

# Data Lakes – Möglichkeiten und Herausforderungen für eine effiziente Erkenntnisgewinnung



Tomas Chroust



Josef Nemecek



Sandro Kürschner

*Die Menge an Informationen, die von Unternehmen gespeichert, verarbeitet und ausgewertet werden, nimmt zu wie nie zuvor. Dies in den Dimensionen Volumen, Geschwindigkeit und Diversität. Die Wertschöpfung aus der Information muss dabei Schritt halten mit dem Aufwand des Managements der Information. Aufgrund einer stetig steigenden Erwartungshaltung der Fachbereiche muss die IT vermehrt entsprechende innovative Technologien bereitstellen, die ein breites Spektrum an Anwendungsszenarien abdecken. Eine Basiskomponente kann dabei ein sogenannter «Data Lake» oder auch «polyglotter Datenspeicher» sein, der die Ablage unterschiedlichster Daten in verschiedensten Formaten des Originalzustands – ohne vorbestimmte Datenstruktur – ermöglicht. Viele bedeutende Versicherungsunternehmen beginnen mit «Data Lakes» das Potenzial ihrer Daten zu nutzen, sich damit einen Vorsprung gegenüber der Konkurrenz zu sichern und dies für ihren unternehmerischen Erfolg einzusetzen.<sup>1</sup>*

## Forderung nach Agilität

Traditionelles Data-Warehousing und Business Intelligence werden den Forderungen nach Agilität und gleichzeitiger Nutzeffizienz zunehmend nicht mehr gerecht. Zentral sind in diesem Zusammenhang Entscheidungen, Implementierungen und Ressourcen. Um Entscheidungen agiler zu treffen, sollte die Bereitstellung von Analyseergebnissen auf Basis von vollständigen, hochwertigen und zuverlässigen Informationen zeitnah erfolgen. Implementierungen können agiler gestaltet werden, wenn die Entwicklungsprozesse in einzelnen, bewusst klein gehaltenen iterativen Implementierungsschritten stattfinden, wobei auch kurzfristige Änderungen und Anforderungen berücksichtigt werden können. Durch eine erhöhte Produktivität kann die Nutzung

von Ressourcen agiler gestaltet sowie niedrigere Infrastrukturkosten und optimierte Datenintegration adressiert werden.

Hemmnisse auf dem Weg zur Agilität sind die in traditionellen Architekturen vorherrschenden ETL-Prozesse – sprich Extraktions-, Transformations- und Ladeprozesse – sowie vordefinierte Schemata. Der Zwang zur Ex-ante-Modellierung schliesst gegebenenfalls die Unterstützung unerwarteter Businessanfragen aus. Neue explorative Ansätze erzwingen ein grundsätzliches Hinterfragen traditioneller Vorgehensmodelle und Architekturen.

Bisher wurden Daten in traditionellen Data-Warehouse-Systemen bewirtschaftet, um vordefinierte Auswertungen vornehmen zu können. Jedoch schränken diese den Anwender selbst bei der Auswertung von Ad-hoc-Analysen ein. Völlig neuartige Auswertungen sind, wenn überhaupt, lediglich mit langen Vorlaufzeiten möglich, da die Daten strukturiert und verarbeitet werden müssen. Diese Verarbeitung zum Zeitpunkt des Speicherns ist zwar effizient, jedoch bei unstrukturierten Daten oder der investigativen Datenauswertung ungeeignet. Auf diese Weise geht ein exploratives Durchforsten der Rohdaten nach neuen Zusammenhängen und Erkenntnissen durch das Strukturieren und Normalisieren der Daten verloren.

