Table of Contents

[Introduction 2](#_Toc4011789)

[Requirements 2](#_Toc4011790)

[Architecture 2](#_Toc4011791)

[Assumptions 2](#_Toc4011792)

[Flow 2](#_Toc4011793)

[Technologies 2](#_Toc4011794)

[Design 3](#_Toc4011795)

[Environment 3](#_Toc4011796)

[Partitioning 3](#_Toc4011797)

[Extendibility 3](#_Toc4011798)

[Performance 3](#_Toc4011799)

[Templates 3](#_Toc4011800)

[DTO 3](#_Toc4011801)

[Schema 4](#_Toc4011802)

[API 4](#_Toc4011803)

[Folder Hierarchy 4](#_Toc4011804)

[Usage 5](#_Toc4011805)

[Input 5](#_Toc4011806)

[Console Output 6](#_Toc4011807)

[File System Output 6](#_Toc4011808)

# Introduction

The main purpose of this application is to develop a general solution for high performance writing in parallel to multiple parquet files, where partitions are based on a specific time-based interval. The application should be flexible to support any schema provided by the user.

# Requirements

* Convenient to use from java
* Generic enough so adding support for yet another schema (DTO) is easy
* Microbenchmark

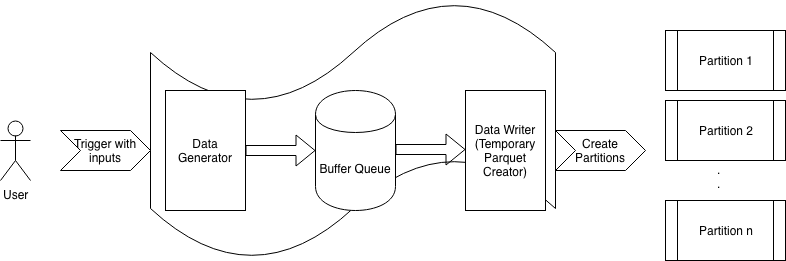
# Architecture

## Assumptions

* Schema must be including timestamp field as the first field
* Only primitive types will be used

## Flow

This application consists of three main part. Data generation, data persistence and partition creation. Data Generation and Data Writer parts are working together until there is no record left. Data Writer starts after a short period of latency to be sure minimum interruotion while reading from buffer, which is seeded by the Data Generator. After all of the records are persisted to the temporary file Apache Spark based partitioner read the parquet file and create partitions based on time interval. Apache Spark also takes care of the ordering. A visualized view of the flow as below.



## Technologies

* Apache Spark: Star of the application
* Mojo: Using for maven exec plugin
* Junit: Testing purposes
* Log4j: Logging purposes
* Apache cli: Command line options parser

# Design

Producer-Consumer Pattern used for to fill the buffer without blocking data generation and prevent to break the synchronization. Design can be split under three main steps;

• Producer: Data generation part. This part used for generating data according to the schema and fill in the buffer for the consumer. Schema preferred rather than using DTO. By this way, flexibility increased and can be used without touching any code. However, using schema brings some disadvantages. The major effect of it is limitation about using custom partitioner and needs to add a column for interval calculation.

• Consumer: Data writer part. This part is used for consuming generated record from the buffer and persist to a temporary parquet writer implementation of API interface. Temporary parquet file used to use this application with limited resources. Data can be persisted to disk rather than memory to save resources.

• Partitioner: Partitioning part. This part is used to get temporary parquet file, prepare it according to time intervals provided and generate partitions to the file system. While partitioning Apache Spark is ordering the records. That’s why prevented from repartitioning and preferred to use temporary parquet file as the seeder.

## Environment

* This application designed to run on single JVM.
* Parquet partition files are written to local file system.

## Partitioning

* First column of the parquet file must be named as timestamp and type of long. This field contains data in milliseconds from 1970 with UTC time-zone. Name of the partition files are prepared according to this column and must be human readable format.
* Partition intervals must be selected according to timestamp column. This interval shall be configurable as like hourly, per 6 hours etc.

## Extendibility

* Input schema can be provided easily
* Input schema supports the primitive types of the parquet file format
* Input data must be generated according to given schema

## Performance

* API calls should be thread safe
* Buffers, threads and memory usage shall be configurable
* Producer must wait whenever the buffer is full, until the writer consumes the records

## Templates

### DTO

public abstract class SampleDTO {  
 protected long timestamp; protected float a;  
 protected float b;  
 protected String c;  
 protected String d;  
}

### Schema

message sample {  
 required int64 timestamp;  
 required binary binary\_field;  
 required int32 int32\_field;  
 required int64 int64\_field;  
 required boolean boolean\_field;  
 required float float\_field;  
 required double double\_field;  
 required int96 int96\_field;  
}

## API

public interface PartitionWriter<DTO\_WITH\_TIMESTAMP> {  
 public void write(DTO\_WITH\_TIMESTAMP rec) throws IOException;  
}

# Folder Hierarchy



# Usage

## Input

Clone to your local repository: git clone https://github.com/rdemirkoparan/ParquetGenerator.git

