

Aufgabenstellungen "Advanced Data Management"

WS 2020/2021

Ziel und Abgabemodalitäten

Ziel der Seminararbeiten soll es sein, die angegebenen Softwarelösungen – soweit möglich – mit den entsprechenden Lösungen aus der Vorlesung zu vergleichen. Dabei soll jeweils herausgearbeitet werden, wo Gemeinsamkeiten und Unterschiede bestehen und welchen Möglichkeiten oder Limitationen sich daraus ableiten lassen.

Zeigen Sie anhand eines nachvollziehbaren Beispiels die Besonderheiten der jeweiligen Lösung auf und geben Sie ebenfalls an, wie die Software in Kubernetes installiert wurde, um von Ihrem Jupyter-Notebook aus verwendet zu werden.

Die Abgabe erfolgt als Jupyter-Notebook (vorzugsweise im ausführbaren Format .ipynb). Der Umfang soll einer inhaltlich vergleichbaren Seminararbeit entsprechen. Deadline ist der 12.04.2020.

Themenliste Dennis Pfisterer

- Thema 1. Apache Spark GraphX (<https://spark.apache.org/graphx/>)
- Thema 2. Apache Hive (<https://hive.apache.org/>)
- Thema 3. Apache Pinot (<https://pinot.apache.org/>)
- Thema 4. Apache Druid (<https://druid.apache.org/>)
- Thema 5. Apache Kylin (<http://kylin.apache.org/>)
- Thema 6. Apache Superset (<https://superset.apache.org/>)

Sollte noch Infrastruktur notwendig sein (wie z.B. ein Hadoop-Cluster, etc.), so wenden Sie sich bitte an Dennis Pfisterer (dennis.pfisterer@dhbw-mannheim.de).