

Cloud Computing als Fundament für Analytics

Big Data und BI verschmelzen zu einer analytischen Architektur auf hybrider Infrastruktur

Im Zuge der Digitalisierung der Geschäftsmodelle verschmelzen die eher in die Vergangenheit schauende klassische Business-Intelligence-Welt und das in die Zukunft gerichtete Analyseverständnis von Big Data zu einer gemeinsamen Analytics-Plattform.

Diese neue „analytische Plattform“ zeichnet sich durch technologische Vielfalt und stetige Veränderung aus und verbindet in verstärktem Maße unterschiedliche Cloud-Ansätze zu einer hybriden analytischen Infrastruktur. Der Artikel betrachtet die Motive für die Verschmelzung von BI und Big Data und beschreibt einen Blueprint für eine analytische Architektur. Anschließend geht er näher auf hybride Cloud-Ansätze ein und zeigt, wie Geschwindigkeit, Flexibilität, Skalierbarkeit und Change schon in der Architektur implementiert werden können.

Die Zukunft von BI und Big Data

Reale und virtuelle Welt gehen heute immer mehr ineinander über. Eine radikale Transformation unseres beruflichen und privaten Umfeldes ist bereits im vollen Gange (1). Gerade die Ex-post-Sicht der Business Intelligence (BI) gerät damit unter Druck, denn der Wert einer Information nimmt in Relation zurzeit stetig ab (2). Geschwindigkeit ist gefragt. Nichts ist schließlich so alt, wie die Zeitung von gestern!

WEB-TIPP:

www.opitz-consulting.com

Zwar hat die IT in den letzten Jahren BI-Lösungen geschaffen, die die Zeitspanne zwischen Verarbeitung und Kommunikation von Informationen reduzieren können. Die Digitalisierung verändert die Anforderungen jedoch noch einmal dramatisch durch ihre Forderung nach Near-Realtime-Analysen und -Nutzung. Somit entfällt der zeitliche Versatz, der den aktuellen BI-Architekturen innewohnt. Dazu kommt, dass der Blick der BI-Welt eher auf die Vergangenheit gerichtet ist. Die Vergangenheit ist bei digitalen Geschäftsmodellen jedoch oft nur ein Teilaspekt: Next-Best-Offer, Predictive Maintenance oder Deep Learning sind als Analyseansätze integraler Bestandteil von operativen Prozessen. Sie verwenden Prognoseergebnisse der Gegenwart zur Steuerung von Entscheidungen.

Parallel dazu verspricht Cloud Computing „IT Power aus der Steckdose“. Welche IT träumt nicht davon? Die IT-Kosten beziehen sich damit auf die tatsächliche Nutzung und auf Preismodelle, die nutzungsbezogen abrechnen. Betriebswirtschaftlich ist somit eine Wandlung der Kapitalbindung durch IT-Infrastruktur (CAPEX) in operative Betriebskosten (OPEX) möglich.

Analytics bei der Digitalisierung

Im Folgenden geht es kurz um die grundlegenden Bausteine eines digitalen Geschäftsmodells. Anhand von diesen Bausteinen, zeigt sich, wie dringend ein Regelkreis aus Big Data und BI benötigt wird. Hieraus motiviert sich am Ende die Zusammenlegung zu einer analytischen Infrastruktur.

Ein digitales Geschäftsmodell stellt nur eine Ausprägung eines klassischen Geschäftsmodells dar, mit der Besonderheit, dass die Leistungserbringung von der Nutzung der Informationstechnologie abhängig ist und der Mehrwert für den Kunden meist ein digitales und somit virtuelles Produkt ist (3). Doch ganz gleich, welche Klassifikation ein angestrebtes digitales Geschäftsmodell besitzt, es wird in seiner Ausprägung grundlegende fachliche Komponenten besitzen. Diese Komponenten sind in Bild 1 näher aufgeführt.

