



Braincourt

Corporate Performance Management

Zukunftssichere DWh-Entwicklung

Die Bedeutung von Transparenz und Nachvollziehbarkeit

Teil 2: Nachvollziehbarkeit

Klaus-Rüdiger Wind, Manager
Braincourt GmbH

Braincourt GmbH

Managementberatung
& Informationssysteme

www.braincourt.com

Meisenweg 37
70771 Leinfelden-Echterdingen
Telefon +49 711 758580 0
Telefax +49 711 758580 80
info@braincourt.com

Rather Straße 110b
40476 Düsseldorf
Telefon +49 211 877420 0
Telefax +49 211 877420 11
info.duesseldorf@braincourt.com

Weitere Artikel finden Sie auf:
www.braincourt.de/presse-news.html

© Braincourt GmbH



Inhaltsverzeichnis

1	Einleitung	3
2	Lineage – Wege aus dem Labyrinth: Schnüre spannen und Spuren lesen	4
2.1	Die Beherrschung des Wandels durch Dokumentation	4
2.2	Lineage als Teil der durch das Anforderungsmanagement getriebenen Änderungsprozesse	6
2.3	Lineage in der Datenarchitektur.....	8
3	Schlusswort zu Teil 2	14



1 EINLEITUNG

Viele Data Warehouse (DWh)-Projekte scheitern und viele der DWhs, die aus einem Projekt in einen geregelten Betrieb überführt werden, haben lediglich eine kurze Lebensdauer. Warum ist das so? Ist doch nicht nur die Einführung eines DWhs eine i.d.R. sehr kostspielige, schätzenswerte Investition, sondern auch der Nutzen eines DWhs unbestritten.

Der Autor sieht folgende zwei wesentliche Gründe:

1. Die Transparenz gängiger DWh-Lösungen genügt nicht.
2. Die Nachvollziehbarkeit der DWh-Evolution und auch der Inhalt eines DWh selbst haben einen zu geringen Stellenwert.

Dies ist der zweite Teil einer dreiteiligen Artikelserie.

Teil 1 befasste sich mit der Situationsdarstellung und der Frage, warum die konsequente Umsetzung transparenter Prozesse, Architekturen, Daten und Strukturen eine wesentliche Voraussetzung für eine erfolgreiche, strategisch tragfähige DWh-Entwicklung und den DWh-Betrieb ist.

Dieser zweite Teil befasst sich mit der Nachvollziehbarkeit der DWh-Entwicklung und der Inhalte und in diesem Sinne also auch mit der Herkunft von Änderungen, die ein DWh im Laufe der Zeit erfährt und der Änderungsverfolgung. Mit der Verwendung von Lineages, die als Metadaten in einem System hinterlegt werden und als Spuren verstanden werden können, die der Nachvollziehbarkeit unterschiedlichster Sachverhalte dienen, wird ein Lösungsweg aufgezeigt, der dabei hilft, ein DWh zukunftssicher zu entwickeln bzw. weiterzuentwickeln.

Gegenstand des dritten Teiles wird ein umfassendes Resümee über die Möglichkeiten der Zukunftssicherung bei der Entwicklung eines Data Warehouses sein.



2 LINEAGE – WEGE AUS DEM LABYRINTH: SCHNÜRE SPANNEN UND SPUREN LESEN

Die Anforderungen an das Business steigen. Internet und Social Media Kanäle ermöglichen dem Konsumenten mehr Transparenz über Produkteigenschaften und Preise. Die erhöhte Volatilität in einigen Märkten, wie z.B. bei den Rohstoffen, erfordert mehr Wachsamkeit und verstärkter Wettbewerbsdruck benötigt strategische Flexibilität. Diese Anforderungen gilt es auch system-technisch nachzuvollziehen, Auswertungen entsprechend flexibel zu gestalten, Planungen jederzeit anpassbar zu machen und auch mit unterschiedlichsten Szenarien zu spielen. Die Beherrschung dieser Dynamik und Flexibilität setzt die Beherrschung der steuerungsrelevanten IT-Systeme voraus. Wesentliche Voraussetzungen um ein System – hier ein Data Warehouse – zu beherrschen sind, neben dem Wissen über die Architektur, das Datenmodell und die Prozesse, insbesondere Transparenz bzgl. der Änderungsprozesse und auf dieser Basis deren uneingeschränkte Beherrschung.