Ein entscheidender, stark kostenreduzierender Faktor bei einem Data Lake ist der Vorteil, dass die originären Daten nicht wie im klassischen Data-Warehouse in mehreren persistierten Schichten gespeichert werden müssen. Dies ermöglicht zusätzlich eine vollständige Sicht auf die ursprünglichen Applikationsdaten und stellt sicher, dass keine Informationen auf dem Weg der Datenverdichtung verloren gehen – einer der Gründe, weshalb grosse Datenbankhersteller auf In-Memory / Tiering-Speichertechnologien setzen. Ei-

## Die Autoren

Tomas Chroust ist Senior Manager bei BearingPoint Switzerland AG und verantwortet den Bereich Information Management mit Fokus auf Big Data & Smart Analytics.

Josef Nemecek ist Technical Architect bei BearingPoint Switzerland AG und fokussiert auf Massive Parallel Processing Systeme sowie In-Memory Technologien.

Sandro Kürschner ist Senior Consultant, bei BearingPoint Switzerland AG und fokussiert auf innovative Datenvirtualisierungskonzepte sowie Daten Integration mittels Semantischen Layers.

nerseits erzielen sie damit eine höhere Performance der Datenverarbeitung, andererseits können dadurch auch sämtliche verfügbaren transaktionalen Daten berücksichtigt werden. Dabei verschmelzen die bislang klar bekannten Grenzen zwischen einer transaktional orientierten (OLTP) und einer analytisch orientierten (OLAP) Systemwelt immer stärker. Ein wesentlicher Faktor für die effektive Nutzung eines Data Lakes ist die semantische Integration der zugrunde liegenden verschiedenen Informationsobjekte. Die bislang mit sehr viel Aufwand verbundene Aufgabe der Normierung / Homogenisierung der Daten im Data-Warehousing wird in eine Metaebene verlagert. Dabei werden Informationen aus verschiedenen Domänen des Unternehmens miteinander semantisch kombiniert.

### Erweiterung der Vielfalt von Datenquellen und Datenarten

Ein Hemmnis bei der Erkenntnisgewinnung aus unternehmenseigenen und externen Datenquellen ist sowohl deren Vielfalt als auch die Verwendung unterschiedlicher Datenarten. Heute wächst der Anspruch, verschiedenste, unterschiedlich strukturierte Datenquellen in diversen Datenformaten und Datenerfassungstechnologien für eine Beantwortung von Geschäftsfragen zu integrieren. Relevante Informationen liefern sowohl klar strukturierte relationale Datenbanken als auch unstrukturierte ereignisorientierte Informationsobjekte, beispielsweise für die Fraud Detection, oder öffentlich verfügbare Informationen, wie etwa bei der Risikobewertung von sozialen Netzen und dem Internet.

Dem Unternehmen stellen sich zwei wichtige Fragen: Wie lassen sich anhand von Big Data & Smart Analytics aus der Vielfalt der Datenquellen und Datenarten wirtschaftlich messbare Geschäftsmehrwerte generieren? Und wie löst man das seit Langem bekannte Problem der Auflösung von Informationssilos mittels Kombination und semantischer Verknüpfung von verschiedenen Informationsobjekten?

Auch heute werden Daten teilweise noch durch Menschen erfasst und verwaltet. Jedoch werden vermehrt elektronische Hilfen eingesetzt, die Daten automatisch und fortlaufend sammeln. Smart Watches überwachen beispielsweise permanent ihre Träger, sammeln deren Vitalwerte sowie Aktivitätsdaten und legen jene dann in der Cloud ab. Auch Bordcomputer moderner Autos generieren enorme Datenmengen und integrieren Informationen über den Zustand des Fahrers in die Vorschlags- beziehungsweise Entscheidungsprozesse. Diese Daten können in den Umgebungs-kontext (Vehicle to Infrastructure) durch vernetzte Integration von mobilen Geräten mit Infrastrukturelementen eingebunden werden. Zudem ist es möglich, diese Daten permanent an eine zentrale Stelle zu übermitteln. Die Auswertung dieser automatisch gesammelten Daten erlaubt eine bislang vollkommen neue Art der Individualisierung von Dienstleistungen und Angeboten.

