Symposium on Software Performance

8th November 2022

The Role of Performance in Streaming Analytics Projects: Expert Interviews on Current Challenges and Future Research Directions

Johannes Rank, Andreas Hein, Helmut Krcmar

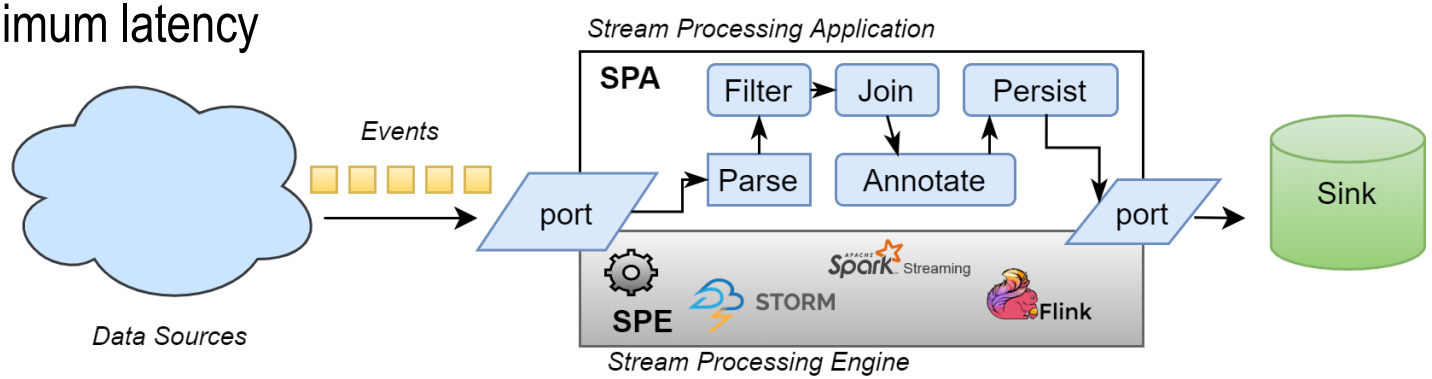
Chair for Information Systems: Lab Krcmar

Technische Universität München

johannes.rank@tum.de

Stream Processing Systems

- Stream Processing Systems (SPS) are processing vast numbers of events with minimum latency



- Examples for SPS include market feed processing, infrastructure monitoring and fraud detection among others(Stonebraker, M., et al. 2005)

- At the same time analytical stream processing is becoming more and more frequent (Hanif, M., et al., 2019) e.g. for business scenarios such as predictive maintenance (Apiletti et al., 2018).

Motivation

- **Research claims that the performance of SPS is of particular importance**
  - For SPS **performance** is not only a quality of service aspect, but **vital** for the whole business scenario to succeed (Stonebraker, M., et al. 2005)
  - **Crucial need for building scalable systems** to enable the processing of vast amounts of streamed data (Bedini et al. 2013)
- Since performance is such a crucial aspect of SPS, do industry implementation projects treat performance in a special way, especially in the context of streaming analytics which may increase the complexity of these systems?

➢ **What is the role of performance in streaming analytics projects?**

Goal

We wanted to understand:

- How is performance is treated in streaming analytics implementation projects?
  - Project Management Perspective
  - Performance Engineering Approaches
- What challenges is the industry currently facing?
- What future research directions can help to solve current issues?

Approach (1)

- Semi-structured interviews with eight experts
  - **Target group**: Software Developers, Solution Architects and Project Managers in the area of streaming analytics projects
  - **„Expert“** = at least 5years working experience

Expert	Business Sector	Job Profile	Experience
A	IT-Consulting	Architect	>5 years
B	IT-Consulting	Project Manager	>10 years
C	Manufacturing (Industry Automation)	Project Manager	>10 years
D	Manufacturing (Industry Automation)	Developer	>5 years
E	Manufacturing (Automobile)	Project Manager	>5 years
F	Banking	Developer	>10 years
G	IT-Consulting	Architect	>5 years
H	Telecommunication	Developer	>10 years

Approach (2)

Interview-guide with 13 questions

1) Warmup Question: “Future Development of Streaming Analytics from a business perspective”

2) Content Questions A – Project Management:

- Does Performance receive special attention as part of the project management?
- Definition of Performance KPIs?
- Who is responsible for performance?
- Performance experience of developers and project members?

