Scheduling edge and in-transit computing resources for stream processing applications

Ali Reza Zamani Zadeh Najari
Rutgers University

5/10/2019 at 10:00 am
CoRE A 301

Abstract

Recent advances in technology and science have caused an exponential growth in data sources, data volumes, and data production rates which have outpaced our ability to transport and deliver the data in a reliable and timely manner. Due to the excessive cost of data movement and wide distribution of data sources, traditional centralized approaches that require data to be first collected and compiled into a central location for processing are not feasible anymore. Instead, new approaches that provision and leverage resources at the proximity of data sources (edge processing) and en-route between geo-distributed resources (in-transit processing) should be explored. Since these approaches (edge, in-transit and clouds) are typically decoupled from each other, designing a framework that uses them together in an integrated manner as part of an application workflow is challenging. In this thesis, we are going to address this challenge by developing a mathematical scheduling model to support edge and in-transit data analysis in cloud federation. In addition, we design and develop a subscription-based data streaming framework to deploy and execute stream oriented workflows, and a runtime management layer that leverages edge and in-transit resources to adjust the resolution and provide end-to-end QoS using on-demand control loops. The validation of this work is done through a series of experiments on three main applications: (1) EnergyPlus, (2) Image processing for surveillance cameras, and (3) Ocean Observatory Initiatives (OOI). The results demonstrate the effectiveness of our approach on increasing job acceptance ratio and utilization of the resources by almost 25% and 16%, respectively.

Defense Committee: Prof. Manish Parashar (Chair), Prof. Ulrich Kremer, Prof. Srinivas Narayana, Prof. Adrien Lebre (IMT Atlantique)