Summary Paper of *Don't Hold My Data Hostage – A Case For Client Protocol Redesign*

Authors of Original Paper: Mark Raasveldt and Hannes Mühleisen

Hannes Pohnke
pohnke@tu-berlin.de
TU Berlin
Berlin, Germany

ABSTRACT

Traditionally, database query processing and ML tasks are executed on separate, dedicated systems, but the current trend goes towards integrated data analysis pipelines that combine both tasks. In state of the art systems, orchestration of those two tasks still is inefficient due to expensive data transfer and missed global optimization potential. The paper we are summarizing, "Don't Hold My Data Hostage – A Case For Client Protocol Redesign" by Mark Raasveldt and Hannes Muhleisen, addresses this problem by investigating the high cost of transferring large data from databases to the client programs, which can be much more time consuming than the actual query execution. The authors explore and analyse current serialization methods, that are used in database systems and identify their inefficiencies through various experiments. They also introduce a new columnar serialization method that can significantly enhance data transfer performance. By improving the data transfer, this approach could be a step towards efficiently combining database and machine learning systems.

KEYWORDS

Databases, Client Protocols, Data Export

ACM Reference Format:
Hannes Pohnke. 2018. Summary Paper of *Don't Hold My Data Hostage – A Case For Client Protocol Redesign*: Authors of Original Paper: Mark Raasveldt and Hannes Mühleisen. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 1 page. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

2 SUMMARY

3 ANALYSIS

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX