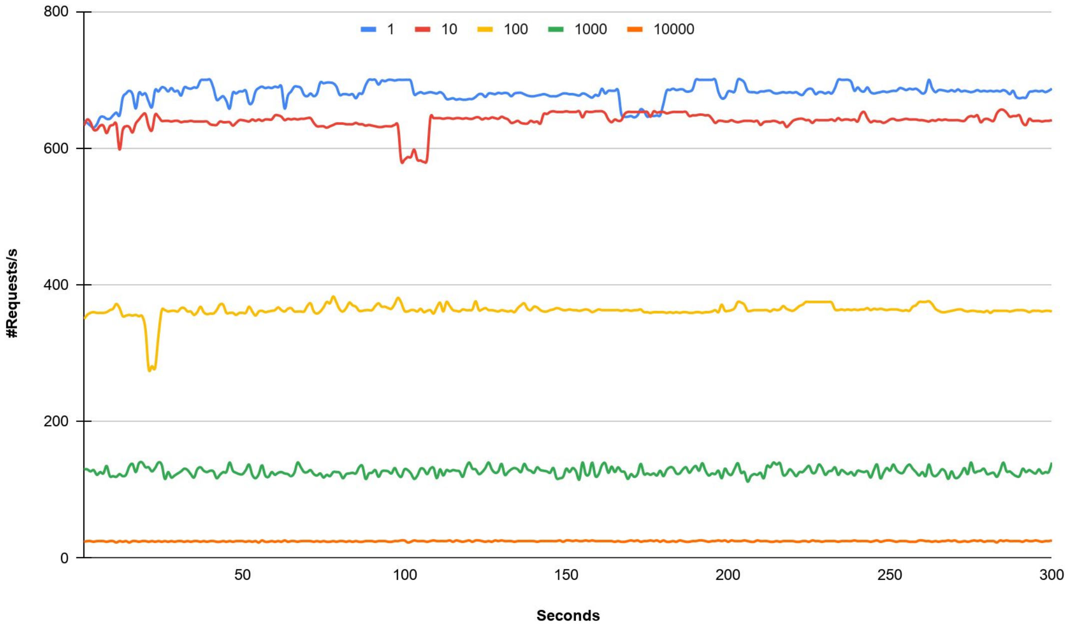


```
InfluxDBSource<DataPoint> influxDBSource =
            InfluxDBSource.<~~>builder()
                    .setDeserializer(new BmDeserializer())
                    .build();

// First source query -> throughput
env.fromSource(influxDBSource, watermarkStrategy())
        .addSink(new DiscardingSink<>());

// Second source query -> latency
env.fromSource(influxDBSource, watermarkStrategy())
        .filter(new FilterDataPoints(10000))
        .map(new AddTimestamp())
        .sinkTo(createFileSink(path));
```