3) Content Questions B – Performance Engineering:

- Are streaming analytics projects more complex regarding performance management
- When do you test performance?
- Can the future workload be accurately estimate?
- Do you use performance simulation or planning tools?
- Are SPS performance benchmarks known or used?
- How do you deal with performance problems?

4) Closing Question: “What would help to make performance management easier/more efficient?”

Future Development of Streaming Analytics

- All interviewees agreed that the importance of streaming analytics applications and the **number of implementations continues to grow**
- Some experts were of the opinion that a **massive increase** was still to come
- As the major drivers for this development the experts mostly named **Industry 4.0** use cases, but also trends such as **electro mobility**

> „*A second topic for us is that new requirements are arising for battery management in the context of electromobility. For example,* ***we are working on displaying the status of batteries in real time*** “

Project Management – Special Attention on Performance?

The expert's opinions differed regarding the question if **performance** receives **special consideration** in the context of **streaming analytics projects**

- **Opinion A:** performance is a general requirement of any system and is not treated differently in the case of streaming analytics

- **Opinion B**: Streaming analytics deserves special attention
  - Those experts stated that they ensure performance already at an early design stage
  - "*we are talking about a data volume of **200 million data records within a very short time** [...] make sure **from the beginning** that our applications are designed in a p**erformance-optimized** way*"

➢ **C1:** Performance is not or insufficiently considered in the project organization

Project Management – Are KPIs Defined?

We also received mixed opinions on the question if performance KPIs or SLOs are defined as part of the streaming analytics project

- **Opinion A:** No hard KPIs, soft expectations at best

  "There are usually no hard limits that are defined in the project. Of course, there is a certain expectation of performance depending on the intended use"

- **Opinion B**: The definition of KPIs is important. This was mostly the case in the context of cloud deployments

  "Yes, such goals are firmly defined at the beginning of a project"

➢ **C1:** Performance is not or insufficiently considered in the project organization

Project Management – What Performance Metrics?

Regarding performance metrics, most experts named resource utilization and cost efficiency as major performance criteria next to latency, especially in cloud deployments

> "*Performance is often a cost issue. The customer wants reasonable response times, but the instance should not be too expensive or over dimensioned*"

> *"Scaling a streaming system in the cloud is not difficult. But it's not cheap(..) customers sometimes ask whether the system can also run on a smaller instance"*

- **C2:** Streaming analytics systems are complex and performance problems are multi-layered

Project Management – Responsibility and KnowHow?

- All eight experts agreed that the responsibility for performance never lies with any one person but that each developer is responsible for the proper performance of his component

  "It is important for us that every developer is responsible for performance so that the topic of performance is already considered in the conception (..)"

- However, regarding how the expert would rate the know-how of the project members in performance engineering, we received mixed experience

  "Especially young colleagues tend to choose the first working design and think about performance when it is too late"

- Gap of people responsible for performance but without the experience to ensure it
  - **C1:** Performance is not or insufficiently considered in the project organization
  - **C3**: Lack of Performance Engineering experience

Performance Engineering – Complexity Streaming Analytics?

- According to all experts, analytical SPS are complex regarding their performance management

  "*These systems are mostly distributed and also the configuration is important e.g. parallelization (..). The search for the performance bottleneck can become very time-consuming"*

- One expert also added that the system with which the SPS is integrated requires special performance attention

  "The performance issues we are currently working on are related with the systems with which the streaming system is integrated"

➢ **C2:** Streaming analytics systems are complex and performance problems are multi-layered

Performance Engineering –Workload Estimation?