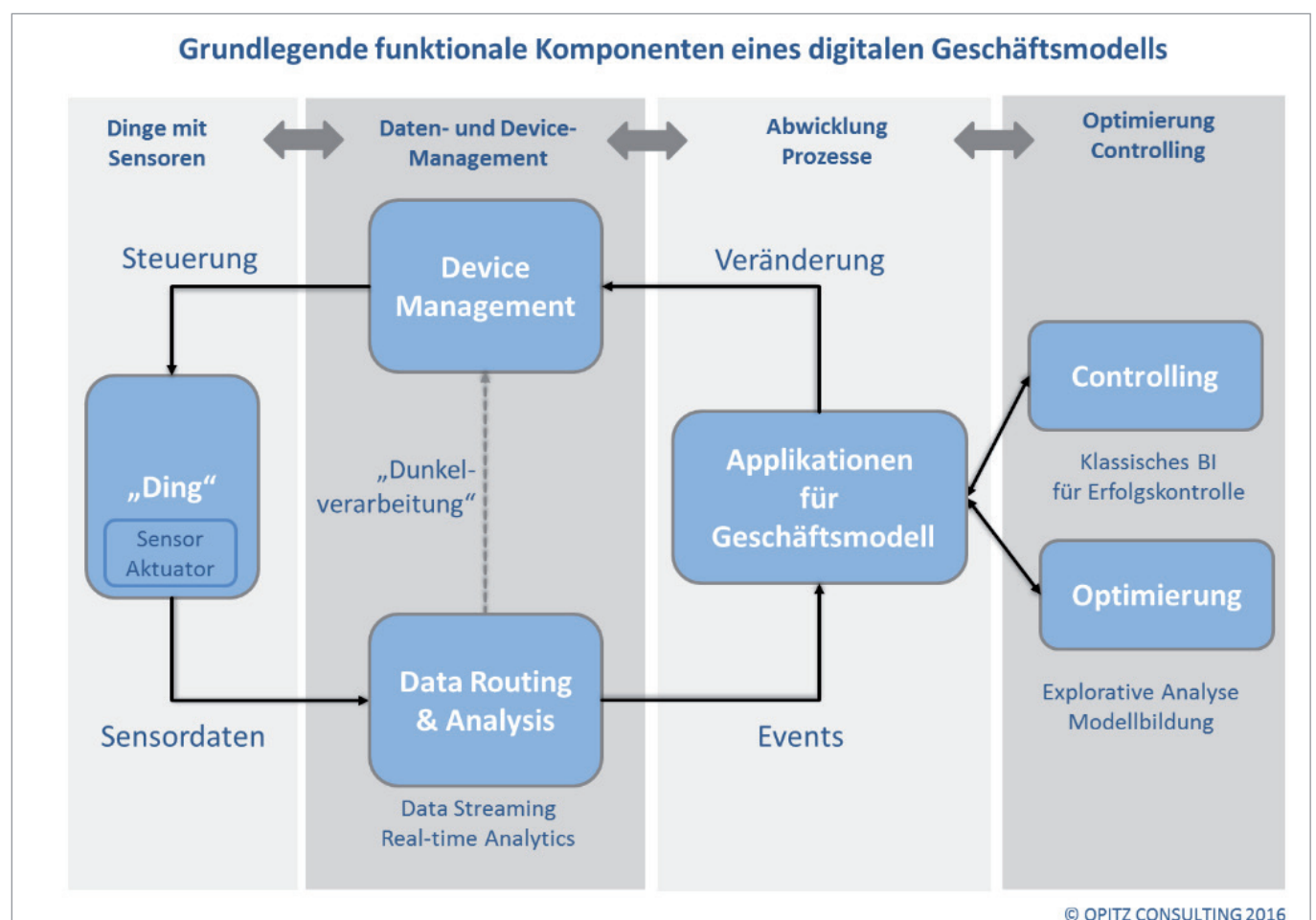


Bild 1: Fachlichen Komponenten eines digitalen Geschäftsmodells.

Der Wertschöpfungsprozess eines digitalen Geschäftsmodells beinhaltet einen „lernenden Regelkreis“. Dieser beginnt und endet mit der Nutzung des Gegenstands oder des Endgeräts. Verhalten und Zustand des Produkts werden durch die Sensorik seiner Daten nach außen sichtbar (4).

Die übermittelten Sensordaten der Endgeräte werden im Rahmen des Datenmanagements aufbereitet. Aufgrund der großen Menge und Variabilität der Sensordaten, wird die Verarbeitung in der Regel als Fast-Data (5) implementiert. Die notwendige Geschwindigkeit bei der Datenaufnahme und Verarbeitung wird über Architekturen erreicht, die sich über mehrere Ebenen des Netzwerks erstrecken. Was das Datenvolumen angeht, helfen die Architekturkonzepte des Big Data weiter. Die Streaming-Daten werden in Near Realtime gefiltert und analysiert, wobei bereits eine lokale Intelligenz zu einer Veränderung der Nutzung führen kann. (Siehe auch „Dunkelverarbeitung“ in Bild 1.)

Im Mittelpunkt vieler digitaler Geschäftsmodelle steht oft ein Verbund an Endgeräten, etwa Beleuchtungseinheiten, Parkplätze, Mietobjekte, Connected Car etc., deren Verhalten gemeinsam an die veränderten Umweltbedingungen angepasst werden muss. Hierzu werden fachliche Nachrichten an ein lokales oder zentralisiertes Device Management versandt. Das Device Management ist nun in der Lage, die logischen Events der unterschiedlichen Endgeräte aufzubereiten und im Netzwerk zu propagieren, um das Verhalten der Endgeräte durch eine Ansteuerung der Aktuatoren zu beeinflussen. Diesen Regelkreis zusammen mit dem Datenmanagement stellt der Kasten „Daten- und Device Management“ in Bild 1 dar. Die Frontend- und Backend-Applikationen des digitalen Geschäftsmodells nutzen anschließend die aufbereiteten Daten zur Etablierung des neuen Angebots oder eines Mehrwerts für den Kunden.

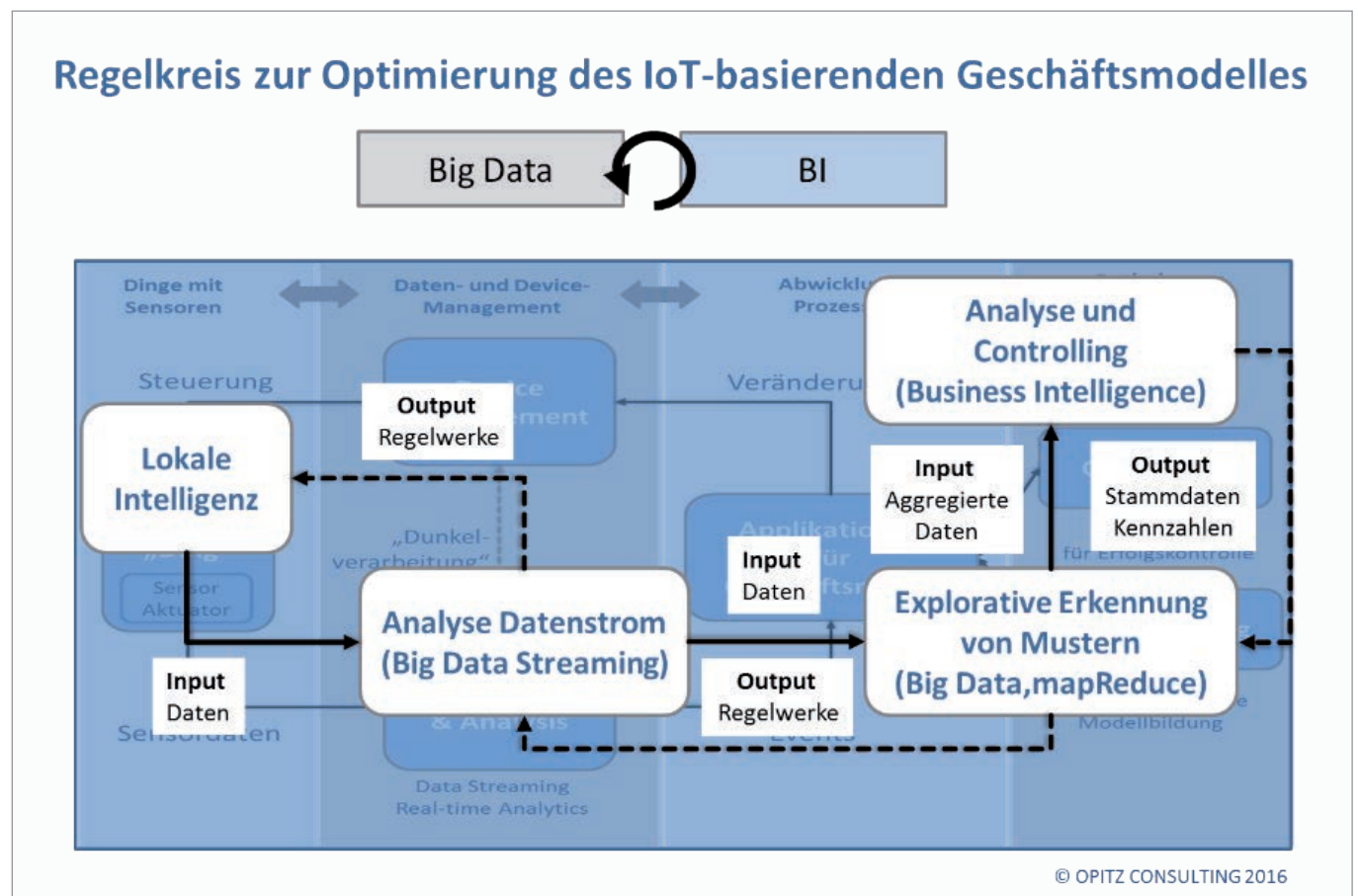


Bild 2: Der analytische Regelkreis.