Vor dem Hintergrund komplexer Unternehmensanwendungen die bspw. auf Technologien der Unternehmen Oracle, SAP, Siebel, PeopleSoft und JD Edwards basieren und welche mit ihren herstellereigenen Datenarchitekturen heutige heterogene IST-Architekturen prägen, sind diese Anforderungen an Änderungsprozesse noch dringlicher. Hierfür werden nicht nur Daten und die Modellierung der Zustandsübergänge bei Datenänderungen benötigt, sondern darüber hinaus die Metadaten und das Wissen über deren Abhängigkeiten und Zustandsübergänge.

Es sind regelmäßige Abgleiche (engl. Alignments) zwischen der Umsetzung der Daten- bzw. Informationsarchitektur und den Anforderungen an dieselbe zu leisten.

Lineages sind dabei technische Eigenschaften (i.S.v. Metadaten) zur Herstellung und dauerhaften Gewährleistung von Transparenz und können als Schnur verstanden werden, die zwischen der Datenherkunft und der Datenverwendung gespannt wird.

2.1 Die Beherrschung des Wandels durch Dokumentation

Data Warehouses stehen vor der Aufgabe, den Prozess des Wandels beherrschbar und nachvollziehbar zu machen. Für die „richtige“ Analyse von Unternehmensdaten ist die langfristige Perspektive und die Datenhistorie bzw. die Ursprungsquelle von Daten und Informationen von Bedeutung. Vor diesem Hintergrund wird das Design der Daten- und Informationsarchitektur im DWH prinzipiell vor folgende grundlegende Fragen gestellt:

- **Woher kommen wir?**
Welche Ausgangssituation hat dazu geführt, das Data Warehouse zu initiieren und aufzubauen?

Grundlegende Fragen
der Daten- und
Informations-architektur



- **Wohin gehen wir?**
Welche Anforderungen stellen sich heute und in Zukunft? Wie kann die bestehende BI-Landschaft für die Zukunft ausgerichtet werden? Wie wirkt sich dies auf das DWh aus?
- **Auf welchem Weg sind wir gegangen?**
Welche Änderungen sind, in welcher Abfolge, wo eingeflossen und werden durch welche Strukturen repräsentiert? Ist im Voraus erkennbar, welche Auswirkungen Änderungen auf das System, die Daten und die Informationsqualität haben?
- **Wird der Weg beschrieben?**
Werden Änderungen, unabhängig von ihrer Veranlassung, vollständig protokolliert? Ist die Nachvollziehbarkeit gleichermaßen für beabsichtigte Änderungsplanungen und -prozesse wie auch für unbeabsichtigte Änderungsereignisse gegeben? Sind historische Daten dauerhaft verfügbar? Inwieweit können DWhs ihr Gedächtnis verlieren (Vorratsdatenspeicherung, Anonymisierung, politisches Interesse an Intransparenz)?
- **Was werden wir erreichen?**
Wie stark zahlt bspw. eine neu eingeführte bzw. geänderte Funktion in die Erreichung eines konkreten Ziels ein? Welche Informationsbedürfnisse werden befriedigt? Welcher Nutzen erzielt?
- **Welcher Weg führt in die Zukunft?**
Werden wir dem Anspruch auf fachliche und technische Skalierbarkeit gerecht? Welchen Einfluss haben die Änderungen der Vergangenheit und die derzeit geplanten Änderungen an die zukünftige Flexibilität des Systems? Könnten zukünftige Änderungsanforderungen gegen die Flexibilitätsanforderungen wirken?

