

Chancen und Grenzen: Warum gerade jetzt Big Data

Alfred Schlaucher, Oracle Deutschland B.V. & Co. KG

Wer sich heute mit Big Data beschäftigt, sollte sich von all dem Hype freimachen, der rund um dieses Thema gemacht wird. Dann sollte es gelingen, den passenden (nicht nur technischen) Einstieg zu finden, um die Potenziale von Big Data nicht zu verpassen.

Der erste Aspekt ist schnell abgearbeitet: Der Begriff „Big Data“ wird heute oft so schwammig verwendet, dass alles, was mit Software zu lösen ist, als „Big Data Use Case“ gefeiert wird. Produzieren wir nicht schon seit vielen Jahren Software und lösen damit Business-Probleme? Halten wir uns mit diesem Missverständnis nicht länger auf und wenden uns dem wesentlich spannenderen zweiten Aspekt zu. Viele der mit Big Data diskutierten neuen Lösungen enthalten im Kern schon recht alte und bekannte Ideen. Doch um die Dimensionen und das Potenzial von Big-Data-Lösungen schätzen und begreifen zu lernen, sollten wir uns zunächst mit den geänderten gesellschaftlichen und sozialen Rahmenbedingungen und – komplementär dazu – mit den heute existierenden sogenannten „Enabler-Technologien“ befassen. Lebensgestaltung und technischer Fortschritt bedingen sich gegenseitig. Aus dieser Wechselwirkung entstehen immer wieder auch neue Geschäftsideen, ganze Berufszweige, neue Unternehmen oder gar Branchen.

Heute hat sich das Internet zum Kommunikationsmedium schlechthin entwickelt. Es ist es eine riesige Quelle von Fakten und Informationen über Beziehungen, über Trends oder gesellschaftliche und politische Bewegungen geworden. Mehr noch: Was für das Internet Text, Sprache oder Bilder sind, das sind für die Dinge und Gebrauchsgegenstände um uns herum Abermillionen Sensoren und Messstationen. Betroffen sind fast alle Produkte, die Umwelt, der Verkehr, aber auch Produktions- und Handelsprozesse. Das

Gemeinsame ist die Möglichkeit, alles, was passiert, digital zu analysieren, zu bewerten, neue Schlüsse daraus zu ziehen und entsprechend zu handeln.

Das Einbauen eines Sensors bedeutet allerdings lange noch nicht, dass Daten auch ausgewertet werden. Während es meist noch an Ideen dazu fehlt, ist die „Enabler“-Technik zum Auswerten heute im Gegensatz zur Vergangenheit da. Da sind Hadoop, In-Memory, extrem schnelle Hardware und leistungsfähige Entwicklungs- und Abfragesprachen wie Scala, Python oder R. Sensor- und Internet-Technologie demonstrieren exemplarisch die Entstehung des Phänomens der „Big-Data-Daten“. Die Datenmenge ist jedoch nur ein Aspekt. Hinzu kommen die spontane, wenig planbare, fast beiläufige Entstehung der Daten sowie die Format-neutrale Speicherung beliebiger Datenarten, in der Big-Data-Diskussion als die drei „V“s Data Velocity, Data Variety und Data Volume bekannt. Skalierungsverfahren wie Hadoop oder auch In-Memory ermöglichen die Speicherung auch von spontan auftretenden riesigen Datenmengen und auf diesen Daten eine bis dahin nicht geahnte Verarbeitungs- und Analysegeschwindigkeit.

Wir haben also nicht nur zu fast allem, was uns umgibt, Daten, wir können diese Daten auch nutz- und gewinnbringend einsetzen. Wir müssen es auch tun. An diesem Punkt stehen heute viele Unternehmen, denn es gilt Ideen zu finden, um diese neuen Chancen wahrzunehmen und Potenziale auszuschöpfen. Big Data zu nutzen, ist demnach eher die besondere Aufgabe des Erkennens des Potenzials

der neuen technischen Möglichkeiten und die Umsetzung in neue Geschäftsideen. Dazu ist technisches Verständnis gepaart mit Business-Fokus nötig.