Bild 2 zeigt den Regelkreis, der durch Analytik-Maßnahmen geschlossen wird. Diese Maßnahmen umfassen neben der Erfolgsmessung, die in Form eines klassischen Controllings über einer BI-Welt stattfindet, auch eine explorative Erkennung von Mustern, etwa des Kunden- und Produktverhaltens. Ziel ist eine Optimierung des digitalen Geschäftsmodells. Die Optimierung erfolgt meist in einem explorativen Ansatz, der zeitlich eher unkritisch ist – einem sogenannten Batch-orientierten Big-Data-Ansatz. Hierzu zählen statistische Prognosemodelle ebenso wie die Ansätze des Deep Learning (6). Die Modelle werden fortlaufend anhand von Testfällen aus der Praxis überprüft und somit kontinuierlich verbessert. Die Erkenntnisse daraus können verwendet werden, um entsprechende Regelwerke auf den Streaming-Daten für eine Near-Realtime-Filterung und -Prüfung zu verwenden. Das eher klassische Controlling rundet den Regelkreis ab und liefert in einer Ex-post- und betriebswirtschaftlichen Sicht Informationen zum Geschäftsmodell selbst.

Die analytische Plattform

Bild 3 verdeutlicht das Zusammenspiel der Ansätze von Big Data und klassischer BI-Welt und beschreibt die Architekturkomponenten.

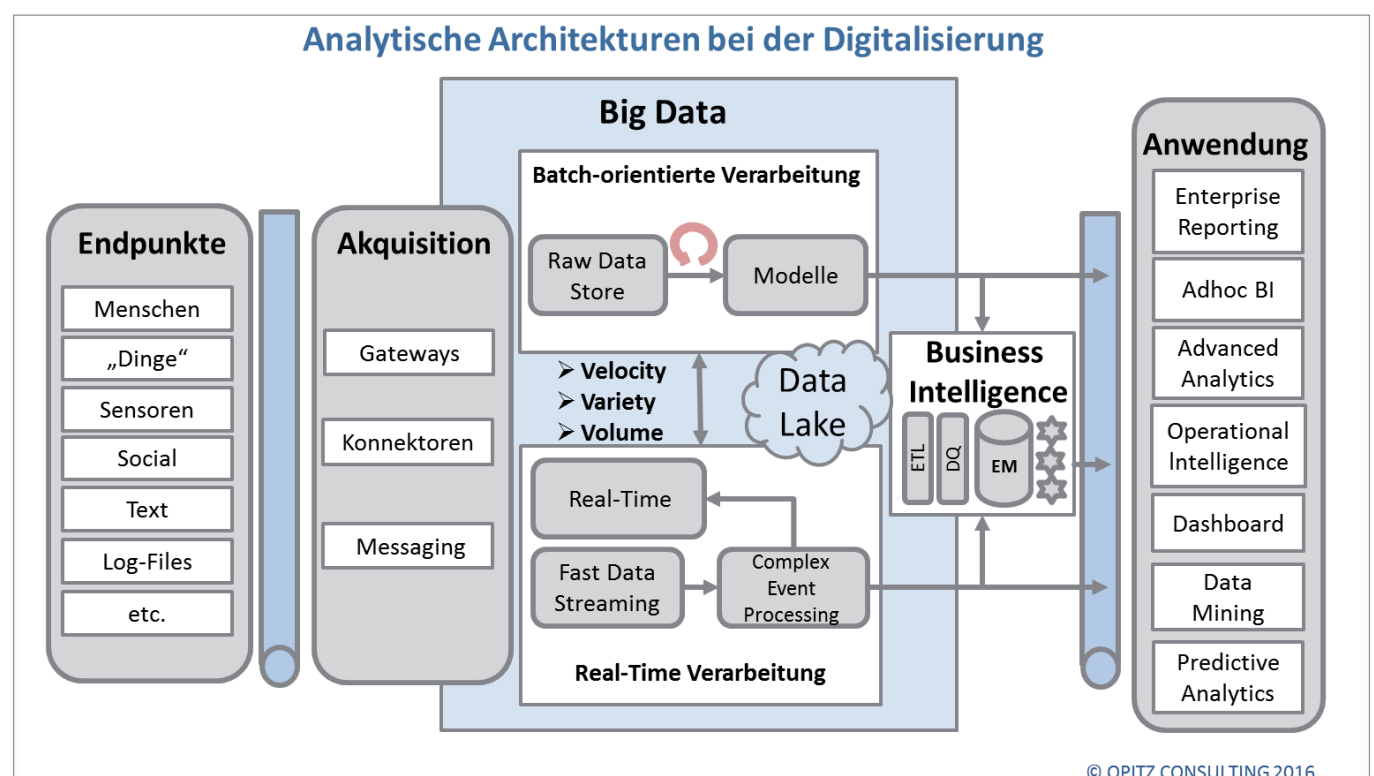


Bild 3: Analytische Architekturen bei der Digitalisierung.

