

DIE RELEVANTEN FRAGEN STELLEN

POTENZIALE AUS BIG DATA ERFOLGREICH NUTZEN

von Ursula Deriu

Die Hürden sind hoch gesteckt: «Technologien zur Lösung von Datenverarbeitungsfragen, bei welchen die bisherigen Technologien versagen» – das ist die Definition vieler IT-Spezialisten für Big Data. Grosse Datenvolumen sind damit ebenso gemeint wie schnelle Verarbeitung, grosse Formatvielfalt und mannigfaltige Analysemöglichkeiten. Dieser Beitrag beleuchtet den Begriff Big Data und zeigt, wie KMU zunehmend von diesen neuen Technologien profitieren können.



Neue Geschäftsmodelle eröffnen sich.

Die heutigen Internet-Riesen mussten als Erste rasch wachsende Datenmengen verarbeiten. Mit den herkömmlichen Technologien stiessen sie bald an die Grenzen. Not macht erfindarisch und beflügelt Software-Ingenieure. So erfanden die Internet-Riesen die neuen Big-Data-Technologien, welche sie vorerst für sich selbst einsetzten.

Gleichzeitig fragte sich die übrige IT-Welt, wie sie mit ihren stetig wachsenden Datenbeständen zurechtkommen soll. Die Hersteller der kommerziell üblichen Technologien, vorab relationaler Datenbanken, stellten immer noch grössere relationale Datenbanken her und konstruierten spezielle Appliances. Im Gegensatz dazu

funktionieren Big-Data-Technologien auf normalen Servern, auf Commodity Hardware und sind erst noch auf Hochverfügbarkeit und Ausfallsicherheit des Gesamtsystems ausgelegt.

GROSSE FORMATVIELFALT

Das altbewährte relationale Format speichert Daten in Tabellen mit einer Informationseinheit pro Zelle, also einer Zahl oder einer Bezeichnung. Die Abfragesprache SQL verknüpft diese Tabellen, setzt sie also in Relation. Das relationale Format ist für viele moderne Anwendungen ungeeignet. NoSQL-Datenbanken sind als Nischenprodukte auf die Verarbeitung eines spezifischen Datenformats spezialisiert. Einige NoSQL-Datenbanken sind auch für

die Verwaltung und Verarbeitung riesiger Datenbestände optimiert. Die Grenze zwischen NoSQL und Big Data verwischt.

Suchmaschinen sind ein Beispiel: Nicht nur bei Google werden ganze Text-Dokumente zur schnellen und zielgenauen Suche aufbereitet, sondern auch auf nahezu jeder Homepage, in vielen Web-Shops und in Enterprise-Search-Systemen, welche die stetig wachsende Menge unternehmensinterner Dokumente durchsuchbar machen. «Hadoop» ist ein Open-Source-File-System, welches auf die Verwaltung riesiger Datenbestände ausgelegt ist. Die Architektur-Idee stammt ursprünglich von Google, wo sie für den Aufbau der Suchmaschine realisiert wurde.

Ein anderes Beispiel sind Wide Column Stores. Sie wirken wie riesige Excel-Sheets, bei welchen jederzeit problemlos neue Zeilen und Spalten hinzugefügt werden können. Facebook, Google, Amazon entwickelten Wide Column Stores. So wurde «Cassandra» ursprünglich von Facebook entwickelt und später der Open-Source-Community gestiftet.

Diese neuen Technologien erreichen jetzt einen marktauglichen Reifegrad. Neue Geschäftsmodelle spriessen wie Pilze aus dem Boden. Und die Hersteller relationaler Datenbanken versuchen, die Technologien in ihre Systeme zu integrieren. Grosse Datenbestände rücken in allgemeine Griffnähe, und die Cloud löst kostengünstig die Infrastrukturfrage.

GROSSE VERARBEITUNGSGESCHWINDIGKEIT

Unternehmensforscher sind sich einig: Schon 2020 sollen mehrere zig-Milliarden Sensoren ihre Messungen via Internet übertragen. Sensoren werden überall verbaut: in Autos, Zügen und Flugzeugen, in Medizingeräten, Robotern und Wetterstationen, in Fitnessgeräten, im Haushalt und bald auch in Kleidungsstücken.

Heute schon wäre es technisch möglich, Verkehrsüberlastungen auf Autobahnen vorauszu sehen oder die Passagierauslastung auf dem Schienennetz kurzfristig zu optimieren. Dazu könnten die Positionsdaten der Smartphones der Verkehrsteilnehmer für Vorhersagen ausgewertet werden. Solche Prognosen funktionieren ähnlich wie Wetterprognosen, müssen jedoch sehr schnell aus Messdaten erfolgen, um rechtzeitig Verkehrsoptimierung in die Wege zu leiten.

