Linnéuniversitetet
Linnaeus University

Master Thesis (level)

Thesis Topic

Fully Decentralized Data Stream Processing

Motivation

Data Stream Processing (DSP) application is a Big Data technology that exploits data produced from multiple, distributed sources and timely processes it in order to extract information/knowledge as soon as the data is collected. Departing from the traditional static processing of data, DSP focuses on flow processing - i.e., data is not stored but continuously processed as it flows -, and timeliness - i.e., time constraints are crucial for DSP applications.

Using traditional Cloud infrastructures to support DSP has some drawbacks: (1) the network bandwidth is expensive and might be scarce, the latency from edge nodes/sensors to the Cloud node can be relatively high and severely affect the overall application performance, also preventing to achieve end-to-end QoS goals; (2) there has been a huge increase in the number of connected devices (50 billion devices by 2020) and the inherent high amount of data to be transferred to the Cloud for elaboration can significantly increase the core network traffic and degrade the network performances, and (3) data might have legal constraints related to a specific jurisdiction and cannot be moved to a distant node located on a different jurisdiction.

Therefore, there is a need to avoid the data movement, to reduce the network latency and the communication over the network. To this end, one of the most prominent solutions is to adhere to the Pure Edge Computing paradigm and implement DSP as a fully decentralized system, where nodes run different functions/operators on incoming data and produce new complex streams. Indeed, DSP is modelled as a Peer-to-Peer system, where nodes represent functions/operators and links data dependencies among them.

Tasks

1. Perform a Systematic Literature Review
2. Design a framework/middleware addressing the requirements
3. Implement a prototype as extension of Apache Storm
4. Evaluate the performances of the developed system

Prerequisites *

- Software Engineering (2DV603) and Software Architecture (2DV604)
- Architectures for Service-based Systems (4DV609)
- Advanced Java
- Computer Networks

Supervisor(s)

- Mauro Caporuscio (mauro.caporuscio@lnu.se)
- Mirko D'Angelo (mirko.dangelo@lnu.se)

* All the course codes, like e.g. 1DV101, refer to courses here at DFM. Similar documented experience from other places will do just as well.