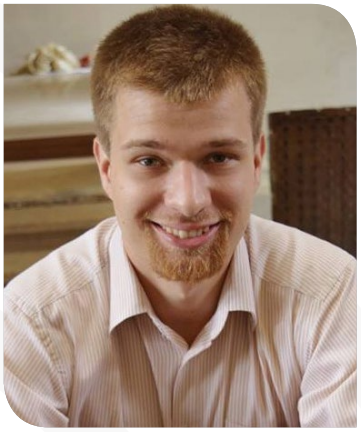
Márton Balassi

EIT Digital Budapest DTC



PhD topic: Distributed computational architectures and algorithms

PhD supervisor: Dr. András Benczúr, SZTAKI

Industrial partner: Kostas Tzoumas, data Artisans

Contact: balassi.marton@gmail.com

'EIT Digital enables me to transfer the results of my research to industry. It gives both the opportunity to work closely with commercial partners and the essential knowledge to navigate in the world of business.'

Achievements & further plans

Currently I am a second year PhD student in Computer Science at Eötvös Loránd University of Sciences and I am focusing on distributed data-intensive processing engines and their applications. During my first year I and my co-authors managed to publish a study in a peer-reviewed journal on computing metrics on datasets exceeding the storage capacity of a single machine. After investigating metrics, my primary interest turned towards the architecture of **Big Data systems** and I made sizeable contribution to the open source **Apache Flink** data platform. I am a co-author of a recently submitted peer-reviewed paper on the architecture of Flink.

Distributed computing architectures enable a wide variety of business applications, ranging from classic ETL to recommender systems or sentiment analysis, producing results in near real time at 1-5% of the cost of well-established proprietary tools.

Educational status at Spring semester of 2016:

RA	OR	BMD	GH	Mobility	BDExp.
✓	⚙	✓	⚙	⚙	⚙

Reserach topic

The field of my research is distributed data-intensive processing engines and their applications. Working with hundreds of machines poses a variety of challenges for both the designers and the users of the framework. There is still a number of open research questions regarding fault tolerance, consistency and the programming model in general.

Recently, the field of distributed data processing systems has turned its focus to near real time or stream processing. In many cases, the shift from bounded datasets to unbounded streams poses semantical challenges and provides support for a set of new applications. The proper way of providing fault-tolerant streaming applications with support to flexible data-driven windowing semantics is still partially unclear.

The challenges I have recently worked on are call drop prediction and real-time network monitoring in collaboration with Ericsson Hungary. This work is in line with the goal of the Budapest DTC, which is 'Excellence in communication software and system performance'. To match the requirements of my PhD studies and the ones posed by the industrial partners, my work has two seemingly separable objectives, core research and software development applying my results.