

HANSER



Leseprobe

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.

1

Management Summary

Bei „Big Data“ geht es um das „Was?“, gefolgt von „Wie?“ und umgesetzt mit Hilfe von „Womit?“.

„Big Data“ beschäftigt sich mit dem Sammeln von Daten und der Gewinnung wertvoller und nützlicher Informationen aus diesen Daten. Und auch mit der Findung und Schaffung sowie Nutzung von weiteren bzw. neuen Datenquellen, um noch mehr nützliche Informationen daraus zu gewinnen. Dabei können die Daten in beliebiger Form und Menge vorliegen.

Wenn Sie die besagten Informationen haben und weiter anreichern, können Sie daraus Schritte zur Verbesserung Ihres Business ableiten. Oder Sie können darauf ganze Strategien aufbauen. Oder, wenn Sie klein genug denken, einfach Ihren Kunden in sozialen Netzwerken nachspionieren und abhängig von ihrem Verhalten dort deren Scoring anpassen.

Oder wenn Sie groß genug denken, können Sie diese Information dazu nutzen, die Welt zu einem besseren Ort zu machen, das Leben Ihrer Mitmenschen zu vereinfachen oder gar Epidemien zu bekämpfen oder Verbrechen vorzubeugen.

„Big Data“ ermöglicht es den Menschen mit höherer fachlicher Kompetenz, Entscheidungen so einfach und qualifiziert wie möglich zu treffen. Das menschliche Gehirn ist eine sehr gute Maschine, aber es ist sehr schlecht darin, aus Milliarden von aufeinander aufbauenden oder unabhängigen Faktoren und Ereignissen eine qualifizierte Entscheidung abzuleiten. Denn die linke Hirnhälfte ist nun mal langsam und rechnet sequentiell¹.

Computer sind wiederum anders. Sie können nur mit sehr großer Mühe unabhängige visuelle und akustische Faktoren parallel zu einem inneren Bild zu verarbeiten, wie es das menschliche Gehirn in der rechten Hälfte tut. Dafür

¹ <http://www.hanser-fachbuch.de/buch/Pragmatisches+Denken+und+Lernen/9783446416437>

kann aber der Computer wahnsinnig schnell und fehlerfrei logische Ketten mit unzähligen Faktoren durchrechnen.

Allein dem Computer das Treffen von qualifizierten und weitreichenden Entscheidungen zu überlassen, ist mehr als falsch. Er hat keinen Bauch, in dem er Gefühle hat. Er kann sich nicht in die aktuelle Situation mithilfe von Intuition hineinversetzen, denn er hat keine. Er ist sehr dumm darin, menschliche Emotionen zu verstehen. Und wenn er es könnte, wäre es mehr als langweilig. Zumaldest für mich persönlich.

Was das „Big Data“ so spannend und nützlich macht, ist die Möglichkeit, die Computerfähigkeiten mit dem menschlichen Gehirn zu kombinieren, ohne dabei Cyborgs zu produzieren. Der Computer errechnet aus Milliarden von Faktoren ein paar Entscheidungsoptionen – und das immer und immer wieder, und der Mensch trifft die finale Entscheidung mit all seiner Intuition, dem Situationsgefühl und seiner Erfahrung.

Ich werde in dem Buch viel über die Datenzentriertheit sprechen, allerdings nicht um jeden Preis. Man darf nicht in die falsche Bequemlichkeit verfallen, Maschinen das Denken zu überlassen, denn damit kann man nur scheitern².

„Big Data“ ist also die Informations- und Technologiegrundlage für die Empfehlungssysteme bzw. Entscheidungsunterstützungssysteme. Die Technologie, also das „Womit?“, ist dabei weit nachgelagert. Als Erstes suchen Sie nach dem „Was?“ in irgendwelchen Daten, die Sie bereits haben oder erst benötigen. Wenn Sie wissen, „was“ Sie wissen wollen, kommt das „Wie?“ ins Spiel, welches Ihnen aus der analytischen bzw. generell wissenschaftlichen Perspektive heraus hilft, das „Was?“ zu beantworten. Und das „Womit?“ ergibt sich dann automatisch, weil Sie technische Unterstützung dafür benötigen werden.