Sowohl Near-Realtime-Entscheidungsfindung, klassisches BI als auch Modellbildung für Prognosen fließen beim Aufbau einer analytischen Architektur mit ein. Es wird also deutlich, dass sowohl eine klassische BI/DWH-Sicht, als auch eine reine Big-Data-Sicht für sich genommen jeweils unzureichend wären. Auf eben dieser Annahme baut die beschriebene Referenzarchitektur auf. Dieser „analytischen Architektur“ liegt die Annahme zugrunde, dass mittelfristig die klassische, eher Ex-post-orientierte Business Intelligence mit Data-Warehouse-Ansätzen und die neuen Big-Data-Ansätze zu umfassenden Architekturen verschmelzen werden. Auch wenn beide Ansätze hinsichtlich der technischen Implementierung äußerst verschieden sind, so ist die betriebswirtschaftliche Aufgabenstellung bei beiden identisch: Beiden geht es darum, aus Daten Informationen zu gewinnen und das Geschäft durch Analysen besser steuern zu können.

Am Beispiel eines Geschäftsmodells zum Verleih von Fahrrädern wird diese Verschmelzung gut sichtbar: In Near Realtime analysiert ein Unternehmen die Sensordaten seiner Fahrräder und erkennt Muster, die auf eine notwendige Reparatur schließen lassen. Hierzu generiert das System einen Serviceauftrag an das Backend-Service-Modul, das als SaaS-Lösung implementiert sein kann. Des Weiteren gelingt es, mit zeitlichem Versatz in Batch-orientierten statistischen Verfahren die Positionen der Fahrräder mit ihrer Mietdauer sowie mit zukünftigen Events und Großereignissen zu korrelieren. Aus dieser Korrelation ergeben sich sogenannte „Lokalisierungsaufträge“, die generiert werden, um Fahrräder an bestimmte geographische Punkte in die Nähe potenzieller Kunden zu bringen und damit bessere Ausleihquoten zu erreichen.

Letztendlich fließen die Fahrradinformationen nebst Daten aus den betriebswirtschaftlichen Systemen (MaWi, FiBu, Service etc.) aggregiert in ein Data Warehouse und ermöglichen so ein umfassendes finanzielles Controlling. Ferner werden die Batch-orientierten Ansätze verwendet, um neue Muster zu erkennen und diese, so möglich, auf dem Stream für eine Near-Realtime-Verarbeitung zu implementieren. Das Fallbeispiel zeigt, wie klassische BI-Ansätze und neue Big-Data-Ansätze mit den Spielarten der Batch-orientierten explorativen Simulation mit statistischen Modellen und der Near-Realtime-Analyse des Datenstroms erst in ihrer Kombination zu wesentlichen Erfolgsfaktoren werden.

Die geschilderten fachlichen Komponenten benötigen unterschiedliche Systeminfrastrukturen. Daher bilden hybride Infrastrukturansätze unter Einbeziehung von Cloud-Lösungen ein passendes Fundament für die analytische Architektur.

Hybride Cloud-Lösungen

Die meisten Unternehmen nutzen bereits eine Vielzahl an Cloud-Lösungen mit unterschiedlichen Liefer- und Servicemodellen. Dies gilt für BI-Branchenlösungen, die vermehrt als SaaS-Lösungen angeboten werden, aber auch für PaaS-Lösungen für Big Data oder BI-Suiten. So ist es nicht verwunderlich, dass ein ganzheitlicher Ansatz für analytische Lösungen notwendig wird. Der ganzheitliche Ansatz verschafft hybriden Ansätzen bei analytischen Lösungen bei allem Wildwuchs eine ordnende Sicht.

Anstatt auf die sich so entwickelnde Dezentralität mit Bestrebungen zur Zentralisierung und Standardisierung zu antworten, und somit wieder Geschwindigkeit und Innovation zu hemmen, kann die IT im Rahmen Hybrider Cloud-Lösungen einen „Design for Change“-Ansatz verfolgen und die analytische Architektur auf einer reaktiven Infrastruktur implementieren.

Bei dezentralen Ansätzen treten schnell Herausforderungen für die Datenqualität auf. Insbesondere sind Inkonsistenzen bei der Nutzung von Stamm-, Meta- und Referenzdaten zu beobachten. In der Folge sinkt das Vertrauen in die Analysen und die dezentralen Plattformen geraten unter Druck. An dieser Stelle liefert die klassische BI-Welt einen Mehrwert als Lieferant von qualitätsgesicherten Daten. Voraussetzung hierfür sind allerdings eine übergreifende Governance beim Datenmanagement sowie die Transparenz über die Nutzung relevanter zentraler Daten.