#### Welche neuen Ansätze und Methoden nutzt man und was versteht man unter Data Lakes?

Bei «Data Lakes» handelt es sich um Technologien, die IT-affinen Analysten den Zugang zu riesigen Datenmengen ermöglichen. Damit sind diese Analysten in der Lage, Daten tiefgreifender auszuwerten und vorhandene Datenquellen mit neuen Quellen zu kombinieren, um daraus Antworten auf komplexe Fragestellungen zu gewinnen.

Sinkende Kosten für Speichermedien und cloudbasierte Big-Data-Technologien erlauben es, Rohdaten aller Art und Grösse auf einfache und effiziente Weise zu speichern, zu analysieren und mit Metadaten zu erweitern. Dabei werden die Daten auf einem verteilten Dateisystem abgelegt, um die Skalierbarkeit der Datenmenge und der Zugriffsgeschwindigkeit durch hochgradige Parallelisierung zu ermöglichen. Gleichzeitig stehen Algorithmen für die Analyse der Daten auf den verteilten Systemen zur Verfügung. Moderne Tools können diese Algorithmen erweitern, was

beispielsweise neuartige semantische Textanalyse historischer Dokumente ermöglicht.

Petabytes von Informationen in Data Lakes können in logische Data-Warehouses integriert werden, was eine umfassende Analyse von strukturierten und unstrukturierten Daten ermöglicht. Data Lakes haben den Vorteil, dass ihre Kosten gegenüber einem Data-Warehouse um mehrere Grössenordnungen (Faktor 10–100) geringer sind und sich diese deshalb für unstrukturierte Rohdaten eignen. Dieses Potenzial von Data Lakes wird mit weiteren Schlüsseltechnologien erschlossen, insbesondere Data Virtualization sowie semantische Layer.

### Data Virtualization

Data Virtualization ist eine neue Technologie und Methode zur virtuellen Integration von Daten aus verschiedenen, stark heterogenen internen und externen Datenquellen. Hierbei werden hochgradig parallelisierbare Middleware-Infrastrukturkomponenten eingesetzt, welche die Verbindung zu den Datenquellen herstellen, Abfragen ausführen, die erhaltenen Daten mit zusätzlichen Daten kombinieren sowie weitere Abstraktionen und Transformationen durchführen. Data Virtualization belässt die Originaldaten in ihren Datenquellen und benutzt eine definierte Menge von Sichten sowie Data Services, um die Daten den Endanwendern und Applikationen zur Verfügung zu stellen.

Als Datenquellen können dabei existierende Data-Warehouses, Operational Data Stores sowie direkt transaktionale operative Anwendungen dienen. Allerdings entfaltet der Ansatz von Data Virtualization sein volles Potenzial erst, wenn weitere interne und externe Datenquellen miteinbezogen werden. Dazu zählen Enterprise Applications, Cloud Applications, NoSQL data stores, XML, Flat Files, weitere unstrukturierte Dokumente wie PDFs, E-Mails oder gar Stream-Daten, die auch neartime angebunden werden können. Solange keine Endanwen-

derung eine Kopie anfordert, wird keine dieser Datenquellen physisch kopiert und der Zugriff erfolgt immer rein virtuell.

Während der Implementierung eines Data Virtualization Layers werden hauptsächlich Sichten und Datenzugriffsdienste definiert. Um diese Datenzugriffsdienste für eine bestimmte Informationsdomäne bereitzustellen wird für diese Domäne ein Informationsmodell erstellt und im Data Virtualization Layer abgelegt.

### Semantic Layer

Eine der wesentlichsten Datenquellen eines Data Virtualization Layers ist das eingebundene Master-Data-Management-System (MDM-System). Die Verwendung des MDM, gemeinsam mit den im Data Virtualization Layer hinterlegten Informationsmodellen sowie Datenbereinigungs- und Datenanreicherungsmethoden (Cleansing und Enrichment), stellt die lexikalische und syntaktische Integration von Daten sicher. Für viele zu integrierende Rohdaten müssen jedoch noch aufwendigere Kontextinformationen bereitgestellt werden, welche semantische Modelle liefern können.

