

Exposé: Implementing an imperative language to do machine learning on web-scale data within parallel data flow systems.

Master's thesis

Till Rohrmann
Technical University of Berlin, Germany
till.rohrmann@mailbox.tu-berlin.de

1. INTRODUCTION

- Data analytics and web-scale data
- Parallel programming and challenges for users
- Parallel data flow systems (Stratosphere, Spark)

2. THESIS APPROACH

- Language design & parser
- Runtime
- Optimization of generated programs
- Evaluation of approach

3. RELATED WORK

- Spark and Stratosphere
- Pegasus
- SystemML

4. CONCLUSION

5. REFERENCES