

Final Project Deliverable

Team DEBS - 1

April 14, 2022

Manos, Vivek, Jax, Karan

Detecting Trading Trends in Streaming Financial Data

→ **DEBS 2022** Grand Challenge - Tracking the Exponential Moving Average (EMA) and identifying Breakout patterns.

→ Implemented using **Java and Apache Flink**

- ◆ Custom tumbling window operators
- ◆ Distributed Flink Cluster (3 Virtual Machines)
- ◆ Optimized parallelism and buffer timeout microbenchmarks

→ Financial streaming dataset: **~60M event notifications across 10k batches**, provided by Challenge benchmarking platform

- ◆ Timestamped pricing data by equity symbol

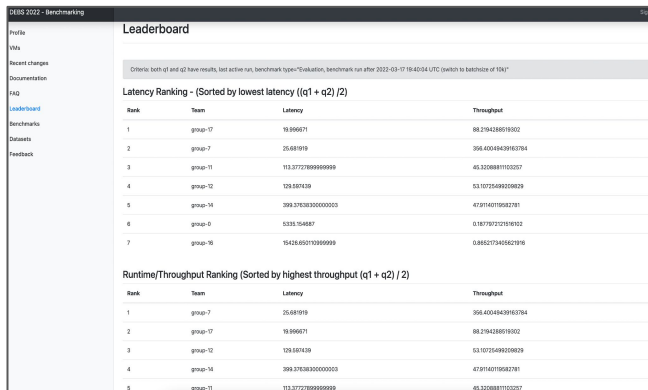
→ Performance

- ◆ **Throughput: 45 batches/s**
- ◆ **Latency: 120 ms**
- ◆ Leaderboard: Top 3 out of 7 teams!

$$EMA_{w_i}^j = \left[Close_{w_i} \cdot \left(\frac{2}{1+j} \right) \right] + EMA_{w_{i-1}}^j \left[1 - \left(\frac{2}{1+j} \right) \right]$$

$$\text{Bullish Breakout} = (EMA_{w_i}^{38} > EMA_{w_i}^{100}) \wedge (EMA_{w_{i-1}}^{38} \leq EMA_{w_{i-1}}^{100})$$

$$\text{Bearish Breakout} = (EMA_{w_i}^{38} < EMA_{w_i}^{100}) \wedge (EMA_{w_{i-1}}^{38} \geq EMA_{w_{i-1}}^{100})$$



DEBS 2022 - Benchmarking

Leaderboard

Criteria: both of q1 and q2 have results, last active run, benchmark type="Evaluation", benchmarks ran after 2022-03-17 10:40:04 UTC (switch to baseline of 10k)

Latency Ranking - (Sorted by lowest latency ((q1 + q2) / 2))

Rank	Team	Latency	Throughput
1	group-17	19.896671	68.2194388519302
2	group-7	25.681919	366.40040439163764
3	group-11	113.37727899999999	45.3208881103257
4	group-12	129.581638	53.10725498209639
5	group-14	289.37636300000003	47.8740719562781
6	group-0	5235.154887	0.1867191212191612
7	group-16	16426.692710999999	0.8832173455621919

Runtime/Throughput Ranking (Sorted by highest throughput (q1 + q2) / 2)

Rank	Team	Latency	Throughput
1	group-7	25.681919	366.40040439163764
2	group-17	19.896671	68.2194388519302
3	group-12	129.581638	53.10725498209639
4	group-14	289.37636300000003	47.8740719562781
5	group-11	113.37727899999999	45.3208881103257