Semantische Modelle gestatten es Fachanwendern, Informationen mit Hilfe möglichst verständlicher Sprache abzufragen. Während ein Versicherungsunternehmen zum Beispiel fünf bis zehn verschiedene Administrationssysteme zur Verwaltung von Policen oder Verträgen hat, spielt es für den Fachanwender keine Rolle, wie diese im System abgelegt werden. Die tatsächliche Anfrage eines Anwenders ist zum Beispiel: «Welche Schäden werden durch diese Police abgedeckt?» oder «Welche Informationen haben wir über den Inhaber dieser Police?».

In traditionellen Integrationsmodellen müssen potenziell mögliche Abfragen festgelegt und anschliessend modelliert werden. Auch semantische Modelle müssen erstellt werden. Die Verwendung von modernen Datenbankkonzepten, beispielsweise Graph-Datenbanken, gestattet eine deutlich höhere Flexibilität der Informa-

tionsmodellierung. In einer Graph-Datenbank werden semantische Netze abgebildet, die Beziehungen zwischen Entitäten der realen Welt untereinander ablegen, ohne die stringenten Grundregeln relationaler Modellierung beachten zu müssen. Somit sind in semantischen Modellen die Beziehungen zwischen den Daten Teil der Daten selbst und nicht des Schemas.

Die Beziehungen eines Semantic Layers werden in sogenannten Triplestores abgelegt. Die Bezeichnung Triplestore wird vom Prinzip abgeleitet, dass jede neue semantische Beziehung als Triple, bestehend aus «Subjekt-Prädikat-Objekt», abgebildet werden kann. Beispielsweise können Beziehungen wie «Claim Nr. 2129304 gehört zu Pricing Nr. 98423434» oder «John Doe ist Fallschirmspringer» abgelegt werden.

#### Wie sieht eine Referenz-Architektur für die Berücksichtigung eines Data Lakes aus?

Data Lakes sind keine isolierten Komponenten einer Analyselandschaft, sondern fügen sich als eine Basiskomponente in bestehende Integrations- und Analyselandschaft ein. Es existieren bereits ETL-Prozesse zur Extraktion und Transformation vorhandener strukturierter Daten. Unstrukturierte Daten und kleinststrukturierte Daten, die zum Teil als Stream-Daten eintreffen, können durch einen kostengünstigen und skalierbaren Data Lake erfasst werden. So können beispielsweise Ereignisdaten aus sozialen Netzen von verschiedensten Sensoren und Maschinen relativ kostengünstig und ohne vordefinierte Struktur abgelegt werden.

Eine Destillationsschicht gestattet die Transformation und Überführung in gewohnte Datenstrukturen, wodurch zuvor unstrukturierte Daten für die Weiterverarbeitung in Endanwendungen bereitgestellt werden. Durch Anbindung der Destillationsschicht an das MDM können neu gewonnene Masterdaten effizient und direkt übernommen werden. Für Extraktionstools ist es nicht nur schwierig, Kon-

textinformation zu erkennen und Informationsobjekte Taxonomien zuzuordnen, sondern auch die Lern- und Verbesserungsprozesse der Erkennung kontinuierlich weiterzuentwickeln. Dies gilt insbesondere für relativ spezifische Fachbereiche, welche nicht über entsprechende Taxonomien verfügen.

Gleichzeitig wird es jedoch Data Scientists ermöglicht, direkt im Data Lake Analysen durchzuführen. Zusammengeführt mit Daten im vorhandenen Enterprise-Warehouse können somit neue Beziehungen aufgedeckt und Rückschlüsse gezogen werden. Die Analysedienste können als Service ausgeführt werden, welche durch Endanwendungen transparent genutzt werden. Diese müssen die technologische Speicherbasis der Quelldaten nicht mehr kennen, um Geschäftsfragen beantworten zu können (Abbildung 1).

