

BI Spektrum

EINE PUBLIKATION DES TDWI E.V.

BI und Analytics in der Cloud: Wichtiger Schritt in Richtung digitale Transformation ab Seite 8

Digitale Autoindustrie
Data Governance
als Beschleuniger

Seite 32

Agiles Datenmanagement
Umgehen mit poly-
strukturierten Daten

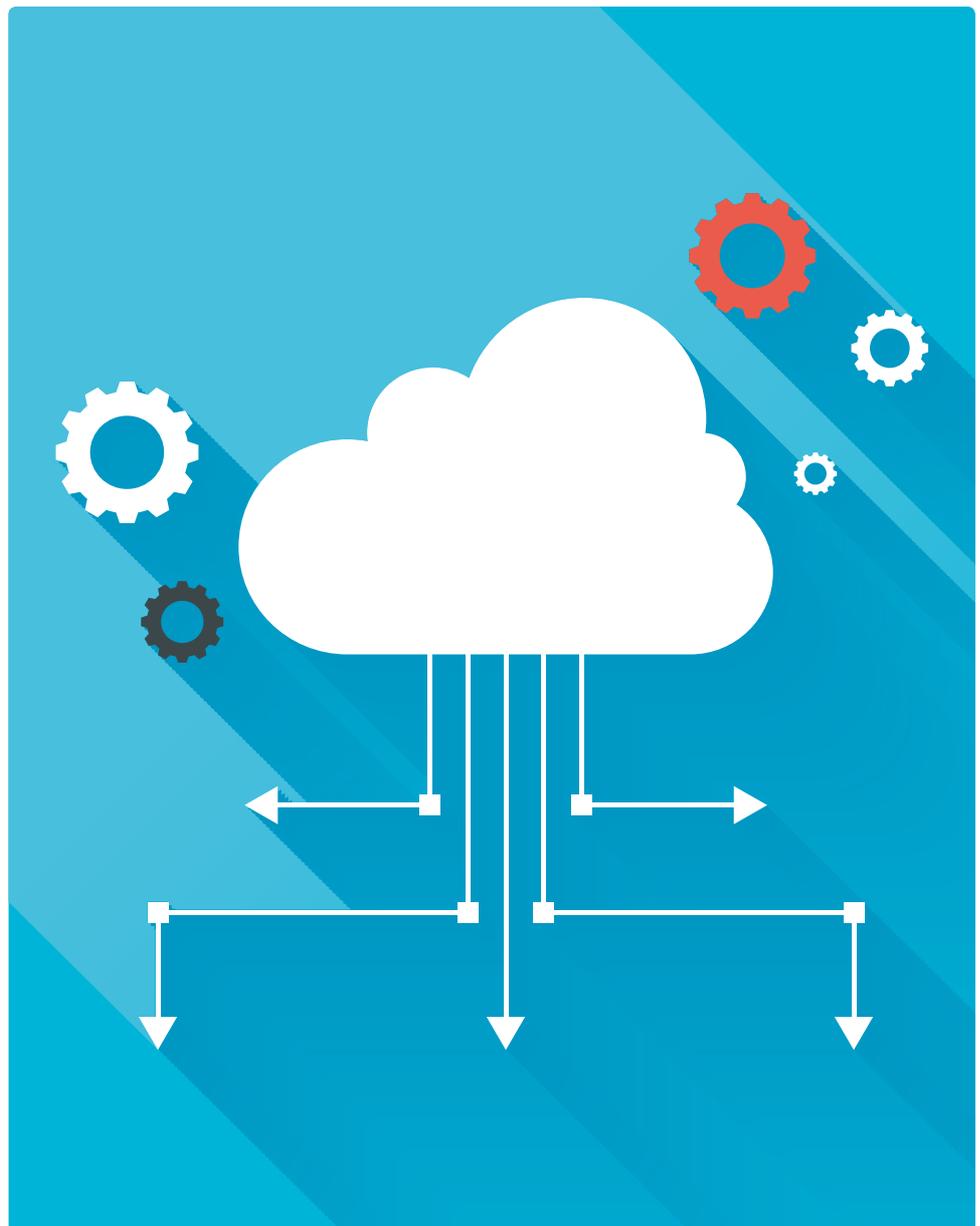
Seite 36

Interview
Vom Datenspeicher
zum Business-Case

Seite 30



Frank Weber,
Hitachi Vantara



[zum Inhalt](#)