Bild 4 stellt die unterschiedlichen Implementierungsansätze inklusive der Cloud-Ansätze für analytische Lösungen in einem Blueprint für eine hybride Infrastruktur der analytischen Welt dar. Hier gibt es kein Richtig oder Falsch, sondern Entscheidungen müssen situativ auf Basis aktueller Anforderungen getroffen werden.

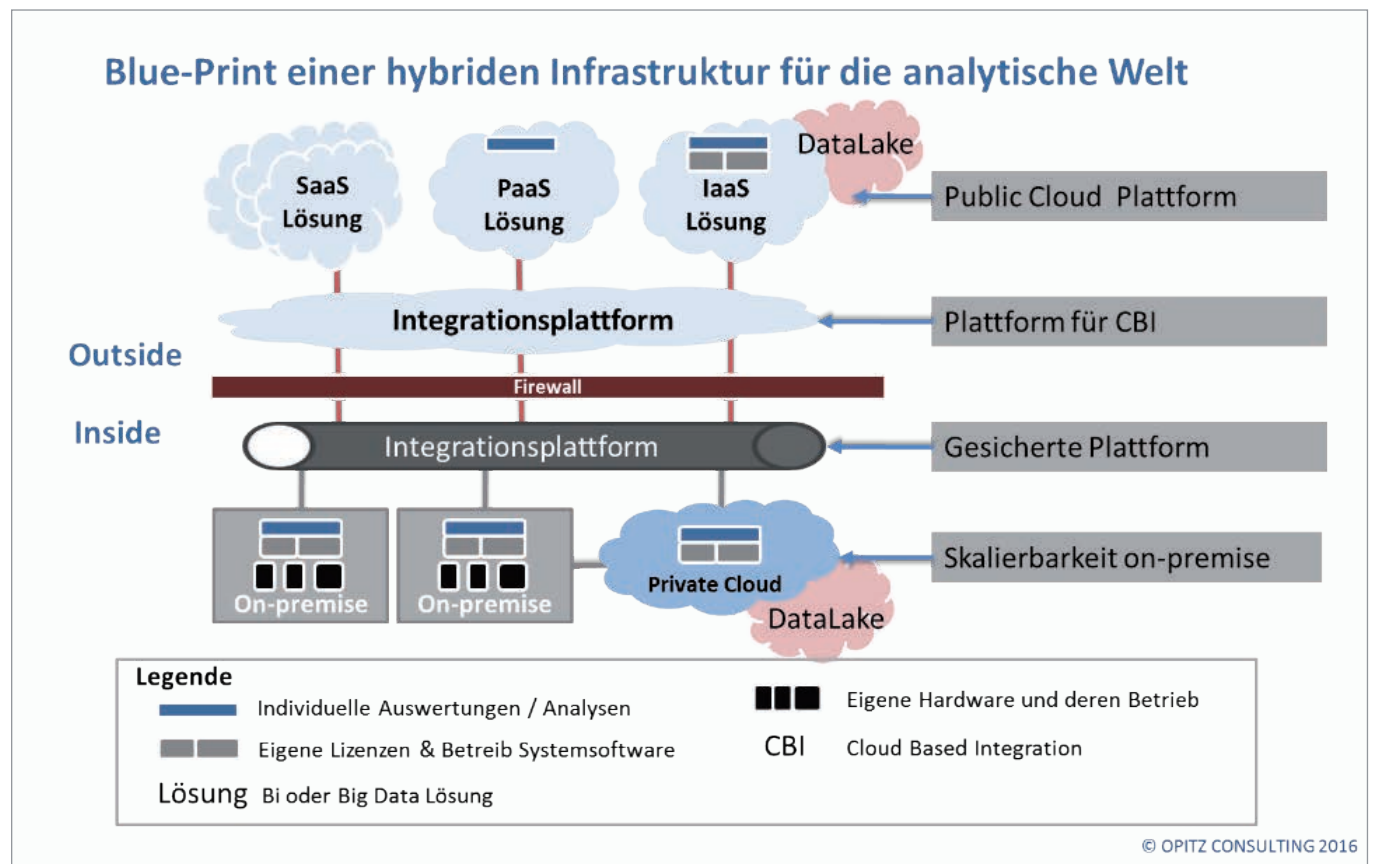


Bild 4: Zukunftsweisende hybride Infrastrukturen.