#### Welche praktischen Erfahrungen sind in der Versicherungsindustrie bekannt?

Insbesondere in der Versicherungsindustrie, wie beispielsweise in der Verwaltung von Versicherungspolicen mit Millionen von Datensätzen aus medizinischen Patientenberichten mit hochauflösenden Bildern aus Radiologie, semistrukturierten medizinischen Berichten als auch komplett unstrukturierten Notizen sowie von unzähligen Statistiken aus dem finanziellen und aktuariellen Bereich, ist ein Einsatz eines Data Lakes sinnvoll. Bei einer traditionellen Strukturierung, die vordefiniert und meistens einen Kompromiss zwischen Vereinheitlichung und Abstriche bei der Datengranularität eingeht, gelangt man bei der Datenmodellierung und kontinuierlichen Datenintegration von neuen Bedürfnissen an technische als auch wirtschaftliche Grenzen. Abhilfe schafft beispielsweise ein auf Hadoop-Technologie aufgesetzter Data Lake, der über verteilte und hochgradig parallelisierte Abfragemechanismen mit anerkannten Open-Software-Standards und über eine massiv parallelsierbare Middleware-Infrastruktur verfügt.

Hadoop erlaubt die verschiedensten Datenformate in ihrer Originalform zu speichern, um diese zu einem späteren Zeitpunkt zu analysieren und zu gliedern. In der traditionellen Data-Warehouse-Datenspeicherung wird hingegen eine «Alles-oder-nichts»-Integration erzwungen. Daten in ihrer ursprünglichen Form abzulegen, bringt auch Vorteile. Zudem lässt sich die Datenherkunft und Wiedergabe-genauigkeit besser gewährleisten, womit unterschiedliche Analysen für verschiedene Kontexte angewendet werden können. Ein Data Lake macht mehrere Datenanalysen für verschiedene Anwendungszwecke möglich, ohne dass diese mehrfach persistiert werden müssen.

## Risikoeinschätzung

Einige grosse Versicherungsunternehmen wählen im Kontext der Big Data & Smart Analytics-Initiativen neue Wege mit Kunden und Datenaggregatoren, um zusätzliche extern verfügbare Daten in guter Qualität zu erhalten. Damit möchten sie nicht nur zusätzliche historische Erkenntnisse

gewinnen, sondern auch zukunftsorientierte Risikomodelle besser mit Realitätsdaten beziehungsweise Szenarien bereitstellen. Zu den bedeutendsten, aber auch schwierigsten Anwendungsfällen zählen Risikoereignisse, potenzielle Katastrophen sowie deren Kettenreaktionen auf die Industrielieferketten. Um etwaige Folgekosten von Betriebsunterbrüchen frühzeitig zu erkennen, arbeiten heutige auf Big Data & Smart Analytics spezialisierte Abteilungen mit Hochdruck daran, solche Szenarien mit einer Vielzahl von verschiedenen Datenquellen im Kontext der Lieferkettenprozesse zu erkennen und darauf rechtzeitig die betroffenen Unternehmen zu informieren und entsprechende Vorsichtsmassnahmen zu treffen.

kleine Anzahl unterschiedlicher technologischer Ansätze sowie eine geringe Vielfalt und Auswahl für die Lösungsansätze zur Verfügung. Auch beginnen viele IT-Abteilungen erst langsam sich mit diesen Technologien auseinanderzusetzen. Dabei fehlen oftmals Know-how und Erfahrungen. In vielen Unternehmen werden zahlreiche Pilotprojekte und Prototypen aus den Big Data & Smart Analytics-Bereichen erstellt, jedoch sind diese oft nicht operativ und produktiv integriert. Ein gutes Zusammenspiel traditioneller Data-Warehouse-Architekturen und neuartiger Technologien, wie Data Lakes, benötigen ein angepasstes Betriebsmodell. Es sollte klar definiert sein, welche strategischen fachlichen Ziele verfolgt werden, wie die neuartigen Technologien in bestehende Systemumgebungen integriert werden und mit welchen fachlichen Fähigkeiten die spezifischen Expertenteams der Big Data & Smart Analytics-Abteilung ausgestattet sein soll.