Das Protokollieren der Datenherkunft ist eine kritische Funktion bei unternehmenswichtigen IT-Systemen und Prozessen wie DWh, Enterprise Application Integration und Master Data Management (MDM). Der Datenfluss muss inhaltlich widerspruchsfrei aus allen unterschiedlichen Quellen und entlang der Transformationsschritte durch die Anwendungen und Geschäftsbereiche bis zur Auswertung nachvollzogen werden können. Die Dokumentation und die Kommunikation sind demzufolge Schlüsselaktivitäten insbesondere bei der Entwicklung und im Betrieb. Doch auch bereits bei der Anforderungsbeschreibung durch einen Fachanwender, der z.B. durch die Definition einer neuen Kennzahl seine Anforderung beschreibt, ist es wertvoll einen methodischen Standard anzuwenden, der alle Anforderungen an eine ganzheitliche, fachliche und technische Dokumentation erfüllt (Daten und Metadaten).



Die Beherrschung des Wandels drückt sich darin aus, dass Änderungsprozesse ...

- ... auditierbar sind.
Audits stellen sicher, dass die in Berechnungen verwendeten Daten richtig ausgewiesen werden. Wenn beim Auditing Fehler oder Schwächen in der Datenqualität entdeckt werden, muss der exakte Pfad der Datenverarbeitung entweder ausgehend von der Quelle oder zurücklaufend zur Quelle verfolgt werden können (track down / trace).
- ... anhand der Datenmodellierung und der erfolgten Protokollierung nachvollziehbar sind.
- ... regelmäßig einem Review unterzogen werden, so dass sich aus diesem Prozess Erkenntnisse ergeben, aus denen Schlüsse für die Zukunft gezogen werden können.
- ... standardisiert, effektiv und effizient geplant und umgesetzt werden.
- ... in ihren Durchlaufzeiten kontinuierlich beschleunigt werden.

2.2 Lineage als Teil der durch das Anforderungsmanagement getriebenen Änderungsprozesse

Wird das Paradigma „IT follows business“ zu Grunde gelegt, so beeinflusst die Beschaffenheit der strategisch-fachlichen Aspekte die Gestaltung der operativ-technischen Aspekte.

Das Anforderungsmanagement als Unterstützungsprozess des (auf ein DWh bezogenes) Änderungsmanagements zu etablieren, ist hier von besonderer Bedeutung. Die Änderungsprozesse im DWh können mehrere Funktionsbereiche betreffen:

- **Requirement Management im DWh Projekt**
Betrifft die Steuerung und Verwaltung von Anforderungen an das DWh inkl. der Kommunikation mit den Anforderungsabsendern und -adressaten. Funktionsbereiche, aus denen die Anforderungen kommen
- **Management of Change (ITIL)**
Betrifft den Prozess zur formalen Umsetzung und Abwicklung von technischen Änderungsanforderungen innerhalb einer DWh-Landschaft nach Rollout; dies erfolgt i.d.R. über Change Requests an produktiven Systemen im Rahmen der Anwendungsbetreuung (Application Management)
- **Deployment von Software-Patches**
Eine durchgängige Prozesssicht von der Anforderungsbeschreibung bis hin zur technischen Realisierung ist vor allem beim Lösen von Softwarefehlern und dem in der Folge vorzunehmenden Einspielen von



Patches von Bedeutung: Neue Software oder -komponenten lassen sich konkretisieren, indem man ihren Herstellungsprozess beschreibt. Die Ursache dafür, dass ein Patch entsteht, liegt im Auftreten eines Problems, das eine Reihe möglichst klar definierter Analyse-, Programmierungs-, Test- und Deployment-Aktivitäten auslöst. Ein Patch definiert sich über das Ausführen dieser Schritte. Als Ergebnis wird ein Software-Patch eingespielt, der beschreibt, welches konkrete Problem (z.B. via Trouble Ticket Nummer) gelöst wird.

- **IT Demand Management**

Betrifft die Aufgaben bzw. Prozesse im IT-Management und IT-Betrieb, welche sich mit der Aufnahme, Bündelung, Bewertung und fachlichen Umsetzung von Geschäftsanforderungen (business requirements) an das DWh befassen. Als fortlaufender Prozess ist das IT Demand Management nicht projekt- oder systemgebunden und insofern auch dem hier angesprochenen konkreten DWh-System übergeordnet. Es ist inhaltlich am weitesten gefasst und greift die Geschäftsanforderungen strategisch ab – unabhängig davon, ob und wie diese in der IT umgesetzt werden.

