



## HAUPTSEMINAR RECHNERARCHITEKTUR UND PROGRAMMIERUNG

Thema:

### Schnelleres Data Analytics? Apache Spark und RDMA

Inhaltliche Schwerpunkte und Hinweise:

- Standard Data Shuffle bei Apache Spark, Unterschiede zu einer RDMA-Lösung
- Aufsetzen von RDMA-Spark oder Spark mit Crail
- Suche eines geeigneten Testprogramms; Laufzeitmessungen
- Vergleich des gewählten Ansatzes mit Spark ohne Erweiterung

Literatur:

LU, Xiaoyi, et al. Accelerating spark with RDMA for big data processing: Early experiences. In: *High-performance interconnects (HOTI), 2014 IEEE 22nd annual symposium on*. IEEE, 2014. S. 9-16.

STUEDI, Patrick, et al. Crail: A High-Performance I/O Architecture for Distributed Data Processing. *IEEE Data Eng. Bull.*, 2017, 40. Jg., Nr. 1, S. 38-49.

Apache Spark <http://spark.apache.org/>

RDMA-Spark <http://hibd.cse.ohio-state.edu/#spark>

Apache Crail <https://crail.apache.org/>

Betreuender Hochschullehrer: Prof. Dr. Wolfgang E. Nagel

Betreuer:

Jan Frenzel  
jan.frenzel@tu-dresden.de  
Raum: FAL 253