Neue Daten, neue Chancen

Daten sind die besonderen Mittel, aus denen heraus viele neue Ideen rund um Big Data entstehen. Die technische Innovation der letzten Jahre hat den Horizont der nutzbaren Daten erweitert. Wir können diese neuen Daten entsprechend ihrer Verwendung in mehrere Datenarten untergliedern. Die Quelle für diese Kategorisierung in Datenarten stammt aus einer Sammlung mit generischen Big Data Use Cases zu unterschiedlichen Industrien, die man in einem separaten Whitepaper nachlesen kann (siehe [„http://oracledwh.de/downloads/AutoIndex-2.2.4/index.php?dir=Artikel/&file=Big%20Data%20im%20Kontext%20von%20Data%20Warehouse.pdf“](http://oracledwh.de/downloads/AutoIndex-2.2.4/index.php?dir=Artikel/&file=Big%20Data%20im%20Kontext%20von%20Data%20Warehouse.pdf)):

- Daten wie betriebswirtschaftliche Ergebnisse, Verbrauchsdaten, Umsätze, Gewinne, abgesetzte Waren, Menge der Kunden etc. sind klassische Informationen, wie sie in Data-Warehouse-Systemen schon lange vorgehalten und gemessen werden. Das sind überschaubare Datenmengen. Auch in Big-Data-Zeiten nutzen wir diese Daten als Grundlage für viele Entscheidungen. Diese Daten verschwinden nicht, sie werden um die nachfolgenden Datenarten ergänzt.
- Die nächste Gruppe sind Daten, die im Verlauf von Abläufen oder Geschäftsprozessen auftreten, also Log-Informationen oder Sensordaten aus Maschinen.

Es sind Daten mit einem unmittelbaren Zusammenhang zu den verursachenden Aktionen. Man kann direkt Rückschlüsse auf den Verlauf der jeweiligen Aktion schließen. Solche Daten kommen der Vorstellung von Massendaten im Big-Data-Zeitalter sehr nahe.

- Im Gegensatz dazu steht die Gruppe der mittelbaren Daten. Das sind beispielsweise Wetter- und Umweltdaten (Klima) oder Verkehrsmessungen. Solche Daten können einen gewissen Einfluss auf Geschäftsprozesse haben. Vorstellbar ist die Optimierung von Transportrouten per LKW, Bahn oder Schiff. Der Zusammenhang ist jedoch nicht absolut und nur beschränkt vorhanden, aber er ist messbar. Diese Daten sind nicht neu. Sie wurden bislang nur noch zu wenig in Verbindung mit Geschäfts- oder Planungsszenarien gebracht.
- Eine weitere Gruppe sind sogenannte „weiche Daten“ wie Meinungsäußerungen in den sozialen Medien, aber auch Presseartikel und Nachrichten. Hier lassen sich über „Sentiments“, also Textanalysen, Einstellungen und Vorlieben erkennen. Die Nutzung dieser Datenart ist recht neu.
- Eine wichtige neue Gruppe sind die Dokumenten-orientierten Daten, wobei der Begriff „Dokument“ als Platzhalter für alles in Textform Geschriebene, für Bilder, Filme und Tonaufnahmen steht. Wissen über Zusammenhänge entsteht hier durch Vergleiche von vielen ähnlich angelegten Datenobjekten, es ist die Suche nach unbekanntem Auffälligkeiten. Hier hat die Technik in der jüngsten Zeit sehr viele Fortschritte gemacht. Während es heute schon große, mit elektronischen Mitteln verwaltete Medien-Archive gibt, fehlt es immer noch an Anwendungen, die diese Datenart etwa mit den Daten unter Punkt 1 verknüpfen.
- Eine Sonderrolle spielen Bewegungsdaten. Hier ist der Orts- und Zeitbezug der entscheidende Aspekt. Wann treten Ereignisse wo auf? Gibt es Häufungen? Gibt es räumliche oder zeitbezogene Verbindungen, Abhängigkeiten etc.?
- Es fehlt noch die Gruppe der Umfeld-Informationen. Das sind Daten über Wettbewerber oder eine allgemeine Marktlage. Hier handelt es sich ebenfalls um klassische Daten, die immer schon genutzt wurden. Auf solche Informationen

kann man heute durch neuere Techniken wesentlich leichter zugreifen, etwa durch Web-Zugänge auf freie öffentliche Datenquellen oder durch permanentes Monitoring von Web-Auftritten oder Nachrichtenportalen.

