Enterprise Computing WS18/19

Technische Universität Berlin

ISEngineering

Parallel Batch Processing Apache Spark VS Apache Flink

SUTs

Benchmark approach: client perspective

SUTs:
- Flink 1.6.2
- Spark 2.4.0

Cluster settings:
- Provider: AWS
- One master; Two slaves
- EC2 type: m5.large

Workload/Metrics

Data structure: typical book store

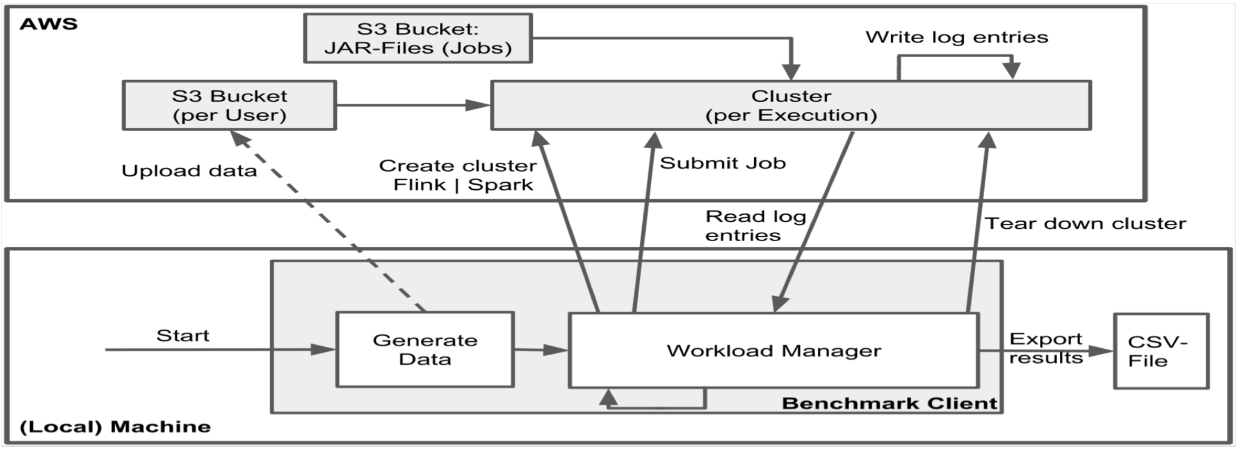
Workloads for each metric and SUT:

a) GroupBy: genre, count per group
b) Sorting: number of pages (ascending)
c) Aggregation: price (max)

Metrics:

1. Runtime
2. Throughput

Benchmarking Tool: Architecture



Result: GroupBy

1. TODO Plot
   1. Flink data
   2. Spark data

1. TODO: Plot interpretation

Florian Muchow, Domenic Bosin, Muhammad Taha, Tristan Schneider