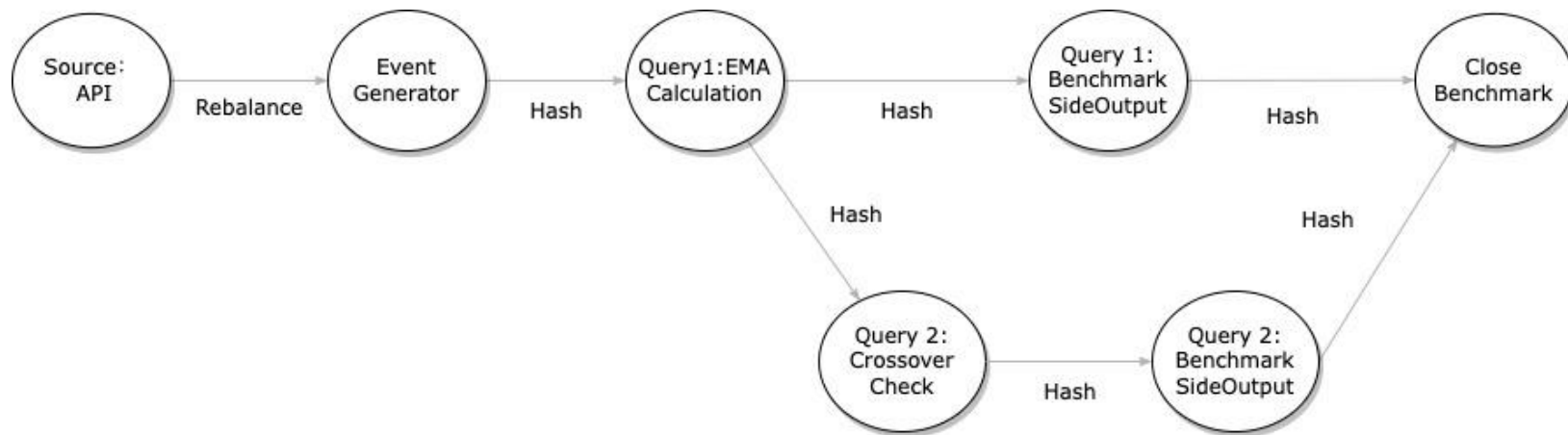
Dataflow Architecture Overview

Event Generation: Parallelize the input event stream.

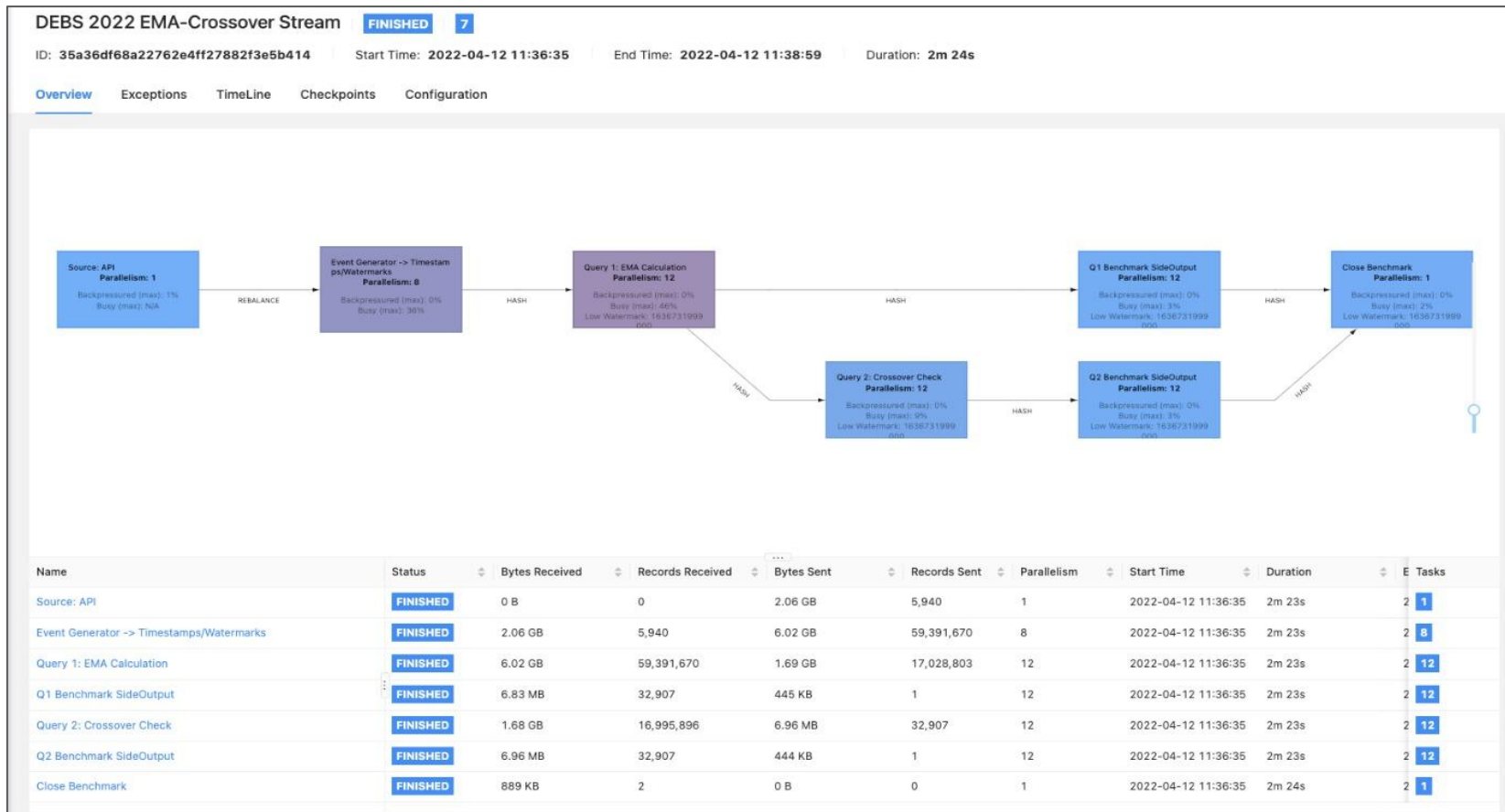
Query 1: Track the Exponential Moving Average (EMA).

