



**Westfälische
Hochschule**

Gelsenkirchen Bocholt Recklinghausen

Shoppers-Challenge

Projektdokumentation

im Masterstudiengang Verteilte Systeme

vorgelegt von: Daniel Hardes, Dennis Miller,
Fabian Paus, Christian Schlütter,
Lutz Kalkofen, Marvin Weck,
Johannes Döing

Modul: Fortgeschrittene Datenbanktechniken (FDB)

Gutachter: Prof. Dr. Benrhard Convent

Abgabetermin: 7. Mai 2015

Inhaltsverzeichnis

1. Einleitung	2
1.1. Aufgabenstellung	2
1.2. Projektmanagement	2
2. Einarbeitung	2
2.1. Datenmodell	3
2.2. Data-Mining-Verfahren	3
2.3. Technologieentscheidung	3
3. Implementierung	3
3.1. Iteration 1	3
3.2. Amazon Web Services	3
3.3. Iteration 2	3
4. Fazit	3
A. Anhang 1	4

1. Einleitung

- Datenbanken Fach -> Projektaufgabe "Big-Data Hadoop"

1.1. Aufgabenstellung

- Kaggle Projekt SShoppers Challenge"

1.2. Projektmanagement

- Phasen - Einarbeitung - Implementierung (iterativ) - Projektabschluss
- Meilensteine - Technologieentscheidung treffen - Erste Submission Wechsel von Lokal VM nach AWS - "Finale Submission Dokumentation abgeschlossen - Präsentation abgeschlossen

TODO: terminieren

- ggf. verwendete Werkzeuge (Git, Drive)

2. Einarbeitung

- Hadoop (Hortonworks VM) - Map / Reduce

2.1. Datenmodell

2.2. Data-Mining-Verfahren

2.3. Technologieentscheidung

3. Implementierung

3.1. Iteration 1

3.2. Amazon Web Services

3.3. Iteration 2

4. Fazit

A. Anhang 1