Data Warehouse-Organisationen, die Neu- und Änderungsanforderungen an das DWh verwalten, bedienen sich einer anforderungsbezogenen Lineage. Als Quelle für die Bestimmung der Herkunft der Anforderungen müssen Verweise auf allen Abstraktionsebenen der Anforderungsbeschreibungen und der fachlichen sowie technischen Spezifikationen vorhanden sein. Dies sollte idealerweise bis in die einzelnen, implementierten Objekte weiter geführt werden.

Eine "Traceability Matrix"¹ (s. Abb 1) stellt dar, welche Systemanforderung (Product Requirement) woher stammt (Requirement Source), worauf sie sich sowohl in der Grobspezifikation (HLD, High Level Design), als auch in der Feinspezifikation (LLD, Low Level Design) bezieht und welche Software- bzw. Programmteile von der (Neu- oder Änderungs-) Anforderung betroffen sind (Code Unit, User Test Specification, System Test Specification). Ebenso wird auf das konkrete Kapitel des Anwender-Handbuches (User Manual) hingewiesen, in dem sich die Anforderung niederschlägt.

¹ Linda Westfall „Bidirectional Requirements Traceability“ / www.westfallteam.com / 2006



Requirement Source	Product Requirements	HLD Section #	LLD Section #	Code Unit	UTS Case #	STS Case #	User Manual
Business Rule #1	R00120 Credit Card Types	4.1 Parse Mag Strip	4.1.1 Read Card Type	Read_Card_Type.c Read_Card_Type.h	UT 4.1.032 UT 4.1.033 UT 4.1.038 UT 4.1.043	ST 120.020 ST 120.021 ST 120.022	Section 12
			4.1.2 Verify Card Type	Ver_Card_Type.c Ver_Card_Type.h Ver_Card_Types.dat	UT 4.2.012 UT 4.2.013 UT 4.2.016 UT 4.2.031 UT 4.2.045	ST 120.035 ST 120.036 ST 120.037 ST 120.037	Section 12
Use Case #132 step 6	R00230 Read Gas Flow	7.2.2 Gas Flow Meter Interface	7.2.2 Read Gas Flow Indicator	Read_Gas_Flow.c	UT 7.2.043 UT 7.2.044	ST 230.002 ST 230.003	Section 21.1.2
	R00231 Calculate Gas Price	7.3 Calculate Gas price	7.3 Calculate Gas price	Cal_Gas_Price.c	UT 7.3.005 UT 7.3.006 UT 7.3.007	ST 231.001 ST 231.002 ST 231.003	Section 21.1.3

Eine Traceability Matrix dient zur Schaffung von Transparenz zwischen Anforderung, Deliverable und Deployment

Abbildung 1: Beispiel einer Traceability Matrix

Die in der Traceability Matrix zur Anwendung kommenden IDs (z.B. R00120, UT 4.1.032, ST 120.020) werden von einem zentralen Repository automatisch vergeben. Auf diese Weise entsteht eine ganzheitliche sowohl aus fachlicher als auch aus technischer Sicht nachvollziehbare Dokumentation der Stellen, an denen sich Veränderungen im System und den dazugehörigen Dokumenten auswirken. Darüber hinaus zeigt diese Methodik alle Abhängigkeiten zwischen den vorgenommenen oder geplanten Änderungen auf, welches gerade im Falle einer Fehlersuche von erheblichem Nutzen sein kann. Letztere wird auch durch das Vorwärts- und Rückwärts-Tracking unterstützt, welches durch die konsequente Verwendung eindeutiger und konsistenter IDs ermöglicht wird.

