



Inhaltsverzeichnis

Pavlo Baron

Big Data für IT-Entscheider

Riesige Datenmengen und moderne Technologien gewinnbringend nutzen

ISBN (Buch): 978-3-446-43339-7

ISBN (E-Book): 978-3-446-43392-2

Weitere Informationen oder Bestellungen unter

<http://www.hanser-fachbuch.de/978-3-446-43339-7>

sowie im Buchhandel.

Inhalt

Vorwort	XI
1 Management Summary	1
2 Was?	7
2.1 Mein klassisches Business ist konkurrenzlos, nicht wahr?.....	8
2.1.1 Der heimliche Markteroberer: Amazon	8
2.1.2 Der klare Fall: Google.....	11
2.1.3 Der zunehmende Lebensinhalt: Facebook.....	13
2.1.4 Nachrichtenquelle Nr. 1: Twitter.....	15
2.2 Die wilden Kleinen	16
2.2.1 Das unscheinbare Orakel: Prismatic	17
2.2.2 Der stille Mitlauscher: Boundary	18
2.3 Big Data ist das neue Öl, nicht wahr?	18
2.4 Technologie steht im Vordergrund, nicht wahr?	20
2.5 Die Essenz von Daten ist angewendete Information	21
2.6 Die „Big Data“-Schmerzen und -Symptome	22
2.6.1 Die 3 V's – ganz einfach, oder?	23
Velocity	23
Volume	24
Variety	24
2.6.2 Das lokale Denken: Business-Verbesserung	25
Bessere Angebote durch mehr Wissen	26
Menschen zusammenbringen	27

2.6.3	Das kostenbewusste Denken: Kostenreduktion	28
	Geld mit günstiger Maschinerie und geeigneten Technologien sparen	28
	Kosten durch Analyse drücken.....	29
2.6.4	Die Wachsamkeit: Die Gefahren aus dem bösen Internet	30
	Enterprise-Level Stalking	30
	Fraud Detection.....	31
2.6.5	Das globale Denken: Weltverbesserung	32
	Öffentlichen Verkehr endlich funktionsfähig machen.....	32
	Precogs implementieren.....	33
3	Wie?	35
3.1	Typische falsche Hoffnungen	36
3.1.1	Große Daten sind wie alle anderen Daten, nicht wahr?	37
3.1.2	Tools machen es für mich, nicht wahr?	38
3.1.3	Meine Prozesse müssen sich nicht ändern, nicht wahr?	39
3.2	Angewandte Wissenschaft ist der entscheidende Marktvorteil	40
3.2.1	Mathe pauken	40
3.3	Ihr Staff – die einzig sinnvolle Investition	42
3.3.1	Ich kann Big Data mit normalen Entwicklern, nicht wahr?	43
3.3.2	Also, alle Entwickler zurück in die Schule schicken?	46
3.3.3	Die neuen Berufsgruppen	47
3.3.4	Polyglott, auf jeder Ebene	48
3.4	„Big Data“-Schmerzen lassen sich erzeugen	50
3.4.1	Alles protokollieren, was Nutzer tun	50
3.4.2	Nutzer solange wie möglich beim Online-Angebot halten	52
3.4.3	Nutzer anregen, mehr zu tun und Daten zu hinterlassen	54
3.4.4	Soziale Netzwerke anzapfen.....	55
3.4.5	Crowdsourcing	58
3.4.6	Wissen, wer wo ist	59
3.5	IT-Strategie im „Big Data“-Licht.....	61
3.5.1	Daten bzw. Informationen sind gemeinsame Sprache von Business und IT	61
3.5.2	Daten sind Ihr Gold	63

