

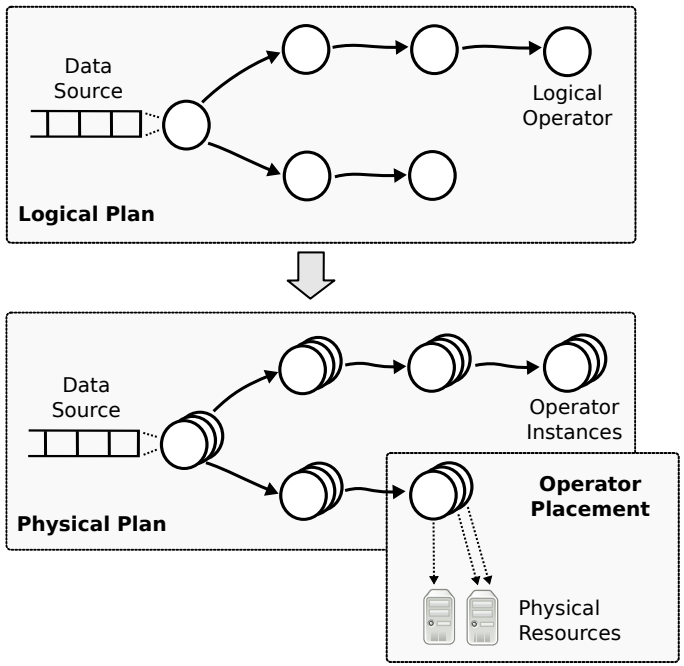

STREAM PROCESSING IN EDGE AND CLOUD INTERPLAY

Alexandre Veith

Advisors: Laurent Lefevre and Marcos Dias de Assuncao

LIP - ENS Lyon — March 6, 2019

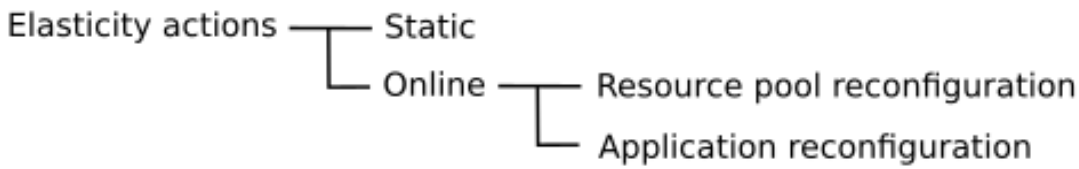
- **Data Streams** - input data that arrives at high rate
- **Frameworks** - Directed Acyclic Graph of processing elements
- **Logical plan** - logical abstraction for specifying operators and how data flows between them
- **Physical plan** - parallelisation hints or specify how many instances of each operator should be created
- **Auto-scaling** - system capacity should be modified



PROBLEM STATEMENT

- Most of existing work on multi-operator placement considered network metrics such as latency and bandwidth whilst proposing decentralised algorithms, without taking into account that **the network can be programmed and capacity allocated to certain network flows**
- Push **analytics tasks to edge resources** [Cheng et al., 2016]

- When considering the operator DAG based solutions
  - **Static**: comprise optimisations to **modify the original graph** to improve task parallelism and operator placement, optimise data transfers, among other goals [Hirzel et al., 2014]
  - **Online** (Dynamic): comprise both actions to modify the pool of available resources and dynamic optimisations carried out to **adjust applications dynamically** to utilise **newly allocated resources**



- Place data processing tasks in **Edge Computing** whilst minimising the use of network resources and latency, **efficient methods to manage resource elasticity**
- **Optimise the physical plan** with or without previous knowledge of the network architecture or behavior (*i.e.*, SDN)

- Create a model
  - Analyze the possible techniques and models
  - Point out the metrics
  - Restrict the environment variables
- Evaluate existing frameworks, infrastructures and algorithms
  - Apache Flink
  - Apache Storm
  - Apache Kafka
  - Apache Edgent
  - Mosquitto
  - Vivaldi

THANK YOU!

Alexandre Veith
ENS-Lyon — LIP
alexandre.veith@ens-lyon.fr
alexandre.da-silva-veith@inria.fr

BIBLIOGRAPHY

Cheng, B., Papageorgiou, A., and Bauer, M. (2016). Geelytics: Enabling on-demand edge analytics over scoped data sources. In *IEEE International Congress on Big Data (BigData Congress)*, pages 101–108.

Hirzel, M., Soulé, R., Schneider, S., Gedik, B., and Grimm, R. (2014). A catalog of stream processing optimizations. *ACM Computing Surveys*, 46(4):1–34.

Kulkarni, S., Bhagat, N., Fu, M., Kedigehalli, V., Kellogg, C., Mittal, S., Patel, J. M., Ramasamy, K., and Taneja, S. (2015). Twitter Heron: Stream processing at scale. In *ACM SIGMOD International Conference on Management of Data*, SIGMOD '15, pages 239–250, New York, USA. ACM.

Netto, M. A. S., Cardonha, C., Cunha, R., and de Assuncao, M. D. (2014). Evaluating auto-scaling strategies for cloud computing environments. In *22nd IEEE Int. Symp. on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2014)*, pages 187–196. IEEE.