Einzelne Datenarten mögen schon seit längerer Zeit in den Unternehmen zur Verfügung stehen. Die mögliche Verknüpfung der Daten untereinander, das nahezu unbegrenzte Fassungsvermögen der Datenspeicher und die Anwendung automatisierter Algorithmen ist das Neue.

In-Memory – heute eine der wichtigsten „Enabler“-Techniken für Big Data

Am Beispiel der In-Memory-Verarbeitung erkennt man gut, in welche Richtung die technische Innovation voranschreitet. Viele der Big-Data-Anwendungsfälle produzieren aus ihrer Natur heraus permanent Einzel-Informationen, die sich in ihrer Summe kaum noch ökonomisch sinnvoll speichern und verwalten lassen. Die Beispiele sind bekannt, ob Sensor- oder Kommunikationsdaten, Bewegungsprofile von vielen Millionen Personen und Fahrzeugen oder Verbrauchsdaten von Energie. Nehmen wir nur Wetterdaten hinzu und schon potenzieren sich die Datenmengen.

Die Hadoop-Technologie mit dem HDFS-Dateisystem ermöglicht dieses massenhafte Speichern von Daten nach heutigen Erfahrungen unbegrenzt. Man nutzt einzelne Knotenrechner (CPU, Memory und Plattenspeicher) zum Speichern von relativ großen Datenblöcken. Wenn die Kapazität der Rechner ausgereizt ist, schaltet man neue Rechner (Knoten) dazu. Um zu wissen, welche Daten (Blöcke) auf welchen Knoten liegen, nutzt man wiederum weitere Rechner. Es entsteht ein riesiges Netzwerk aus Datenknoten und Verwaltungsrechnern. Es ist dieses Konzept, das Skalierung garantiert, während die bislang übliche Aufteilung in SAN und separaten Serverpark beim Kostenvergleich hinterherhinkt.

Das Speichern von Massendaten ist jedoch nur ein erster Schritt. Die Verarbeitung, das Lesen und Analysieren dieser Daten ist der eigentlich wichtige zweite Schritt. Hier gab es gewaltige technische Fortschritte in den letzten Jahren und In-Memory spielt dabei eine zentrale Rolle.

Das ursprüngliche sogenannte „MapReduce-Verfahren“, bei dem Java-Programmcode parallel auf die vielen Datenknoten geschickt wurde („Mapper“-Phase), um die Einzelergebnisse danach über Konsolidierungsprozesse wieder zusammenzufassen („Reducer“-Phase), gelten mittlerweile schon als überholt. Die aktuell viel diskutierte Spark-Technik übernimmt das Skalierungs- und Verteilungsprinzip der vielen Rechner von MapReduce. Spark nutzt dagegen konsequent die Hauptspeicher der jeweiligen Datenknoten und verteilt „In-Memory“ sogenannte „RDD-Objekte“ (Resilient Distributed Dataset) auf die Rechner. Die parallelisierte Verarbeitung schreibt Zwischenergebnisse nicht wieder auf die Datenknoten zurück, wie bei MapReduce, sondern belässt sie als RDD-Bestandteil im Hauptspeicher und reicht sie in Pipeline-Manier an nachfolgende Verarbeitungsschritte weiter. Das Ergebnis sind Performance-Gewinne um Faktoren im Hunderter-Bereich.

Hohe Rechengeschwindigkeit auf riesigen Datenmengen

Ein eindrucksvolles Beispiel für diese extreme Performance-Verbesserung ist die Umstellung der beiden Oracle-R-Mining-Algorithmen „Generalisierte Lineare Modelle“ (GLM) und „Neural Networks“ (MLP) von MapReduce auf Spark. Solche Analysen erfordern hohe Rechenleistungen. In einem Testlauf auf 100 Millionen Sätze konnte man GLM um den Faktor 114 beschleunigen und eine Neural-Network-Analyse auf eine Milliarde Sätze um Faktor 217. Das macht das Ganze natürlich sehr attraktiv, zumal Data Mining durch seinen explorativen Charakter für Big-Data-Analysen genau passt.