Query 2: Identify Bullish and Bearish Breakout Patterns.

Benchmarking Output: Send results for evaluation.

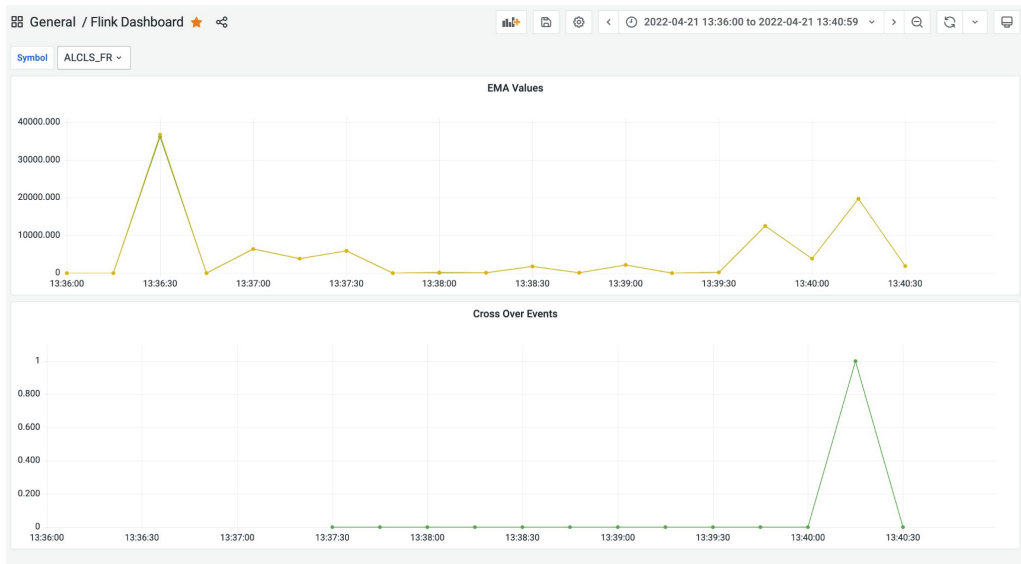


Design Overview - Flink Dashboard



Visualization

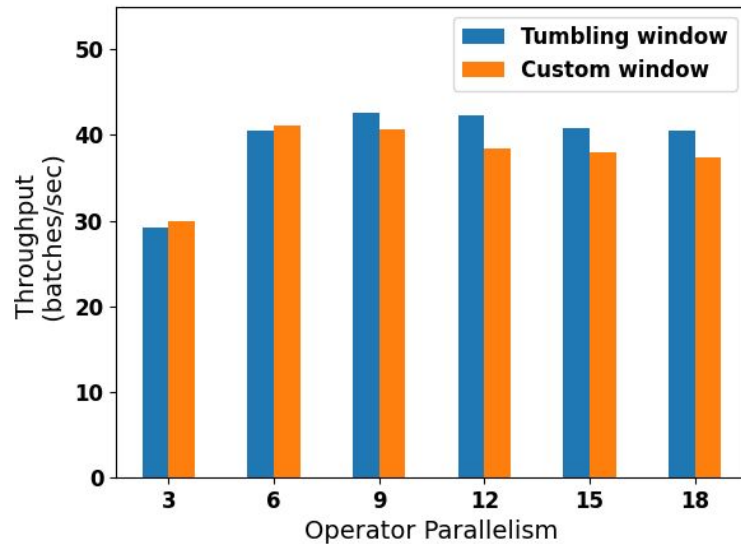
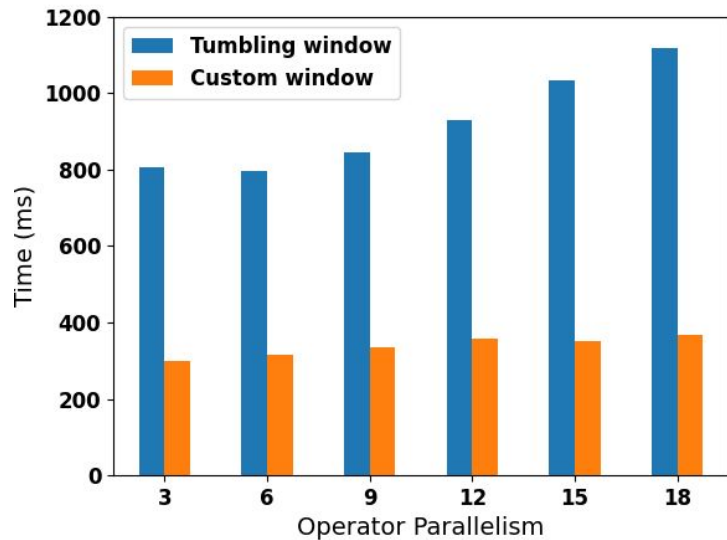
- Uses Flink's inbuilt metrics registry to register EMA and crossover events
- The metrics are pulled by Prometheus and stored in the time series database
- Grafana is used to visualize the data based on the selected symbol



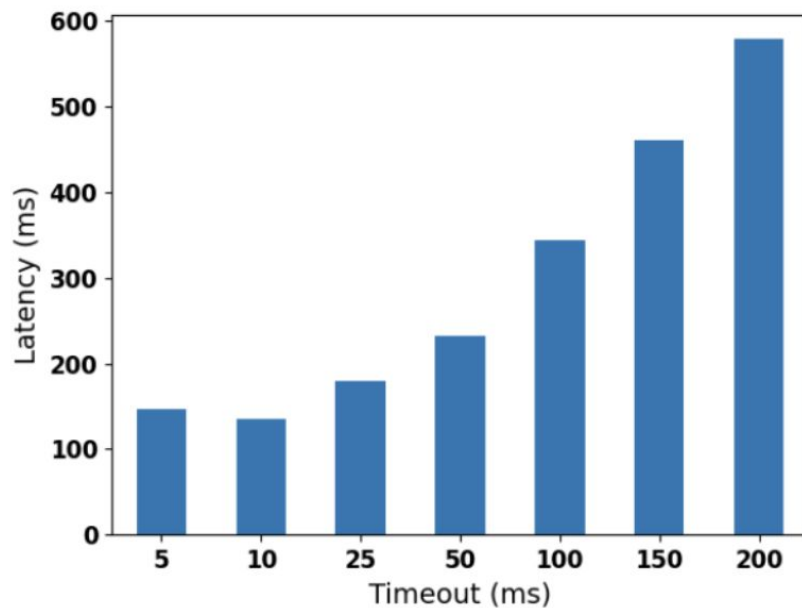
Experimental Section

- Setup
 - Deploy experiments on 3 virtual machines (VMs).
 - Each VM consists of 8GB of memory RAM and 4 CPU cores.
 - Evaluation Dataset
 - Leaderboard as our Benchmarking Platform
- Experimental Metrics
 - Throughput
 - Latency
- Experiments
 - Compare Custom vs Tumbling Window
 - Explore Buffer Timeout Hyperparameter

Custom vs Tumbling Window



Exploring Buffer Timeout Hyperparameter



Lessons Learned

- Custom windowing decreases latency up to **3x**.
- Reducing size of intermediate objects increases throughput and drops latency by **10x**.
- Event generator parallelism boosts throughput up to **3x**.
- Buffer timeout reduction minimizes latency up to **2x**.

Experiences/Challenges

- Learning about Flink and understanding the documentation
- Change in deadlines for DEBS submission
- Running the code on a cluster and managing static variables

Appendix

Github Repository: <https://github.com/kvombatkere/DEBS22-Group11>

Detailed code walkthrough: <https://youtu.be/UU6BBR7Xo20>

DEBS 2022 Grand Challenge: <https://2022.debs.org/call-for-grand-challenge-solutions/>

Evaluation platform and Leaderboard: <https://challenge.msrg.in.tum.de/leaderboard>