Wie werden Daten klassischerweise behandelt?	64
Wie werden große, chaotische Daten behandelt?	66
Wie leitet man Informationen aus vielen und chaotischen Daten ab?	69
Wie hält man Unmengen chaotischer Daten sauber?	71
Wie kombiniert man mehrere Datenquellen?	71
3.6 Was bedeutet es, datenzentriert zu arbeiten?	73
3.6.1 Statistiken lügen nicht	73
3.6.2 Der entscheidende Unterschied zwischen schnell und gleich .	74
3.6.3 Die Macht der Suche	77
3.6.4 Die Macht der Empfehlung.....	79
3.6.5 Die Macht des Bildes	80
3.7 IT-Instrumente richtig nutzen oder richtige Instrumente nutzen	82
3.7.1 Commodity Hardware nutzen	83
3.7.2 Alles aus der Technik herauspressen.....	85
3.7.3 Dort in den Wolken.....	85
3.7.4 Die Open Source Welt hat klar die Nase vorn	87
3.8 „Big Data“-Perspektiven.....	89
3.8.1 Wie speichert man große Datenmengen?	89
Der schwierige Weg der RDBMS	89
Tipps und Tricks	91
Der Preis der Verteilung.....	93
Und noch ein paar Randthemen	95
3.8.2 Wie bereitet man große Datenmengen auf?	97
3.8.3 Wie stellt man große Datenmengen bereit?.....	98
3.8.4 Wie analysiert man große Daten in Echtzeit?	100
3.8.5 Wie visualisiert man große Datenmengen?	102
3.9 Von der Hand in den Mund	103
3.10 Segen für Europa, Fluch für Big Data: Datenschutz	105
4 Womit?	109
4.1 Willkommen im Zoo der „Big Data“-Technologien.....	117
4.2 Einige theoretische Grundlagen	117
4.2.1 Verteilte Systeme bzw. verteilte Data Stores.....	117

Hashing und Delta-Tracking	119
Replikation und Redundanz	121
Objekte versionieren	126
Sharding	130
Herr der Ringe	134
Anti-Entropy und Garbage Collection	136
Infrastrukturinformationen aktuell halten	138
CAP – nicht die Kappe	138
Eventual Consistency	140
4.2.2 Caching bzw. In-Memory-Datenhaltung	144
4.2.3 Graphen – ein Sonderfall	148
4.3 Analytics sind das A und O. Aber womit machen?	149
4.4 Was ist NoSQL?	151
4.4.1 Key/Value Stores	152
4.4.2 Document Stores	155
4.4.3 In-Memory Stores	156
4.4.4 Graphen, mal wieder ein Sonderfall	157
4.4.5 ... und noch weitere Hundert Stores	159
4.4.6 Und was ist mit der Cloud?	159
4.5 Was ist NewSQL?	160
4.6 Wo kommen klassische RDBMS ins Spiel	161
4.7 Was ist MapReduce?	162
4.8 Und welche Programmiersprache ist nun die beste für „Big Data“? ...	166
4.8.1 Kann/muss ich alles mit Java machen?	166
4.8.2 Ist JavaScript nicht etwa nur für Script-Kiddies?	167
4.9 Analyse von Logs	167
4.10 Time-Series-Daten	169
4.11 Umgang mit Commodity Hardware	170
4.12 Wie kann ich es richtig flott machen?	172
4.12.1 Was ist (Fast-)Echtzeit?	172
4.12.2 Die superschnelle Technik	173
4.13 Suche, Index und Secondary Index	175
4.14 Die Antwort der Dinosaurier: die Appliances	176
4.15 Kann ich das alles mit einem einzigen Tool machen?	177

4.16	Visualisierung	179
4.17	Wohin mit meinen aktuellen Errungenschaften?	181
4.17.1	Wohin mit meiner Enterprise-Architektur?	181
4.17.2	Wohin mit meinem Enterprise-Datenmodell?	182
4.17.3	Wohin mit meinem MDM?	182
4.17.4	Wohin mit meinem Enterprise-Technologie-Stack?	184
4.17.5	Wohin mit meiner SOA?	185
4.17.6	Wohin mit meiner BI-Lösung?	186
4.17.7	Wohin mit der Cloud?	188
4.18	Ist „Big Data“ also Hadoop mit einem Schuss MongoDB?	190
5	Way Forward	193
5.1	Aktuelle Trends	194
5.1.1	NoSQL und der Rest der Welt	194
5.1.2	Hadoop und der Rest der Welt	196
5.1.3	Hauptspeicher und Plattenspeicher werden immer billiger	197
5.1.4	Die Monster wachen auf	198
5.2	Wie starte ich mit Big Data?	200
5.3	Ich habe schon Big Data. Wie mache ich weiter?	201
5.4	ROI, TCO und dergleichen und worauf es ankommt	203
5.5	Wo und wie kann ich mich weiter informieren?	205
5.6	Persönlicher Blick in die ferne Zukunft	208
	Stichwortverzeichnis	211