Das zweite Beispiel für eine hochperformante In-Memory-Verarbeitung ist das Graphen-Modell, das sich für die Analyse von Zusammenhängen und Beziehungen ideal eignet. Der Anwendungsfall sind Analysen über die Kommunikationsbeziehungen in sozialen Netzwerken. Ein Graph ist, verkürzt gesagt, die Beschreibung der Beziehungen (Kanten) von zwei Objekten (Netzknoten). Einen solchen Graphensatz nennt man Triple. Die drei Aussagen „Meier kennt Müller“, „Müller kennt Schmidt“, „Schmidt kennt Schuster“ ergeben danach ein Graphen-Modell mit der Information „Meier kennt indirekt Schuster“. In sogenannten

```

1: library(twitteR) # Bibliothek für Twitterzugriffe
2: library(tm) # Bibliothek für Textmining
3: tweets <- userTimeline("source", n = nnn) # Lesen von nnn Anzahl Tweets
4: tweets.df <- twListToDF(tweets) # Tweets zum Bearbeiten vorbereiten
5: myCorpus <- Corpus(VectorSource(tweets.df$text)) # In-Memory Objekt "myCorpus" erstellen
6: myCorpus <- tm_map(myCorpus, tolower) # alle Buchstaben klein machen
7: myCorpus <- tm_map(myCorpus, removePunctuation) # alle Punkte, Striche etc. entfernen
8: myCorpus <- tm_map(myCorpus, removeNumbers) # alle Ziffern entfernen
9: myCorpus <- tm_map(myCorpus, stripWhitespace) # Alle Leerzeichen entfernen
10: myCorpus <- tm_map(myCorpus, stopwords("english")) # Entfernen der Stoppwörter, (ich du.)
11: library(SnowballC) # Bibliothek für Grundwort-Bildung
12: myCorpus <- tm_map(myCorpus, stemDocument) # Rückführung aller Wörter auf Grundform
13: tdm <- TermDocumentMatrix(myCorpus) # Matrix mit der Verwendung der Wörter in den Tweets, auf diesem Objekt
wird in der Folge analysiert
14: library(graph) # zusätzliche Netz-Graphik-Bibliothek
15: library(Rgraphviz) # zusätzliche Netz-Graphik-Bibliothek
16: term.freq <- subset(term.freq, term.freq >= 4)
17: plot(tdm, term = freq.terms, corThreshold = 0.05, weighting = T) #Netzgraphik erstellen

```

Listing 1

„Property Graph Data Models“ lassen sich die Beziehungen zusätzlich noch gewichten beziehungsweise attributieren. Leicht vorstellbar, dass solche Netzmodelle gewaltige Ausmaße annehmen können, wenn etwa das Beziehungsgeflecht von Twitter-Nutzern beschrieben wird. Solche Modelle lassen sich in der Oracle-Datenbank, in der Oracle-NoSQL-Datenbank und über HBase in HDFS realisieren.

Oracle liefert heute 35 Algorithmen zur Analyse dieser Netzwerke, etwa zum Herausfinden von Hotspots in Netzen, Finden von Twitters Who-to-follow, Clustern von Netzwerken, Finden von Pfaden und zu vielem mehr. Alle diese Algorithmen nutzen in besonderem Maß die Big-Data-Cluster-Infrastruktur aus, sie laufen auf vielen Knoten parallel und vor allem: Sie laufen ausnahmslos im Hauptspeicher. Um auch besonders große Netze im Hauptspeicher abarbeiten zu können, zerlegt Oracle große Netze in Teilnetze und lädt diese komplett in den Hauptspeicher.

Das erklärt, warum die jüngste Oracle Big Data Appliance Maschine (X5-2) bis zu 13 TB Hauptspeicher pro Rack mitbringt, der von 648 Cores gleichzeitig bearbeitet werden kann. Genau das ist die jüngste Erfahrung der Entwickler der Big Data Appliance: Hauptspeicher und die Anzahl der Rechenkerne werden für Big-Data-Anwendungen immer wichtiger. So ist hohe parallele Rechenleistung durch viele Cores möglich und die zu bearbeitenden Massendaten liegen als Datenobjekt im Hauptspeicher (In-Memory Objekte).