Die aufgezeigte Methodik wird natürlich nicht nur dazu verwendet transparent zu machen, an welchen Stellen sich eine Anforderung auswirken wird. Vielmehr dient sie auch als Plattform dafür, den jeweiligen Bearbeitungsstatus zu dokumentieren und aufzuzeigen. Somit ist jederzeit klar, in welchem Entwicklungsstadium sich ein System und die dazugehörigen Dokumente in spezifischen Umgebungen (Entwicklung, Test, Produktion) befinden oder zu beliebigen Zeitpunkten in der Vergangenheit befunden haben.

Das beschriebene „Trace Tagging“ wird, nebst allen dazugehörigen Prozessen und Teilprozessen (z.B. Vergabe von IDs, Monitoring, Tracking) software-unterstützt durchgeführt.

2.3 Lineage in der Datenarchitektur

Datenarchitektur ist eine spezifische Teilarchitektur der Informationssystem-Architektur und beschreibt die Datenstruktur eines Informationssystems. Die Aufgabe eines DWhs ist es (im Zusammenspiel mit darauf aufsetzenden BI-Tools) Informationen aus Daten bereitzustellen.

Ein Definitionsansatz für den Begriff Information ist im Zusammenhang zwischen Ursache und Wirkung begründet (Causa efficiens = wirkende Ursache), d.h. ein Sachverhalt wird definiert, indem dargelegt werden kann, wie er entstanden ist.

Allgemein stellen Lineages Bezüge zwischen Informationsobjekten, auf der Grundlage einer Abhängigkeits-Relation her. Bezüge können sich dabei bspw. durch die zeitliche Abfolge darstellen oder durch eine Vorgänger-Nachfolger-Beziehung.

Data Lineage beantwortet daher in einem DWh-System die Frage nach der Datenherkunft, also wie zu gegebenen aggregierten Datensätzen die ursprünglichen Datensätze bestimmt werden können, aus denen sie entstanden sind. Dies führt zu Aufzeichnungen über Abstammungen, historische Verzweigungen, Herkunft von Daten und Strukturen, Informationen und Anforderungen, Visualisierung von Bewegungsprofilen der sich ändernden Daten.

Im Rahmen der Datenarchitektur kann Lineage folgendermaßen klassifiziert werden:

a) Content-bezogene Lineage

Content-bezogene Lineage ist insbesondere relevant bei Data Warehouses, die auf Basis temporaler Datenbank-Modellierung die Änderungen aller fachlichen Sachverhalte (z.B. fast oder slowly changing dimensions) speichern, um Historienvergleiche der Datenänderungen durchführen zu können: „as is“- versus „as was“-Betrachtungen.

Mögliche fachliche Fragestellungen:

- Zu welchen Ergebnissen würde die Analyse der heutigen Kennzahlen auf Basis der Strukturen der Vergangenheit führen? Bspw. bezogen auf die geänderte Organisationsstruktur (s. Abb. 2).

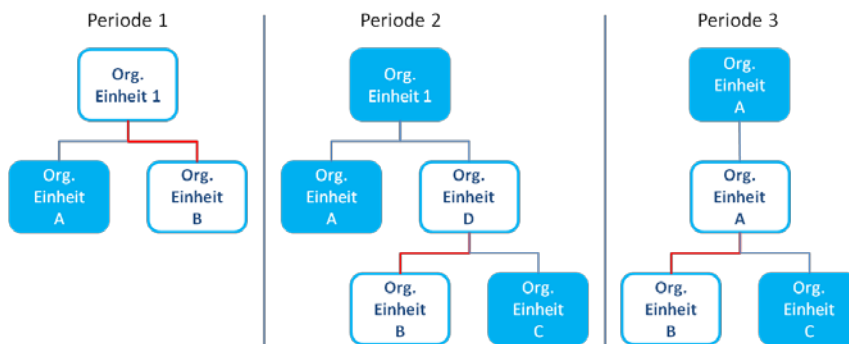


Abbildung 2: Periodenabhängige Änderungen von Organisationseinheiten

- Zu welchen Ergebnissen würde die Analyse der Kennzahlen auf Basis der Rechenregeln der Vergangenheit führen? Im Kennzahlenbaum wird die Abhängigkeit der Kennzahlen untereinander dokumentiert (s. Abb. 3): Die Ableitung von Kennzahlen, die sich mittels Rechenvorschriften aus den originären Kennzahlen bzw. Berichtspositionen ergeben, wird festgehalten. Die Rechenvorschriften müssen historisiert werden, damit etwaige Anpassungen an denselben bzgl. der jeweiligen Auswirkung rückverfolgt bzw. über eine Abweichungsanalyse quantifiziert werden können.