Bild 4 zeigt unterschiedliche Szenarien für die Implementierung einer BI- oder Big-Data-Lösung:

- Nutzung von Public-Cloud-Lösungen ohne eigene Lizenzen
 - als SaaS-Lösung
 - als PaaS-Plattform für BI oder Big Data
- Nutzung eigener BI- und/oder Big-Data-Lizenzen
 - in einer Public-IaaS-Lösung
 - in einer Private-Cloud-Lösung
 - On-Premise auf der eigenen Infrastruktur
- Nutzung einer Integrationsplattform für die Datenbewirtschaftung
 - in einer Public Cloud (Cloud Based Integration)
 - On-Premise

Für eine Cloud-Lösung sprechen einige Vorteile: Beim Betrieb einer Cloud-BI-Lösung mit eigenen Lizenzen im Rahmen einer IaaS-Lösung ist die Unternehmens-IT mit den gleichen IT-Service-Prozessen belastet wie bei einer On-Premise-Lösung. Im Gegensatz dazu entfallen bei einer PaaS-Lösung mit nutzungsabhängiger Abrechnung die direkten Lizenzkosten ebenso wie die Verantwortung für wesentliche IT-Service-Prozesse. Jedoch müssen bei PaaS (wie auch bei SaaS) die internen Verrechnungsalgorithmen angepasst werden.

Bei Pilot-Projekten oder einem Proof-of-Concept (PoC) hat ein Cloud-Ansatz Vorteile für die Time-to-Value (7). In vielen Fällen ermöglichen die Anbieter von Cloud-Lösungen

- eine schnelle Bereitstellung der benötigten Systemressourcen inklusive der Software selbst,
- eine umfassende Testumgebung,
- eine nutzungsabhängige Abrechnung,
- die leichtere Einführung von modernen Lean-Startup-Ansätzen nach dem „Fail fast, fail early“-Prinzip und deren Nutzung für den PoC.

Unternehmen beginnen verstärkt mit der Implementierung von Data Lakes (8)(9). Eine Vielzahl von neuartigen Daten werden in die Analysen einbezogen: Sensordaten, Lokalisierungsdaten, Geodaten, Text Messages, Log File etc. Die Heterogenität dieser Daten, die Variabilität der Datenstrukturen und die schnellen Zyklen der Formatänderungen sind mit den klassischen Datenbewirtschaftungsprozessen aus der BI-Welt schwerlich realisierbar.

Ferner lassen diese Daten sich nur mit hohem Aufwand in die notwendigen relationalen Strukturen wie zum Beispiel Star-Schemata überführen. An dieser Stelle sind NoSQL-Datenhaltungen (10) auf dem Vormarsch und die Data-Lake-Konzepte lassen die Ablage der Daten in heterogenen Datenhaltungen zu.

Neben den betriebswirtschaftlichen Überlegungen hinsichtlich Investitionsschutz und Abschreibungszeiträumen sind am Ende vor allem zwei Kriterien bei der Entscheidung für oder gegen die Cloud-Lösung ausschlaggebend: Der Datenschutz und die Kosten.



Rolf Scheuch
Managing Director and Chief Strategy
Officer bei Opitz Consulting

Datenschutz

Beim Datenschutz ist es notwendig, die gewachsenen, oft schon veralteten Datenschutzbestimmungen und Regelungen zu hinterfragen, um die Entscheidung gegen eine Cloud-Lösung nicht vorschnell zu treffen. Dabei empfiehlt es sich, keine zentrale übergreifende Entscheidung für alle BI- und Big-Data-Lösungen zu treffen, sondern zunächst die relevanten Datenobjekte für den analytischen Bereich zu definieren und daraufhin eine Entscheidung für das jeweilige Sub-Set zu treffen. Abgesehen davon sind alle in Bild 4 aufgeführten Spielarten der Implementierung realistisch und sollten geprüft werden.

Kostentreiber

Ein wesentlicher Kostentreiber ist der Preis für die Datenspeicherung und den Datenexport von einer Cloud-Infrastruktur in eine andere Infrastruktur. Je größer die Datenmenge ist, die aus der Cloud in BI- und Big-Data-Systeme fließt, desto preiswerter wird eine Inside-Lösung im Unternehmensnetzwerk sein. Aktuell sind die Preismodelle der Cloud-Anbieter nur für konkrete Anwendungsfälle vergleichbar.

Eine Besonderheit besteht in der Near-Realtime-Verarbeitung durch Streaming-Lösungen, Fast-Data-Ansätze, Complex Event Processing und Realtime-Analytik, wie in Bild 3 aufgeführt. Diese Lösungsansätze findet man häufig beim Internet der Dinge mit einer Vielzahl an heterogenen Sensordaten. Die IoT-Plattformen selbst liegen in der Regel schon außerhalb des Unternehmensnetzwerks und die Security bezieht sich auf einen Zugriffschutz zu den Endgeräten und das Einspielen von fehlerhaften Daten auf die IoT-Plattform. In der Regel besteht kein ausgeprägter Datenschutz. Für den Fall eines Datenverlusts bietet sich für das Backup eine Cloud-Lösung als Plattform an.

