



## Master thesis student - Performance improvement of Hive workloads

### Engineering | Uppsala

We are looking for students who are excited to improve the efficiency of our heavyweight data transformations!

#### *Background*

To extend credit to its consumers, Klarna performs various assessments in real-time, such as consumer identification, credit assessment, and fraud prevention checks. To provide adequate data to these assessment systems, Klarna combines and transforms input data from several sources on a regular basis. These transformations are computationally intensive.

To improve resource utilization, it is interesting to improve the execution efficiency of these transformations, by changing how they are performed without affecting the result. To that end, it is important to carefully validate any changes introduced.

#### *Work includes*

- Study and profile heavyweight queries.
- Propose and implement changes. Ideas for efficiency improvements include, but are not limited to, materialization of intermediate results, query optimizer tuning, and alleviation of data skew.
- Compare the original transformation with the new transformation, with respect to characteristics such as transformation result, resource utilization (CPU, memory disk), and wall clock execution time.
- Write a thesis showing the technical content of the work and the results.

#### *Relevant courses*

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

#### *Desired experience*

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](#)