UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

DEPARTEMENT D'INFORMATIQUE
DEPARTEMENT FÜR INFORMATIK

INFORMATICS COLLOQUIUM

Speaker:

Dr. Carlo Curino, Microsoft

Tensor Query Processing: Neural Network $$ to speed up Databases and Classical ML!

Abstract:
Massive market interest in AI has driven unprecedented investments in Special HW and runtimes for Neural Networks. Tensor computations are emerging as the de-facto API for all these special hw and runtimes. In this talk, we show how we can automatically transform and optimize relational queries and Classical ML pipelines into tensor computations, and run on special hardware. Interestingly the performance we obtain significantly outperform classical systems and even custom-build GPU DBMSs. At the same time, this approach retains very low engineering costs, thanks to a minute code footprint (<10k LoC) and free portability---as we piggyback on tensor runtimes getting ported to all the new HW coming out. We conclude touching on further research directions that emerge once both queries and ML models are uniformly represented as tensors computations.

Bio:
Carlo Curino is the lead of Gray Systems Lab (GSL), and applied research group working at the intersection of Databases/Systems/Machine Learning. Before this Carlo was a Principal Scientist in Cloud and Information Services Lab (CISL), working on large-scale distributed systems, with a focus on scheduling for BigData clusters; this line of research was co-developed with several team members and open-sourced as part of Apache Hadoop/YARN. Intrinsically, this research work enables us to operate the largest YARN clusters in the world (deployed on 250k + servers within Microsoft). Prior to joining Microsoft he was a Research Scientist at Yahoo!; primarily working entity deduplication and scale and mobile+cloud platforms. Carlo spent two years as a Post Doc Associate at CSAIL MIT working with Prof. Samuel Madden and Prof. Hari Balakrishnan, working on relational databases in the cloud. At MIT he also served as the primary lecturer for the course on databases CS630, taught in collaboration with Mike Stonebraker. Carlo received a Bachelor in Computer Science at Politecnico di Milano. He participated in a joint project between University of Illinois at Chicago (UIC) and Politecnico di Milano, obtaining a Master Degree in Computer Science at UIC and the Laurea Specialistica (cum laude) in Politecnico di Milano. During the PhD at Politecnico di Milano, Carlo spent two years as a visiting researcher at UCLA.

Date and time: *Tuesday August 16th, 2022, 10.00 am*
Location: *Pérolles 21, room G230, Bd de Pérolles 90, Fribourg*
Contact person: *Prof. Philippe Cudré-Mauroux*

The colloquium is free and open to the public.