Der Zoo an Technologien und Anbietern im Bereich von „Big Data“ ist schier riesig und unüberschaubar geworden. Und am Ende des Tages hören Sie nur Hadoop, MongoDB und Business Intelligence aus dem Hause EMC. Und Facebook, natürlich. Und dann schreiben Sie in Ihre Strategiefolien an irgendeiner recht zentralen Stelle „Big Data“. Und das höhere Management fragt Sie ununterbrochen und mit Nachdruck, warum Sie immer noch kein Hadoop machen. Habe ich recht? Ehrlich, habe ich nicht recht?

Nein, ich habe unrecht. Eigentlich erwartet man von Ihnen oder Sie erwarten es von sich selbst, Innovation ins Unternehmen zu bringen und diese voranzutreiben. Man erwartet von Ihnen, dass die IT Ihres Unternehmens dem Business dabei hilft, bei Marktentwicklungen immer eine Nase vor allen anderen zu sein. Man erwartet von Ihnen, dass sie die Business-Bedürfnisse mit IT-

²<http://timvangelder.com/2009/05/07/dangers-of-datacentrism/>

Lösungen umsetzen. Und man erwartet von Ihnen, dass Sie Ihre Kosten unter Kontrolle haben.

Ich helfe Ihnen mit diesem Buch. Ich sortiere für Sie den besagten Zoo aus- einander, kaue Ihnen einzelne Teilgebiete vor und mache es Ihnen leichter, Ihre qualifizierten und tragfähigen Entscheidungen qualifiziert und tragfähig zu treffen. Und sich Ihren Vorgesetzten und/oder Kunden gegenüber als kompetenter Partner darzustellen. Ich zeige Ihnen hier, wie Sie mithilfe von „Big Data“ innovativ und konkurrenzfähig werden und bleiben. Und kostenbewusst.

Der Preis dafür sind ein paar Euro für das Buch und ein wenig Zeit und Interesse.

Ende der Management Summary.

Nun, lieber IT-Entscheider mit vollem, mehrdimensionalem Terminkalender und dünnem Geduldsfaden, wenn Sie möchten, können Sie an dieser Stelle gerne mit dem Lesen aufhören. Denn genau das hätten Sie höchstwahrscheinlich gemacht, wenn Sie mich mit dem Schreiben beauftragt hätten. Wir Techniker haben dafür ein Kürzel: „tl;dr“. Es ist Internet-Slang und heißt „too long; didn't read“, zu Deutsch also „zu lang; hab's nicht gelesen“. Wir wissen sehr wohl, welches Schicksal in der Regel die restlichen Seiten eines Dokuments jenseits der Management Summary erwartet.

Aber Sie haben doch nicht wirklich gedacht, ich mache es Ihnen so leicht, oder? Sie müssen sich schon durch das ganze Buch „quälen“, das erspare ich Ihnen nicht. Vielleicht als kleiner Racheakt für all die Management Summaries, die ich in meinem Leben schreiben musste und immer noch schreiben muss. Aber vielmehr deswegen, dass es nun wirklich an der Zeit ist, das „Big Data“-Haus in Ordnung zu bringen, vor allem aus der Managementperspektive.

Ich habe für Sie in diesem Buch alles zusammengefasst, was Sie aktuell und in Zukunft über „Big Data“ wissen und verstehen sollten bzw. eigentlich müssen, um qualifizierte Entscheidungen treffen zu können. Denn genau dafür werden Sie bezahlt, nicht wahr? Genau. Und ich werde dafür bezahlt, Sie bei diesen Entscheidungen zu unterstützen. Nicht durch die Tantiemen. Denken Sie ruhig größer.