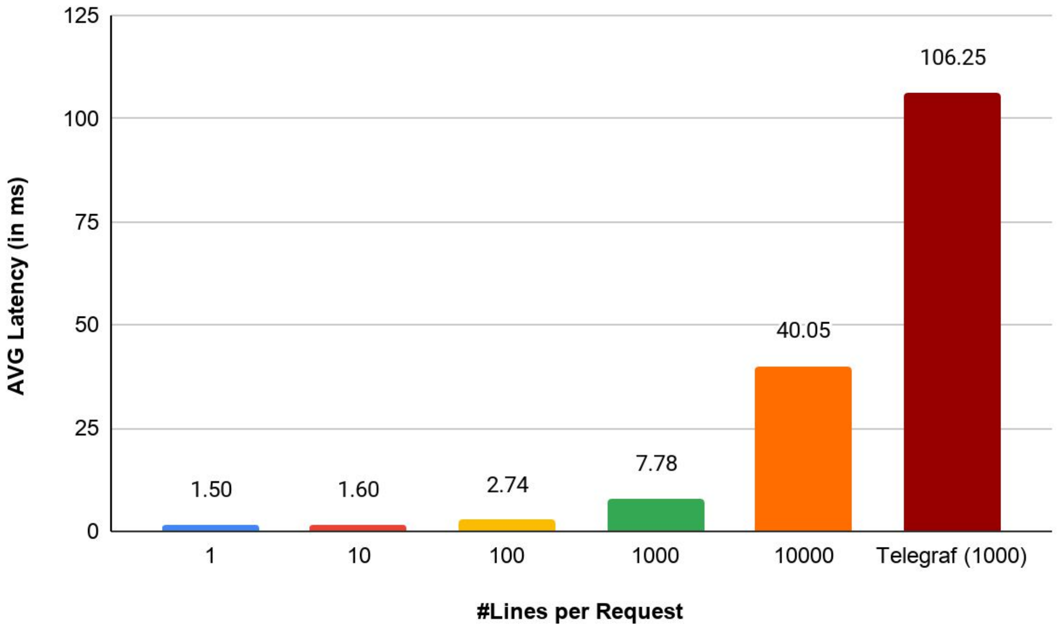
Event Throughput



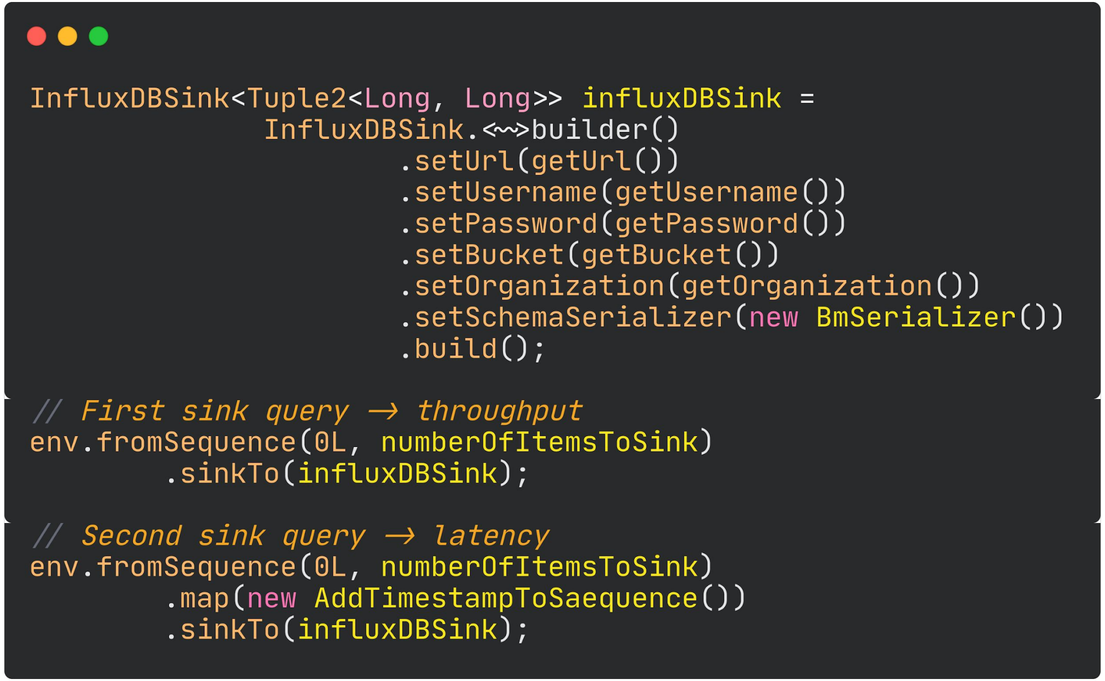
Request Throughput



Event Latency

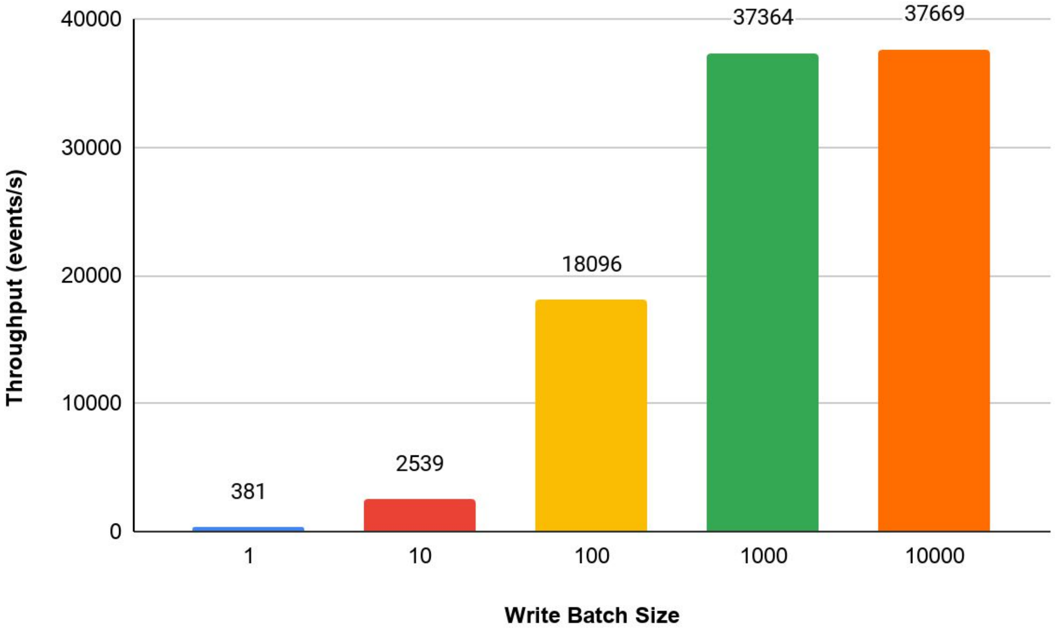


Sink Queries

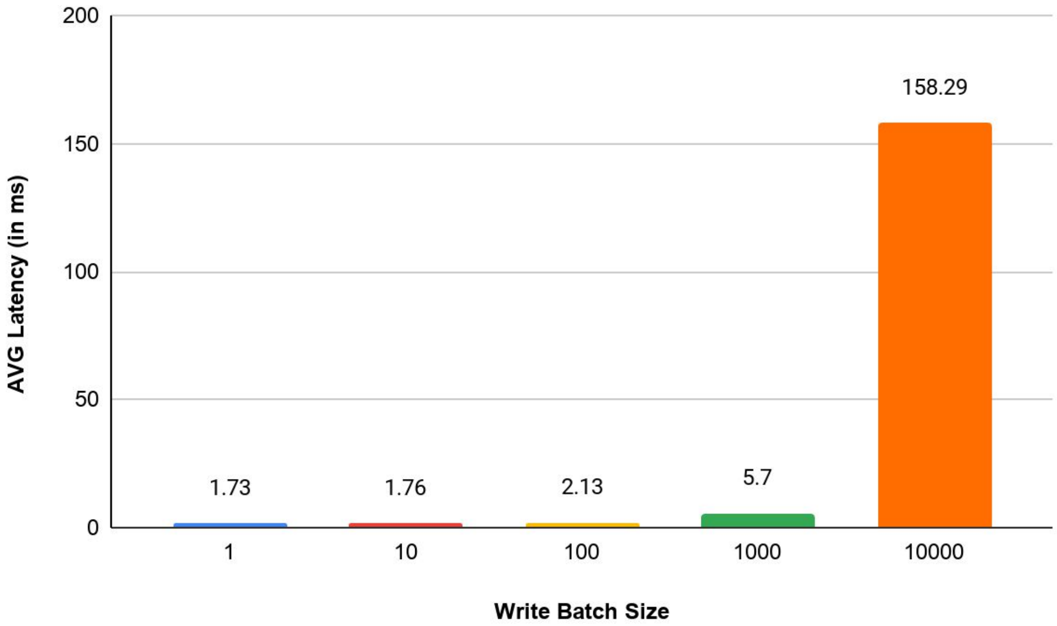


```
InfluxDBSink<Tuple2<Long, Long>> influxDBSink =
              InfluxDBSink.<~>builder()
                      .setUrl(getUrl())
                      .setUsername(getUsername())
                      .setPassword(getPassword())
                      .setBucket(getBucket())
                      .setOrganization(getOrganization())
                      .setSchemaSerializer(new BmSerializer())
                      .build();

// First sink query → throughput
env.fromSequence(0L, numberOfItemsToSink)
        .sinkTo(influxDBSink);

// Second sink query → latency
env.fromSequence(0L, numberOfItemsToSink)
        .map(new AddTimestampToSaequence())
        .sinkTo(influxDBSink);
```

Event Throughput



Event Latency


AVG Latency (in ms)
Write Batch Size