Kompaktes Prototyping, kompaktes Analysieren und schnelle Ergebnisse

Die Spark-Technologie liefert noch eine zweite spannende Entwicklung: das kompakte und funktionsorientierte Arbeiten. Die ersten MapReduce-Projekte waren meist von der Java-Programmierung geprägt und schreckten Unternehmen damit eher ab. Zu befürchten waren „never-ending“, technisch ausgerichtete Programmierprojekte mit kaum ersetzbaren Experten als Projektmitarbeiter. An schnelles, Business-orientiertes Prototyping oder gar Sandboxing für Fachmitarbeiter war nicht zu denken. Mit neuen Programmierumgebungen und -sprachen wie Scala, Python oder auch R muss man heute zwar immer noch programmieren, die angebotenen Sprachen-Konstrukte, Bibliotheken und darin enthaltenen Funktionen bewegen sich jedoch auf einem höheren Level, es sind speziell vorbereitete Aufrufe. In diesen Sprachen schickt man High-Level-Funktionen gegen In-Memory-Datenobjekte. Die Skript-Erstellung ist effizient und sehr flexibel.

Ein einfaches R-Text-Mining-Beispiel zur Sprach-Analyse von Twitter-Feeds demonstriert diese typisch kompakte Arbeitsweise (siehe Listing 1). Man nutzt wenige knappe Funktionsaufrufe auf ein Datenobjekt im Hauptspeicher. Das Datenobjekt („myCorpus“) ist das Ensemble aller Tweets, die als Einheit dem Mining-Vorgang unterzogen werden. Das Skript sucht diejenigen Wörter, die gemeinsam in den Tweets genutzt vorkommen. Tre-

ten Begriffe besonders häufig mit anderen Begriffen gemeinsam auf, wird dies erkannt und als „besonders intensive Beziehung“ vermerkt. Die Grafik in *Abbildung 1* zeigt dies als fette Beziehungsstriche. Aus der Analyse lassen sich letztlich Themen-Komplexe ableiten, also Diskussionsaspekte in der Twitter-Gruppe.

Das Ergebnis ist eine netzartige Darstellung über gemeinsam in den Feeds genutzte Wörter. Bei großen Netzen erkennt man über „Hotspots“ besondere Themen-Komplexe. Über eine anschließende LDA-Topic-Model-Analyse lassen sich sogar fachliche Diskussions-Schwerpunkte identifizieren.

Ein solches Ergebnis ist in der Kürze wohl kaum mit anderen Mitteln erreichbar. Um die Bedeutung des Beispiels voll zu erfassen, sollte man sich die Datenmenge von Hunderten Millionen von Tweets vorstellen, die im Verlauf eines Tages im englischsprachigen Raum zu finden sind. Das „myCorpus“-Objekt, in dem alle Tweet-Texte vorgehalten sind, liegt komplett im Hauptspeicher. Eine ideale Anwendung für einen Hadoop-Cluster mit riesigem Hauptspeicher und vielen parallel arbeitenden Rechenkernen in den Cluster-Knoten.

JSON und NoSQL – pragmatischer geht es kaum

Das Text-Mining-Beispiel führt zu einem weiteren wichtigen Big-Data-Aspekt. Die analysierten Tweets liegen zwar im Hauptspeicher, für weitere künftige Analysen muss man sie jedoch dauerhaft speichern. An diesen Massendaten interessieren letzt-