### Welche Grenzen beziehungsweise Risiken und weiterführenden Chancen sind bekannt?

Nur wenige Anbieter liefern das volle Spektrum der Data-Lakes-Technologien. Daher steht den Unternehmen eine relativ

Eine weitere nicht zu vernachlässigende Aufgabe beim Aufbau und Betrieb eines Data Lakes sind die Sicherstellung von

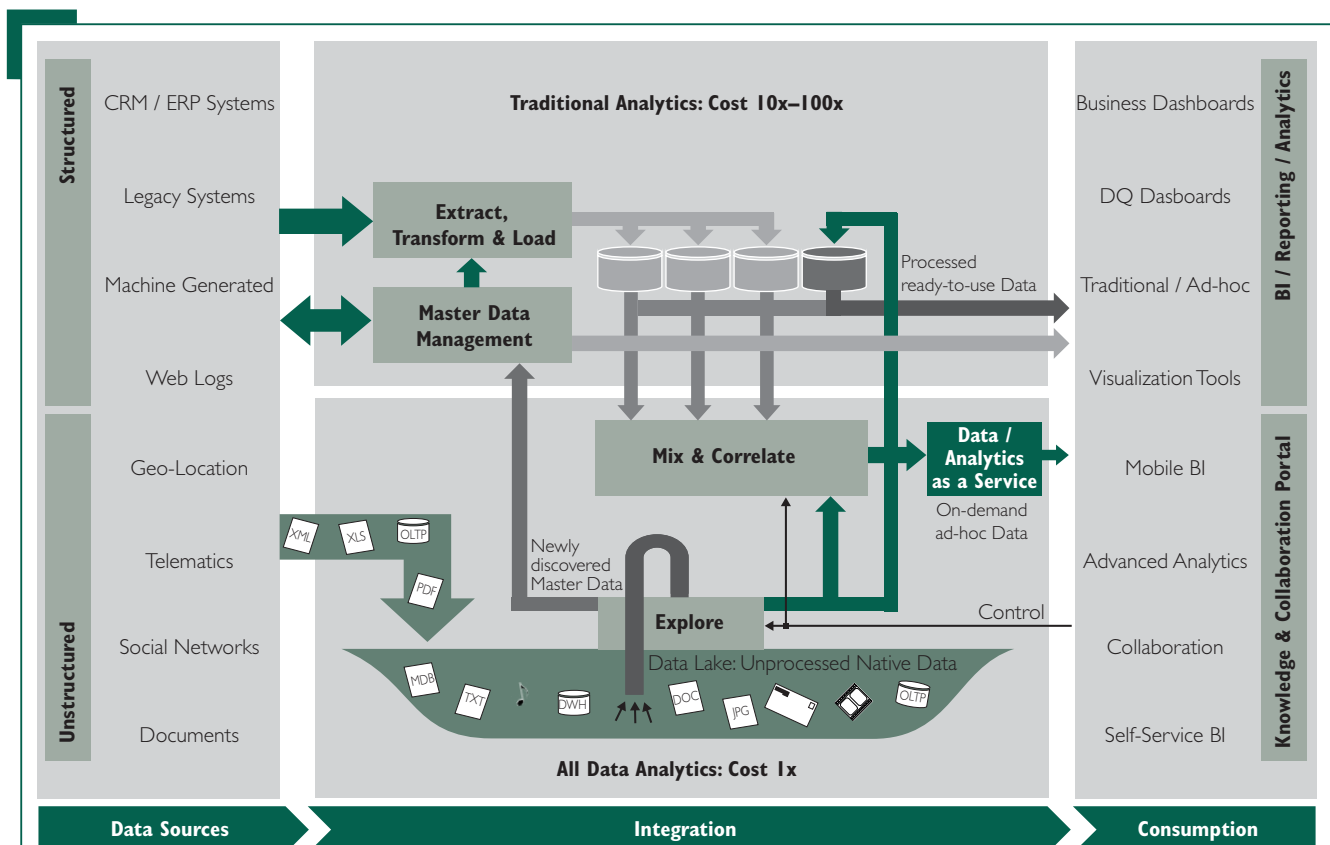


Abb. 1: Data Lakes als Ergänzung einer bestehenden Architekturlandschaft

Datenqualität, der Nachweis der Datenherkunft (Data Lineage) sowie die Sicherheits- und Zugriffskontrolle. In der Grunddefinition stellt ein Data Lake keinerlei Data Governance und MDM-Mechanismen bereit. Ohne einer deskriptiven Metadatenebene und einem Mechanismus, die Metadaten zu verwalten, riskiert ein Unternehmen den Data Lake in einen «Daten Sumpf» zu verwandeln.

### Chancen für «Early Adopters»

Es gibt sicherlich Unternehmen, die sich als «Early Adopters» zusätzliche Chancen und Erkenntnisse durch neue Auswertungen aneignen und sich so auf neue Geschäftsmodelle einstellen. Dies ist insbesondere in der Versicherungsindustrie ein möglicher Ansatz, um sich einen Vorsprung gegenüber der Konkurrenz zu verschaffen. Dabei werden auch praktische Erfahrungen bezüglich Big Data & Smart Analytics-Anwendungsfälle gewonnen. Zudem fordern die Kunden der Versicherungsunternehmen verstärkt Individualisierung und Ausrichtung ihrer Versicherungsprämien an dem tatsächlich angefallenen Bedarf, wie aus zahlreichen Umlageverfahren – so zum Beispiel «pay-as-you-drive» in der