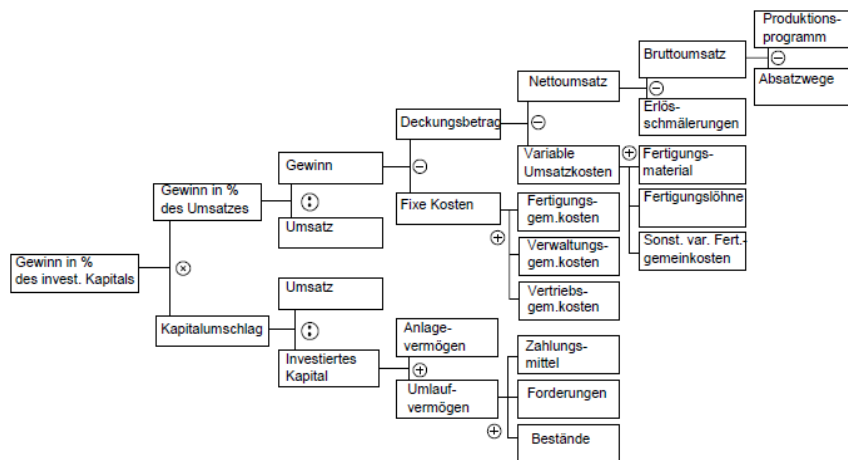


Abbildung 4: Beispiel eines Kennzahlenbaums

b) Konzeptionstransfer-bezogene Lineage

Diese hat insbesondere Bedeutung bei Data Warehouses, in denen die Historie des Transfers zwischen Fachlichkeit und technischer Umsetzung verwaltet und gespeichert wird. Die Lösung stellt Transparenz über den umgesetzten Konzeptionstransfer her, insbesondere bei iterativem Durchlaufen der Abstraktionsebenen und der dafür vorgesehenen Modellierungen.

Bei konsequentem Design-Ansatz (top down) können konzeptionstransfer-bezogene Lineages als Quelle für die Bestimmung der Herkunft bei der semantischen bzw. logischen und physischen Datenmodellierung dienen (s. Abb. 4).



Abbildung 3: Top-Down Ansatz bei der Datenmodellierung (fachlich→technisch)

Die Nachvollziehbarkeit der Anforderungsumsetzung über die unterschiedlichen Design- und Entwicklungsstufen hinweg wird über eindeutige Codes für die Anforderungen (IDs) gewährleistet. Diese sind im semantischen Datenmodell an den modellierten Informationsobjekten hinterlegt. Der Übergang aus dem semantischen Datenmodell in das logische Datenmodell ist ebenfalls über eine Codierung zu versorgen, so dass entweder direkt im logischen (Entity Relationship-) Datenmodell oder mittelbar über das semantische Datenmodell erkennbar ist, welche Anforderung bspw. zu welcher Entität, zu welchem Attribut oder zu welcher Relation geführt hat. Das physische Datenmodell definiert u.a. die Tabellen, Partitionen, Tablespaces etc. Auch hier ist der Rückschluss auf die Objekte aus der logischen Modellierung mittels eindeutigen Bezeichnern herzustellen.

Vor dem Hintergrund fehlender integrierter Tools für alle Abstraktionsebenen, ist der Konzeptionstransfer häufig Gegenstand von Informationsbrüchen. Deshalb ist für das Legen von Änderungsspuren, also für Lineage gerade auch der Übergang von dem einen in das andere Modell interessant.

c) Datenherkunft (Data Lineage)

Data Lineage ermöglicht das Nachvollziehen der Datenherkunft durch mehrere Technologiestacks und Data Warehouse Layers hindurch. Dies basiert auf der lückenlosen Anzeige des Datenflusses von der Datenquelle durch verschiedene Systeme und Datenebenen hindurch bis zur Ziel-Verwertungs-/Auswertungsebene (vgl. Abb. 5).