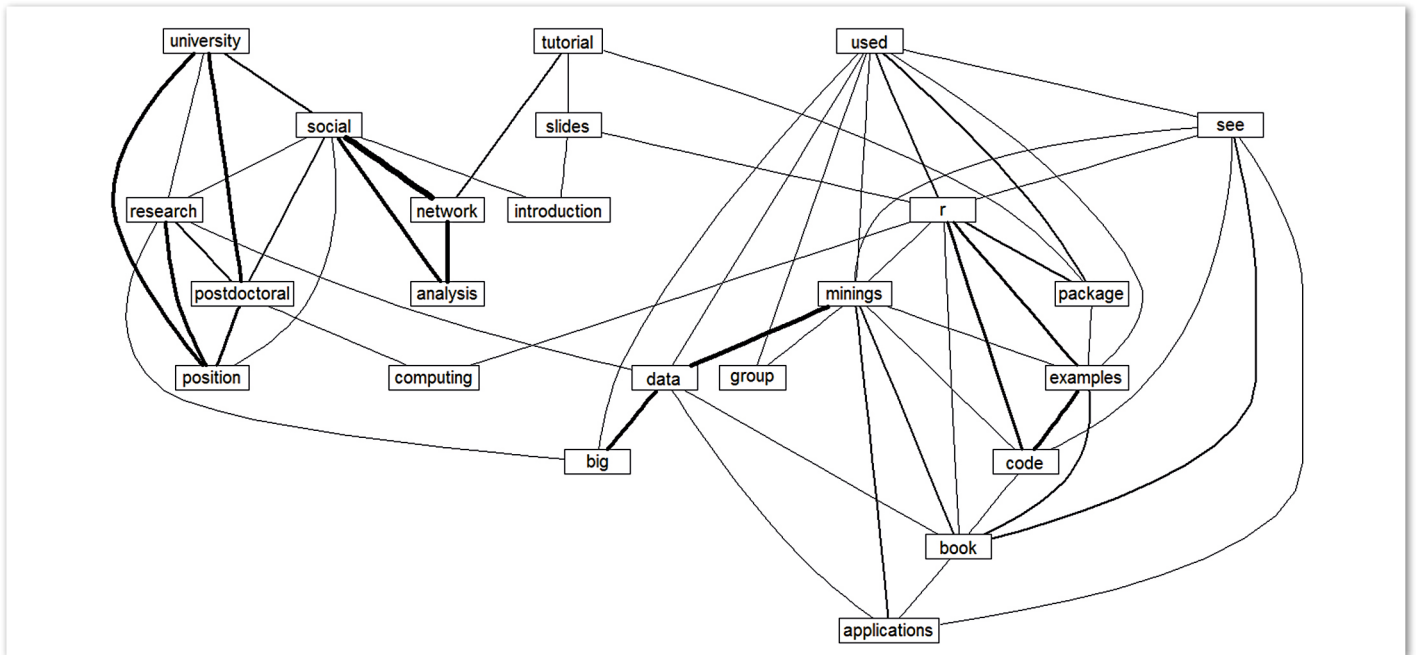


Abbildung 1: Netzgeflecht zur Häufigkeit gemeinsam genutzter Wortpaare

lich nur der Text und wenige Metadaten-Merkmale. Genau für diesen sehr einfachen Zweck eignet sich das zunehmend häufiger genutzte JSON-Format. Jeden Tweet kann man auch als solchen JSON-Satz ablegen (siehe Listing 2).

Das gleiche Skript kann die Tweets auch aus der Oracle-NoSQL-Datenbank (Key Value Store) lesen. Man müsste nur die Zeilen 1 und 3 ändern. Die einzige wichtige Funktion, die die NoSQL-Datenbank zu übernehmen hat, ist schnell und einfach und zig Milliarden Mal einen solchen Twitter-Tweet-Satz abzulegen. Das Strukturieren und Analysieren übernimmt das R-Skript, also die Anwendung. Das hier gezeigte Beispiel funktioniert mit vielen anderen Anwendungsfällen genauso, bei Log-Daten, Sensordaten oder Geopositionen.

HDFS, NoSQL-Datenbank, RDBMS

Damit sind wir bei den Empfehlungen für die Wahl der Speichertechnik. Über die innere Struktur der oben genannten Tweets muss sich das Speichermedium keine Gedanken machen. Es reicht zu wissen, dass ein Tweet ein Datensatz ist, ideal für eine NoSQL-Datenbank. Diese ist kostengünstiger im Betrieb und bei Schreib- und Lese-Operationen sehr performant. RDBMS-Systeme benötigen wir dagegen für sensible und schützenswerte Daten und, bezogen auf Data-Warehouse-Systeme, für komplexere Abfragen, bei denen das Datenhaltungssystem bei der Struk-

```
{ "id": 347272731663945729,
  "retweetCount": "3",
  "Text": "Introduction to R for Data Mining, a one-hour video by Revolution Analytics http://t.co/fo5KVU11k9",
  "longitude": "9.993682",
  "latitude": "53.551085",
  "statusSource": "web",
  "created": "2015-10-18 21:11:07" }
```

Listing 2

turierung der Daten mithilft. Sie entwickeln ihre Stärken in großen Multi-User-Systemen mit vielen gleichzeitigen Lese- und Schreib-Operationen und auch bei heterogenen Zugriffen durch unterschiedliche Werkzeuge, wie sie im Business-Intelligence-Umfeld üblich sind. SQL steht im Vordergrund.