Bild: Shutterstock

Synergieeffekte zwischen Cloud-Computing, Data Science und Prozessberatung

Entscheidungen aus der Cloud

Ein Beitrag von
Antonio Bartulovic,
Torben Hanf,
Sebastian Kunert
und Markus Volkmar

„Neukunden kosten Geld, Bestandskosten bringen Geld.“ Das ist eine der ältesten Marketingweisheiten. Beispielsweise kostet es gegenüber der Neukundenakquise in einem Telekommunikationsunternehmen nur etwa die Hälfte, einen Kunden zu halten. Die Entscheidung, welchen Kunden gezielte Angebote gemacht werden, um sie vom Bleiben zu überzeugen, ist somit allenfalls tendenziell zu treffen. Ziel sollte also sein, die Entscheidungsfindung zu verbessern und sich damit vom Prinzip der Gießkanne zu entfernen, mit der Angebote ineffizient auf alle Kunden gleichsam verteilt werden. Durch Datenanalysen können Kunden möglichst optimal nach passenden Attributen klassifiziert werden, sodass dem Kunden ein maßgeschneidertes Produkt angeboten werden kann. Darüber hinaus kann ein gutes Modell den Zeitpunkt des Wechsels bestimmen.

Die Praxis zeigt zwei Herausforderungen bei der datengetriebenen Entscheidungsfindung: (1) Datenanalysen finden oft losgelöst von den Geschäftsprozessen statt. In vielen Unternehmen mangelt es bisher an einem Vorgehensmodell und einer entsprechenden Notationsgrundlage, mit der eine Integration auf einfache Art und Weise gelingt. (2) Zudem fehlt es an flexiblen Architekturen, um auch große Datenmengen mit komplexen Algorithmen analysieren zu können. Die Migration zu Cloud-Computing-Modellen kann hier Abhilfe schaffen. Im Folgenden möchten wir einen Lösungsansatz diskutieren, der sowohl Prozesse (1) als auch Technologien (2) berücksichtigt. Die The-

men Prozessberatung, Data Science und Cloud-Computing sind im Einzelnen betrachtet bereits umfangreiche Themenfelder. Den Fokus möchten wir daher nicht darauf legen, einen Bereich in der Tiefe zu illustrieren, sondern anhand eines Beispiels darzustellen, wie die drei Themenfelder miteinander kombiniert werden können.

Herausforderungen in der Datenanalyse

Betrachten wir dafür beispielhaft einen Callcenter-Mitarbeiter, der im Gespräch mit einem Kunden die richtige Produktempfehlung geben möchte. Dafür müssen Kunden- und Transaktionsdaten analysiert

werden. Ein Data Analyst in der Telekommunikationsbranche steht dabei vor wesentlichen Herausforderungen, charakterisiert durch die 4 Vs – in der Literatur spricht man auch von den 3 Vs (ohne Veracity) oder den 6 Vs (plus Variability und Value):

1. **Volume:** Die großen Mobilfunkanbieter in Deutschland haben zwischen 30 und 44 Millionen Kunden [vgl. Sma18]. Wo und wie soll die wachsende Datenmenge gespeichert werden?
2. **Variety:** Informationen zur Abwanderungswahrscheinlichkeit von Kunden können aus internen wie externen Quellen sowie aus Daten verschiedener Schemata entnommen werden. Wie können heterogene Daten zu einem gesamtheitlich analysierbaren Core zusammengefügt werden?
3. **Velocity:** Jeder Kunde produziert permanent Daten. Darüber hinaus können auch kontinuierlich einlaufende Daten aus sozialen Netzwerken und Callcentern Hinweise zur Abwanderungswahrscheinlichkeit liefern. Wie können in Echtzeit einlaufende Daten in die Analyse einbezogen werden?
4. **Veracity:** Kunden wie Mitarbeiter sind Menschen. Und Menschen machen falsche oder fehlerhafte Angaben. Wie kann die Qualität der Daten und der darauf aufbauenden Analyse sichergestellt werden?

