

Inhalt

Vorwort des Reihenherausgebers — V

- 1 Der Begriff Big Data — 1**
 - 1.1 Hands on — 4
 - 1.1.1 Virtualisierung — 5
 - 1.1.2 Cloudlösungen — 5
 - 1.1.3 Lokale Installationen — 5
 - 1.2 Die Dimensionen von Big Data — 6
 - 1.2.1 Volume – Größe — 6
 - 1.2.2 Velocity – Schnelligkeit — 7
 - 1.2.3 Variety – Vielfältigkeit — 8
 - 1.3 Wrap-Up — 9

- 2 Das Hadoop File System und Map-Reduce — 11**
 - 2.1 Hadoop Distributed File System — 12
 - 2.2 Map-Reduce — 15
 - 2.3 Yet Another Resource Negotiator — 18
 - 2.4 Hello World in Big Data — 18
 - 2.5 Den Planeten erkunden — 29
 - 2.5.1 XML-Input — 31
 - 2.5.2 Counter — 42
 - 2.6 Wrap-Up — 44

- 3 Jetzt aber Flink — 45**
 - 3.1 Die Natur von Streams — 45
 - 3.2 Flink WordCount — 46
 - 3.3 Textdaten — 53
 - 3.4 Eventverarbeitung — 58
 - 3.4.1 Komplexe Eventverarbeitung — 61
 - 3.4.2 Verknüpfung von Events — 65
 - 3.5 Machine Learning — 68
 - 3.6 Wrap-Up — 69

- 4 Die Vielfalt macht's — 71**
 - 4.1 Loggrabber — 72
 - 4.2 Planet OSM — 80
 - 4.3 Flink-NiFi-Connector — 89
 - 4.4 Performance Monitoring — 95
 - 4.4.1 MiNiFi — 106
 - 4.5 Wrap-Up — 106

VIII — Inhalt

5	Was es noch benötigt — 109
5.1	Big-Data-Architekturen — 109
5.2	Laufzeitverhalten — 113
5.3	Datenbanktypen — 122
5.4	Serialisierung — 127
5.4.1	Wrap-Up — 136
A	Appendix — 137
A.1	Maven — 137
A.2	Lizenzen und Urheberschaft — 137
A.2.1	OpenStreetMap — 137
A.2.2	Apache-Lizenz 2.0 — 138
A.2.3	JSON Lizenz — 145
Literatur — 147	
Stichwortverzeichnis — 149	