

# Aufgabenstellung „Laborarbeit“ (Big Data Engineering, W3M20027)

## Zielsetzung

Das Ziel dieser Laborarbeit ist es, eine Anwendung im Big Data Engineering umzusetzen, die auf einer Kappa-Architektur basiert und Apache Spark als zentrale Technologie verwendet. Implementieren Sie den Ingestion-Prozess (z.B. auf Basis von Apache Kafka) wobei Daten in Echtzeit verarbeitet werden sollen (optional). Dabei soll der Umfang der beiden Veranstaltungsteile als Grundlage verwendet werden. Alternative Architekturen und Absprachen sind möglich.

Erarbeiten Sie eine detaillierte Systemarchitektur für die geplante Anwendung. Berücksichtigen Sie dabei die Komponenten einer Kappa-Architektur wie Ingestion Layer, Stream Processing Layer und Serving Layer. Der Serving Layer kann dabei die Daten z.B. in einer MariaDB speichern.

Verfassen Sie eine Dokumentation, die den Aufbau, die Implementierung und die Funktionalitäten der Anwendung beschreibt. Fassen Sie die Ergebnisse, Herausforderungen und Lernerfahrungen in einem Abschlussbericht zusammen.

## Umfang der Abgabe / Mindestanforderungen

- Ein PDF pro Gruppe, das alle Bestandteile der Abgabe enthält
- Deckblatt nennt die Gruppenteilnehmer
- Alle notwendigen Komponenten sind vorhanden und werden genutzt
- Daten werden entweder in das System gestreamt oder wiederholt per Batch abgearbeitet
- Die Ergebnisse der Berechnungen werden in einer Datenbank gespeichert

## Wesentliche Bewertungskriterien

- Dokumentation der Anwendung ist vorhanden und Dokumentation enthält
  - o Idee der Anwendung,
  - o Architektur,
  - o Entwurf,
  - o Screencast eines Demosin der Datei README.md
- Vollständigkeit im Sinne der Aufgabenstellung
- Code der Anwendung ist enthalten (im Anhang, beschränkt auf die wesentlichen Teile)
- Dokumentation der Architektur
- Code ist kommentiert, gut formatiert und verständlich
- Nachvollziehbarkeit ist gegeben