



Fakultät Informatik, Mathematik und
Naturwissenschaften
Studiengang Informatik Master

**Abstrakt zum Oberseminar
Datenbanksysteme: Aktuelle Trends**

Data Lakes

Autor: Raphael Drechsler
Matrikelnr.: 69872
Abgabedatum: ??.??..2018

1 MOTIVATION – TAUCHEN SIE EIN!

DRT:

Begriff ist dann und dann gehyped
 Anschließend gab es Kritik bis hin zu einem Artikel über Fake News?
 Zum Zeitpunkt der Recherche kein deutsch-Sprachiger Artikel auf Wiki
 Motivation besteht darin diese Unklarheiten zu beleuchten, klären was ein Data-Lake ist und sich mit der Frage "Fake-News" auseinanderzusetzen.

2 DEFINITIONSFRAGE "DATA LAKE"

DRT:

Kein Akademischer Ursprung. Dixon hat's gemacht.

2.1 Definition nach Dixon

2.2 Pentaho-Architektur 2010

2.3 Begriffs-Chaos

3 WIE FUNKTIONIERT EIN DATA LAKE?

3.1 Aufbau

3.2 Workflow

Allgemein und die Rolle des Data Scientist

STORAGE

INGESTION

PROCESS

CONSUMPTION

MONITORING

DATA GOVERNANCE

4 DATA SWAMPS: KRITIK AM DATA LAKE

Mögliche Darstellungen die es so gibt:

- Sumpf
- Finnland
- Flohmarkt

Gartners wesentliche Punkte

Battle: Gartner vs. Dixon und DL revisited → ungenaue Definition ein Problem zu kritisieren oder nur ein Kritikpunkt?

In Praxis dann Sean Martin zitieren.
Trend des Vorsichtiger werdens und die Flut kommen sehen. Paradigmenwechsel.

5 FAKE-NEWS! EXISTIEREN DATA LAKES ÜBERHAUPT?

Blogeintrag nur nennen und erste Zeile zitieren.

Nach Recherche lassen sich da schon ein paar Firmen finden, die Data Lake Lösungen anbieten. → auflisten

Nach weiterer Recherche auch success-stories auffindbar. Also irgendwie schon.

Die wesentliche Frage ist allerdings die Definitionsfrage. Selber Schluss im Blogeintrag. Lösung die dem Paradigma grundlegend folgen gibt es. Jetzt im Auge des Betrachters ob man das Kind beim Namen nennt oder nicht.

LITERATUR/QUELLEN

- [1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition, 2009.
- [2] Jonathan Hartley. Time complexity - python wiki. <https://wiki.python.org/moin/TimeComplexity>. Zuletzt bearbeitet: 05.06.2017, Zugriff: 01.02.2018.
- [3] Spotify uris and ids - spotify web api user guide. <https://developer.spotify.com/web-api/user-guide/#spotify-uris-and-ids>. Zugriff: 01.02.2018.
- [4] A java wrapper/client for the spotify web api. <https://github.com/thelinmichael/spotify-web-api-java>.
- [5] The python standard library documentation - 26.4.3. profile and cprofile module reference. <https://docs.python.org/2/library/profile.html#module-cProfile>. Zugriff: 01.02.2018.
- [6] Snakeviz - a browser based graphical viewer for the output of python's cprofile module. <https://jiffyclub.github.io/snakeviz/>.
- [7] The python standard library documentation - 15.3. time — time access and conversions. <https://docs.python.org/2/library/time.html#time.time>. Zugriff: 02.02.2018.