Grand Challenge: Automatic Anomaly Detection over Sliding Windows

Zaarour Tarek, Pavlopoulou Niki, Hasan Souleiman, ul Hassan Umair and Curry Edward

Insight Centre for Data Analytics, National University of Ireland, Galway
Lero - The Irish Software Research Centre, National University of Ireland, Galway

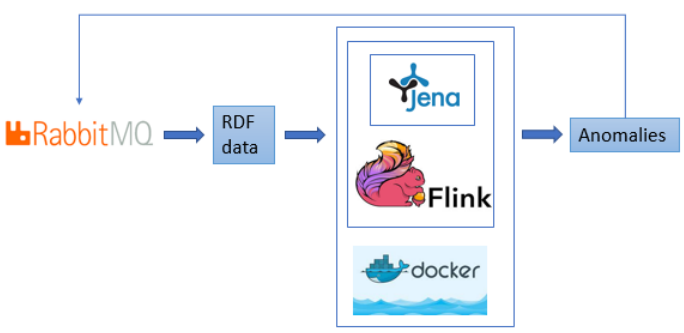
Centre for Data Analytics

Insight

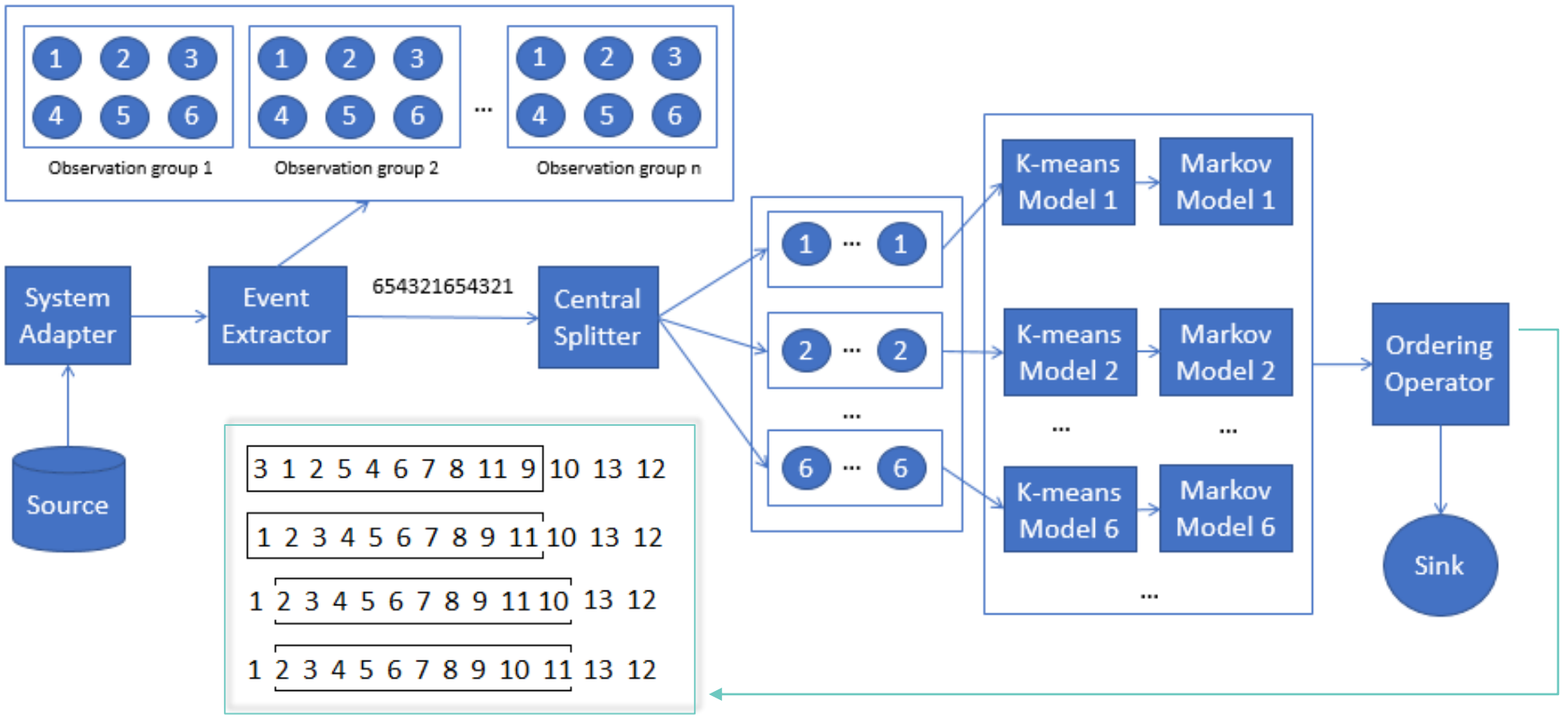
Problem

- Real-time anomaly detection based on the observation of a stream of events.
- Data parallelism requires consistent stream partitioning and efficient task allocation
- Parallel computations result in out-of-order captured anomalies.

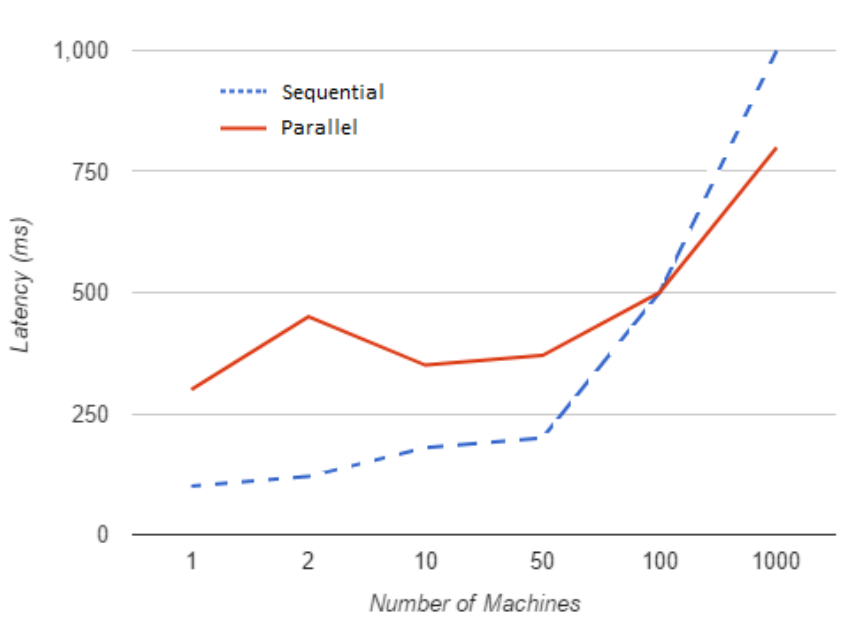
Technology stack

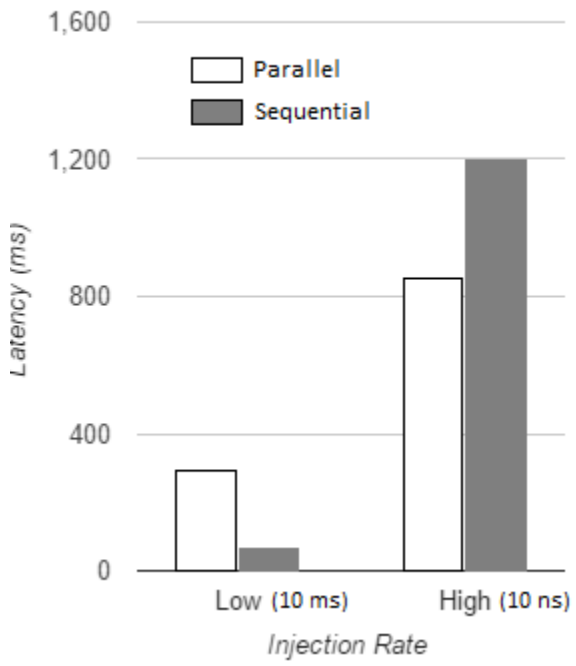


Architecture



Results





- Sequential approach performs better at low injection rates (10ms)
- Mean latency increases as the number of machines goes near 1000
- The ordering operator increased the mean latency at low injection rates to 359 ms
- Decrease in overall latency from 1.2 seconds for the sequential approach to 850 ms for the distributed approach

Conclusions

- Parallel and sequential version performed better at high and low injection rates respectively
- Future directions are sorting out-of-order streams and logical efficient stream partitioning

This project has been funded in part by Science Foundation Ireland grants 13/RC/2094 and SFI/12/RC/2289 and in part with the European Union's Horizon 2020 research programmes Transforming Transport (TT) grant No 731932 and Big Data Value ecosystem (BDVe) grant No 732630.