

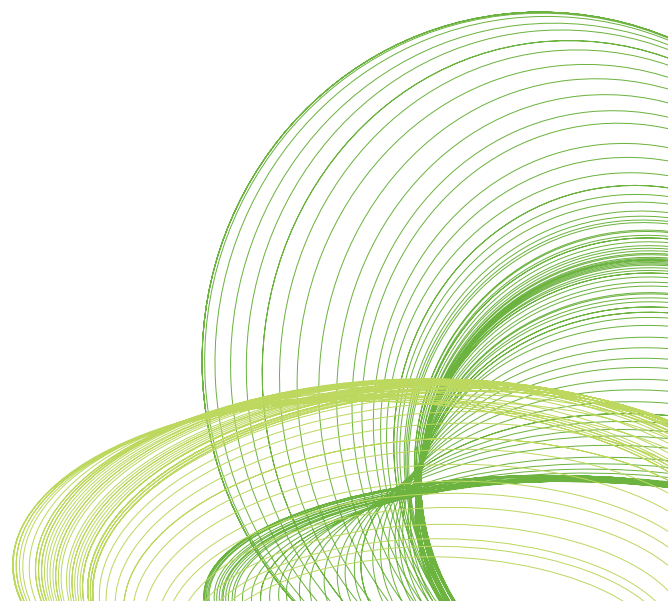


QLIKVIEW UND BIG DATA

Ein QlikView-Technologie-Whitepaper

Juli 2012

qlikview.com



Einleitung

Das Interesse am Thema ‚Big Data‘ ist derzeit immens: Für viele Unternehmen ist die Nutzung großer Datenmengen bereits Realität. Sie können riesige Berge unterschiedlicher Daten speichern und für erfolgskritische Analysen heranziehen. Mitarbeiter sind damit in der Lage, neue Umsatzchancen zu entdecken und bisher nicht lösbare Probleme aus der Welt zu schaffen. Für viele andere Unternehmen ist Big Data ein aktueller IT-Trend, den es nüchtern zu bewerten gilt und dessen Relevanz erst einmal losgelöst vom ganzen Medienwirbel betrachtet werden muss. In diesem Whitepaper wird die Rolle der QlikView-Business-Discovery-Plattform als primäre Analysebasis für eine Big-Data-Lösung beschrieben. Der Inhalt richtet sich nicht nur an IT-Fachleute. Das Whitepaper ist auch für Fachbereichsleiter interessant, die wissen wollen, wie sich eine Big-Data-Implementierung am wirkungsvollsten einsetzen lässt, um ihren Mitarbeitern einen sinnvollen Zugriff und eine einfache Auswertung großer Datenmengen zu ermöglichen.

Warum Big Data so wichtig sind

Allein in den USA gibt es ungefähr neun Millionen Flüge im Jahr. Jeder erzeugt quasi sekundlich Daten zu Hunderten von Parametern: aus dem Flugzeug, der Radarerfassung und anderen Quellen. Zusätzlich sind mit jedem Flug auch unstrukturierte Daten verbunden, beispielsweise die Sicherheitsberichte und Meldungen der Piloten und Co-Piloten.ⁱ Die NASA (National Aeronautics and Space Administration) durchsucht diese Daten mithilfe spezieller Analysetechniken, um Einblicke zu gewinnen, die Störungen auf den Start- und Landebahnen und andere Unfälle verhindern, bevor sie eintreten.

