

Data Lakes

RAPHAEL DRECHSLER

HTWK Leipzig

Fakultät Informatik, Mathematik und Naturwissenschaften

Studiengang Informatik Master - Matrikelnr. 69872

30.30.2018

Zusammenfassung

Begriff ist 2010 entstanden und in den letzten Jahren gehyped (Quellen 1,2,3) Anschließend gab es Kritik bis hin zu einem Artikel über Fake News? (Quelle 4) 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.

I. DEFINITIONSFRAGE "DATA LAKE"

Jetzt gehts los. **DRT:** Kein Akademischer Ursprung. Dixon hats gemacht. [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15]

i. Definition nach Dixon

Quellen: 5,6

ii. Pentaho-Architektur 2010

Quellen: 6

iii. Begriffs-Chaos

Quellen: 7

II. WIE FUNKTIONIERT EIN DATA LAKE?

i. Aufbau

QUELLEN : 8, 9

ii. Workflow

Allgemein und die Rolle des Data Scientist
QUELLEN : 8, 9

Storage QUELLEN : 8, ?

Ingestion QUELLEN : 8, 10, 11

Process QUELLEN : 8, 12

Consumption QUELLEN : 8, 12

Monitoring QUELLEN : 8, ?

Data Governance QUELLEN : 8, 12

III. DATA SWAMPS: KRITIK AM DATA LAKE

Mögliche Darstellungen die es so gibt:

- Sumpf
- Finnland
- Flohmarkt

Gartners wesentliche Punkte

Battle: Gartner vs. Dixon und
QUELLEN : 3, 12, 13
DL revisited → ungenaue Definition ein Problem zu kritisieren oder nur ein Kritikpunkt?
Quellen: 14,15,

In Praxis dann Sean Martin zitieren.
Trend des Vorsichtiger werdens und die Flut kommen sehen. Paradigmenwechsel. Quelle: 1

IV. FAKE-NEWS! EXISTIEREN DATA LAKES ÜBERHAUPT?

Blogeintrag nur nennen und erste Zeile zitieren.

Quelle: 4

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

Nach weiterer Recherche auch successstories 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.

V. METHODS

Maecenas sed ultricies felis. Sed imperdiet dictum arcu a egestas.

- Donec dolor arcu, rutrum id molestie in, viverra sed diam
- Curabitur feugiat
- turpis sed auctor facilisis
- arcu eros accumsan lorem, at posuere mi diam sit amet tortor
- Fusce fermentum, mi sit amet euismod rutrum
- sem lorem molestie diam, iaculis aliquet sapien tortor non nisi
- Pellentesque bibendum pretium aliquet

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultricies augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet

Tabelle 1: *Example table*

Name		
First name	Last Name	Grade
John	Doe	7.5
Richard	Miles	2

aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Text requiring further explanation¹.

VI. RESULTS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultricies augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$e = mc^2 \quad (1)$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultricies augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit

¹Example footnote

amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

LITERATUR

- [1] Alan Morrison Brian Stein. Data lakes and the promise of unsiloed data. Technical report, PricewaterhouseCooper, 2014.
- [2] James Ovenden. Say goodbye to your data lake in 2017. <https://channels.theinnovationenterprise.com/articles/say-goodbye-to-your-data-lake-in-2017>. Veröffentlicht: 10.01.2017, Zugriff: 29.04.2018.
- [3] Rob van der Meulen Janessa Rivera. Gartner says beware of the data lake fallacy. <https://www.gartner.com/newsroom/id/2809117>. Veröffentlicht: 28.07.2014, Zugriff: 29.04.2018.
- [4] Uli Bethke. Are data lakes fake news? <https://sonra.io/2017/08/08/are-data-lakes-fake-news/>. Veröffentlicht: 08.08.2017, Zugriff: 29.04.2018.
- [5] James Dixon. James dixon's blog: Pentaho, hadoop, and data lakes. <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>. Veröffentlicht: 14.10.2010, Zugriff: 29.04.2018.
- [6] James Dixon. Pentaho hadoop series part 1: Big data architecture. https://www.youtube.com/watch?v=tR_yLsr87Uk. Upload: 24.10.2012, Zugriff: 01.02.2018.
- [7] Lance Weaver. Why companies are jumping into data lakes. <https://blog.equinix.com/blog/2016/11/10/why-companies-are-jumping-into-data-lakes/>. Veröffentlicht: 10.11.2016, Zugriff: 29.04.2018.
- [8] Christian Mathis. Data lakes. *Datenbank-Spektrum*, 17(3):289–293, 2017.
- [9] Bhushan Satpute. Enterprise data lake: Architecture using big data technologies. https://www.youtube.com/watch?v=hsq4s_19ZDM&t=380s. Upload: 28.03.2016, Zugriff: 01.02.2018.
- [10] Nathan Marz. How to beat the cap theorem. <http://nathanmarz.com/blog/how-to-beat-the-cap-theorem.html>. Veröffentlicht: 13.10.2011, Zugriff: 29.04.2018.
- [11] Jay Kreps. Questioning the lambda architecture. <https://www.oreilly.com/ideas/questioning-the-lambda-architecture>. Veröffentlicht: 02.07.2014, Zugriff: 29.04.2018.
- [12] Alex Gorelik. How to build a successful data lake: Talk at hadoop summit 2016. <https://www.youtube.com/watch?v=zHokpz3qNJ8&t=610s>. Upload: 29.06.2016, Zugriff: 01.02.2018.
- [13] Martin Willcox. What is a data lake, anyway. <https://www.youtube.com/watch?v=N00r452uQM0&t=835s>. Upload: 10.02.2015, Zugriff: 01.02.2018.
- [14] James Dixon. James dixon's blog: Data lakes revisited. <https://jamesdixon.wordpress.com/2014/09/25/data-lakes-revisited/>. Veröffentlicht: 25.09.2014, Zugriff: 29.04.2018.
- [15] James Dixon. Pentaho hadoop series part 5: Big data and data warehouses. <https://www.youtube.com/watch?v=1CG01JmKp2Y&t=2s>. Upload: 24.10.2012, Zugriff: 01.02.2018.