Betrachten Sie also das ganze Buch als Management Summary, denn wenn ich Ihnen hier die gesamte Materie in epischer Breite beschreiben würde, lägen jetzt anderthalb Meter Brockhaus-Enzyklopädie vor Ihnen. Stattdessen halten Sie einen recht dünnen Band in weichem Umschlag in der Hand. Machbare Aufgabe, oder? Denke ich auch.

Das Buch richtet sich an alle, die in der IT Entscheidungen treffen dürfen bzw. müssen: IT-Leiter, CIOs, CTOs und Architekten, Lead Developer und wie Sie alle da draußen heißen mögen. Sie können das Buch sogar Ihrem Top-

Management und Ihren Kunden geben – dann haben Sie schneller eine gemeinsame Gesprächsbasis geschaffen und können in die konkreten Details einsteigen.

Das Buch arbeitet sich entlang der drei Fragen „Wie?“, „Was?“ und „Womit?“ und gibt Antworten darauf – im Kontext von „Big Data“. Die Antworten beziehen sich vor allem im Technologischen auf alle drei Fragen, doch dieses schrittweise Denken führt meiner Erfahrung nach ganz gut zum Ziel, und das nicht nur bei „Big Data“.

„Was?“, „Wie?“ und „Womit?“ also. Ja, das ist definitiv meine Absicht, das Ganze ein bisschen nach „Wer? Wie? Was?“ oder den fünf „W“s der ersten Hilfe im Straßenverkehr klingen zu lassen. Soviel, um wütenden E-Mail-Fragen vorzubeugen. Ich spiele den Besserwisser, der Ihnen in Ihrer verantwortungsvollen Position etwas über das Leben erzählt. Ja, das tue ich. Nutzen Sie den Vorteil des Lesers, denn ich habe mich jahrelang mit dem Thema auseinander gesetzt und gebe nun komprimiert meine Erkenntnisse und Erfahrungen an Sie weiter.

Die drei besagten Fragen werden also in den jeweiligen Buchkapiteln eingehend diskutiert und beantwortet. Am Ende des Buches findet sich der Ausblick in die Zukunft und Handlungsempfehlungen. Mehr ist zu diesem Thema auf oberster Strukturebene nicht zu sagen – da müssen Sie schon in die einzelnen Inhalte einsteigen. Auch hier halte ich gegen die typische, gesichtslose Kurzform der Management Summary.

Ich bin aus moralischen und einfach menschlichen Gründen nicht mit allem einverstanden, was ich Ihnen als Möglichkeiten in dem Buch aufzeige. „Big Data“ ist ein dünnes Eis, denn früher oder später wird es mehr oder weniger sensible und persönliche Daten von Menschen betreffen. Ich spiele einfach einen motivierenden Berater für Sie, der Sie auf die Reise durch dieses stellenweise verminzte Feld mitnimmt. Die journalistische Pflicht zwingt mich zur Vollständigkeit und Neutralität in alle Richtungen, was allerdings doch nicht immer möglich ist.

Außerdem werde ich im ganzen Buch durchgehend „Big Data“ in Anführungszeichen setzen, weil ich diesen Begriff für unsinnig und inzwischen allzu gehypt halte. Um Ihnen das Verständnis zu erleichtern, bleibe ich jedoch bei „Big Data“, gebe Ihnen aber mit den Anführungszeichen immer das Signal, nicht auf den Begriff selbst, sondern auf die Inhalte und Herausforderungen zu achten. Und ich hoffe, am Ende der Lektüre verwerfen Sie gedanklich den Begriff und nennen die Sache bei dem Namen, der für Sie und Ihr Unternehmen am komfortabelsten ist.