Fahrzeugversicherung – bekannt ist. Durch die immer weiter wachsende Vernetzung und Interaktion verschiedener Unternehmen innerhalb der Versicherungsbranche kommt dem Kunden neben zahlreichen neuen «Gadgets» auch ein zunehmender Wettbewerb zugute. Bei diesen Anwendungsfällen werden beispielsweise die Art der Lebensweise einzelner Kunden bei der Prämienberechnung berücksichtigt und positiv beeinflussende Faktoren unmittelbar belohnt.

### Fazit und Ausblick

Die Nutzung aller zur Verfügung stehenden Informationsquellen wird zu einem entscheidenden Erfolgsfaktor für Unternehmen. Für schnell agierende Unternehmen wird es möglich sein, einen Vorsprung durch den zusätzlichen Erkenntnisgewinn zu erzielen und die Vorteile neuer Technologien voll auszuschöpfen. Unternehmen, die Trends nicht voraus-

sehen, sondern diesen folgen beziehungsweise erst spät auf Veränderungen reagieren, können den Vorteil der technologischen Innovation nicht ausnutzen, sondern müssen sich tendenziell aus einer Verteidigungsposition heraus einem immer höheren Implementierungsdruck aussetzen.

Durch die Berücksichtigung von neuen technologischen Basiskomponenten lassen sich traditionelle Data-Warehouse-Architekturen besser auf fachliche Anforderungen der Datenanalyse und der explorativen Erkenntnisgewinnung ausrichten. Gleichzeitig können erhebliche Kostenvorteile durch den Einsatz von Virtualisierungstechniken erzielt werden, womit teure und zeitaufwändige ETL-Systeme durch schlanke logische Data-Warehouses ersetzt werden können.

