

Abschlussaufgabe zur Vorlesung „Scalable Data Management“ (Sommersemester 2021)

## Qualitative Charakterisierung und Klassifizierung ausgewählter BigData-Systeme

Im Rahmen der Vorlesung werden Konzepte für skalierbare Datenmanagementsysteme besprochen. Die Diskussion erstreckt sich dabei von den Eigenschaften unterschiedlicher Datenmodelle über Partitionierungsschemata und Synchronisationmodelle in Shared-Nothing- und Shared-Disk-Ansätzen bis hin zu Replikation und verteilter Anfrageausführung. Parallel werden in der Übung die drei Systeme **Redis**, **Mongo** und **Spark** im Detail vorgestellt und auch praktisch evaluiert.

Zur Vorbereitung auf die Prüfung soll als Abschluss der Vorlesung sollen die gesammelten Erfahrungen in eine qualitative Beschreibung und Einordnung münden. Bei der Beschreibung der Systeme ist auf Merkmale einzugehen:

- Einsatzbereich und ggf. kurze Historie
- Datenmodell: welches Datenmodell wird unterstützt? Gibt es modellinhärente Zusicherungen? (z.B. referentielle Integrität im relationalen Datenmodell) Gibt es eine explizite Unterstützung für zeitliche Veränderungen des Datenbestands
- Datenbankanfragesprache: Wie kommuniziert die Anwendung mit dem System? Welche Mächtigkeit besitzt die Anfragesprache? Kann Anwendungslogik „zu den Daten“ geschickt und dort ausgeführt werden (a la „user defined functions“?)
- Architekturansatz: Wird ein Shared-Disk oder Shared-Nothing-Ansatz verfolgt? Gibt es Besonderheiten bei der zugrundeliegenden Rechnerinfrastruktur? Gibt es Besonderheiten bei der Ausführung bzw. Planung von Anfragen?
- Speichermodell: Welche Speicherformate der Datenbestände werden unterstützt? Können externe Daten (z.B. CSV-Dateien) direkt verarbeitet werden? Welche Strategien für Partitionierung werden eingesetzt?
- Replikationsmodell: Unterstützt das System eine Replikation von Datenbeständen? Welche Form der Konsistenz realisiert das System? Wie wird „Availability“ im Fehlerfall realisiert?

Bitte verwenden Sie ein 2-spaltiges Format und nutzen Sie entweder das ShareLatex der TU-Dresden ([tex.zih.tu-dresden.de](http://tex.zih.tu-dresden.de)) oder einen (kostenfreien) Overleaf-Account ([overleaf.com](http://overleaf.com)). Die Formatvorlage laden Sie bitte von der PVLDB-HomePage (<http://vlldb.org/pvlldb/vol15-formatting/>). Für jedes System reservieren Sie bitte ein eigenes Kapitel und planen Sie ca. 1 Seite ein. Um das Arbeiten mit Tabellen zu erlernen/üben, versuchen Sie als Zusammenfassung eine tabellarische Kurzübersicht über die Eigenschaften der jeweiligen Systeme zu erstellen.

Termin für die Einreichung der Ausarbeitung ist jeweils 1 Woche vor Ihrer Prüfung!

Dokument (PDF) bitte an [wolfgang.lehner@tu-dresden.de](mailto:wolfgang.lehner@tu-dresden.de) mit „[SDM] ...“ im Subject schicken