Ich habe das Buch komplett in meiner Freizeit und basierend auf meiner eigenen, neutralen Meinung geschrieben. Es repräsentiert nicht unbedingt an allen Stellen die Meinung und die Strategie meines aktuellen Arbeitgebers oder

eines meiner ehemaligen oder auch künftigen Arbeitgeber. Mir ist es wichtig, Ihnen eine herstellerunabhängige Sicht zu vermitteln, frei von Zwängen, die das Thema „Big Data“ zu einem Hype gemacht haben. Falls diese Sicht Sie in irgendeiner Weise auf die Palme bringt, so richten Sie bitte Ihren Wut direkt gegen mich und gehen Sie mit mir in eine konstruktive Diskussion – ich freue mich darauf! So, das wäre jetzt auch erledigt.

Und damit Sie wissen, wer Sie hier so frech berät, hier sind ein paar Takte über meine Wenigkeit.



Ich bin seit über 20 Jahren in der IT-Industrie unterwegs. Und ich berühre bis heute die Tastatur. Nicht (nur), um hübsche Slides zu erstellen, sondern um konkret zu programmieren und Systeme zu administrieren, Messungen durchzuführen, automatisierbare Tests zu implementieren etc. Und das in verschiedenen Sprachen und Technologien. Und darauf bin ich stolz.

Wenn es sich für Sie aber allzu sehr nach „einfachem Entwickler“ anhört, hilft Ihnen vielleicht die Information, dass ich u.a. Enterprise-Architekt von Sixt und Chief IT-Architekt von PlanetHome (UniCredit/HypoVereinsbank) war. Auch Entwicklungsleiter in diversen Unternehmen, z.B. bei M-net (Telekommunikation). Auch eine eigene Firma habe ich gehabt. Aktuell bin ich Lead Architekt der codecentric AG. Ich spreche regelmäßig auf nationalen und internationa-

len Konferenzen, wovon zahlreiche Slides und Videos zeugen, die Sie von mir im Web finden können.

Meine aktuellen Schwerpunkte sind verteilte Systeme, Hochskalierbarkeit, Machine Learning, Natural Language Processing. Und das alles mündet in „Big Data“. Ich bin viel unterwegs, um in konkreten Kundenprojekten große „Big Data“-Lösungen auf Basis von Riak, MongoDB und Neo4j zu bauen. Dabei kombiniere ich diese mit vielen weiteren Tools wie Nginx, Redis, RabbitMQ etc., von denen ich Ihnen in diesem Buch berichten werde. Und natürlich fehlt auch große MapReduce-Erfahrung nicht – ich contribute z.B. Code zu Disco und habe mehrere Jahre konkrete Hadoop-Erfahrung zu vermelden, auch in Randbereichen.

Ich habe bereits drei Bücher in Deutschland veröffentlicht, die allesamt ganz unterschiedlich sind: „Pragmatische IT-Architektur“³, „Fragile Agile“⁴ und „Erlang/OTP“⁵. Ich bin generell sehr sarkastisch, was man in den Büchern, in meinen Vorträgen und auch z.B. in meinem „Geek's Guide To The Working Life“⁶ feststellen kann. Eine mehr oder weniger aktuelle Aufstellung meiner Aktivität lässt sich auf meiner Webseite⁷ finden. Die Wahrscheinlichkeit ist aber höher, dass meine Twitter-Meldungen immer die aktuellsten Informationen und Überlegungen enthalten. Dort bin ich als `@pavlobaron` zu finden.

Dessen aber genug. Legen wir los mit dem eigentlichen Thema. Fangen wir mal mit dem „Was?“ an.

³ http://entwickler-press.de/ep/psecom,id,2,buchid,184,p,0,_language,de.html

⁴ <http://www.hanser-fachbuch.de/buch/Fragile+Agile/9783446422582>

⁵ https://www.opensourcepress.de/index.php?226&tt_products=327

⁶ <http://it-republik.de/jaxenter/artikel/Geek%92s-Guide-to-the-Working-Life-2857.html>

⁷ <http://www.pbit.org>