Das vollständige Potenzial von Data Lakes wird entfaltet, wenn auch integrale Bestandteile einer IT-Architektur wie beispielsweise das MDM mit eingebunden werden. Semantische Layer ermöglichen ein halb- bis vollautomatisches Nachführen von Masterdaten über beinahe unbegrenzt viele Datenquellen hinweg.

Ein vollständiges Konzept umfasst die technologische Speicherbasis (Data Lake), das Zugriffskonzept (Data Virtualization) und ein MDM-Konzept auf Basis von semantischen Modellen. Zu berücksichtigen sind auch Governance-Fragen, Daten- und Zugriffsschutz sowie der Lebenszyklus von Informationen. Unternehmen, welche mit Data Lakes ihre Architekturen und Big Data & Smart Analytics-Initiativen beleben, sollten diese Konzepte und Fragen vorab anwendungsfallspezifisch aufbereiten und berücksichtigen.

### Anmerkung

- 1 Source: BearingPoint Institute Report Issue 004 – «The smart insurer: embedding big data in corporate strategy».

## Glossar

### Hadoop-Technologie

ist ein in Java programmiertes Software-Framework, mit dessen Hilfe Anwender rechenintensive Prozesse mit grossen Datenmengen auf Server-Clustern abarbeiten können. Applikationen können mit Hilfe von Hadoop komplexe Computing-Aufgaben auf tausende von Rechnerknoten verteilen und Datenvolumina im Petabyte-Bereich verarbeiten, so das Versprechen der Entwickler.

### In-Memory-Technologien

basieren auf einem schnellen Hauptspeicherzugriff und ermöglichen, dass operative Applikations-Quelldaten in komprimierter, nicht relationaler Form geladen werden. Neben extrem schnellen Antwortzeiten reduziert oder beseitigt In-Memory-Analytics die Notwendigkeit für die Indizierung und Speicherung von voraggregierten Daten in OLAP-Würfeln oder aggregierten Tabellen. Somit haben Anwendungen vollständigen und hochgradig performanten Zugriff auf operative Daten und Datenanreicherungsregeln können ad hoc definiert und gebildet werden.

### Middleware-Infrastruktur

ist eine zusätzliche Schicht in einer komplexeren Software-Struktur, deren Aufgabe es ist, die Zugriffsmechanismen auf unterhalb angeordnete Schichten zu vereinfachen und die Details deren Infrastruktur nach aussen hin zu verbergen. Dazu stellt die Middleware Funktionen zur Verteilung sowie Dienste zur Unterstützung der Anwendung bereit. Dahingehend ist das Ergebnis einer Middleware die Entlastung der Anwendungsprogramme und ausserdem, durch eine höhere Produktivität, die Optimierung des Entwicklungsprozesses.

### Online Analytical Processing (OLAP)

ist ein effizientes Instrument zur analytischen Bearbeitung von multidimensional aufbereiteten, sehr grossen Datenmengen, die aus Data Marts, aus dem Data-Warehouse (DWH) oder auch aus operativen Systemen stammen können. OLAP, das eng verbunden ist mit Data Warehouse, strukturiert Daten auf hierarchische Weise und erlaubt in der Regel sowohl Einzel- als auch Trendanalysen und ermöglicht es Anwendern, schnell und einfach an entscheidungsrelevante Informationen zu kommen.

### Online Transaction Processing (OLTP)

befasst sich mit der Verarbeitung von operativen transaktionsorientierten Datenbeständen und zeichnet sich dadurch aus, dass es kleinere Datenmengen schnell erfassen und verarbeiten kann. Bei den Onlinetransaktionen werden diese nacheinander in Stapelverarbeitung abgearbeitet, wobei der Client ständig mit dem Server verbunden ist und die Kommunikation aufrechterhält. Sobald der Server die Transaktion abgeschlossen hat, erhält der Client die Ergebnisse.