Understanding distributed dataflow systems and applications

Ioannis Liagouris
ETH Zurich

Host: Krish Chatterjee

Distributed dataflow systems have become the de-facto systems for large-scale data processing. So far, most efforts from academia and industry focus on improving the performance, scalability, and reliability of these systems. However, as their complexity increases, explainability emerges as a first-class concern.In this talk I will present my work on understanding the behavior of distributed dataflow systems and applications. First, I will describe a framework for explaining the semantics: Why and how does a dataflow return certain results? To answer such questions, the framework leverages ideas from database provenance to provide output explanations that are both sufficient and concise. Second, I will focus on understanding performance: Why is a dataflow execution slow and which are the bottlenecks in the pipeline? My work in this area generalizes existing approaches on critical path analysis methods to dynamic and continuous computations. I will conclude the talk with a discussion on the challenges of explaining emerging AI applications and an overview of my future research agenda.

Monday, March 18, 2019 10:00am - 11:00am
Mondi Seminar Room 2, Central Building

This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station. Please find a schedule of the ISTA Shuttle on our webpage: https://ista.ac.at/en/campus/how-to-get-here/ The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.