Reine HDFS-Hadoop-Systeme sind kostengünstige Massen-Schreibsysteme. Große Datenmengen lassen sich in kurzer Zeit in einem günstigen Cluster speichern. Security, parallele Nutzung der Daten und performante Leseoperationen sind weniger gut entwickelt (siehe Abbildung 2).

Visuelles Analysieren oder muss man doch programmieren?

Programmieren muss nicht unbedingt sein. Von Oracle gibt es heute mit Big Data Discovery ein interaktiv bedienbares Werkzeug, mit dem man Hadoop-Daten auch explorativ bearbeiten kann. Zunächst

verschafft man sich durch „Profiling“-artige Stichproben eine erste Übersicht zu dem angebotenen Datenmaterial. Fehlen zur weiteren Analyse wichtige Merkmale in den Daten, etwa Schlüsselwerte für eine Verknüpfung mit anderen Datenbeständen, so kann man diese interaktiv mit diversen Wizards (ähnlich wie klassisches ETL) ergänzen. Am Ende stehen Grafiken zur visuellen Darstellung von Zusammenhängen.

Allerdings wird auch ein solches Werkzeug immer nur ein Teil in einer größeren Werkzeugsammlung für die Analyse sein. Man muss mit Sprachen wie R direkt auf den Daten arbeiten können, man benötigt Batch-Skripte zum Automatisieren von Arbeitsschritten, man braucht Third-Party-Bibliotheken für Zugriffe auf Remote-Datenbestände etc. Das Text-Mining-Beispiel macht deutlich, wie oft zusätzliche Bibliotheken für spezielle Analyse-Methoden

