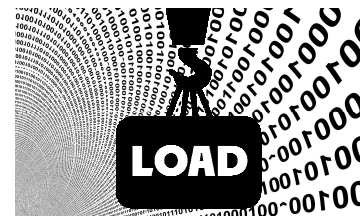


### **Warum ist das interessant?**

Big Data Applikationen versuchen Datenmengen zu verarbeiten, die sich mit traditionellen Applikationen nicht mehr handhaben lassen. Allerdings können sich die Datenmengen zur Laufzeit drastisch ändern, so dass eine dynamische Anpassung der Datenverarbeitung erforderlich ist. Eine Möglichkeit der Anpassung ist das gezielte Verwerfen von Daten, auch admission control oder load shedding genannt, um Überlasten zu verhindern und dennoch die Datenverarbeitung auf einem (statistisch) guten Niveau aufrecht zu erhalten. Diese Arbeit erfolgt im Kontext des EU Projekts QualiMaster, in dem eine adaptive Big Data Analyse Infrastruktur entwickelt wird. Die Techniken können aber in ähnlichen Anwendungsgebieten angewendet werden.



### **Was soll ich tun?**

In dieser Arbeit sollen Sie einen existierenden Ansatz zur lastabhängigen Planung und Steuerung von Tatbul et al. [1] für die QualiMaster-Infrastruktur anpassen und mit verschiedenen load-scheduling Methoden [2, 3] erproben. Sie arbeiten sich in den Ansatz und die QualiMaster-Infrastruktur ein, erarbeiten ein Konzept zur Integration des Ansatzes von Tatbul et al., implementieren diesen als Teil der Infrastruktur und führen Experimente zur Messung und Validierung der Effekte durch. Die Infrastruktur realisiert bereits verteilte Verarbeitung von Datenströmen, Monitoring und Adaptionmöglichkeiten. Zudem existieren Pipelines zur Verarbeitung von Daten im Bereich der finanziellen Risikoanalyse.

### **Welche Ergebnisse werden erwartet?**

Ihre Arbeit diskutiert das Design und die Integration des Ansatzes von Tatbul et al. für/in die QualiMaster-Infrastruktur und umfasst eine prototypische Realisierung sowie insbesondere die Validierung ihres Prototyps. Dabei sollen Experimente mit verschiedenen load-scheduling Methoden in verschiedenen Lastszenarien dokumentiert und diskutiert werden.

### **Was bringt mir das?**

Neben den Kreditpunkten und der Note für den oben genannten Typ dieser Arbeit:

- Einblick in aktuelle Forschung im Bereich Big Data
- Einblick in aktuelle Forschung im Bereich Adaptiver Systeme
- Vertiefung ihrer Kenntnisse der Programmiersprache Java

[1] N. Tatbul, U. Cetintemel, S. Zdonik, Staying FIT: Efficient Load Scheduling Techniques for Distributed Stream Processing, VLDB'07, 159-170

[2] S. Chakravarty and Q. Jiang, Stream Data Processing: A Quality of Service Perspective Modeling, Scheduling, Load Shedding, and Complex Event Processing, Springer, 2009

[3] N. Tatbul, QoS-Driven Load Shedding on Data Streams, Workshops XMLDM XML-Based Data Management and Multimedia Engineering (XMLDM '02), 566-576, 2002.

### **Kontakt**

Dr. Holger Eichelberger  
eichelberger@sse.uni-hildesheim.de

### **Aufteilung der Arbeit**

| Theorie | Implementierung | Literatur |
|---------|-----------------|-----------|
| 20%     | 60%             | 20%       |