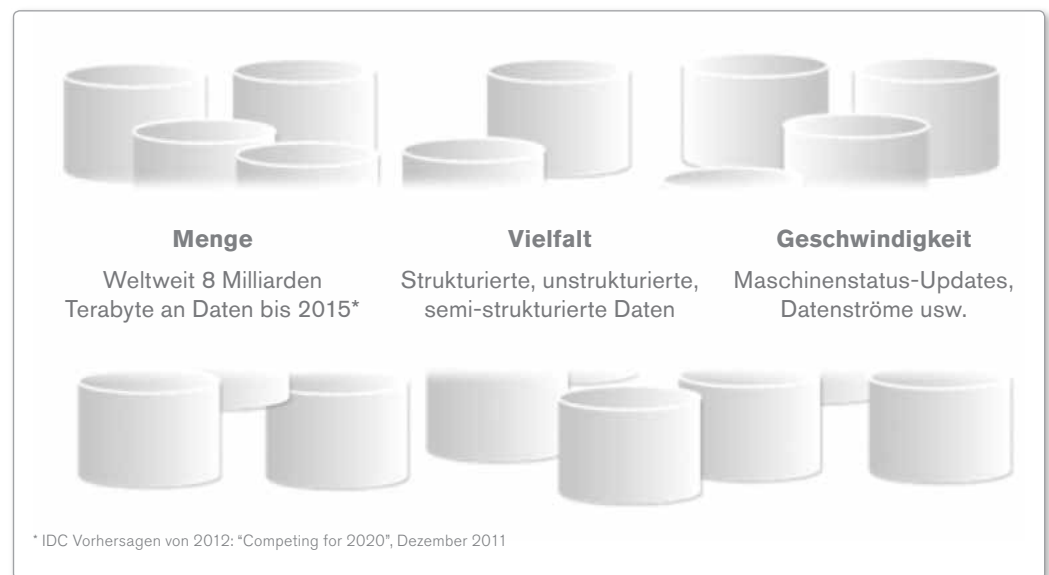
Ein anderes Beispiel ist der E-Commerce-Gigant eBay. Dort werten die Entscheider soziale Daten mit einem BI-Programm aus. Dadurch können sie die Zielgruppen des Unternehmens, Multiplikatoren und Meinungsführer sowie die Wettbewerbsposition besser einschätzen und den Kundenservice entsprechend optimieren. Stand Juni 2012 hat eBay mehr als 40 Millionen Blogs und Foren (60 Milliarden Posts, 10.000 pro Sekunde!) indiziert – insgesamt 65 Terabyte an Daten. Ein globales Social-Analytics-Team arbeitet mit verschiedenen Gruppen unternehmensweit zusammen, um relevante Informationen in diesen Daten zu finden und weiterzuleiten.ⁱⁱ

Für privatwirtschaftliche Unternehmen wie eBay heißt dies: Je mehr Daten im Vergleich zur Konkurrenz verwaltet und analysiert werden können, desto größer sind die Wettbewerbsvorteile. Für öffentliche Organisationen wie die NASA bedeutet dies: Je mehr Daten verarbeitet und ausgewertet werden können, desto genauer lassen sich Vorhersagen treffen. Und dies sind nur zwei Beispiele, warum Big Data so wichtig sind.

Was versteht man unter „Big Data“?

Laut McKinsey Global Institute und auch anderen Quellen bezeichnet man mit dem Begriff „Big Data“ Daten-Sets, die so umfangreich sind, dass sie von herkömmlicher Datenbank-Software nicht mehr in angemessener Zeit erfasst, organisiert und verarbeitet werden können. Abhängig von der Branche, können diese Daten-Sets einige dutzend Terabyte oder bis zu mehreren Petabyte groß sein. Außerdem umfasst der Begriff Big Data nicht nur die schiere Masse an Daten, sondern auch deren Vielfältigkeit (beispielsweise der Datentyp, strukturiert oder unstrukturiert) oder Geschwindigkeit (die Dynamik und das Verhalten von Daten, wenn in einem System neue zu den vorhandenen hinzukommen). Oder einfach ausgedrückt: Wenn man Daten mit Standard-Software nicht mehr in den Griff bekommt, dann handelt es sich um Big Data (siehe Abbildung 1).

Abbildung 1: Was versteht man unter Big Data?



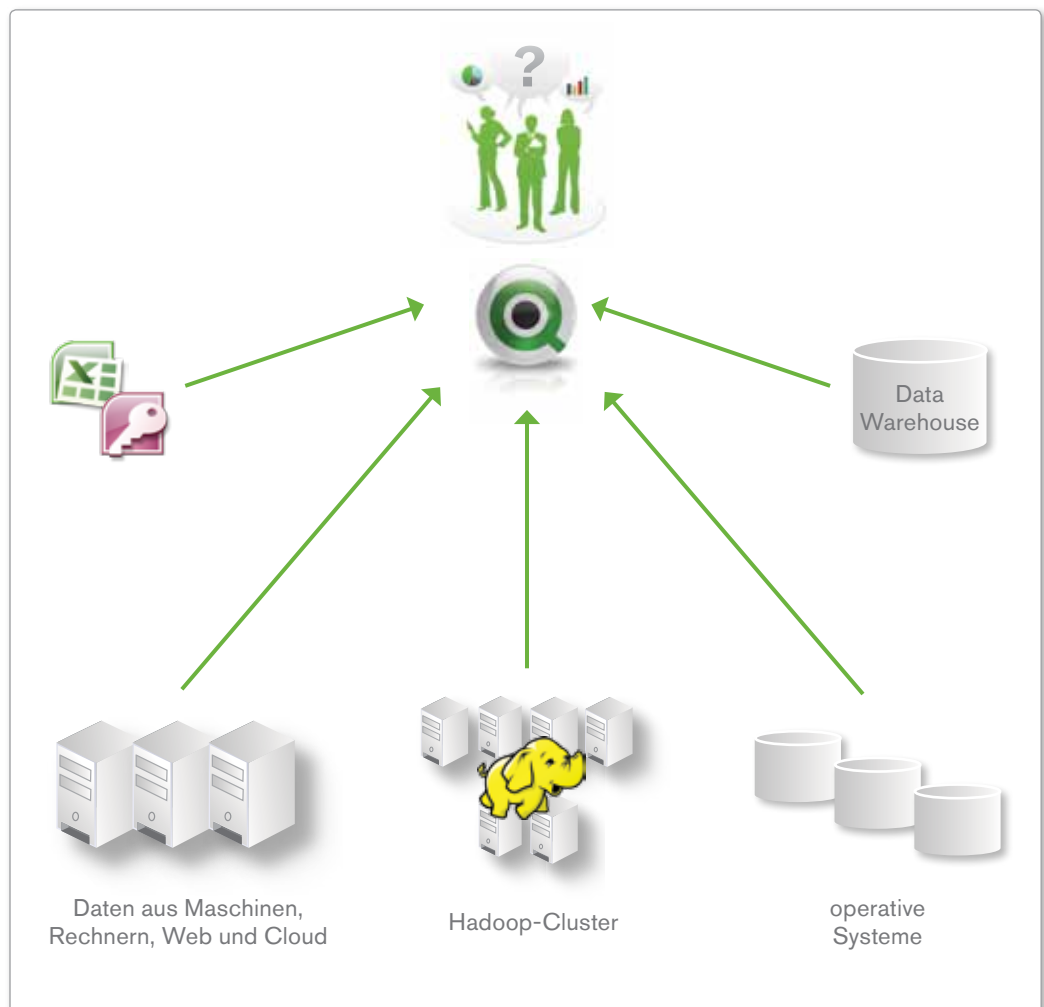
In den letzten Jahren haben wir den Aufstieg von Big-Data-Lösungen wie Hadoop und Google BigQuery erlebt. Solche Systeme nutzen verteilte Rechnernetzwerke oder Massivparallelrechner (Massively Parallel Processing = MPP), um große Datenmengen zu speichern und zugänglich zu machen. In diesem Whitepaper sollen nicht die zugrunde liegenden Technologien der Big-Data-Anbieter besprochen werden. In einem der folgenden Abschnitte finden Sie jedoch einen kurzen technischen Überblick.

Es gehört zu den wichtigsten Merkmalen von Big Data, dass dabei oft unstrukturierte Daten (beispielsweise Web-Content wie Online-Besprechungen, Texte, Inhalte aus Social-Media-Plattformen) gespeichert und verarbeitet werden. Zusätzlich zu strukturierten (jedoch extrem umfangreichen) Maschinendaten wie von Sensoren erfassten Daten (zum Beispiel Stromzähler), automatisierte Computersysteme (wie Rechner-Logs oder algorithmische Aktienhandelssysteme). Damit sich derartige Daten praktisch nutzen lassen, wollen Unternehmen diese mit den vorhandenen strukturierten Daten aus ihren internen OLTP- (Online Transaction

Processing)-Systemen, Data Warehouses und Unternehmensanwendungen wie CRM (Customer Relationship Management) und ERP (Enterprise Resource Planning) verknüpfen. Durch die Verbindung dieser Daten und die Möglichkeit, Muster und Beziehungen zu erkennen, lassen sich Kundenzufriedenheit, Verbraucherverhalten, Produktqualität, Erfolge klinischer Versuchsreihen, Sicherheitsrisiken und vieles mehr auswerten.

Im Rahmen von Big-Data-Installationen spielt QlikView zwei wichtige Rollen. Zum einen verfügen die Mitarbeiter damit über eine schnelle und flexible Analysemöglichkeit. Zum anderen lassen sich Daten aus unterschiedlichen Quellen (beispielsweise aus Big-Data-Systemen, Data Warehouses, Abteilungsdatenbanken und Spreadsheets) in einer interaktiven Analyseebene integrieren (siehe Abbildung 2).

Abbildung 2: Der wahre Nutzen von Big Data erschließt sich erst durch die Kombination mit anderen Daten



Big Data als Ergänzung vorhandener Informationsarchitekturen

In der Regel sollen Big-Data-Systeme die bestehende Dateninfrastruktur eines Unternehmens erweitern und nicht ersetzen. Dafür gibt es folgende Gründe:

- Die zentral und vor Ort eingesetzten Data Warehouses stellen eine bedeutende Investition dar.
- Es gibt Bedenken bezüglich Datenschutz und Datensicherheit und man ist nicht bereit, sensible oder wertvolle Daten in einer Cloud-basierten Umgebung abzulegen.
- Die für den Einsatz und die Verwaltung von Big-Data-Systemen wie Hadoop vorhandenen Software-Tools sind zu komplex. Die aktuelle Werkzeuggeneration zum Management großer Datenmengen ist noch nicht so weit entwickelt wie die Standard-SQL-Werkzeuge für den Datenzugriff.
- Die Abfragen werden nicht dynamisch, sondern im Batch-Modus verarbeitet. Die damit verbundene hohe Latenz (schlechte Antwortzeit) ist für die Analyseanforderungen von Business-Usern nicht akzeptabel.
- In der IT fehlen die Kenntnisse für Big-Data-Management-Systeme wie Hadoop, NoSQL und Google BigQuery.
- In vielen Fällen wäre der Einsatz von Big-Data-Systemen wie Hadoop, NoSQL und Google BigQuery schlicht überdimensioniert. Für diese Herausforderungen reichen gewöhnliche relationale Datenbanksysteme voll aus.
- Die Technologien für das Management großer Datenmengen sind neu und viele sind Open Source und/oder erst im Beta-Stadium.

Die Rolle der Analysen bei Big Data

Für sich genommen, bringt eine Big-Data-Lösung einer Organisation keine Vorteile. Sie wird erst dann interessant, wenn die Daten aus diesem System in die internen Entscheidungsprozesse einfließen können. Über die Technologien, die das Speichern und den Zugriff auf große Mengen und verschiedene Arten von Daten unterstützen, wird viel geschrieben und gesprochen. Dabei liegt der praktische Nutzen für ein Unternehmen nicht in den verteilten Rechenleistungen, sondern in der Datenanalyse. Hier entsteht der wahre Mehrwert. Das gilt natürlich für Daten jeder Art und Größe, aber für Big Data ist dies besonders relevant. Wie Wayne Eckerson von TechTarget in seiner Studie ‚Big Data Analytics: Profiling the Use of Analytical Platforms in User Organizations‘ betont: „Ein wichtiges Merkmal ‚großer‘ Daten ist, dass diese mehr Muster und interessante Anomalien enthalten als ‚kleine‘ Daten. Daher ist es für Organisationen lohnender, große Datenmengen auszuwerten als kleine.“

Daten aufbereiten für die Analyse: Relevanz ist Trumpf

Bei Big-Data-Lösungen sollte man eines nicht vergessen: Es ist unvermeidbar, dass diese - quasi per definitionem - größtenteils wertlose Daten enthalten. Diese „Durchgangsdaten“ werden einfach erfasst, weil man sie eben erfassen kann. In den meisten Fällen macht es daher wenig Sinn, eine Analyseebene bereitzustellen, um alle Daten aufzunehmen. Meist ist es viel nützlicher, einen Aggregationsmechanismus einzurichten und die relevantesten und interessantesten Daten aus der Big-Data-Quelle zu extrahieren und für die Analyse vorzubereiten.ⁱⁱⁱ

Die Big-Data-Lösung sollte als Teil der Informations-Management-Architektur betrachtet werden und parallel zu vorhandenen Data Warehouses, Rechnern für Complex Event Processing (CEP), eigenständigen Analyseumgebungen („Analytical Sandboxes“), OLTP-Systemen usw. arbeiten. „Heute verwenden einige Unternehmen Hadoop als Aufbereitungsumgebung für unstrukturierte und semi-strukturierte Daten bevor sie diese in ein Data Warehouse laden“, führt Eckerson bereits in seiner Studie aus. Das Big-Data-System hält alle Detaildaten, während locker zusammengefasste Daten-Sets in der Analyseebene bereitgestellt werden.

In diesem Framework fungiert QlikView entweder als direkter Empfänger der Daten aus Hadoop oder anderen Big-Data-Systemen oder setzt auf dem Data Warehouse auf – oder beides. In jedem Fall bietet die Einfachheit, mit der QlikView Daten aus unterschiedlichen Quellen für eine assoziative Analyse zusammenstellen kann, einen deutlich spürbaren Vorteil.

QlikView und Big Data

Die Mitarbeiter der Fachbereiche stehen vor einer ständigen Herausforderung: Sie sollen Daten effizient zusammenstellen, filtern, analysieren und daraus neue Erkenntnisse gewinnen. Analyselösungen, die Spezialkenntnisse erfordern, sind dabei nicht gefragt. Diese Anwender benötigen einfache, bessere Möglichkeiten, um im riesigen Datenschwungel die für sie relevanten Informationen zu finden. Sie brauchen konkrete Antworten auf ihre Fragen, damit sie schnell und sicher geschäftliche Entscheidungen treffen können.

Es werden immer mehr MPP-Lösungen (Massenparallelrechner) zur Bewältigung der stetig wachsenden – strukturierten und unstrukturierten - Datenmengen eingeführt. Dadurch steigt auch die Nachfrage an Analysewerkzeugen, mit denen die Mitarbeiter in den Fachbereichen Big Data zur Gewinnung von neuen Erkenntnissen nutzen können.

QlikView antwortet auf diese Herausforderung mit einem zweigleisigen Ansatz.

Erstens dreht sich bei QlikView schon immer alles um die Anforderungen der Anwender und was diese von ihren Analysen erwarten. Es geht nicht um das sture Durchsetzen einer möglicherweise ungeeigneten Lösung. Die Verfügbarkeit der zu ihrem Anwendungsfall passenden Daten ist für die User viel nützlicher als die permanente Bereitstellung aller

Daten. Beispielsweise interessieren den Niederlassungsleiter einer Bank nicht die Zahlen des deutschlandweiten Filialnetzes der Bank. Er möchte sich lediglich einen Überblick über die Vertriebsaktivitäten, die Kunden und die Marktdynamik in seinem Einzugsbereich verschaffen. Berücksichtigt man diese Fakten, ändert sich der Fokus der ganzen Big-Data-Diskussion. Statt über der Menge zu sprechen, sollten vielmehr die Relevanz und der Nutzen der Daten im Mittelpunkt stehen.

Der Personenkreis, der in einem Unternehmen große Datenmengen analysieren muss, ist eher klein. Eine Privatkundenbank hat beispielsweise Tausende von Niederlassungen, aber es gibt nur hundert Business Analysts mit einer zentralen, unternehmensweiten Funktion. Während die Spezialisten Zugriff auf viel mehr Daten benötigen, reicht dem einzelnen Filialleiter ein Datenausschnitt mit seinen Zahlen völlig.

Mit QlikView lassen sich beide Anwendungsanforderungen umsetzen. Die Mitarbeiter können sich auf die Daten konzentrieren, die besonders relevant und nützlich für ihre Aufgabe und ihren Bereich sind. Durch Bereitstellung des passenden – großen oder kleinen – Datensegments fungiert die QlikView-Analyseapplikation als Downstream-Plattform für die Datenquellen und ermöglicht Business Analysts und weniger technisch versierten Mitarbeitern, den Einblick in die für sie interessantesten Daten.

Zweitens ist die sinnvolle Nutzung großer Datenmengen schon immer ein Schwerpunkt von QlikView gewesen. Big Data ist eine neue Herausforderung, auf die QlikView gut vorbereitet ist. Wir sorgen dafür, dass QlikView-Anwendungen stets die von allen Mitarbeitern im Unternehmen benötigten Datenmengen verarbeiten können und Applikationen nichts an Relevanz einbüßen – und zwar so:

- Es liegt im Trend, auch auf Standard-Hardware immer leistungsfähigere Intel-Prozessoren und Arbeitsspeicher einzubauen. Dadurch kann QlikView noch mehr Daten im Memory bearbeiten (und den Anwendern blitzschnelle, interaktive Auswertungen ermöglichen).
- Wenn sehr große Datenmengen als Basis herangezogen werden sollen, haben QlikView-Best-Practices gezeigt, dass eine Architektur-basierte Implementierung am meisten Sinn macht. Der richtige Einsatz von verteilten Servern in einer Cluster-Umgebung, der Aufbau zielgruppengerechter Anwendungen, die Verwendung leistungsfähiger Reload-Engines zum Einspielen der Daten und – wo nötig - die Verkettung von Dokumenten tragen dazu bei, dass sich aggregierte Ansichten effizient mit Detailansichten verknüpfen lassen und dass das Potential der Hardware-Ressourcen optimal ausgeschöpft werden kann.
- Über eine Reihe von APIs (Application Programming Interfaces) können Entwickler das von QlikView bereitgestellte offene Datenprotokoll (QVX oder QlikView-Datenaustausch) als Schnittstelle zu den APIs von Hadoop- und anderen Anbietern von Big-Data-Systemen nutzen. Das QVX-Protokoll von QlikView lässt sich über zwei verschiedene Methoden mit Hadoop-Systemen verbinden:
 - Festplatten-basierte QVX-Dateiextrakte aus Hadoop (Push-Prinzip)
 - dem „Named Pipe“-QVX-Konnektor für Hadoop (Pull-Prinzip)

- Der Zugriff auf Big-Data-Quellen wie Attvio, DataRoket und Informatica erfolgt über die Produkte von QlikView-Partnern. Es ist ein spezielles QVX-SDK erhältlich, mit dem die Entwickler von Drittanbietern passende Konnektoren für jedes System mit einer offenen API erstellen können.
- Durch die Partnerschaft mit Google verfügt QlikView über ein grafisches Front-End für die Google-Lösung BigQuery.
- Im Juni 2012 hat QlikTech das Unternehmen Expressor Software übernommen. Der neue QlikView Expressor Server bietet Funktionen für Metadateninformationen und anspruchsvolle Datenintegration.^{iv}

Technische Überlegungen zum Einsatz von Hadoop als Analysebasis

Bei Hadoop handelt es sich um ein Open-Source-Projekt, bei dem es hauptsächlich um eine massive rechnerische Skalierbarkeit geht. Diese Funktionalität geht jedoch meist auf Kosten der Bedienbarkeit und Performance. Man sollte immer im Hinterkopf behalten, dass Hadoop keine Datenbank ist, sondern ein Framework zur verteilten Berechnung einer sehr großen Datenanwendung. Daher gibt es keinen nativen Mechanismus, der die direkte Abfrage von Daten vorsieht. Für jede Query muss ein „Einmal“-Programm geschrieben werden, mit dem MapReduce, ein Framework für die Verarbeitung großer Datensets, angesprochen wird.

Aus diesem Grund haben andere Open-Source-Projekte das Hadoop-Kernprogramm um Module ergänzt, die die Nachteile der extremen Skalierbarkeit ausgleichen sollen. Das bekannteste davon ist wahrscheinlich Hive, ein Data Warehouse, das auf Hadoop Distributed File System (HDFS) und MapReduce aufsetzt. Durch HQL (Hibernate Query Language), einer SQL-ähnlichen Abfragesprache, wird der Zugriff auf Hadoop-Daten erleichtert. Hive ermöglicht eine einfache Verbindung zu Hadoop-Daten über Konnektoren für ODBC (Open Database Connectivity) und JDBC (Java Database Connectivity).

Dadurch wird die Bedienung und Integration zwar einen erheblichen Schritt weiter gebracht, aber in puncto Datenanalyse ist Hive nach wie vor eine unvollständige Lösung, besonders für Anwender ohne technische Vorkenntnisse. Hive stellt jedes Datenanalysesystem ohne In-Memory-Verarbeitung vor das gleiche Problem: Diese Lösungen sind nicht dafür geeignet, OLTP-Arbeitslasten zu übernehmen und bieten keine Echtzeit-Abfragen oder Updates auf Spaltenebene. Sie können Business-Usern keine assoziativen Auswertungen bieten.^v Außerdem sind Unternehmen damit nicht in der Lage, Big Data mit den vorhandenen Unternehmens- oder Cloud-Daten zu verknüpfen. Hive eignet sich am besten für die Batch-Verarbeitung großer Daten-Sets, an die neue Daten nur angehängt werden.

Die Datenintegration mit Hive ist ein anderes, potentiell schwieriges Thema. Hive wurde als Data Warehouse für Hadoop entwickelt. Daher werden Daten aus anderen Quellen nicht akzeptiert. Zum Zeitpunkt der Erstellung dieses Dokuments hatte gerade ein Projekt namens SQOOP die erste Entwicklungsphase abgeschlossen. Es soll eine Möglichkeit bieten,

Daten aus relationalen Datenbanken in Hive zu integrieren. Allerdings steckt SQOOP noch in den Kinderschuhen: Die Funktionalität ist begrenzt und die Stabilität muss sich erst noch beweisen. Die Möglichkeiten von QlikView, eine native Verbindung zu unterschiedlichen Datenquellen herzustellen und logische Verbindungen zu definieren, ist sicher eine bessere Lösung zur Datenintegration von Hadoop-basierten Big-Data-Systemen. Nicht zu vergessen, dass die Anwender nur mit QlikView Daten assoziativ auswerten können.

QlikView und Hadoop: Ein Erfahrungsbericht

King.com ist ein Online-Spiele-Anbieter mit Hauptsitz in Skandinavien. QlikView wird mit einem Hadoop-basierten Big-Data-System eingesetzt und ermöglicht Anwendern aus dem King.com-Marketing einen schnellen Einblick in das Verhalten der Kunden. Die Daten werden über die Nutzung der Spiele erfasst. Jede Aktion des Kunden, vom Browsen durch die Site bis hin zu den Eingaben in jedem Spiel und noch viele weitere Kennzahlen werden in einem vor Ort installierten Hadoop-basierten System erfasst. QlikView setzt auf der Hadoop-Lösung auf und bietet King.com die Business-Discovery-Funktionen, mit denen sich neue Kunden, Spiele, Angebote und mehr gezielt adressieren lassen.

Die Menge der gesammelten Daten ist beeindruckend: Jeden Tag werden 1,6 Milliarden neue Datenzeilen erzeugt und im Hadoop-System gespeichert. King.com setzt QlikView für die Analyse aggregierter Daten aus Hadoop ein. Bei Milliarden von Permutationen spielt der Aggregationseffekt nach Aussage der Verantwortlichen statistisch kaum noch eine Rolle. Wenn etwa 211 Millionen Zeilen in QlikView ausgewertet werden, kann man sicher sein, dass diese Zahlen repräsentativ für das gesamte, in Hadoop gehaltene Daten-Set sind.