Prognosen basieren immer auf sehr grossen Datenmengen, indem sie die darin vorhandenen Muster mit der aktuellen Datenlage vergleichen. Und so raten Big-Data-Spezialisten, alle möglichen Daten aufzubewahren, um eine ausreichende Datenbasis für alle denkbaren und heute noch nicht vorstellbaren Diagnosen aufzubauen. Bei Datenschützern klingeln hier bekanntlich die Alarmglocken.

DER GROSSE ERKENNTNISGEWINN

Business Intelligence (BI) wird seit den 90er-Jahren betrieben. Durch Analyse der Geschäftsdaten werden strategische

Handlungsempfehlungen abgeleitet. Big Data Analytics verfolgt ähnliche Ziele, umfasst jedoch mit «Machine Learning» zusätzlich sehr vielfältige und mächtige algorithmische Methoden für die Analyse sehr grosser und unstrukturierter Datenbestände. Der Data Scientist ist der Nachfolger des BI-Spezialisten: Die Methoden wurden und werden massiv erweitert, die Technologien radikal revolutioniert, die Datenbasis beträchtlich vergrössert und ein neues Berufsbild geschaffen.

Ein Analyse-Beispiel aus der Medizin unterstreicht die These. Der Hersteller von Computer-Tomografen könnte seine Geräte so ausstatten, dass die Tomografien und die dazugehörige Diagnose aus aller Welt im Herstellerhaus gesammelt werden. Sobald eine ausreichende Datenbasis vorliegt, wird ein Machine-Learning-System trainiert. Für eine neue Tomografie wird dieses System eine mögliche Diagnose errechnen können. So käme das Wissen sehr vieler Diagnostiker zusammen, und Diagnosen würden präziser werden.

Beispiel Meinungsbarometer für Politiker: Heute gelingt es immer besser, aus den Twitter-Kurznachrichten zu errechnen, ob eine positive oder negative Stimmung zu einem Thema vorherrscht. Unlängst belegte ein Forscherteam der ETHZ einen ersten Platz bei SemEval, einem jährlich stattfindenden weltweiten Wettbewerb zur Verbesserung dieser Algorithmen.

Und noch ein Beispiel: Liegen genügend Vergangenheitsdaten zum Kundenverhalten vor – sprich Anfragen, Bestellungen, Reklamationen – dann können mit Machine-Learning-Algorithmen diejenigen Kunden identifiziert werden, welchen man mit besten Erfolgsaussichten ein Produkt anbietet.

EMPFEHLUNGEN ERARBEITEN

Der Data Scientist geht explorativ vor: Eine Vermutung wird geäussert, die Unternehmensdaten werden mit Big-Data-Daten verknüpft und analysiert. Ist das Ergebnis inhaltlich plausibel, werden Handlungsempfehlungen abgeleitet.

Ein einfaches Beispiel: Die Mitarbeitenden eines Warenhauses meinen zu beobachten, dass freitags mehr Gesellschaftsspiele verkauft werden, wenn für Samstag schlechtes Wetter angesagt ist.

Werden jetzt die Verkaufszahlen aus der Vergangenheit kombiniert mit den jeweiligen Wetterdaten, dann könnte die Vermutung statistisch und algorithmisch bestätigt oder verworfen werden. Würde sie bestätigt, dann könnten in Zukunft aufgrund der Wettervorhersagen die Gesellschaftsspiele freitags im Warenhaus prominent platziert werden. Zusätzlich könnten andere Zusammenhänge zwischen Wetterprognose und Kaufverhalten aufgespürt und für die Positionierung des Angebots genutzt werden.

Nicht nur Wetterdaten, sondern mannigfaltig viele Social-Media-Daten, Sensor-Messungen, Volltexte können für Analysen und Prognosen hinzugezogen werden.

Die nötigen Daten werden via Schnittstellen angeboten, können von Lieferanten in Paketen erworben werden oder werden als Open Data abgerufen.

IN REICHWEITE VON KMU

Big-Data-Technologien sind Open Source und als solche für jedermann zugänglich. Sie wurden konzipiert für marktübliche Hardware und sind somit auch für ein breites Publikum erschwinglich. Kostengünstige Cloud-Lösungen erleichtern zusätzlich die Infrastrukturfrage.

Immer mehr Daten liegen bereit, können abgerufen und erworben werden. Neue Geschäftsmodelle etablieren sich, und immer mehr Data Scientists bieten ihre Analyse- und Prognose-Dienstleistungen an.

Die Technologie-Hürden sind mehrheitlich bewältigt. Die Kunst ist vielmehr, die richtigen Fragen zu stellen, um aus der Datenviel-falt den verborgenen Schatz zu heben. ■



URSULA DERIU

berät und schult Unternehmen in IT-Architektur-Fragen, mit Schwerpunkt Suchmaschinen und Big Data. Sie ist Dozentin bei der FFHS in den Fachbereichen Data Science und Software-Entwicklung.

www.tirsus.com
www.ffhs.ch