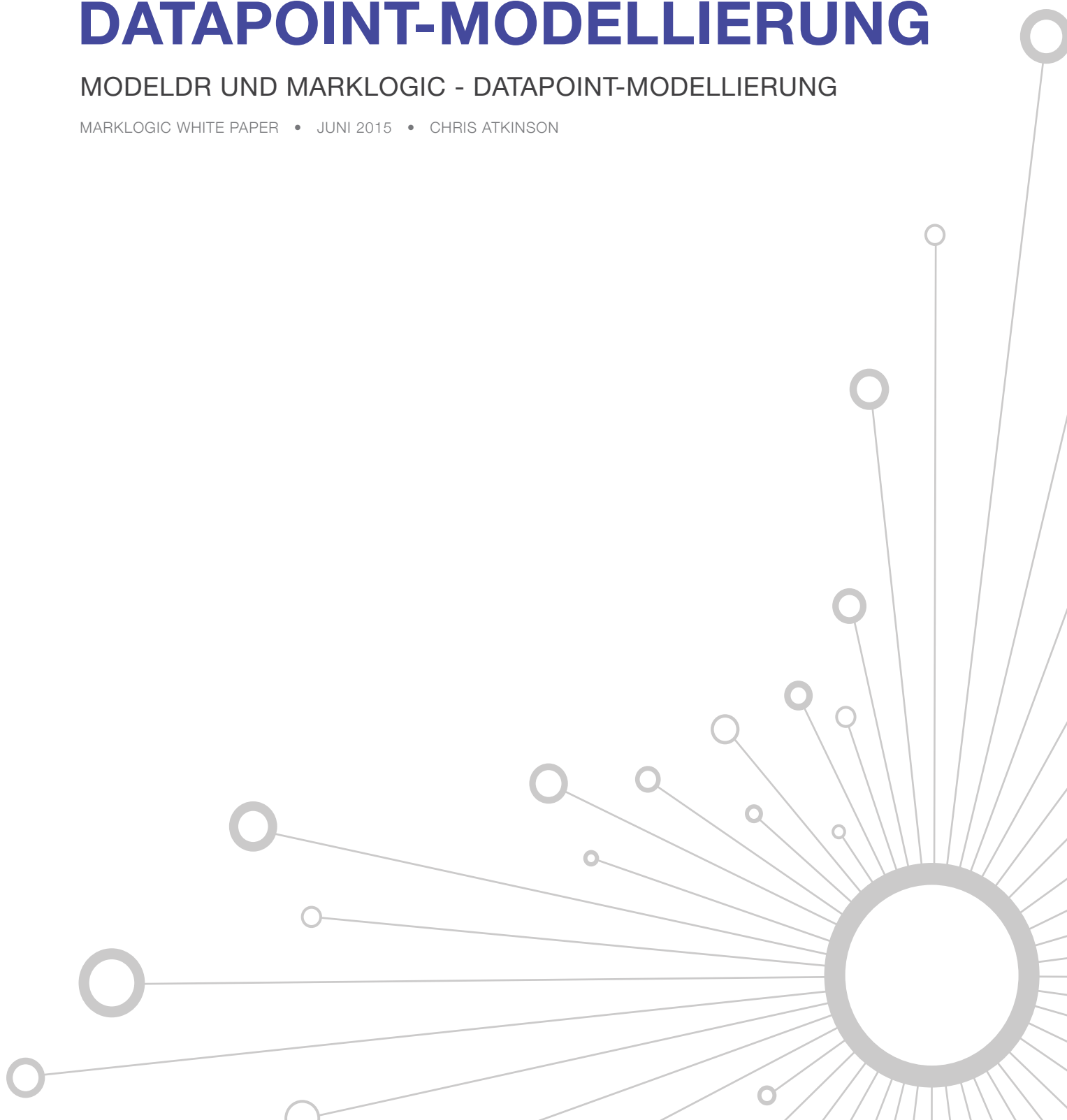


SCHNELLES UND FLEXIBLES REPORTING DURCH DATAPOINT-MODELLIERUNG

MODELDR UND MARKLOGIC - DATAPOINT-MODELLIERUNG

MARKLOGIC WHITE PAPER • JUNI 2015 • CHRIS ATKINSON



Inhalt

Die Einhaltung Regulatorischer Vorgaben wird immer komplexer	3
Einführung in die Data-Point-Modellierung	3
Die Marklogic Datenbank-Plattform	5
Die Architektur der ModelDR Lösung	5
Die Integration von Marklogic in ModelDR	5
Die Vorteile von Marklogic und ModelDR	6
Abschaffung von Datensilos	
Höhere Datenqualität	
Schnellere Informationsbereitstellung	
Bessere Auffindbarkeit von Informationen	
Schnellere Rechnungslegung Dank Bitemporaler Funktionen	
Schnelle Erweiterung des Wissens	
Erkenntnisse über verschiedene Quellen hinweg	
Zusammenfassung	7

DIE EINHALTUNG REGULATORISCHER VORGABEN WIRD IMMER KOMPLEXER

Die zunehmende Regulierung der Finanzbranche führt dazu, dass Banken oft nur noch ihre Kapitalkosten decken können. Viele der führenden Finanzinstitute sind mit den strengeren Compliance-Anforderungen in Bezug auf Datenaggregation, Governance, Architektur und Prozesse überfordert.

Durch die Regulierung sollen das Risikomanagement, das Reporting und die Entscheidungsfindung verbessert werden, doch viele Banken tun sich schwer mit einer effektiven Datenarchitektur und -aggregation zur Einhaltung von Data Governance und Qualitätssicherung.

Aufgrund der häufigen Änderungen der Reporting-Anforderungen und die damit verbundenen Änderungen der Datenmodelle, ist die Bereitstellung einer harmonisierten, standardisierten Sicht auf Daten zu bestimmten Zeitpunkten (häufig als Kongruenz bezeichnet) keine einfache Aufgabe. Viele Banken versuchen, diese Herausforderung durch die Implementierung von ETL-Prozessen (Extraktion, Transformation, Laden) zwischen Silos für spezifische Geschäftsfunktionen zu bewältigen. Dies führt jedoch zu einer Beeinträchtigung der Datenqualität und einer Erhöhung des regulatorischen Risikos. Darüber hinaus steigen aufgrund der teuren ETL-Prozesse auch die Kosten für die Kontenabstimmung sowie die Gesamtbetriebskosten deutlich an.

Um die Abstimmung über verschiedene Silos hinweg zu ermöglichen, werden häufig Teilinformationen in relationalen Data Warehouses gespeichert, wo sie von Benutzern über Abfragen anhand eines starren Schemas ausgelesen werden.

Änderungen an diesen Schemata erfordern den Einsatz umfangreicher betrieblicher Ressourcen: Änderungen müssen getestet, Abfragen validiert und Prozesse für die Änderungskontrolle berücksichtigt werden. Es kann Stunden, Tage oder sogar Monate dauern, die Daten erneut zu verarbeiten, bevor sie den Benutzern bereitgestellt werden können. Darüber hinaus erhalten Benutzer häufig nicht den nötigen Einblick, da sie nicht über die Mittel zur Visualisierung komplexer Datenumgebungen verfügen.

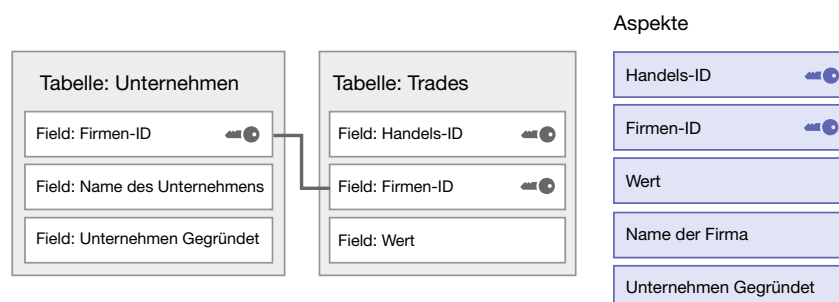
Datenkongruenz wird mit der Zeit immer schwieriger, da die Prozesse für Datentransparenz fest im ETL-Prozess kodiert sind. Durch die kontinuierliche Veränderung der Daten wird die Komplexität weiter erhöht.

EINFÜHRUNG IN DIE DATA- POINT-MODELLIERUNG

Die Data-Point-Modellierung bietet klare Vorteile gegenüber der herkömmlichen relationalen Datenmodellierung. Der auf XBRL (extensible Business Reporting Language) basierte Ansatz entkoppelt das Datenmodell vom Ausgabeformat und bietet damit nicht nur Zugriff auf spezifische Daten, sondern auch Flexibilität. Des Weiteren ermöglicht er die Umwandlung relationaler Datenstrukturen in ein nicht relationales Data-Point-Modell (DPM). Diese Methode umfasst die Aufschlüsselung der Tabellenfelder in eine Reihe von Aspekten (Facetten). Dabei werden die Beziehungen zwischen den einzelnen Aspekten aufgezeichnet, um die Referenz und Integrität zu gewährleisten (Abbildung 1).

Darüber hinaus werden zusätzliche Informationen zu den Aspekten erfasst, um die Kongruenz der Daten aus verschiedenen Quellen zu gewährleisten. Das folgende Beispiel veranschaulicht das Konzept der

Abbildung 1



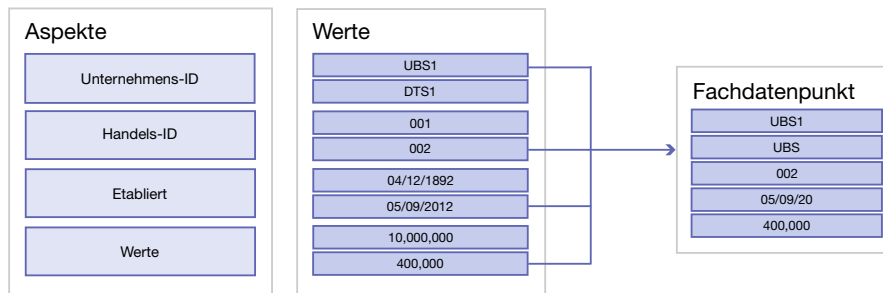


Abbildung 2

Kongruenz: An verschiedenen Orten aufgezeichnete Daten zu denselben Ereignissen können in unterschiedlichen Formen und Datenmodellen gespeichert sein. Um eine kongruente Sicht der Daten zu ermöglichen, werden die Zieldaten an einem zentralen Ort zusammengeführt. Dabei werden die unterschiedlichen Datenmodelle und Formen mithilfe eines ETL-Prozesses in ein einheitliches Datenmodell überführt.

Kongruenz ist erreicht, wenn die Daten aus allen Quellsystemen harmonisiert sind. Das heißt, die Sicht auf die Daten ist vollständig, und es fehlen keine Informationen.

Der Daten Punkt Modellierungs-Ansatz (DPM-Ansatz) bietet die folgenden wichtigen Vorteile:

- Schnelle Kongruenz über ungleichartige Datensätze hinweg ohne komplexe und risikobehaftete Transformationsprozesse
- Bewahrung der Referenz-Integrität aller über das Datenpunktmodell konvertierten Informationen
- Import einer Daten-Taxonomie zur Vermengung oder Verbindung zusätzlicher Daten, die in das Modell importiert wurden
- Schneller Zugriff auf Daten anhand der bevorzugten Daten-Taxonomie anstelle von Abfragen auf Grundlage eines starren Schemas

Trotz der Flexibilität des Datapoint-Modells gab es bisher jedoch auch Nachteile: Daten in nicht relationaler Form lassen sich am besten als RDF-Tripel (Resource Description Framework) darstellen. Dies führte zu verschiedenen Herausforderungen: Bei der Generierung der RDF-Tripel aus den Quelldaten außerhalb der Datenbank wurden die Quelldaten von den abgeleiteten Informationen in den RDF-Tripeln getrennt. Nachträgliche Änderungen an diesen Informationen erforderten daher eine erneute Verarbeitung der Quellen. Gefragt ist also

ein Verfahren, das Quelldaten verarbeiten und RDF-Tripel ohne erneute Verarbeitung nutzen kann. Bei geänderten Reporting-Anforderungen könnten dann einfach neue RDF-Tripel generiert und Abstimmungen unmittelbar durchgeführt werden.

Zum Glück gibt es nun eine Datenbank-Plattform und eine Modellierungslösung, die diesen Herausforderungen gewachsen sind:

- MarkLogic® – eine nicht relationale, Multi-Modell-Datenbank-Plattform, die sowohl eine RDF-Tripel-Datenbank als auch eine Dokumentendatenbank umfasst. MarkLogic speichert die Quelldaten in ihrer nativen Form gemeinsam mit dem generierten RDF-Tripel
- ModelDR – eine Data-Point-Modellierungslösung, mit deren Hilfe die Generierung und Zuordnung von RDF- und Transformationsdesigns in MarkLogic visualisiert und verwaltet werden können

DIE MARKLOGIC DATENBANK-PLATTFORM

MarkLogic ist eine Enterprise NoSQL-Datenbank-Plattform, die speziell für die Daten von heute, einschließlich Dokumenten, Beziehungen und Metadaten, konzipiert wurde. MarkLogic ist die einzige Datenbank, die JSON, XML, RDF und viele weitere Formate nativ speichern und schnell abfragen kann. Sie erhalten somit eine einzige leistungsstarke Plattform für alle Ihre Daten. Das dokumentenzentrierte Datenmodell ist Schema-agnostisch und bietet Flexibilität bei der Datenmodellierung, da neu aufgenommene Daten nicht erst in Tabellenform aufgedgliedert werden müssen. MarkLogic speichert Transaktionen in ihrer ursprünglichen Form. Durch den Wegfall der Datentransformation vermeiden Sie starre ETL-Prozesse und den mit Datenkonvertierungen einhergehenden Verlust von Integrität und

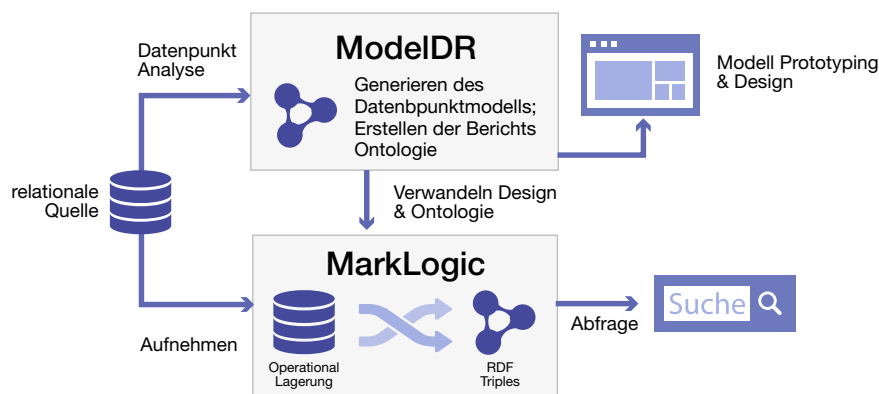


Abbildung 3

Kontext. Zudem werden das Laden von Daten aus verschiedenen Quellen sowie das Anpassen von Daten deutlich vereinfacht.

DIE ARCHITEKTUR DER MODELDR LÖSUNG

ModelDR stellt Benutzern eine grafische Benutzeroberfläche bereit, über die sie ihre relationalen Datenquellen in ein Datapoint-Modell überführen können. Dieses Datenpunktmodell wird als Ontologie dargestellt, und diese wiederum bildet die Grundlage für die Transformationszuordnung zur Erstellung der RDF-Tripel. Die Zuordnung kann von jeder Datenbank erfolgen, die RDF-Ontologien unterstützt. ModelDR bietet Vorlagen auf Grundlage bestehender branchenspezifischer Ontologien, wie z. B. FIBO (Financial Industry Business Ontology). Diese Vorlagen werden mit dem Datapoint-Modell kombiniert, um die Struktur für die Ontologie- und Transformationsschritte zu generieren. ModelDR ermöglicht Benutzern jedoch auch die Verwendung ihrer eigenen Ontologien.

DIE INTEGRATION VON MARKLOGIC IN MODELDR

Die ModelDR Lösung umfasst eine Verarbeitungs-Engine mit einer Front-End-Benutzeroberfläche für die Analyse relationaler Quellen, den Import von Ontologien und das Design von Transformationen für die Konvertierung relationaler Daten in das RDF-Format.

Abbildung 3 stellt den Prozess für die Integration der MarkLogic Datenbank-Plattform in ModelDR dar:

- **Generierung des Datapoint-Modells**
Durch die Kombination von ModelDR und

MarkLogic sind Benutzer in der Lage, relationale Modelle zu analysieren und in ein Datenpunktmodell zu konvertieren. Sie können hierfür die vorkonfigurierten ModelDR Vorlagen für gängige Ontologien aus der Finanzbranche oder bestehende Kunden-Taxonomien verwenden

- **Erstellung einer Designschicht**
In der DPM-Schicht wird die Datenkongruenz erreicht
- **Design der Transformation**
ModelDR generiert ein strukturelles Datendesign für die Transformation auf Grundlage von SPARQL, der Abfragesprache für RDF-Tripel
- **Verarbeitung des Transformationsdesigns**
MarkLogic verarbeitet das Transformationsdesign und erstellt auf den Quelldaten basierte RDF-Tripel. Dieser Prozess kann nachträglich durchgeführt oder in den Verarbeitungsprozess der Datenbank integriert werden
- **Abfrage der Daten**
SPARQL und die generierte DPM-Ontologie ermöglichen Benutzern das Abfragen von Daten

DIE VORTEILE VON MARKLOGIC UND MODELDR

MarkLogic erweitert ModelDR um eine flexible, agile Datenbank-Plattform. Dank seines Schema-on-Read-Ansatzes können viele Herausforderungen im Zusammenhang mit der Aggregation mehrerer Datenmodelle in einem einzigen Operational Data Store vermieden werden. Die Lösung bietet die folgenden wichtigen Vorteile:

ModelDR bereichert MarkLogic um einen intuitiven und

branchenspezifischen Mechanismus zum Generieren und Analysieren der komplexen Beziehungen zwischen Reporting-Ontologien und -Taxonomien und Daten. Dadurch können Benutzer die erforderlichen Ausgabedaten innerhalb kürzester Zeit und mit höherer Genauigkeit bearbeiten, analysieren und generieren.

ABSCHAFFUNG VON DATENSILOS

Durch die Bereitstellung eines einzigen, gemeinsamen Operational Data Store sind Benutzer nicht mehr an ein bestimmtes Datenmodell gebunden. Diese Multi-Modell-Fähigkeit vereinfacht die Handhabung von Änderungen beim Quelldatenmodell. Darüber hinaus bedeutet die Speicherung von Daten in ihrem nativen Format, dass MarkLogic als zentraler Data Hub verwendet werden kann, der ohne die Konvertierung von Daten auskommt.