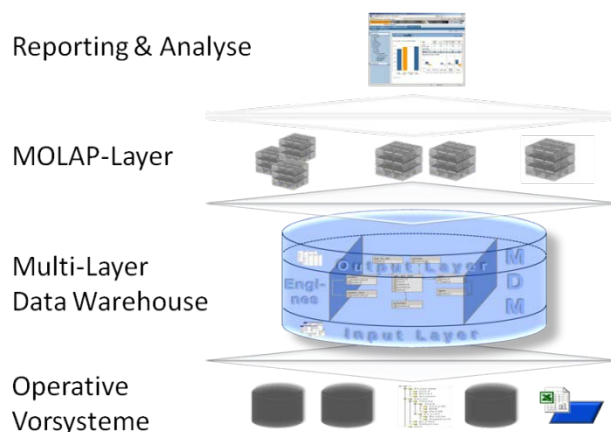


Abbildung 5: Der Datenfluss in einer klassischen DWh-Architektur

Es stellt die Dokumentation bereit, woher die Daten kommen, wie sie umgewandelt und transformiert werden und wo sie abschließend landen. Data Lineage kann bspw. darstellen, ob und wie Daten aus einer Datenbanktabelle eine Kennzahl in einem BI Bericht beeinflussen.



- Data Lineage dient als Quelle für die Bestimmung des strukturellen Aufbaus der ETL-Prozesse über mehrere Technologie-Ebenen hinweg.
- Im DWh sollte der Verlauf der Datentransporte an den Daten ablesbar sein (Metadaten): Hier werden üblicherweise im Rahmen von ETL-Prozessen Daten aus verschiedenen Quellen extrahiert, nach bestimmten Regeln transformiert und zur Analyse bereitgestellt.
- Source Control Systeme zur Verwaltung der Source Code-Versionen.

Data Lineage muss häufig den umgekehrten Weg beschreiten, um von Analyseergebnissen zu den Quellen zu gelangen. Dies findet z.B. dann statt, wenn Daten nicht plausibel erscheinen oder zum Nachweis der Compliance rückverfolgt werden müssen. Dazu werden die Transformationen mathematisch modelliert, um für gegebene Ausgabewerte einer Transformation, die dazu gehörenden Eingabewerte zu bestimmen.

Die Navigation entlang der Datenstammbäume (die Quellen und die hierauf aufbauende, nachfolgende Verarbeitungshistorie) basierend auf Lineages erfolgt in zwei Richtungen, um

- die Herkunftsquellen im Source System zu identifizieren (zurück zu verfolgen) und
- abgeleitete Objekte bis zum Zielsystem zu verfolgen.

d) Historical Lineage (u.a. Strukturelle Historisierung)

Historical Lineage hat besondere Bedeutung für Data Warehouses, die auch die Schema-Informationen bzw. Änderungen an Data Warehouse-Objekten verwalten und speichern. Die Lösung stellt Transparenz über die strukturelle Entwicklung der verschiedenen Umsetzungsvarianten der Datenbereitstellung, -transformation oder -errechnung mittels Business Rules her. Damit ergibt sich z.B. die Antwort auf die Frage, welche Funktion in der Vergangenheit (ggf. auf einem konkreten Patchlevel) genutzt wurde, um eine Kennzahl zu berechnen? Hierbei ist dafür zu sorgen, dass Datenstrukturänderungen (Versionsverwaltung, Branches) und die Historisierung der Schnittstellen mittels Integration von Source Control Systemen beim Deployment im DWh „Spuren“ hinterlassen, dass also ein Datengedächtnis über vergangene Strukturen aufgebaut und auswertbar bereitgestellt wird.

Bei den Marktführern unter den Herstellern relationaler Datenbanksysteme ist es i.d.R. state of the art, nicht nur Datenänderungen, sondern auch Änderungen an den Datenstrukturen zu historisieren und auswertbar verwalten zu können. Dies ist insbesondere für eine qualitätsgesicherte



Deployment-Strategie unabdingbar. Die dazugehörige Design-Entscheidung der Nutzung der vorhandenen Datenbankfeatures ist damit aber selbstverständlich nicht gefällt.