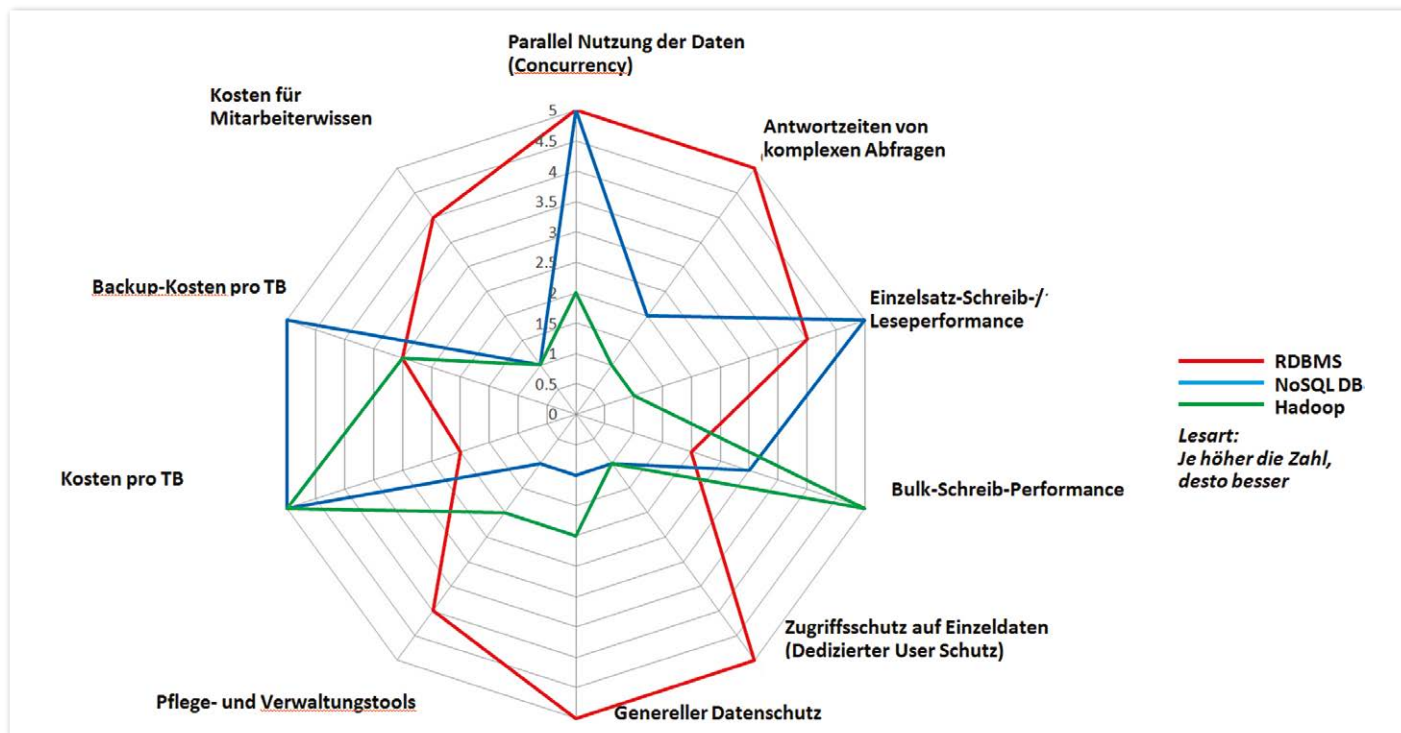


Abbildung 2: Speichersysteme im Vergleich

in den Arbeitsprozess einzubinden sind. Notwendig sind auch Sprachen wie Scala für systemnahe Operationen. Die Datenfachkraft („Data Scientist“) der Zukunft muss sich mit allen Aspekten rund um die Daten beschäftigen. Das sind betriebswirtschaftlicher Kontext, Beziehungen in den Daten, Struktur der Daten, aber auch physikalische Eigenschaften und Art der Speicherung sowie performantes Analysieren. Sich nur auf Business-Aspekte zurückzuziehen und Technik als Blackbox zu betrachten, reicht zur Erfüllung dieser Aufgabenstellungen nicht mehr aus.

Kritische Aspekte und was nebenbei so alles machbar ist

Neben all den technischen Möglichkeiten gibt es natürlich auch eine kritische Sicht. Das oben dargestellte Text-Mining-Skript kann man mit wenigen Änderungen auch zum Monitoring von Callcenter-Mitarbeitern in Kundengesprächen nutzen. Gesprächs-Mitschnitte, also Tonaufnahmen, überführt man in geschriebenen Text. Skripte bereinigen sie, wie oben gesehen und vergleichen den Sprachtext mit einer Liste positiver und negativer Wörter. Hinzu kommen Werte zur Stimmhöhe im Verlauf der gesprochenen Sätze und zu weiteren Stimm- und Sprach-Merkmalen (Satz-Analyse im Gegensatz zur Wort-Ana-

lyse). Das Ergebnis ist eine Gesprächsbewertung über ein positiv beziehungsweise negativ geführtes Kundengespräch – ein idealer Input für Mitarbeiter-Bewertungsgespräche. Zukunftsmusik? Leider nein. Wir können davon ausgehen, dass dieses Szenario heute schon Praxis ist, wenn nicht offen, dann zumindest versteckt.

Das Szenario ist nur ein Beispiel aus der Callcenter-Branche. Diese Art des „Monitorings“ kann uns durch Big-Data-Technologie künftig in fast jeder Lebenslage umgeben. Beim Callcenter könnte man noch (etwas hochnäsig) argumentieren: „Warum hat der Mitarbeiter diesen Beruf gewählt?“. Doch die Biodaten-Erfassung der Freizeitsportler ist das freiwillige Nutzen von Sensortechnik. Auch der permanente „Gesunde-Lebensführungs-Check“ durch elektronische Bio-Tracker der Krankenkassen ist freiwillig. Er wird irgendwann durch einen niedrigeren Kassenbeitrag honoriert. Das ist natürlich immer noch freiwillig, aber jeder wird es tun, weil er sparen möchte.

Diesen Anwendungsfällen kann man immer noch einen positiven Aspekt abgewinnen. Kritischer wird es allerdings, wenn Medienunternehmen die „Sentiments“ der Bürger messen und ihnen nur angepasste Unterhaltung und wohlgefällige Informationen als Bild und

Print anbieten. Unangenehme Wahrheiten, Berichte über Missstände, Aufforderungen zum Nachdenken und Verändern wird es in solchen Medien immer weniger geben.

Steht die Technik zur Verfügung, dann wird sie auch eingesetzt, wenn jemand einen ökonomischen Nutzen für sich darin erkennt. Man sollte sich der gesellschaftlichen Dimension von Big Data bewusst sein.



Alfred Schlaucher
alfred.schlaucher@oracle.com