King.com nutzt ein Cluster mit acht Servern für seine Hadoop-Umgebung. Jedes Benutzerereignis wird protokolliert, verarbeitet und am Ende über einen Hive-ODBC-Konnektor den QlikView-Anwendern zur Analyse bereitgestellt. Der Datenfluss von den Ursprungsspielsystemen bis hin zur QlikView-Analysenwendung ist in Abbildung 3 dargestellt.

Abbildung 3: Datenfluss in der King.com-Implementierung von QlikView und Big Data

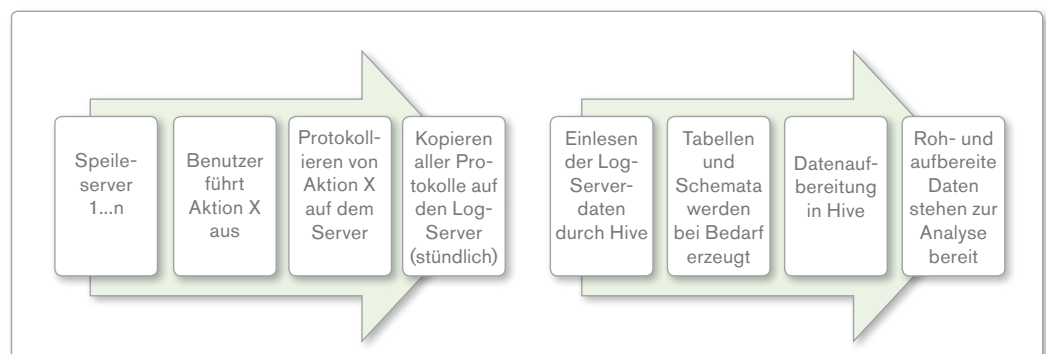
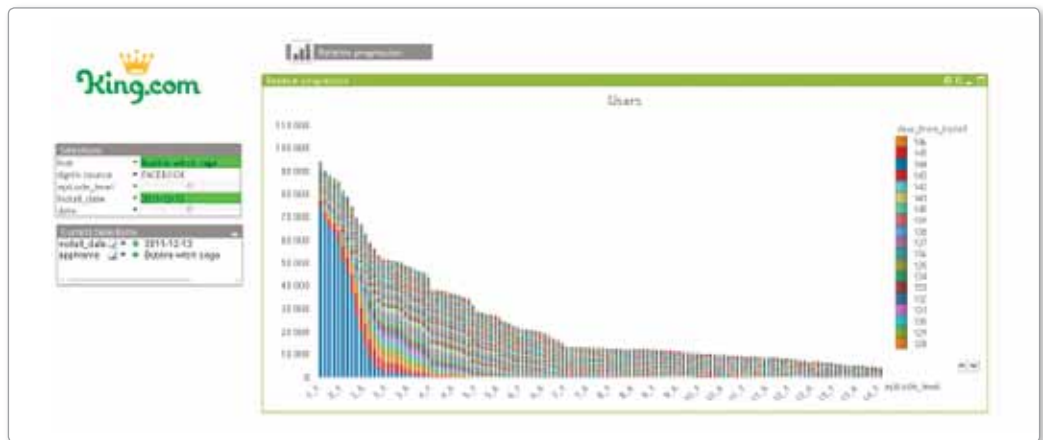


Abbildung 4: Screenshot der QlikView-Anwendung von King.com



QlikView und Google BigQuery

Google BigQuery ist ein Web-Service, mit dem Business User und Entwickler sehr große Daten-Sets mit Milliarden von Zeilen interaktiv analysieren können, ohne dass dafür vorab Hardware- oder Software-Investitionen anfallen. BigQuery ist skalierbar und einfach zu bedienen: Entwickler und Unternehmen können on-Demand leistungsfähige Datenanalysen nutzen.

Die Erweiterungen und benutzerdefinierbaren Konnektoren der QlikView-Business-Discovery-Plattform ermöglichen eine nahtlose Integration in Google BigQuery. Mit ihrem QlikView-Konnektor können sich Anwender BigQuery-Daten in den Arbeitsspeicher laden und die Informationen flexibel und ohne Einschränkung durch vorgegebener Analysepfade untersuchen. Für tiefere Einblicke lassen sich die BigQuery-Daten spontan mixen und im Handumdrehen in neuen Ansichten und Visualisierungen zusammenstellen. Mit der assoziativen Datenanalyse in QlikView können die Anwender BigQuery-Daten ganz nach Belieben durchsuchen und bearbeiten.

Zusätzlich zum benutzerdefinierbaren Konnektor erlaubt die QlikView-Objekterweiterung eine direkte Verbindung von QlikView-Dashboards zu Google BigQuery. Die Mitarbeiter der Fachbereiche können ad-hoc Fragen stellen und erhalten auch bei großen Datenmengen blitzschnell Antworten. Für BigQuery bedeutet das: Die Anwender können auch Daten, die sich nicht im Memory befinden, spontan auswerten. Die Ergebnisse liegen in wenigen Sekunden vor, ohne dass eine einzige Zeile SQL geschrieben werden muss.

