



# Big Data Analytics Organisation

Prof. Dr. Stephan Trahasch  
Hochschule Offenburg

# Vorlesung und Übungen

- Anmeldung zum Diskussionsforum in Moodle  
<https://elearning.hs-offenburg.de/>  
Kurs: Big Data Analytics
- Zugang zu Materialien über Git-Repository bei GitHub. Erstellen Sie einen kostenlosen Account <https://github.com/>
- Repository  
<https://github.com/csog/bigdata19>  
Machen Sie sich mit Git vertraut.
  - <https://rogerdudler.github.io/git-guide/index.de.html>
  - <https://www.atlassian.com/git/tutorials>
- Nutzung des Command Line Interfaces von Git oder bspw. die Anwendung <https://www.sourcetreeapp.com/> von Atlassian.

# Prüfungsleistungen

---

**Vorlesung → Klausur**

**Praktikum, Daniel Müller**

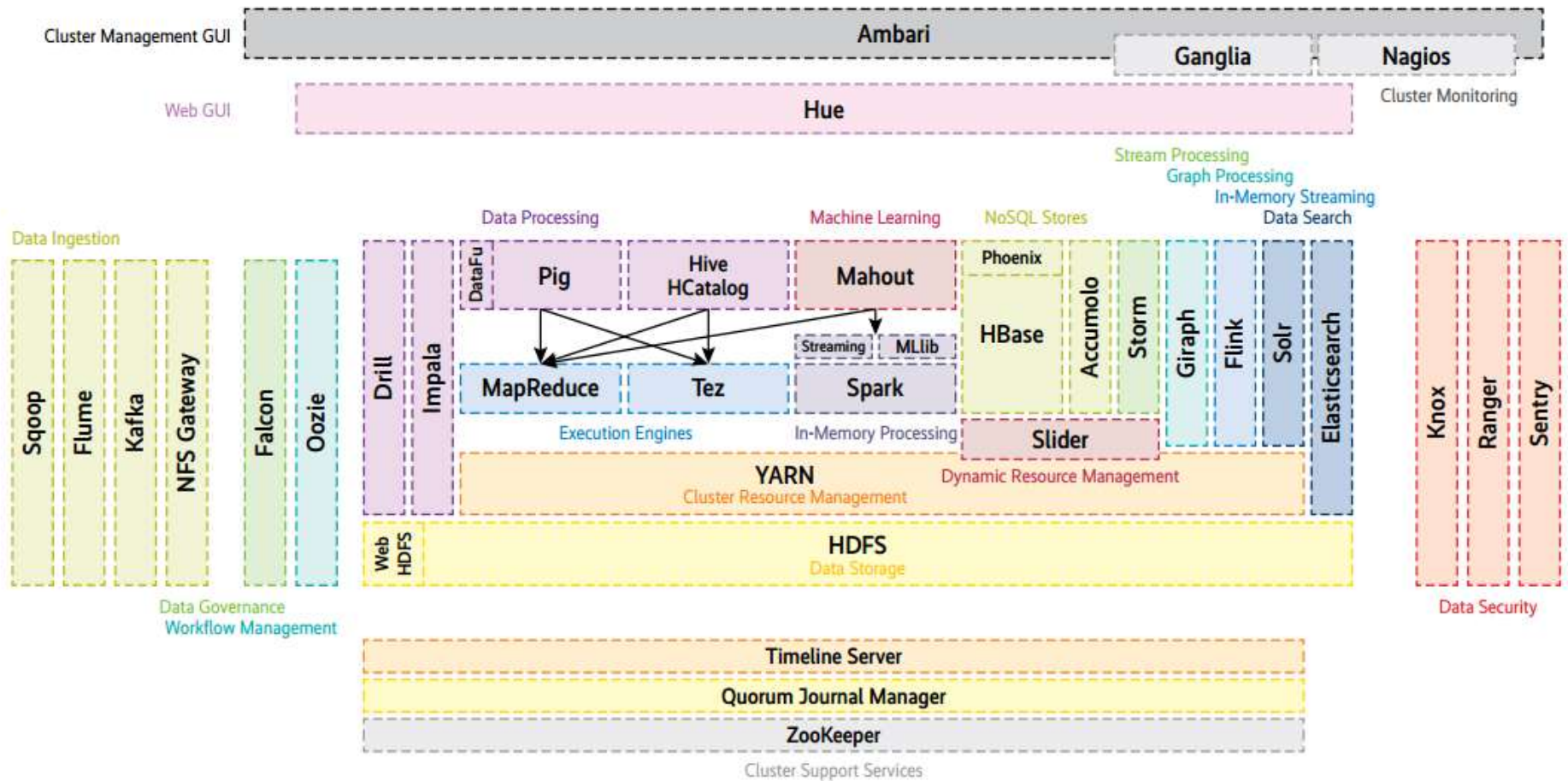
Aufgaben und Termine im Moodle

Es besteht Anwesenheitspflicht

Alle zwei Wochen Übungsblätter (vermutlich 5 insgesamt)

Regelmäßige und sinnvolle Bearbeitung der Übungsblätter

# Themen



# Übungen mit Hadoop-Distribution



# Skript

---