- The experts agreed that workload estimations are quite reliable

  "I don't think you can define it to the fifth decimal place in advance, but you can at least estimate it roughly"

- However, most experts explained that the differences between development and production system cause uncertainties that complicates performance estimations

  "We often have weaker hardware in development systems, but we also often test with smaller amounts of data(..). With the combination of weaker hardware and less data, it's difficult to make predictions"

➢ **C4:** Performance tests are insufficiently carried out

Performance Engineering – Performance Tools?

- Only three of the experts were aware of performance benchmarks for SPS and none of them use one. They were considered not suitable due to limited result transferability and missing advantages over stress tests

  "Yes, I know that such benchmarks are used in research (..)the question is to what extent the results are transferable and what advantage they would bring?"

- None of the experts uses simulation tools either. Some experts consider them as too complex or that measuring is the better alternative

  "We have not yet used performance simulation tools, the complexity of setting up the models seems too great, we rather test the actual behavior with a load generator"

- However, planning tools are used by four experts

➢ **C5:** Lack of complementary performance tools

Performance Engineering – Approaching Performance Issues?

- The experts agreed that it should always be the first step to optimize the software before increasing hardware resources

- The majority stated that the biggest potential for improvement lies in raising developer's awareness for performance and improving their know-how in performance engineering

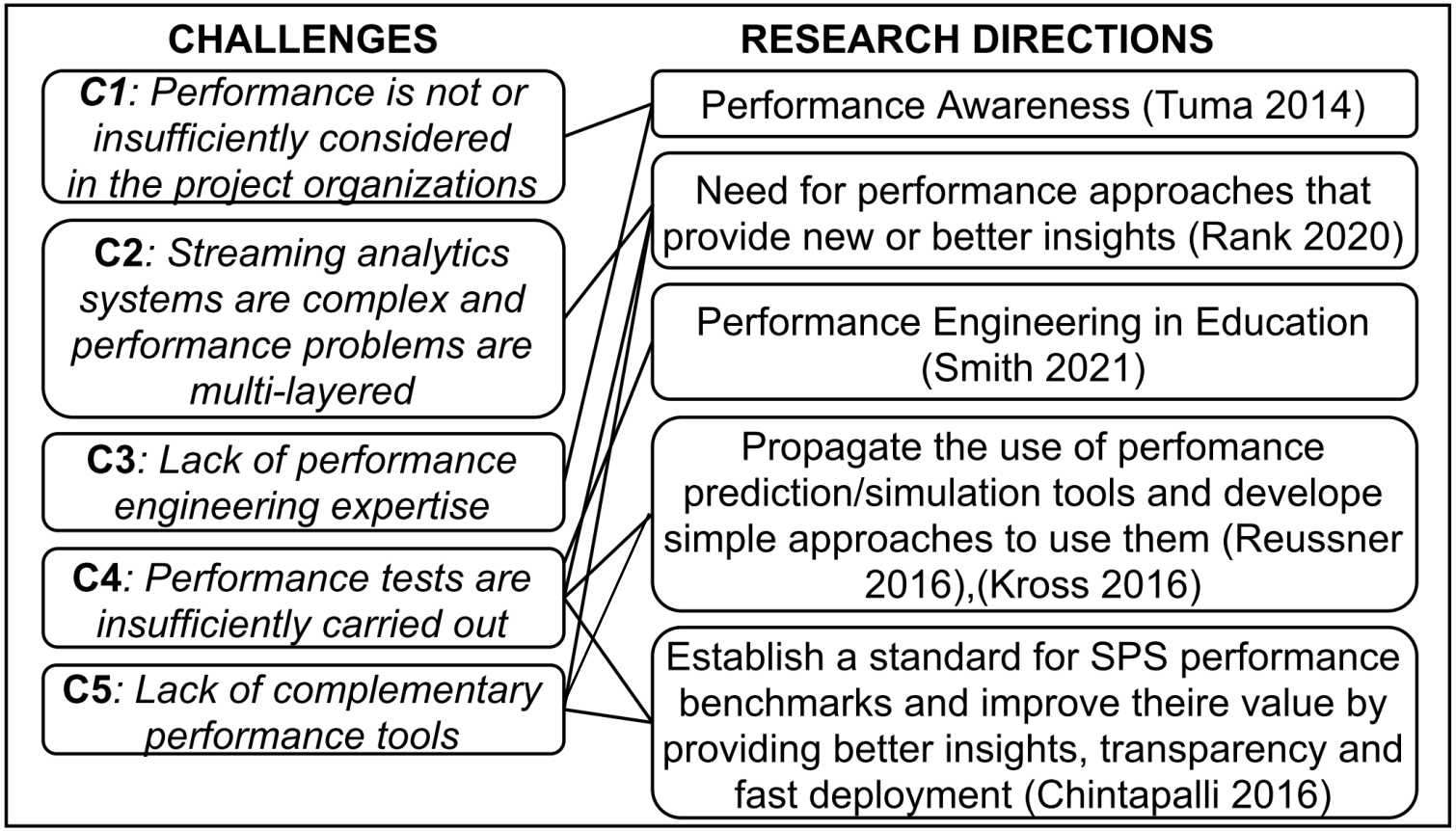
- Several experts also mentioned the need for better measurement tools

➢ **C1:** Performance is not or insufficiently considered in the project organization

➢ **C3:** Lack of performance engineering expertise

➢ **C5:** Lack of complementary performance tools

Challenges and Research Directions



Discussion(1)

C1: Performance is not or insufficiently considered in the project organization

- Performance goals should always be formulated as part of the project.
- From a research perspective a key driver to cope with this challenge is the raise of performance awareness to put an emphasis on proper performance design at early stages of the software lifecycle

C2: Streaming analytics systems are complex and performance problems are multi-layered

- Response time and cost efficiency are important requirements when it comes to industry implementations
- In terms of performance evaluation, research should not only focus on response time, but also on the efficient utilization of resources
- At the same time better tooling support is required to identify performance bottlenecks

Discussion(2)

C3: Lack of performance engineering expertise

- Performance engineering is an important skill for any developer.
- Performance awareness and a greater focus on performance engineering from a educational point of view could address this

C4: Performance tests are insufficiently carried out

- All experts used performance measurement approaches. However, there was uncertainty because the testing environment did not adequately reflect the production environment
- a quality assurance system that reflects the sizing of the production system is cost-intensive
- Model-based prediction tools such as the Palladio Component Model

Discussion(3)

C5: Lack of complementary performance tools

- Performance benchmarks were not applied
- Some experts did not know that such are available in the context of streaming. Others felt that the benchmark results are not transferable.
- Research should focus on establishing an industry standard benchmark
- For the benchmark to have an advantage over self-developed load tests, it should offer more result transparency, new insights and be able to be deployed with little effort

Conclusion

- Streaming analytics projects face several challenges with regards to performance management
- Many of these challenges are similar to those of other systems, but some, such as the lack of an industry benchmark, are specific to this domain
- From a performance management perspective there is plenty of room for future research directions

References

Apiletti, D., et al. (2018). iSTEP, an Integrated Self-Tuning Engine for Predictive Maintenance in Industry 4.0. 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom).

Bedini, I., et al. (2013). Modeling performance of a parallel streaming engine: bridging theory and costs. Proceedings of the 4th ACM/SPEC ICPE. Prague, Czech Republic, ACM: 173-184.

Chintapalli, S., et al. (2016). Benchmarking Streaming Computation Engines: Storm, Flink and Spark Streaming. 2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW).

Kroß, J. and H. Krcmar (2016). "Modeling and simulating Apache Spark streaming applications." STT 2016 36(4).

Rank, J., et al. (2020). "A Dynamic Resource Demand Analysis Approach for Stream Processing Systems." Softwaretechnik-Trends 40(3): 40-42.

Reussner, R. H., et al. (2016). Modeling and simulating software architectures: The Palladio approach, MIT Press.

Stonebraker, M., et al. (2005). "The 8 requirements of real-time stream processing." SIGMOD Record 2005 34(4): 42-47.

Tůma, P. (2014). Performance awareness: keynote abstract. Proceedings of the 5th ACM/SPEC international conference on Performance engineering, ACM.

Thank you for your attention!

Questions?

mail: johannes.rank@tum.de