



**Master thesis student - Scalable data validation tool**

**Engineering | Uppsala**

We are looking for students who are excited to make us faster and smarter!

#### *Background*

To feed its online assessment systems and its business intelligence and analytics, Klarna combines and transforms input data from several sources on a regular basis. Driven by developing business needs, these transformations are continuously updated and refined by Klarna's engineers. For such development work, it is critical to carefully validate any changes in the result produced by different versions of a transformation. Klarna has internally developed a tool for automatic comparison of output data from transformations. As this tool is extensively used, it is interesting to improve its performance and usability, as this will save a lot of (expensive) engineer time. This will make us faster and smarter.

#### *Work includes*

- Study the existing validation tool.
- Propose and implement changes. Ideas for efficiency improvements include, but are not limited to, implementation of UDFs (User Defined Functions) in Hive, improvement of key based comparison based on interaction with metadata, automation of execution, improvement of presentation.
- 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 (most important first)*

SQL. Java. Shell script. Hadoop and Hive. 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](#)

powered by **Jobvite**