HÖHERE DATENQUALITÄT

Daten können in ihrer nativen Form in MarkLogic aufgenommen werden, wodurch die Notwendigkeit umfassender Transformationsprozesse entfällt. In Kombination mit der Reduzierung von Datensilos senkt dies den Bedarf an „Schatten-IT“. Häufig suchen Benutzer eigenständig nach Quellen, führen die erforderliche Transformation durch und speichern die Daten in ihren eigenen Silos. Dieses als „Schatten-IT“ bezeichnete Phänomen ist in erster Linie auf die Einschränkungen in Bezug auf Datenmodelle und die fehlende Flexibilität zurückzuführen, die mit der Bereitstellung zentraler Data Hubs einhergehen. Dank der Multi-Modell-Fähigkeit von MarkLogic können Benutzer einen zentralen Service nutzen, ohne Abzüge bei der Datenmodell-Flexibilität in Kauf nehmen zu müssen. Die von ModelDR generierten Transformationsdesigns kommen im Rahmen des Verarbeitungsprozesses zur Anwendung oder können nachträglich genutzt werden, um die für das Reporting erforderlichen RDF-Tripel zu generieren. All dies erfolgt ohne Änderung der ursprünglichen Quelldaten. Über das bereitgestellte Data Hub kann eine einheitliche Sicht der Quelldaten und ihrer Derivate abgerufen und veröffentlicht werden.

SCHNELLERE INFORMATIONSBEREITSTELLUNG

Zahlreiche Facetten der MarkLogic Datenbank-Plattform beschleunigen den Informationsfluss von der Erstellung bis zur Nutzung. Durch die Vermeidung von Schemaänderungen lassen sich veränderliche Daten, die laufend aus den Informationen neuer Geschäftsprozesse generiert werden, problemlos

handhaben. Dank des Verzichts auf starre relationale Schemata entfällt die Notwendigkeit, Quellen erneut zu verarbeiten. Zudem sind auch keine ETL-Prozesse mehr erforderlich, denn die Einrichtung von Silos zur Gewährleistung geschäftlicher Flexibilität gehört der Vergangenheit an.

BESSERE AUFFINDBARKEIT VON INFORMATIONEN

Benutzer können mithilfe der von ModelDR generierten Ontologie problemlos die gewünschten Ergebnisse finden. Anstatt im Voraus die Facetten der gesuchten Informationen kennen zu müssen, können Benutzer die gewünschten Informationen durch eine freie Suche in der Ontologie abrufen.

SCHNELLERE RECHNUNGSLEGUNG DANK BITEMPORALER FUNKTIONEN

Die Plattform von MarkLogic umfasst eine bitemporale Abfragefunktion, die selbst den anspruchsvollsten regulatorischen Anforderungen gerecht wird. Durch die Abfrage von Informationen aus verschiedenen Datenmodellen und -quellen zum jeweiligen Berichtszeitraum lassen sich die Kosten für die Kontenabstimmung senken.

Gleichzeitig beschleunigt die Verwendung von bestehenden ModelDR Vorlagen oder geschäftsspezifischen Ontologien die Generierung der Datenansichten für Kunden. All dies sorgt für eine deutliche Reduzierung der im Zusammenhang mit der regulatorischen Berichterstattung anfallenden Betriebskosten.

SCHNELLE ERWEITERUNG DES WISSENS

Mithilfe von RDF-Technologien können neue Informationen im Kontext der Daten generiert werden. Die expliziten Beziehungen werden im Rahmen eines Schlussfolgerungsprozesses genutzt, um neue explizite Beziehungen zu erstellen und auf diese Weise das Wissen zu erweitern (Abbildung 4).

ModelDR bietet eine kompakte Modellierungslösung, die Daten ohne vorherige Transformation verarbeitet. Dies ermöglicht eine agilere Benutzererfahrung sowie eine schnellere Amortisierung der Kosten für Datenvorgänge. Der Modellierungsansatz von ModelDR stellt eine perfekte Ergänzung der Funktionen von MarkLogic zum Speichern von Daten in ihrer natürlichen Form dar.

