Master thesis student - Scheduling of heavyweight workflows

Engineering | Uppsala

We are looking for students who are excited to dig into data dependency graphs!

Background
Data is the lifeblood of Klarna. To make data work for Klarna and its customers, Klarna internally maintains several workflows, which combine and transform input data from several sources, specified in substeps. We want to speed up the execution of these workflows. Currently, all substeps of each workflow currently are executed in sequence, whereas running independent steps in parallel may shorten execution time. To test this idea, we want the ability to automatically identify data dependencies in existing workflows and re-write their specifications on parallelizable form.

Work includes
- Study existing tools and current state-of-the-art
- Apply parallelization on one or more workflows
- Measure the improvement of the parallelized workflow and validate the result
- Write a thesis showing the technical content of the work and the results from the experiments

Relevant courses
- Database technology
- Algorithms and data structures. Scalability and complexity analysis.
- Distributed systems
- Mathematical statistics and modelling

Desired experience (most important first)
SQL. Shell script. Hadoop and Hive. Java. Experience in IntelliJ/IDEA (or similar IDE). Use of Maven, Git, and Jenkins.

Confidentiality considerations
To protect Klarna's customers and IPR, all data is considered sensitive and/or confidential and cannot be disseminated outside Klarna. The candidate may undergo background checks prior to the project, and must sign all non disclosure agreements and follow all Klarna internal procedures that may apply.

Apply Now