Für das Balance-Verhältnis zwischen On-Premise-Lösungen und Cloud-Lösungen steht das „Center of Gravity“ (11). Es setzt den Anteil der Cloud-Lösungen in Bezug zu den On-Premise-Installationen. Aus diesem Verhältnis lassen sich wichtige Schlüsse ziehen, denn: Je höher der Einsatz unterschiedlicher Cloud-Lösungen ist, desto eher verschiebt sich auch die Integrationsplattform in die Cloud und rückt damit näher an den Ursprung der Daten.

Fazit

Die Verbreitung von Cloud-Lösungen ist unaufhaltsam. Unterschiedliche Liefer- und Servicemodelle nehmen Einzug in die Unternehmen, hybride Ansätze beim Cloud Computing werden die Regel. Die Client/Server-Revolution hat uns gelehrt, die Verbreitung der Cloud-Lösungen seitens der zentralen IT zu unterstützen, anstatt als „Bremsen“ von Innovation aufzutreten.

Auch der Big-Data-Zug rollt. Und für eine zentrale BI-Einheit im Unternehmen macht es wenig Sinn, die Big-Data-Welle einzudämmen. Wie geschildert ergänzen sich die beiden Facetten analytischer Fragestellungen gerade im Hinblick auf die kontinuierliche Verbesserung von digitalen Geschäftsmodellen. Aus diesem Grunde werden die zentralistische BI-Welt und die eher föderale Big-Data-Welt zu einer neuartigen analytischen Plattform verschmelzen. Das Thema der Governance der analytischen Plattform wird in dem Zusammenhang ein neues wichtiges Handlungsfeld.

Auch etablierte Denkmuster der zentralen Plattform werden sich wandeln. Zu dynamisch ist der Markt und zu groß sind die Chancen neuer Geschäftsmodelle durch neue Technologien. Der „Design for Change“-Gedanke nimmt Einzug in die Denkmuster und Cloud-Computing wird zu einem entscheidenden Hilfsmittel um die Time-to-Value zu reduzieren und Flexibilität zu erhöhen. Cloud Computing wird ein Fundament der neuen analytischen Plattform werden.

ROLF SCHEUCH

QUELLEN UND ANMERKUNGEN

- (1) Tim Cole: Digitale Transformation, Franz Vahlen Verlag 2015
- (2) Dominik Bial: IoT erfordert neues Denken, BI-Spektrum 02-2016
- (3) Thomas Bieger, Stephan Reinhold: Innovative Geschäftsmodelle: Konzeptionelle Grundlagen, Gestaltungsfelder und unternehmerische Praxis. 2011 in: T. Bieger, D. zu Knyphausen-Aufseß, C. Krys (Hg.), Innovative Geschäftsmodelle, Springer Verlag 2011
- (4) Ferner kann der Gegenstand auch ein Smart Device wie etwa auch ein Tablet oder Smartphone sein.
- (5) BITKOM, Frauenhofer (Hg.): Big-Data-Technologien – Wissen für Entscheider, <https://www.bitkom.org/Publikationen/2014/Leitfaden/Big-Data-Technologien-Wissen-fuer-Entscheider/140228-Big-Data-Technologien-Wissen-fuer-Entscheider.pdf> (Download Juni 2015)
- (6) Einen guten Einstieg erhält man auf dieser Website: <http://deeplearning.net/> (Download 10/2016)
- (7) Nello Franco: Time To Value: A Key Metric, <https://blog.nellofranco.com/2013/08/25/time-to-value-a-key-metric/> (Download 10/2016)
- (8) Tamara Dull: Data Lake vs Data Warehouse: Key Differences, <http://www.kdnuggets.com/2015/09/data-lake-vs-data-warehouse-key-differences.html> (Download 10/2016)
- (9) Matthias Reiss und Stephan Reimann: Das Data Lake Konzept: Der Schatz im Datensee, <https://www.it-daily.net/it-management/big-data-analytics/11222-das-data-lake-konzept-der-schatz-im-datensee> (Download 10/2016)
- (10) Eine Übersicht und Einführung findet sich unter <http://nosql-database.org/> (Download 10/2016)
- (11) Jim Harris: The Cloud is shifting our Center of Gravity, <http://www.ocdqbog.com/home/the-cloud-is-shifting-our-center-of-gravity.html> (Download am 28.5.2015)