

Corporate Performance Management

Neue Herausforderungen für Big Data

Erkenntnisgewinn und Entscheidungen in Echtzeit bei einem Energieversorger

Tobias Gorhan, BI Consultant
Marcel Weiß, BI Consultant
Tilman Hagen, Manager
Braincourt GmbH

Inhaltsverzeichnis

1	Einleitung.....	3
2	Neue Big Data-Herausforderungen	3
3	Energieversorger und Big Data	4
4	Die Lambda-Architektur	5
5	Beispielszenario	6
6	Fazit	8

Ansprechpartner & Quellen

I.	Ansprechpartner.....	8
II.	Literaturempfehlung	9

1 Einleitung

Praxisbeispiele und Studien zeigen, dass Big Data heute in den Unternehmen angekommen ist. Doch inzwischen ergeben sich im Umgang mit Big Data neue Herausforderungen: Datenarten, wie personenbezogene Informationen oder Umweltdaten, müssen integriert und Analysen, Erkenntnisse und Entscheidungen in Echtzeit generiert werden. Das Beispiel eines Energieversorgers zeigt, dass diesen Herausforderungen mit dem Konzept der „Lambda-Architektur“ und einer geeigneten Technologieauswahl erfolgreich begegnet werden kann. Somit ist es möglich, im Wettbewerb und einer sich verändernden Energiebranche zu bestehen und dabei gleichzeitig neue Geschäftsmodelle und Wachstumschancen zu entwickeln.

In Unternehmen ist Big Data inzwischen Realität. Eine Entscheidungsfindung in Echtzeit bei Massendaten noch nicht. Die Lambda-Architektur und eine geeignete Technologieauswahl machen dies nun möglich.

2 Neue Big Data-Herausforderungen

In den Anfängen des „Big Data-Zeitalters“ stand im Unternehmenskontext insbesondere die rückblickende („retrospective“), beschreibende („descriptive“) und vorhersagende („predictive“) Analyse von Datenmengen im Vordergrund. Die Daten entstammten z. B. in der produzierenden Industrie Maschinen-Logs, im Handel und dem Dienstleistungsbereich Informationen wie Verkaufsbelegen, Klickpfaden auf Internetseiten oder Online-Shops. Vor dem Hintergrund dieser, der Wertschöpfungskette sehr nahen Datenanalysen, galt es, unter Berücksichtigung der bekannten drei „Vs“ – Volume, Variety und Velocity, den konkreten Business-Nutzen von „Big Data“ zu beweisen. Heute zeigen Studien, Case Studies und Vorträge, dass „Big Data“ diesbezüglich in der Praxis und der Unternehmenswirklichkeit angekommen ist.¹

Aktuelle Big Data-Anwendungsfälle begegnen heute neuen Herausforderungen, welche sich aus den derzeitigen Trends, wie der Massenindividualisierung sowie insbesondere der automatisierten Entscheidungsfindung unter dem Stichwort „Prescription“ ergeben.

Im Kontext der Massenindividualisierung in der Kundenansprache ist insbesondere von „360°-Kundenprofilen“ die Rede, welche oftmals indirekt, auf Basis von Kaufentscheidungen oder Äußerungen in Social Media-Portalen entwickelt werden. Die Entwicklung dieser individuellen Profile wird heute durch einen neuen Trend zusätzlich konkretisiert: Quantified Me. Unter Quantified Me werden personenbezogene Daten verstanden, wie sie z. B. von SmartWatches, vernetzten Automobilen oder intelligenten Stromzählern, sogenannten „Smart Meters“ erzeugt werden.

Neue Trends im Umfeld von Big Data sind „360°-Kundenprofile“ und „Quantified Me“.

Neben diesen Trends, steht die Wahrnehmung, dass in den ersten Big Data-Jahren insbesondere die beiden Vs für „Volume“ (Datenmenge) und „Variety“ (Datenvielfältigkeit) im Fokus der Anwendungsfälle standen. Die durch das dritte V – „Velocity“ – ausgedrückte tatsächliche Verarbeitungsgeschwindigkeit wurde im Kontext der beiden Geschwister-Vs teilweise als notwendiges Übel betrachtet. In

¹ Capgemini und EMC: Big & Fast Data: The Rise of Insight-Driven Business. URL: https://www.de.capgemini.com/resource-file-access/resource/pdf/big_fast_data_the_rise_of_insight-driven_business.pdf

den letzten Jahren hat sich hier ein Wandel in der Erwartungshaltung gegenüber Big Data-Applikationen vollzogen: Aus der reinen Vorhersage von Ereignissen (Prediction) wurde das Finden der optimalen Entscheidung (Prescription). In einer vollständig vernetzten und transparenten Welt resultieren solche automatisierten Entscheidungsprozesse jedoch nur in einem Wettbewerbsvorsprung, sofern diese direkt und in Echtzeit erfolgen – ein Trend der zusätzlich durch neue Hochgeschwindigkeitstechnologien wie „SAP HANA“ gefördert wird.

Hochgeschwindigkeitstechnologien wie „SAP HANA“ unterstützen den Trend hin zur automatischen Entscheidung in Echtzeit.

Dass die beschriebenen Big Data-Herausforderungen in der Praxis in Kombination auftreten und gleichzeitig einen immensen Business-Nutzen darstellen können, zeigt sich am Beispiel eines Energieversorgers.

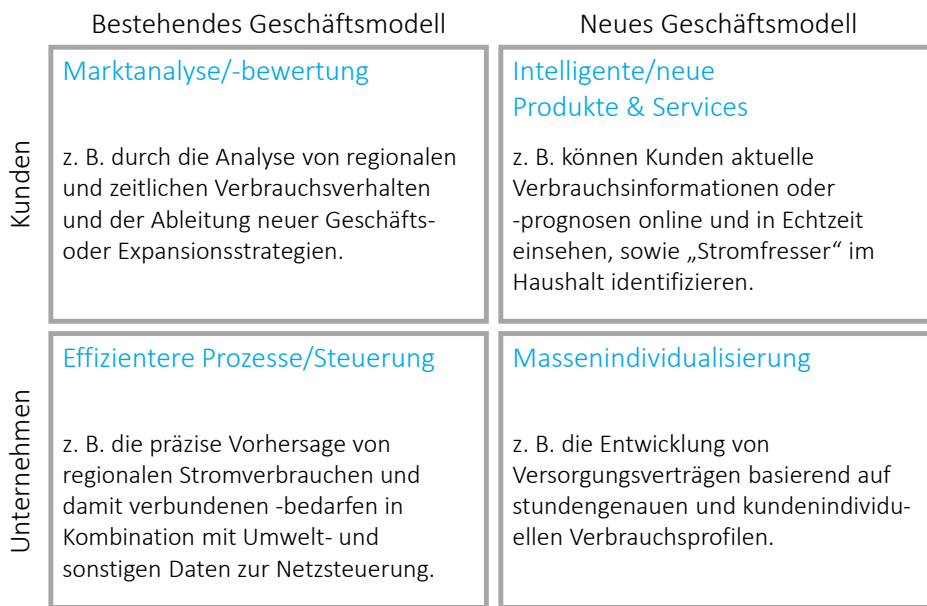
3 Energieversorger und Big Data

Die Grundlage für Stabilität im Stromnetz ist das Gleichgewicht aus Stromerzeugung und -verbrauch. Heute ist es insbesondere durch die verstärkte Nutzung von erneuerbaren Energien und die dezentrale Stromerzeugung weitaus schwieriger, den tatsächlichen Strombedarf zu überwachen. Dieser ist z. B. abhängig von Wetterbedingungen, welche direkten Einfluss auf die Stromerzeugung durch Windräder oder Photovoltaikanlagen haben. Um die Netzstabilität aufrecht zu erhalten, sind intelligente Netze nötig, die den Strom nicht ausschließlich verteilen, sondern ebenfalls Entscheidungen zu automatisieren, die für das Gleichgewicht zwischen Stromerzeugung und -verbrauch sorgen. Als eine Grundlage für die Messung des tatsächlichen Verbrauchs werden Smart Meter eingesetzt, welche eine Kommunikation zwischen Energieverbrauchern und -erzeugern in Echtzeit ermöglichen.

Herausforderung der Energiebranche: Für die Sicherung der Netzstabilität werden intelligente Netze benötigt.

Die erzeugten Verbrauchsdaten werden mithilfe von Big Data-Technologien in Echtzeit gesammelt und analysiert. Auf Basis der Analyseergebnisse werden Entscheidungsempfehlungen gegeben oder direkt automatisiert Entscheidungen getroffen.

Für den Stromversorger selbst bieten sich darüber hinaus, neben den genannten Erfordernissen zur Aufrechterhaltung der Netzstabilität, weitere Anwendungsfälle und Potenziale. Big Data-Anwendungsfälle werden oftmals den vier Quadranten „Marktanalyse/-bewertung“, „Intelligente/neue Produkte & Services“, „Effizientere Prozesse/Steuerung“ und „Massenindividualisierung“ zugeordnet. Eine exemplarische Zuordnung solcher Anwendungsfälle für einen Stromversorger ist in Abbildung 1 dargestellt.



Neben der Bewältigung aktueller Herausforderungen schafft die Nutzung von Big Data neue Chancen.

Abbildung 1: Big Data-Anwendungsfälle für einen Energieversorger

Für einige der genannten Beispiele, wie die Generierung individueller Versorgungsverträge, eignet sich eine regelmäßige aber nicht in Echtzeit erfolgende (Batch-)Verarbeitung. Viele Anwendungsfälle, wie die Steuerung von Stromnetzen, setzen jedoch zwangsläufig eine Datenverarbeitung und Entscheidungsfindung in Echtzeit voraus.

4 Die Lambda-Architektur

Bereits heute existieren technologische Konzepte, welche sich eignen, um diesen Herausforderungen zu begegnen. Ein Ansatz ist die von Nathan Marz und James Warren entwickelte „Lambda-Architektur“. Diese wird allen Anforderungen an Robustheit und Fehlertoleranz, niedrige Latenz, Skalierbarkeit, Allgemeingültigkeit, Erweiterbarkeit, Gestattung von Ad hoc-Abfragen, minimaler Wartungsbedarf sowie Korrigierbarkeit gerecht.²

Die Lambda-Architektur ist ein wesentlicher Baustein zur Bewältigung neuer Big Data-Herausforderungen.

Die Lambda-Architektur verbindet die drei Schichten „Batch Layer“, „Serving Layer“ und „Speed Layer“. Daten werden bei ihrem Eingang in die Lambda-Architektur sowohl im Batch als auch im Speed Layer erfasst. Der Batch Layer verarbeitet den gesamten Datenbestand zu einem definierten Zeitpunkt und schreibt die Ergebnisse in „Batch Views“. Dieser Prozess wird als „Batch Processing“ bezeichnet und weist eine äußerst hohe Latenz auf, wodurch Analyseergebnisse nur verzögert zur Verfügung stehen. Der Serving Layer stellt die Batch Views für Abfragen durch den Anwender oder weitere Applikationen bereit und ermöglicht somit tiefgreifende Analysen innerhalb des gesamten Datenbestands. Aktuelle Daten, die während der Vorverarbeitung gesammelt werden, werden durch den Speed Layer für Abfragen bereitgestellt. Dafür erstellt dieser Layer sogenannte „Real Time Views“, die beim Eintreffen neuer Daten in die Lambda-Architektur

² Marz, Nathan; Warren, James: *Big Data. Principles and best practices of scalable real-time data systems*. New York: Manning Publications 2015

sofort ergänzt werden. Durch den dreischichtigen Aufbau der Architektur wird zu jedem Zeitpunkt die Analyse des gesamten Datenbestands in Echtzeit ermöglicht und die Aktualität der Ergebnisse sichergestellt. Hierfür ist es erforderlich, dass Abfragen sowohl auf die Batch Views im Serving Layer, als auch auf die Real Time Views im Speed Layer zugreifen. Abbildung 2 stellt die beschriebenen Komponenten der Lambda-Architektur und deren Beziehungen untereinander schematisch dar.

Der dreischichtige Aufbau der Lambda-Architektur ermöglicht Echtzeitanalysen auf den gesamten Datenbestand.

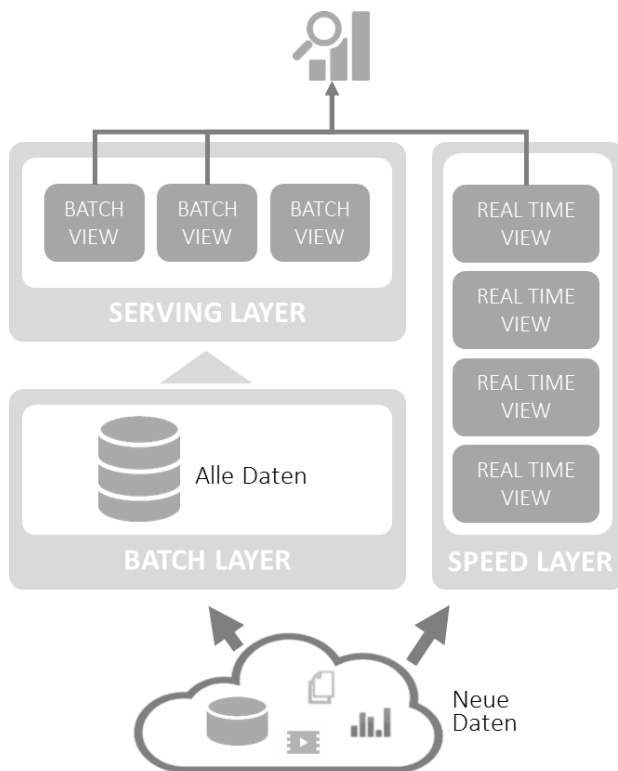


Abbildung 2: Komponenten der Lambda-Architektur

State of the Art-Technologien existieren bereits heute in großer Zahl am Markt, welche sich für den Einsatz innerhalb einer Lambda-Architektur eignen und ihre Praxistauglichkeit bereits bewiesen haben. Aus diesem Technologie-Portfolio lassen sich einzelne zu einer leistungsstarken hybriden Referenzarchitektur kombinieren.

Die Kombination aus SAP HANA und Hadoop ist Basis einer leistungsstarken Lambda-Referenzarchitektur.

In Unternehmen ist zudem eine nahtlose Integration hinsichtlich bestehender analytischer Applikationen notwendig, um Analysen, Erkenntnisse und Entscheidungen auf einer allumfassenden Datenbasis zu generieren. Im konkreten Beispiel werden diese Applikationen durch eine SAP BW on HANA-Landschaft repräsentiert.

5 Beispielszenario

Die angeforderte Real Time-Fähigkeit des Speed Layers erfordert eine hoch performante Technologie, welche Daten schnell verarbeiten und bei Bedarf sofort bereitstellen kann. Im Kontext des Beispielszenarios – einer bestehenden SAP

BW-Landschaft auf Basis der In Memory-Technologie SAP HANA – bietet sich hierfür die Nutzung dieser HANA-Datenbank an. Diese ist vollständig in SAP BW eingebettet und ermöglicht somit Echtzeit-Analysen über alle Daten hinweg. Die HANA-Integration ermöglicht im Beispielszenario, dass Analysen über sogenannte „Virtual Provider“ und bestehende „Multiprovider“ nahtlos in das bestehende Reporting integriert werden. Der Endanwender findet sich somit in der bereits bekannten mehrdimensionalen OLAP-Welt wieder und kann alle bisher genutzten Analysewerkzeuge weiterhin verwenden.

Während SAP HANA auf Geschwindigkeit optimiert ist, eignet sich für den Batch- sowie den Serving Layer insbesondere das bekannte, am Markt bereits etablierte und durch die Verwendung von Commodity Hardware verhältnismäßig kostengünstige „Hadoop Framework“. Dieses Framework ist darauf optimiert, riesige Datenmengen, wie sie innerhalb des Batch Layers typischerweise anfallen, zu sammeln, vorzuhalten und zu verarbeiten. Innerhalb des Frameworks werden Komponenten für unterschiedliche Funktionalitäten zusammengefasst. Z. B. für das Ablegen von Daten in das „Hadoop File System (HDFS)“, für die Ressourcenverwaltung und die Steuerung von Verarbeitungsprozessen „YARN“, für die Abbildung einer Data Warehouse-Schicht „Hive“ oder für das Sammeln, Aggregieren oder Transportieren von Daten „Apache Flume“.

Um integrierte Algorithmen zu entwickeln, welche sowohl für die Verarbeitung im Batch, als auch im Speed Layer geeignet sind, kann die Statistiksprache „R“ genutzt werden. So können beispielweise im historischen Datenbestand (Batch Layer) ermittelte Zusammenhänge und Datenmuster direkt für Prognosen im Speed Layer Anwendung finden.

Am Beispiel des Energieversorgers werden Basisdaten mithilfe von Apache Flume in Echtzeit sowohl an das Hadoop-System, als auch an SAP HANA übertragen. Auf dieser Grundlage und mithilfe der Statistiksprache R werden Verbrauchsprofile ermittelt, welche die Ausgangsbasis für die Prognosen sind. Wichtigste Quelle für diese Profile sind die Smart Meter-Daten, die bereits eine relativ genaue Vorhersage unter Berücksichtigung individueller Tagesabläufe ermöglichen. Die Verbrauchsmuster lassen sich stundengenau für einzelne Wochentage, Werkzeuge und Wochenenden oder beispielsweise für Sommer- und Wintermonate im Detail differenzieren.