Chancen durch Cloud-Technologien

Seit einiger Zeit schon etabliert sich der Ansatz Cloud Analytics, der Antworten auf die genannten Herausforderungen verspricht. Damit ein Mitarbeiter im Callcenter noch während des Gesprächs auf Handlungsempfehlungen zurückgreifen kann, ist er auf Informationen und Analysen in Echtzeit angewiesen. Mit Hilfe von Cloud-Technologien ist eine Skalierung der Performance möglich. Performance ist hierbei die Kombination aus Datenkomplexität und dem Grad an Parallelisierung der genutzten Algorithmen [OEE13].

Komplexe Analysen können zu genaueren und aussagekräftigeren Ergebnissen kommen. Dabei erweist sich die bestehende IT-Landschaft häufig als Flaschenhals. Als Alternative zu On-Premise-Lösungen bietet sich die aus vielen Gründen sinnvolle Auslagerung der Daten in die Cloud an. Leistungsfähigere Algorithmen lassen sich so einbinden, beispielsweise Ensemble-Lernen mit einer großen Zahl an Wiederholungen. Die schnelle Verarbeitung erlaubt Echtzeitanalysen in zuverlässiger Weise. Hervorzuheben ist hier zudem die ökonomische Komponente: Ein Cloud-Nutzer muss keine Vorabinvestitionen tätigen, um die Vorteile der Technologie nutzen zu können. Auch in der Laufzeit kommen niedrigere Betriebskosten auf den Nutzer zu, da keine Zusatzkapazitäten vorgehalten werden müssen. Bezahlt wird nur das, was auch genutzt wird.

Ein weiterer Vorteil ist die bedarfsgerechte Skalierbarkeit. Eine Analyse kann schnell auf einen größeren Maßstab übertragen werden, um größere Datenmengen zu verarbeiten. Durch das Auslagern

ANTONIO BARTULOVIC hat Wirtschaftsinformatik in Venlo & Köln studiert. Er ist Abteilungsleiter bei der msg systems AG und verantwortet die Themen datengetriebene Lösungen und Blockchain. Seit 2008 berät er Kunden mit den Schwerpunkten auf Strategie, Anforderungsanalyse und Projektleitung.

Antonio.Bartulovic@msg.group



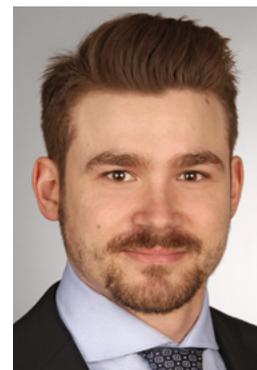
TORBEN HANF hat ein Wirtschaftsmathematik-Studium an der TU Dortmund absolviert und arbeitet inzwischen als IT-Berater bei der msg systems ag. Sein Fokus liegt auf Data Science und insbesondere auf Text Mining.

Torben.Hanf@msg.group



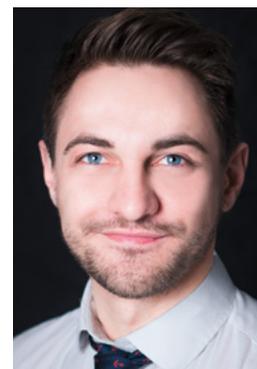
SEBASTIAN KUNERT, Jahrgang 1989, hat Controlling / Information Systems in Lüneburg und San Diego studiert. Seit August 2017 ist er als Unternehmensberater bei der msg systems ag tätig. Die Schwerpunkte seiner Arbeit liegen im Bereich Quantitative Finance und Business Intelligence.

Sebastian.Kunert@msg.group



MARKUS VOLKMAR hat Psychologie, Statistik und Human Resource Management in Koblenz und Maastricht studiert und arbeitet heute als Unternehmensberater für die msg systems ag. Die Schwerpunkte seiner Arbeit liegen in Data Science und Data Warehousing in den Bereichen Public und Health.

Markus.Volkmar@msg.group



der IT-Infrastruktur in die Cloud gibt der Nutzer Geschäfts- und Wartungsrisiken an den Anbieter der Cloud ab. Im Sinne der klassischen Arbeitsteilung haben diese oft eine hohe Expertise im Betrieb oder im Umgang mit entsprechenden Risiken. Traditionell werden Dienstleistungen in der Cloud webbasiert angeboten. Daraus resultiert eine einfache Erreichbarkeit der Services.

Trotz der großen Zahl an Vorteilen, welche die Einbindung von Cloud-Technologien bietet, erwachsen daraus auch einige Herausforderungen. Sicherheit und Kontrollverlust werden hier immer wieder diskutiert. Es müssen nutzerspezifische Lösungen erarbeitet werden, um im Fall eines Verfügbarkeitsausfalls oder eines Sicherheitslecks handlungsfähig zu bleiben.

[zum Inhalt](#)

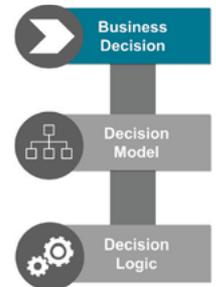
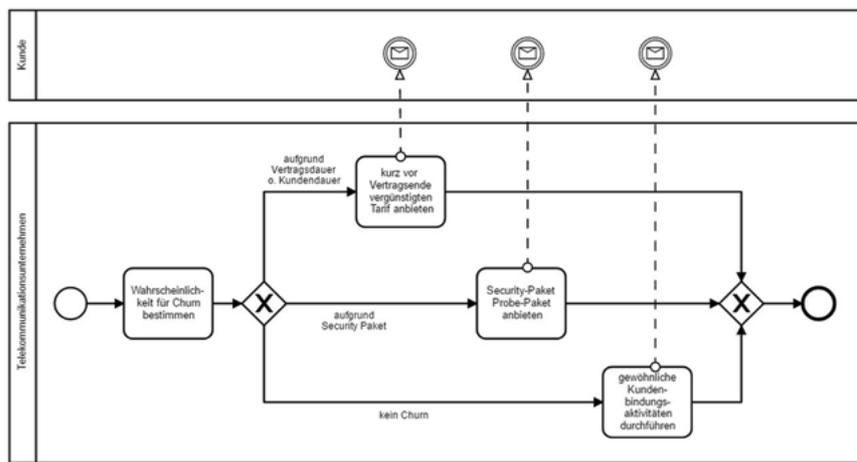


Abb. 1: Entscheidungsidentifikation auf Ebene der Business Decision

In der Nutzung der Cloud existieren zwei Ansätze, die eine unterschiedliche Architektur zur Grundlage haben:

1. **Data Clouds** werden mit dem Ziel entwickelt, hochperformante Analysen bei maximaler Aggregation zu ermöglichen. Gerade auf die Anforderungen Volume und Variety kann so eine zuverlässigere Antwort gefunden werden, als traditionelle Analyseverfahren sie bieten. Data Clouds eignen sich vor allem, wenn die Quellen besser als die Senken bekannt sind, die Architektur also bottom-up entwickelt wird [vgl. Far11].
2. **Utility Clouds** werden mit dem Ziel entwickelt, dem Nutzer granulare Daten zur Verfügung zu stellen. Je nach Anforderung kann in der Cloud auf Software zur Datenanalyse (Software as a Service, SaaS), auf leistungsstärkere Laufzeitumgebungen (Platform as a Service, PaaS) oder auf eine größere Speicherkapazität (Infrastructure as a Service, IaaS) zurückgegriffen werden. Auch nutzerspezifische Kombinationen bis hin zur Einbindung aller Dienstleistungen sind denkbar (Everything as a Service, XaaS). Utility Clouds eignen sich bei größerer Kenntnis der Nutzeranforderungen und werden dementsprechend top-down entwickelt [vgl. Hog11].

Nach der Analyse der Rahmenbedingungen und der Datengrundlage unter Beachtung der 4 Vs und der darauf aufbauenden Wahl der Cloud-Architektur werden Prozessberatung und technische Lösung zur Nutzung der Technologie zusammengeführt.

Prozessberatung

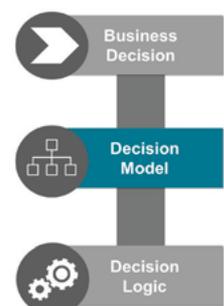
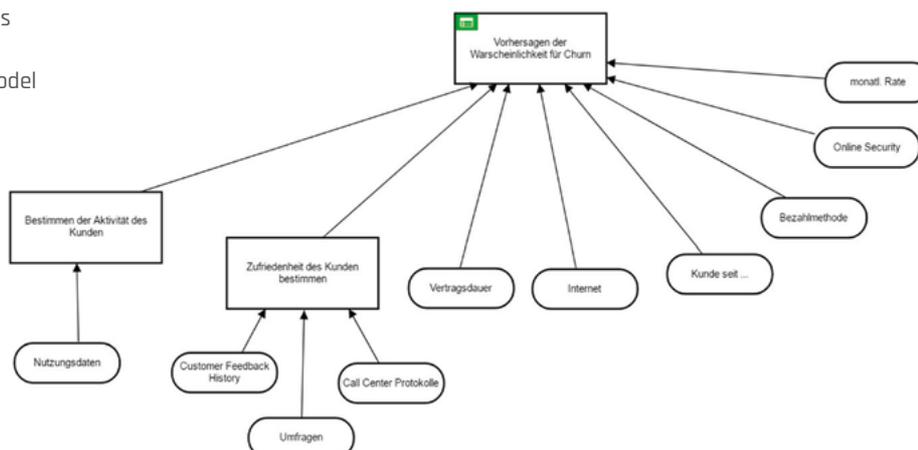
Modellierung ist der beste Weg, um wiederkehrende Entscheidungen im operativen Geschäft eines Unternehmens transparent und nachvollziehbar abzubilden. Dies bietet die notwendige Operationalisierung, die eine Datenanalyse und die Speicherung in der Cloud erst ermöglicht. Hierfür eignet sich der DMN-Standard (Decision Model and Notation) [OMG 18].

Mit Hilfe von DMN kann eine gemeinsame Basis in Unternehmen geschaffen werden, um aus großen Datenmengen Entscheidungsregeln abzuleiten, zu automatisieren und zu dokumentieren. Automatisierte Entscheidungen haben gegenüber menschlichen Entscheidern den Vorteil, dass sie frei sind von kognitionspsychologischen Wahrnehmungs- und Urteilsfehlern.

Über ein dreistufiges Vorgehen wird eine gemeinsame Notationsgrundlage geschaffen, die für alle Geschäftsbereiche leicht lesbar und nachvollziehbar ist:

1. **Business Decision:** Identifizierung einer Entscheidung, um sie in den Geschäftsprozess zu integrieren und hervorzuheben (siehe Abbildung 1). Dabei wird vor allem klar, welche Rolle zu welchem Zeitpunkt eine Entscheidung trifft. Gegenwärtig beschriebene BPMN-Modelle betrachten Entscheidungen nur am Rande.
2. **Decision Model:** Modellierung des Wissensmodells der Entscheidung, das die Eingabevariablen und Subentscheidungen beinhaltet. Damit wird

Abb. 2: Definition eines Wissensmodells auf Ebene des Decision Model



F	Input						Output
	Vertrag	Internet	Kunde seit ...	Bezahl-methode	Online Security	monatl. Rate	Churn
1	1-2 Jahres-vertrag	-	-	-	-	-	Nein
2	Monatlich	Glasfaser	[12..24]	-	Nein	-	Ja
3	Monatlich	Glasfaser	> 24	-	Ja	-	Nein
4	Monatlich	Glasfaser	< 12	-	-	-	Ja
5	Monatlich	DSL, kein	> 12	-	-	-	Nein
6	Monatlich	kein	< 12	-	-	-	Nein
7	Monatlich	DSL	< 12	-	-	≤ 55,4	Ja
8	Monatlich	DSL	< 12	-	-	> 55,4	Nein
9	Monatlich	Glasfaser	> 12	Elektronisch	-	-	Nein
10	-	-	-	-	-	-	Nein

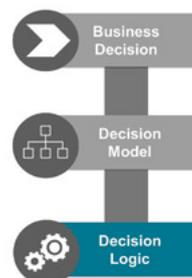


Abb. 3: Definition quantitativer Bedingungen für die Entscheidungsfindung auf Ebene der Decision Logic

deutlich, welche Informationen für die Entscheidungsfindung von Relevanz sind und wie diese in Zusammenhang stehen (siehe Abbildung 2).

3. **Decision Logic:** Konkretisierung der Entscheidungsfindung mit Eingabeparametern, Variablen und einem Regelwerk. Auf diese Weise werden die zu treffenden Entscheidungen messbar und nachvollziehbar. Die genaue Definition von Werten erfolgt automatisiert durch die zugrunde liegende technische Analyse. Somit werden auch komplexe Modelle erklärbar (siehe Abbildung 3).

Im Fall der Abwanderungsprävention kann auf die-

se Weise dargestellt werden, wann und aus welchen Gründen ein Kunde kündigt wird.

Die geschilderte Prozessberatung legt ihren Fokus auf die Entscheidung sowie ein Verständnis dafür, welche Rolle die Entscheidung im Geschäftsprozess spielt. Darüber hinaus wird durch die Verbesserung der Anforderungserhebung die Abhängigkeit von den begrenzten Ressourcen der Fachbereiche reduziert. Eine gemeinsame Notationsgrundlage ermöglicht ein effektiveres und transparentes Zusammenarbeiten in der gesamten Organisation, insbesondere von Management, Analytikern und IT.

Datenmodellierung in BI- und Big-Data-Systemen

10.-11. April.2019, Münster | Referenten: Bodo Hüsemann, Thomas Wörmann

Self Service BI

12. April.2019, Frankfurt a.M. | Referent: Uwe Müller

Visual Business Analytics

13. Mai.2019, Hamburg | Referenten: Dirk U. Proff, Arne-Kristian Schulz

Big Data Technologien

14. Mai.2019, Nürnberg | Referent: Prof. Dr. Jens Albrecht

Data Science Boot Camp

03.-07. Juni 2019, Berlin | Referenten: Prof. Dr. Carsten Felden, Claudia Koschtial

tdwi

AKADEMIE

TDWI SEMINARE 2019

Lernen Sie von führenden Experten.

SIGS DATACOM GmbH

Tel.: +49(0) 22 41 / 23 41-202

E-Mail: julia.seliger@sigs-datacom.de

zum Inhalt

tdwi.eu

JETZT
ANMELDEN!

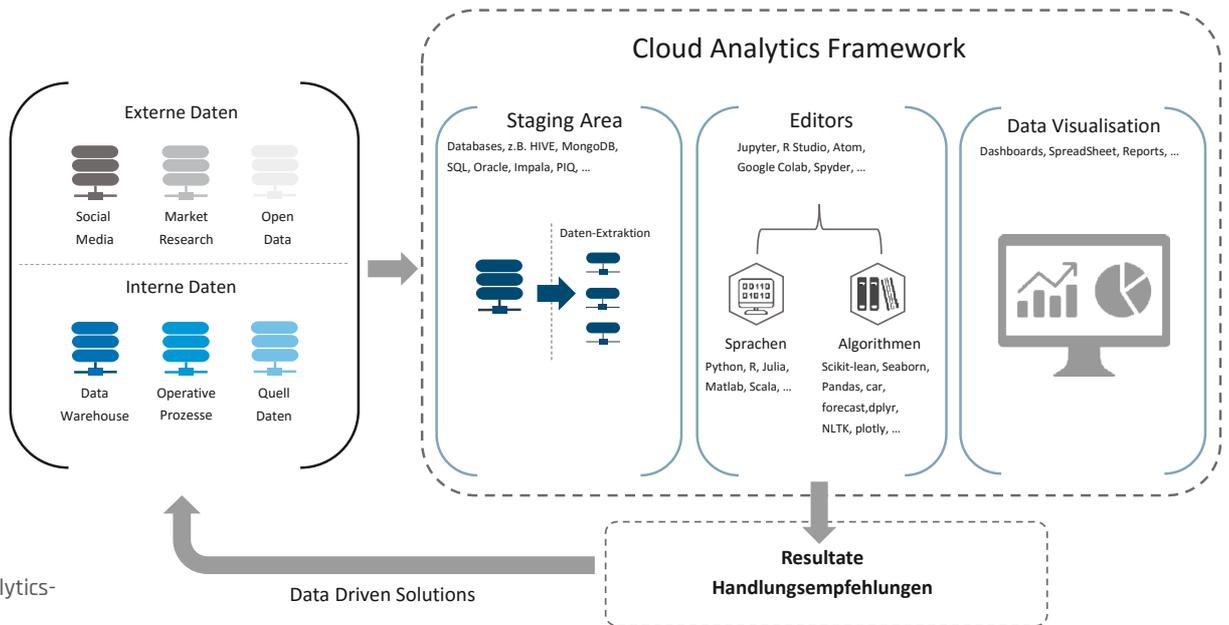


Abb. 4: Cloud-Analytics-Architektur

Technische Lösung

Cloud-Plattformen bieten für Data Analysts ein immer größer werdendes Angebot an Technologien und Tools, um komplexe Analysen und automatisierte Entscheidungssysteme zu erstellen und diese in übersichtlichen Dashboards visuell aufzubereiten. Abbildung 4 stellt eine High-Level-Architektur für die technische Umsetzung dar. In Kombination mit der genauen Kenntnis des Anwendungsfalls bietet sich hier Everything as a Service an, da neben Speicherkapazität und einer performanten Laufzeitumgebung auch Analyse-Software in der Cloud genutzt werden kann. Die Analyse der Daten in der Cloud durchläuft folgende Schritte:

- 1. Data Storage:** Im Fall der Abwanderungsprävention können aufbauend auf der Analyse der Business-Prozesse externe und interne Daten in den Data Storage der Cloud geladen werden. Die Daten werden für die folgenden Schritte bereinigt und strukturiert in den Data Storage integriert. Je nach Anwendungsfall werden die

Daten anschließend in Marts mit den für die Analyse notwendigen Variablen extrahiert.

- 2. Data Analytics:** Bei der Abwanderungsvorhersage handelt es sich um ein Klassifikationsproblem mit binären Entscheidungsoptionen. Daher bieten sich Algorithmen wie Random Forests oder neuronale Netzwerke an. Diese Modelle können große Datenmengen und eine performante Laufzeitkapazität benötigen. Die bedarfsgerechte Nutzung der Cloud eignet sich hierfür ideal.
- 3. Data Visualisation:** Damit der Endnutzer die passende Handlungsalternative schnell begreifen und nutzen kann, sollten die Analyseergebnisse in übersichtlichen Dashboards visualisiert werden.

Was bedeutet das für den konkreten Fall?

Kehren wir zurück zum Callcenter-Mitarbeiter des Telekommunikationsunternehmens. Ein Kunde des

Abb. 5: Kunden-Scorecard als Dashboard für Callcenter-Mitarbeiter

The dashboard, titled "Turn your Churn", provides a comprehensive overview of a customer's profile and churn risk. It includes a sidebar with navigation options like "Customer Scorecard", "Churn Statistics", and "Correlations".

Churn Risiko: A gauge chart shows a risk level of 71.9%.

Strategies: Three strategy buttons are visible: "Jahresvertrag anbieten" (Main Strategy), "tech support anbieten" (Commercial Strategy), and "OK" (Financial Strategy).

Merkmal	Wert
Kunde seit	12 Monaten
Vertrag	Monatlich kündbar
Internet Service	Glasfaser
Monatliche Kosten	86,55 €
Online Backup Package	Nein
Online Security Package	Nein
Geräte-Versicherung	Nein
Tech Support Package	Nein
Streaming Package	Nein
Telefon Service Package	Ja
Bezahlmethode	Elektronische Überweisung

Kundendaten:

Name	Richard Müller
Geburtsdatum	06.08.1977
Adresse	Im Neuenheimer Feld 131 69120 Heidelberg
Telefonnummer	+49 152 648 824 56
E-Mail	Richard.mueller77@yahoo.de

genannten Telekommunikationsanbieters wendet sich telefonisch an das Service Center. Die Daten des Kunden liegen schon in der Cloud. Bereits hinterlegt ist eine auf Basis der vorhandenen Daten kalkulierte individuelle Abwanderungswahrscheinlichkeit. In Betracht gezogen wird auch die Gewichtung einzelner Faktoren, sodass eine auf diese Person zugeschnittene Kundenstrategie vorliegt. Die aus dem Gespräch gewonnenen Daten fließen in Echtzeit in die Datenmenge ein, werden der Analyse des trainierten Modells unterzogen und ändern, vor allem aufgrund ihrer Aktualität, die Strategie. Aufgrund der Performance der Cloud-Architektur liegen Handlungsempfehlungen für den Mitarbeiter noch während des Gesprächs vor. Darauf aufbauend kann er an dieser Stelle Angebote formulieren und die Beratung kundenindividuell zuschneiden.

Fazit

Um Entscheidungen nachvollziehbar und automatisierbar zu machen, werden diese zuerst im Geschäftsprozess identifiziert, daraus ein unternehmensweites Wissensmodell abgeleitet und dessen Inputparameter analysiert. Die Analyse dieser Parameter benötigt große Datenmengen

und baut auf komplexen Algorithmen auf. Um diesen hohen Grad an Komplexität zu ermöglichen, wird eine enorme Rechenleistung benötigt, die durch Cloud-Technologien erreicht werden kann. Cloud Analytics ist der nächste notwendige Schritt in der BI-Welt und vor allem dann erfolgreich, wenn Prozessberatung, Datenanalyse und die gewählte Cloud-Architektur ergänzend ineinandergreifen.

Literatur

[Far11] Farber, M. et al.: Massive Data Analytics and the Cloud. Booz Allen Inc. 2011

[Hog11] Hogan, M. et al.: (2011). Nist cloud computing standards roadmap. NIST Special Publication, 35, 2011, S. 6-11

[OEE13] Osman, A. / El-Refaei, M. / Elnaggar, A.: Towards real-time analytics in the cloud. IEEE, Ninth World Congress on Services, 2013, S. 428-435

[OMG18] OMG (Object Management Group): About the Decision Model and Notation Specification Version 1.2 Beta (DMN). August 2018, <http://www.omg.org/spec/DMN/>, abgerufen am 27.12.2018

[Sma18] Smartweb: Mobilfunk Report Deutschland Q3 2017. April 2018, <https://www.smartweb.de/mobilfunk-report-deutschland-q3-2017>, abgerufen am 27.12.2018

[Sta18] <https://de.statista.com/statistik/daten/studie/237715/umfrage/monatliche-churn-rate-der-deutschen-telekom-im-mobilfunk-in-deutschland/>, abgerufen am 27.12.2018

[ZCB10] Zhang, Q. / Cheng, L. / Boutaba, R.: Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1 (1), 2010, S. 7-18



Künstliche Intelligenz für den Menschen

KONFERENZ FÜR KÜNSTLICHE INTELLIGENZ

24.-26. JUNI 2019 | MOC MÜNCHEN

www.ai4u-konferenz.de

[zum Inhalt](#)

