University of ZurichUZH

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Claudio Brasser

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, July 15, 2019

Master's Basic Module, 3 ECTS
Datenbanktechnologie

Title: A Seamless Integrating of Management and Statistical Computations

There is a significant demand to find solutions that tightly integrate data management and data analytics. A seamless integration would eliminate the time and effort that is required to transform data and move it back and forth between a database and an analytics application.

The goal of this Master's basic module is to investigate in detail selected existing approaches for such an integration.

Tasks:

- Review existing approaches in the literature [1] [5] [4] [3] [2] for integrating data management with data analytics.
- Set up a representative ML use case that you use to compare and evaluate the approaches.
- Summarize the approaches in a report and discuss their respective advantages and disadvantages.

References

[1] Oksana Dolmatova, Nikolaus Augsten, and Michael Böhlen. A relational matrix algebra and its implementation in a column store. In *submitted for publication*, 2019.

[2] Joseph Vinish D'silva, Florestant De Moor, and Bettina Kemme. AIDA - abstraction for advanced in-database analytics. *PVLDB*, 11(11):1400–1413, 2018.

[3] Philipp Große, Wolfgang Lehner, Thomas Weichert, Franz Färber, and Wen-Syan Li.

Bridging two worlds with RICE integrating R into the SAP in-memory computing engine. *PVLDB*, 4(12):1307–1317, 2011.

[4] Jonathan Lajus and Hannes Mühleisen. Efficient data management and statistics with zero-copy integration. In *Conference on Scientific and Statistical Database Management, SSDBM '14, Aalborg, Denmark, June 30 - July 02, 2014*, pages 12:1–12:10, 2014.

[5] R project. The R Project for Statistical Computing. `https://www.r-project.org/`, 2018.

Supervisor: Michael Böhlen

Start date: 1.7.2019

End date: 9.8.2019

Exam: 12.8.2019, 15:00 - 15:30

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen