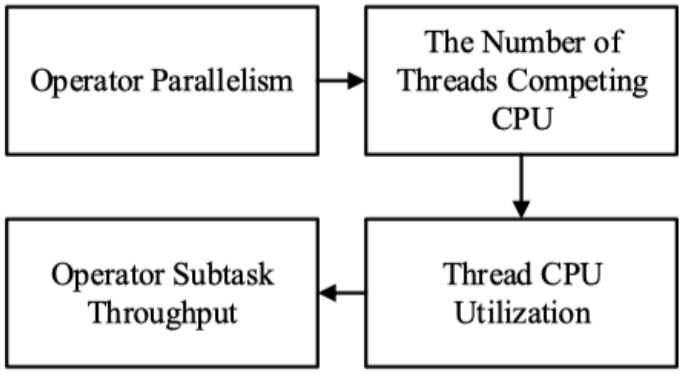
QAAS: Quick Accurate Auto-Scaling For Streaming Processing

Shiyuan LIU, Yunchun LI, Hailong YANG, Ming DUN, Chen CHEN, Huaitao ZHANG, Wei LI

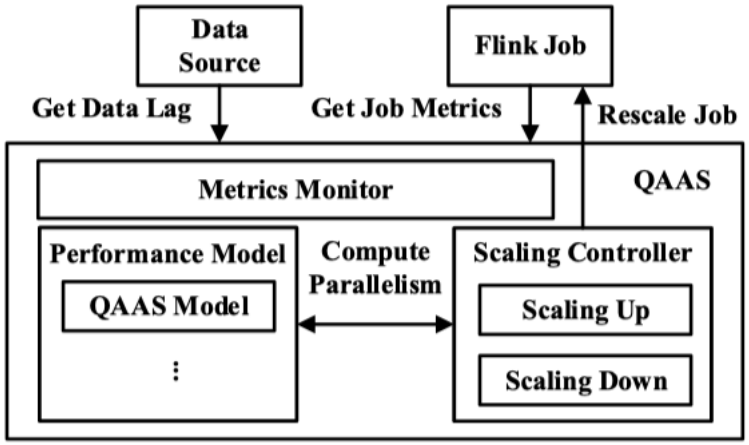
Frontiers of Computer Science, DOI: 10.1007/s11704-022-1706-4

Problems & Ideas

- Problems of resources adjustment in stream processing systems
  - The job auto-scaling strategy leads to task instability
  - Need nonlinear accurately operator performance model
  - Need carefully adjust stream processing job with combination the performance model and strategy
- Ideas: Propose QAAS, a system with multi-stage performance model to adjust stream processing job performance for workload fluctuating



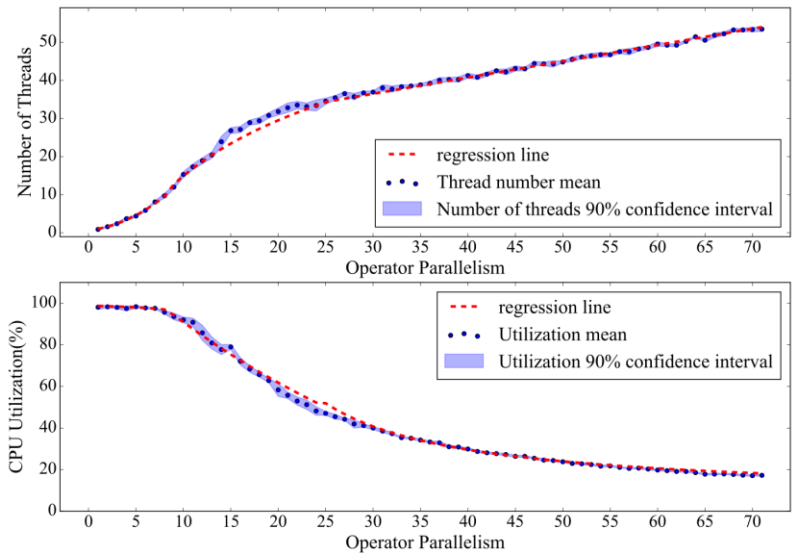
The causal relationship between the parallelism and throughput of operator



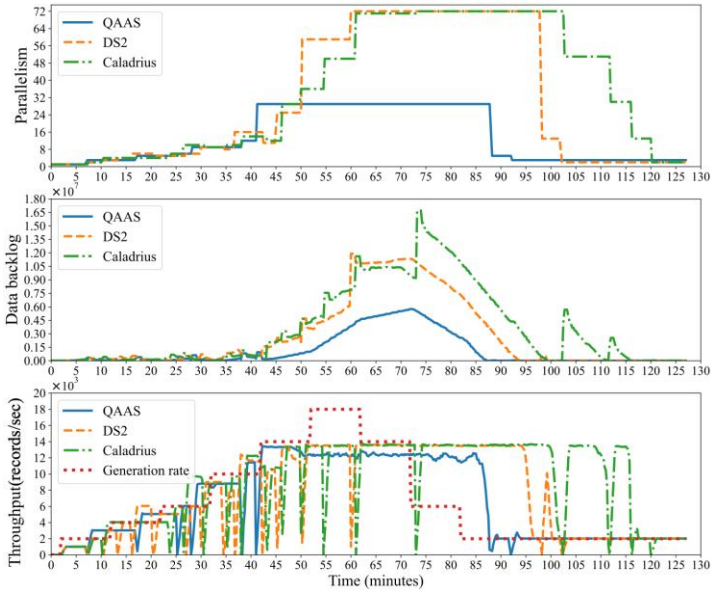
QAAS System Model

Main Contributions

- Contributions:
  - We propose a multi-stage performance model to accurately predict operator throughput under different parallelism.
  - We implement the QAAS system integrated with multi-stage performance model. The QAAS system is built on top of Flink to scale operator parallelism of Flink job automatically.



The Performance Model Evaluation



Comparing the impact of QAAS, DS2, and Caladrius models on operator parallelism, throughput, and data backlog