Folien sind auf Deutsch und Englisch.

Folien teilweise von Cloudera und Hortonworks.

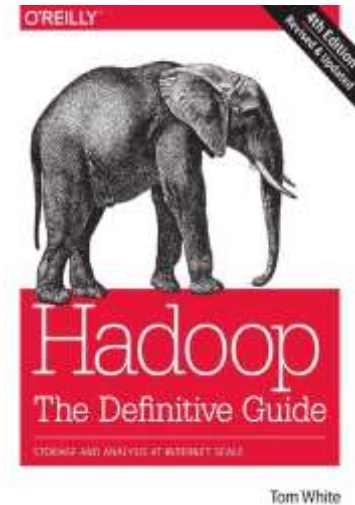
Die Folien sind **kein Skript!**

- Folien sind Basis für Ihre Notizen
- Gemeinsame Basis für Übungen

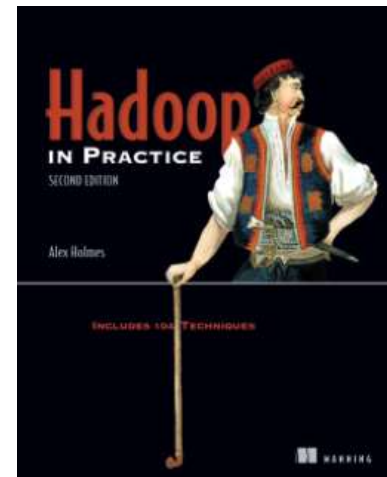
Weitergehende Informationen zum Nachlesen findet man in der angegebenen Literatur bzw. den angegebenen Webseiten.

# Literatur zu Hadoop

White, T. (2015). Hadoop: The definitive guide (4. edition): O'Reilly & Associates.



Holmes, A. (2015). Hadoop in practice (2nd ed.). Shelter Island, N.Y.: Manning.

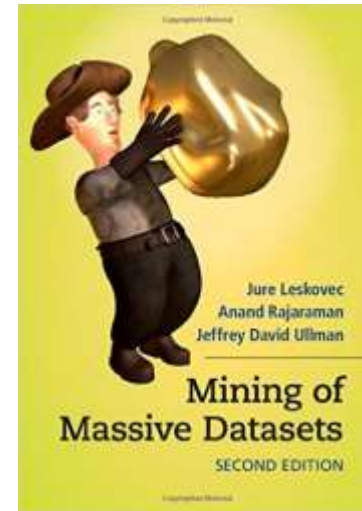


... und Online-Doku zu Hadoop-Tools

# Literatur

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2015). Mining of massive datasets (Reprinted.). Cambridge: Cambridge University Press.

<http://www.mmds.org/>



Marz, N., & Warren, J. (2013). Big data: Principles and best practices of scalable realtime data systems: O'Reilly Media.