Die Prognose des Strombedarfs lässt sich durch die Einbeziehung von Umweltdaten und die Kombination aller Daten miteinander weiter detaillieren. Somit lässt sich z. B. der Einfluss des Wetters (wie Temperatur oder Sonnenscheindauer) oder von TV-Ereignissen (wie die Übertragung wichtiger Fußballspiele) auf den Stromverbrauch mit einkalkulieren. In der nachfolgenden Abbildung 3 wird exemplarisch die Prognose des Stromverbrauchs für einen Haushalt dargestellt. Diese Prognose basiert auf einem individuellen und auf Basis historischer Daten berechneten Verbrauchsprofil. Solche Profile stellen eine Basis für stundengenaue und kundenindividuelle Versorgungsverträge dar. Durch eine z. B. regionale Aggregation dieser Verbrauchsprofile und -vorhersagen lassen sich frühzeitig

SAP HANA ist unter anderem auf den Datenzugriff mit hoher Geschwindigkeit optimiert.

Das Hadoop-Framework auf die Verarbeitung großer Datenmengen.

Mithilfe Apache Flume und der Statistiksprache R lassen sich Muster erkennen, Vorhersagen treffen und somit Entscheidungen automatisieren.

Energieengpässe identifizieren, -engpässe vermeiden und eine lückenlose Versorgungssicherheit gewährleisten.

Eine leistungsstarke hybride Big-Data-Architektur unterstützt Energieversorger beim erfolgreichen Bestehen in einem immer kompetitiveren Marktumfeld.

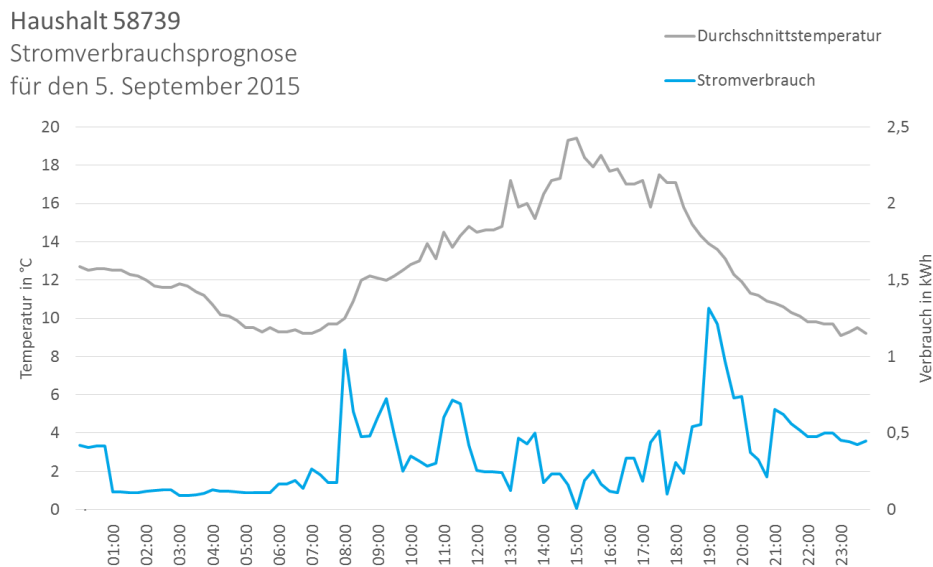


Abbildung 3: Exemplarische Stromverbrauchsprognose für einen Haushalt

6 Fazit

Das Beispiel des Energieversorgers zeigt, dass es durch aktuelle Big Data-Trends und durch einen weiter steigenden Wettbewerbsdruck erforderlich wird, Analysen und Vorhersagen in Echtzeit zu generieren und daraus Entscheidungsempfehlungen abzuleiten oder Entscheidungen automatisiert zu treffen. Um diesen Herausforderungen zu begegnen, wird mit der Lambda-Architektur ein geeignetes Konzept bereitgestellt, für welche z. B. mit dem Hadoop-Framework und der Datenbanktechnologie SAP HANA bereits etablierte und bewährte Tools zur Verfügung stehen.

I. Ansprechpartner

Tilman Hagen
 Manager
 Braincourt GmbH
 Fasanenweg 11
 70771 Leinfelden-Echterdingen
Tilman.Hagen@braincourt.com
 Telefon: +49 711 75 85 80 -54

II. Literaturempfehlung

Autor, Buchtitel, Verlagsangaben

Marz, Nathan; Warren, James

Big Data. Principles and best practices of scalable real-time data systems.

New York: Manning Publications 2015.