Change directory: cd ParquetGenerator/

Compile: mvn compile

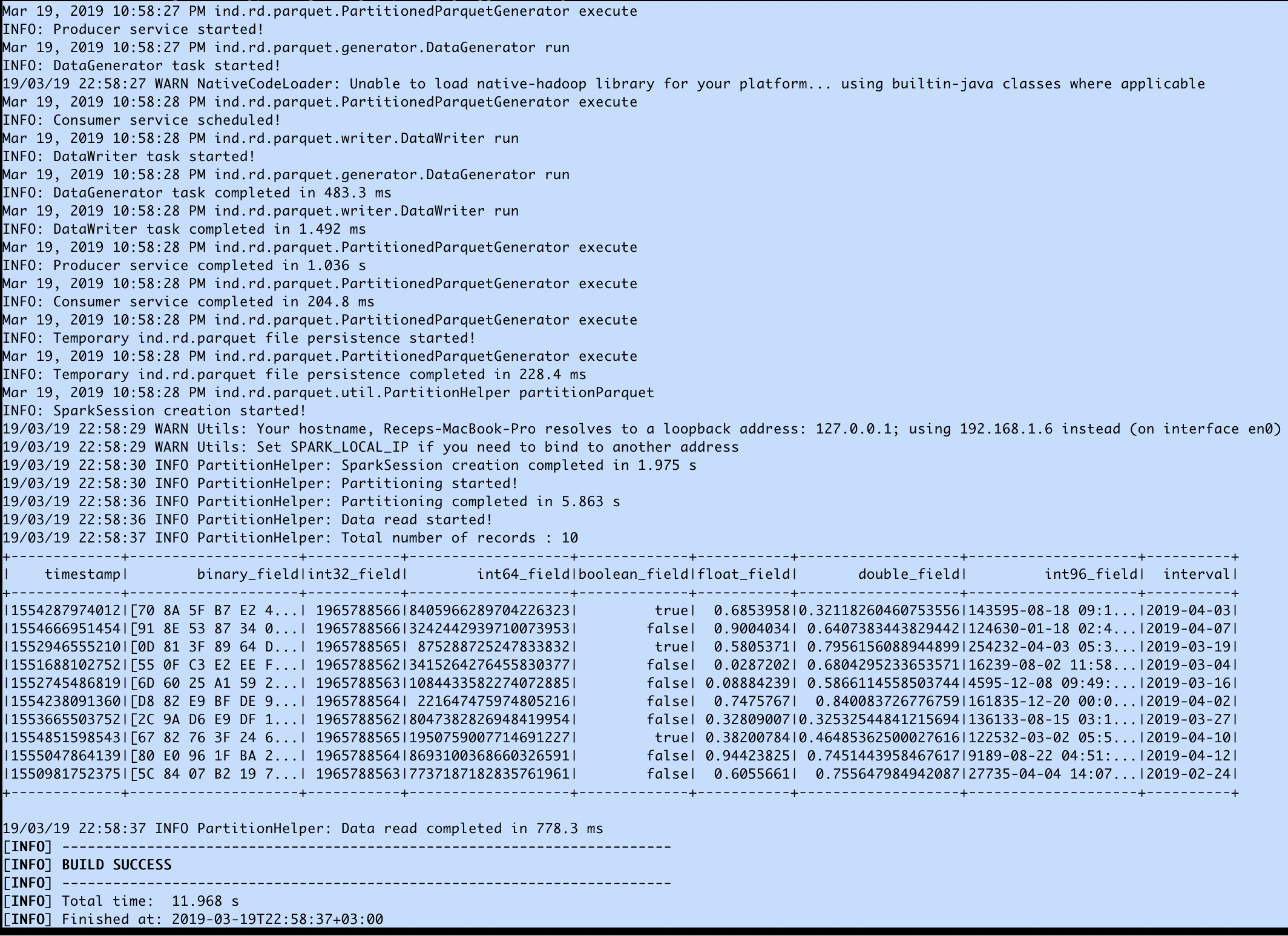
Test: mvn test

Run: mvn exec:java -Dexec.mainClass="ind.rd.parquet.PartitionedParquetGenerator" -Dexec.args="-i doc/sample.schema"

Full list of the input parameters are listed below;

* -i,--inputSchema <arg> Input message schema, requires string as the full path of the schema file
* -l,--bufferLimit <arg> Number of records to buffer, accepts integer (10, 50, etc), default value is 200
* -m,--memoryLimit <arg> Maximum memory buffer while partitioning, requires string as usual memory parameter (1g, 2048m, etc), default value is 1g
* -o,--targetFileName <arg> Output file name, requires string as the full path of the partition directory, default value is /tmp/par.out
* -r,--numberOfRecords <arg> Number of records to generate, accepts integer (10, 9999, etc), default value is 1000
* -s,--partitionInterval <arg> Partitioning time interval as hours, accepts integer as hour (1, 12, etc), default value is 1
* -t,--threadCount <arg> Number of threads, accepts integer (2, 4, etc), default value is 4

## Console Output



## File System Output