Inference Example Trade Datenpunkt

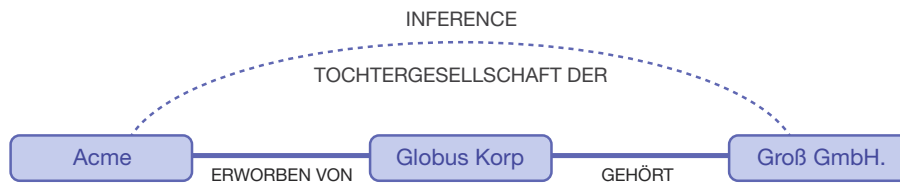


Abbildung 4

Normalerweise führt die Erweiterung des Wissens durch Daten zu einer regelrechten Datenexplosion, da die Daten innerhalb des Datenmodells mehrfach dupliziert werden. Das Datapoint-Modell verhindert diese unkontrollierte Zunahme des Datenvolumens, indem es die Nutzung von Werten durch mehrere Datenpunkte zulässt. Diese Mehrfachnutzung gespeicherter Werte sorgt für die Deduplizierung von Daten.

ERKENNTNISSE ÜBER VERSCHIEDENE QUELLEN HINWEG

Da die Struktur und die Metadaten explizit in die Daten integriert sind, können wissensbasierte Abfragen erstellt werden. Neue Regeln und Beziehungen lassen sich ganz einfach und ohne strukturelle Datenbankänderungen hinzufügen. Diese neuen Beziehungen werden anschließend als RDF-Tripel in der Datenschicht von MarkLogic dargestellt.

Das Datapoint-Modell ermöglicht ModelDR das Verarbeiten jedes beliebigen Datenschemas und das anschließende Laden von Lösungen in die MarkLogic Datenbank-Plattform. MarkLogic erlaubt das Auslesen der Informationen aus dem von ModelDR generierten RDF-Tripel sowie aus der ursprünglichen Quelle über eine einzige Abfrage. Ein Beispiel hierfür ist die Erfassung und anschließende Aufnahme des Datapoint-RDF-Tripels aus einem Finanzbericht im PDF-Format. Das PDF-Dokument wird in seiner nativen Form verarbeitet, und semantische RDF-Tripel werden hinzugefügt. Diese Datenpunkte können anschließend schnell auf ihre Quelldaten zurückgeführt werden, was besonders für die Analyse nützlich ist.

ZUSAMMENFASSUNG

Vielen Banken fällt es schwer, eine effektive Data Governance und Qualitätssicherung zu gewährleisten. Aufgrund der häufigen Änderungen der Datenmodelle in der Finanzbranche ist die Bereitstellung einer harmonisierten, standardisierten Sicht auf Daten zu bestimmten Zeitpunkten (häufig als Kongruenz bezeichnet) keine einfache Aufgabe.

Datapoint Modelling ist ein wirksamer Ansatz zur Bewältigung der vielen mit der Datenarchitektur und -aggregation verbundenen Herausforderungen, die das Risikomanagement, die regulatorische Berichterstattung und die fundierte Entscheidungsfindung erschweren. Allerdings erfordert die Datenpunktmodellierung in einer relationalen Datenbank eine Vielzahl von ETL-Prozessen. Durch die Kombination von RDF-Triples und Dokumentenmodellen lässt sich der Datenmodellierungsaufwand um mehrere Monate – wenn nicht sogar Jahre – reduzieren.

Die MarkLogic Datenbank-Plattform bietet die einzigartige Möglichkeit zum gemeinsamen Speichern von RDF-Tripeln und den ursprünglichen Trade-Data, aus denen die RDF-Tripel generiert wurden. Die ModelDR Lösung ermöglicht die Aufnahme von Daten in ihrer nativen Form und ruft anschließend weitere Daten in Tripeln ab. Gemeinsam schaffen die Technologien von MarkLogic und ModelDR eine hochgradig flexible Lösung für effizientes Reporting über verschiedene Datenmodelle und -quellen hinweg.

Weitere Informationen zu Enterprise NoSQL-Lösungen für Finanzdienstleister erhalten Sie auf unserer [Website](#).



MARKLOGIC GMBH

Theatinerstr. 11, 8. Etage, München 80333 | +49 897 10 42 2151
Skyper Villa, Taunusanlage 1, Frankfurt 60329 | +49 695 05 06 0588
de.marklogic.com | germany@marklogic.com