Lineages können auf zwei grundlegende Arten hergeleitet werden:

a) Explizite Herleitung

Direkte Speicherung der Lineage-Informationen in der Dokumentation und in Systemen für alle interessierenden Sachverhalte:

- Als Teil der Governance-Strategie in der Dokumentation und dem Systemdesign, woraus sich fachlich-technische Abstraktionsebenen, Detaillierungen und Data Warehouse-Roadmaps ableiten lassen.
- Bei der Modellierung der Datenhaltung, wodurch Beziehungen der Daten untereinander (historisch) explizit gespeichert werden.
- Bei der Datenerhebung und Ladestrategie (ETL).

b) Implizite Herleitung

Mittels forensischer Ad-hoc-Analyse oder Profiling, aus einem Sonder-sachverhalt heraus. Hierbei sind Aussagen (insbesondere bei Informationsbrüchen) nur auf Wahrscheinlichkeiten begründet. Sie wird nur bei Bedarf durchgeführt und ist nicht als durchgängige Datenbasis vorhanden.

Reverse Engineering von ETL-Prozessen und Datenmodellen stellt hier eine technisch bereits weit fortgeschrittene Disziplin zur nachträglichen Dokumentation von Lineages dar.



3 SCHLUSSWORT ZU TEIL 2

Ein Data Warehouse muss so dynamisch sein, wie das Unternehmen selbst, dessen Steuerung es unterstützen soll. Das bedeutet, dass es einer hohen Änderungsfrequenz effizient und effektiv gerecht werden muss, ohne an Stabilität zu verlieren.

Um ein solch komplexes System zu beherrschen, ist es wichtig, insbesondere den Umgang mit Neu- und Änderungsanforderungen zu beherrschen und dafür zu sorgen, dass jegliche Änderungsverfolgung möglich ist. Hierfür sind über den gesamten DWh-Entstehungsprozess und über den gesamten DWh-Lifecycle hinweg, entsprechende Vorkehrungen zu treffen. Um die Änderungen transparent und nachvollziehbar zu machen, müssen sie – begonnen bei der Anforderungsspezifikation bis hin zur Beschreibung der technisch umgesetzten Anforderung im Anwender-Handbuch – einem methodischen Standard gehorchend dokumentiert werden. Ein Instrument, welches dabei unterstützen kann, ist die „Traceability Matrix“.

Eine bewährte Methodik, die Spuren (Lineages) in den Dokumenten und im System so zu legen, dass sie im Rahmen einer Änderungsverfolgung leicht wieder auffindbar und nutzbar sind, ist es, sie in Form von Metadaten zu hinterlegen. Sie findet Anwendung auf Daten, Datenstrukturen, Prozesse und die Systemarchitektur.

Dieses, zum Zwecke der Änderungsverfolgung erforderliche, explizite Anlegen von Daten und Metadaten in den Dokumenten und im System führt natürlich zu Mehraufwand, der bei der Budgetierung eines DWh-Projektes berücksichtigt werden muss. Da diese Maßnahme allerdings die Entwicklung, Anpassung und Weiterentwicklung eines DWh signifikant vereinfacht und beschleunigt, beginnt die Amortisation der Investition bereits bei der ersten durchzuführenden Änderung des Systems. Konsequenzen sind auf Dauer günstigere DWh-Projekte und ein günstigerer DWh-Betrieb.

Ihre Ansprechpartner:

Klaus-Rüdiger Wind

Manager

Braincourt GmbH

Meisenweg 37

70771 Leinfelden-Echterdingen

klaus-ruediger.wind@braincourt.com

Telefon: +49 711 75 85 80 - 65

Kosmas Kalpakidis

Geschäftsführer

Braincourt GmbH

Meisenweg 37

70771 Leinfelden-Echterdingen

kalpakidis@braincourt.com

Telefon: +49 711 75 85 80 – 0