Dank der Integration von QlikView und Google BigQuery in der Lage, auch Benutzer ohne technische Vorkenntnisse oder SQL-Know-how Milliarden von Datenzeilen nach eigenen Kriterien auszuwerten und schnell und effizient die für sie relevanten Informationen zu finden.

Abbildung 5: Demo-Anwendung QlikView und Google BigQuery
 (<http://Qlikview.com/bigquery>)



Big Data: QlikView schließt die Analyselücke

Eine der größten Herausforderungen im Bereich Telekommunikation ist die „letzte Meile“. Beim Anschluss von Telefon, Kabel oder Internet ist damit der letzte Abschnitt der Leitung gemeint, die zum Teilnehmerhaushalt führt. Für den Service-Anbieter ist die Ausbreitung des Netzwerks vom Hauptzweig oder Backbone eine teure Angelegenheit. Bautrupps müssen ausrücken, Gräben ausgehoben und Leitungen verlegt werden. Folglich werden die hohen Installationskosten in einigen Fällen an den Endkunden weitergegeben – oder die letzte Meile wird überhaupt nicht geschlossen. Bei Big Data gibt es ebenfalls ein „Letzte-Meile-Problem“.

Die meisten Anbieter, die sich mit dem Thema Big Data beschäftigen, konzentrieren sich auf die Verarbeitung der Daten. Um beim Beispiel Telekommunikation zu bleiben: Deren Fokus ist der Backbone. Was ist mit der letzten Meile? Hier kommt QlikView ins Spiel. Das erklärte Ziel von QlikView lautet: „Simplifying decisions for everyone, everywhere.“ Es liegt gewissermaßen in der Natur der QlikView-Business-Discovery-Plattform, dass das Erlebnis und die Anforderungen der Anwender immer an erster Stelle stehen. Das QlikView-Geschäftsmodell unterstützt also aktiv die letzte Meile, sprich den „Anschluss“ der Business-User an das „Datennetzwerk“. QlikView ist die ideale Ergänzung zu den Funktionen, die von den Anbietern zur Verarbeitung von Big Data entwickelt werden, weil es Anwendern die wirklich wichtige und relevante Komponente von Big Data ermöglicht, sprich Datenanalysen und Erkenntnisse für jedermann.

Quellenangaben

- ⁱ Laut der Behörde für Forschung und innovative Technologie im Amt für Beförderungsstatistik (Bureau of Transportation Statistics Research and Innovative Technology Administration) gab es im Jahr 2012 (endend am letzten Tag im Februar) 9.098.000 Abflüge im Gegensatz zu 9.125.000 Abflügen im Vergleichszeitrum in 2011. Dies ist eine Veränderung von -0,3 Prozent. Mehr Informationen auf <http://www.transtats.bts.gov/>.
- ⁱⁱ Am 13. Juni 2012 waren diese Informationen Inhalt einer Präsentation mit dem Titel „Global Social Analytics @eBay“, die Palm Norchoovech, Social Commerce Analyst bei eBay, auf der Veranstaltung Text Analytics Summit in Boston, Massachusetts zeigte. Mehr auf: <http://bit.ly/GSnH03>.
- ⁱⁱⁱ Zum Erfahrungsbericht: Der QlikView-Kunde King.com lädt 1,6 Milliarden Datensätze täglich in seine Big-Data-Lösung, extrahiert aber lediglich 211 Millionen aggregierter Datensätze für die Analyse. King.com verwendet Clustering und das Stichprobenverfahren, um ein statistisch relevantes Subset der Daten zusammenzustellen und nutzt diesen Datenausschnitt für die Analyse.
- ^{iv} Mehr Informationen zu QlikView Expressor finden Sie auf dieser Webseite: <http://www.qlikview.com/us/explore/products/expressor>.
- ^v Die Arbeitsweise von QlikView entspricht der Arbeitsweise seiner Anwender. Die assoziative Datenanalyse in QlikView ermöglicht auch die Beantwortung von nicht gestellten Fragen. Die Selektionen der Nutzer werden grün und die mit seiner Auswahl verbundenen Daten werden weiß hervorgehoben. Daten ohne Relation sind grau hinterlegt. Bei der Anzahl der Selektionen gibt es keine Beschränkungen. Alle Daten in der Anwendung, die aus unterschiedlichen Quellsystemen kommen können, werden sofort gemäß der getroffenen Auswahl gefiltert. Mehr Informationen zur assoziativen Datenanalyse finden Sie im QlikView-Whitepaper „The Associative Experience“ (<http://bit.ly/hgf12U>).